SOEP The Socio-Economic Panel

825

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel at DIW Berlin 2020

SOEP-Core – 2018: Documentation of the Interviewer Dataset (1984 until 2018)

Martin Gerike

DIW SOEP

Running since 1984, the German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at **http://www.diw.de/soepsurveypapers**

Please cite this paper as follows:

Martin Gerike. 2020. SOEP-Core – 2018: Documentation of the Interviewer Dataset (1984 until 2018). SOEP Survey Papers 825: Series D. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core – 2018: Documentation of the Interviewer Dataset (1984 until 2018)

Martin Gerike*

February 24, 2020

DIW Berlin – Deutsches Institut für Wirtschaftsforschung e.V.
Sozio-oekonomisches Panel (SOEP)
Mohrenstraße 58
10117 Berlin

*based on the 2017 version of Martin Kroh, Luisa Hilgert and Arne Langlet

Contents

1 Overview

The SOEP does not only aim at collecting high-quality data on the living conditions and well-being of households, but –as a by-product of internal quality assurance processes– it lends itself increasingly as an empirical source for survey research. The purpose of the INTERVIEWER file is to provide user convenient access to all available, longitudinal information on the SOEP interviewers. These information on interviewers may be linked to the wave-specific person and household files via the interviewer identifier (INTID) and the year of observation. Information stemming from different sources are brought together in this file. First, we derive basic demographical information and employment spells on interviewers from personnel data of the fieldwork organization. Since 2000, Kantar Public regularly updates these information. Second, in 2006, 2012 and 2016, more than 95 percent of all SOEP interviewers of these years completed a short version of the standard individual questionnaire themselves. This allows, among others, a one to one match between responses of interviewers and responses of interviewees. Third, we generated user-friendly information on the working experience and workload of interviewers from the SOEP person and household interviews conducted by them as well as the fieldwork information they report to Kantar Public.

The INTERVIEWER file is available in long format (interviewer × waves) and covers all interviewers who ever conducted at least one interview in the SOEP and a few more interviewers active in SOEP-related studies and pretests who participated in the interviewer surveys of 2006, 2012 and 2016. In 2018, the INTERVIEWER file contains a total of 2935 interviewers. Many of the variables of INTERVIEWER are by definition time-variant, such as the annual workload. Information on gender, year of birth, the start and end of being an interviewer at the SOEP is time-invariant as well as information on country of birth, mother tongue, and country of birth of the interviewer's parents. The same holds true for classification of interviewers in the face-to-face fieldwork and those working in the centralized call center administrating mailed questionnaires.

2 Contents of INTERVIEWER

This section provides an overview of information included in the INTERVIEWER file as well as the respective sources of these information. In the case of generated variables, this section also describes the coding of this information. The following pages give an overview of all variables included in the INTERVIEWER file. The Appendix provides some additionnal sources of information for different purposes such as useful links to the INTERVIEWER-related parts of SOEP Companion and Paneldata.org or the questionnaires.

2.1 Personnel Records

Kantar Public personnel data[1] are updated every year as of 2000. They contain information on the working experience of interviewers within the fieldwork company, education, and marital status (see section 2.4).

2.2 Interviewer Surveys 2006, 2012 and 2016

In 2006, 2012 and 2016, the interviewers filled in (mailed) questionnaires themselves. The interviewer questionnaire mimics for the most part the regular SOEP person questionnaire. Hence, we obtain not only information on basic demographics of interviewers (gender, year of birth, etc.), but also their occupational history, health status, denomination, family status, country of birth, and mother tongue. While in the surveys of 2006 and 2012 the interviewers were asked about their gender, marital status and education, these questions were omitted in the 2016 survey.

Moreover, we collect standard SOEP data on social and political attitudes, life satisfaction, and the interviewer's worries. Additionally questions relate to their personality drawing on the "Big Five"-inventory, risk aversion, impulsiveness, and trust (see questionnaire at the end of the document).

Besides these SOEP-items, interviewers are asked to answer some questions which directly relate to their work. Thus the interviewers are asked about their motivation to work as an interviewer and whether these expectations are fulfilled and for how long they plan to work as an interviewer for the SOEP. Furthermore, the interviewers are asked to estimate the willingness of different given social groups to participate in the SOEP. Finally, the interviewers report on items of an inventory of social desirable responding (Winkler et al. 2006).[2] In 2012, interviewers also report on their participation in interviewer trainings and working satisfaction. Furthermore, interviewers are asked about their competence in a set of languages and about the country of birth of their parents. Finally, interviewers are asked about their actual and preferred working hours.

Country of birth, mother tongue, and country of birth of the interviewer's parents (all binary "Germany" vs. "not Germany") is time-invariant and therefore extended to all years. The INTERVIEWER file lists all time variant information in the respective survey years 2006, 2012

[1] These data were provided to SOEP users until data version v30 in form of the INTVIEW-files

[2] Winkler, Niels, Kroh, Martin, and Martin Spiess. 2006. Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit. DIW Discussion Paper 579.

and 2016 only. In case of conflicting information between personnel data of the fieldwork organization and the interviewers' self-reports provided in the surveys, we document this in form of a flag-variable.

2.3 Fieldwork Outcomes and Interviewer Activity

Information concerning the work as an interviewer in the SOEP is derived from the $p-, $h-, $pbrutto, and $hbrutto-files for the years 1984 till 2018, the $lela-files for the years 1995 till 2018, and the $youth-files for the years 2000 till 2018. Particularly, we use meta-data from these files on, for instance, the number of contacts before interview, the mode of data collection, date of the interview, duration of the interview, and of course the interviewer who conducted the interview. We aggregate the person and household level data, for instance, to the amount of interviews conducted in one year by an interviewer, the pace of interviewing, the mean number of contacts with households, the region of the primary interviewer allocation, and the size of the communities of the primary interviewer allocation. The following paragraphs document the generation of these variables:

2.3.1 Basic Information

TYPINT (interviewer works in the face to face field or in the centralized call center) The variable TYPINT indicates whether an interviewer works in the face to face field or if they administers mailed questionnaires from a centralized call center.

EXPERIENCE The variable EXPERIENCE gives the cumulative number of years an interviewer worked for the SOEP so far. EXPERIENCE is given for each interviewer and for each year from 1984 until 2018. The variable increases by 1 if the interviewer conducts at least one person, household, or youth questionnaire in one year.

2.3.2 Fieldwork: Period and Region

FIRSTINTM (month of first interview in respective year) The variable FIRSTINTM provides the month of the first personal, youth, and household interview for each interviewer for each year from 1984 until 2018. The generation of the variable FIRSTINTM is based upon the variables $pmonin from the $p-files, $jmonin from the $youth-files and upon the following variables from the $h-files: $hmonin for the years 1984 until 1988, as well as $hhmonin from 1989 onwards.

FIRSTINTD (day of first interview in respective year) The variable FIRSTINTD provides the day of the first personal, youth, and household interview for each interviewer for each year from 1984 until 2018. The generation of the variable FIRSTINTD is based upon the variables $ptagin from the $p-files, $jtagin from the $youth-files and upon the following variables from the $h-files: $htagin for the years 1984 until 1988, as well as $hhtagin from 1989 onwards.

LASTINTM (month of last interview in respective year) The variable LASTINTM provides the month of the last personal, youth, and household interview for each interviewer for each year

from 1984 until 2018. The generation of the variable LASTINTM is based upon the variables $pmonin from the $p-files, $jmonin from the $youth-files and upon the following variables from the $h-files: $hmonin for the years 1984 until 1988, as well as $hhmonin from 1989 onwards.

LASTINTD (day of last interview in respective year) The variable LASTINTD provides the day of the last personal, youth, and household interview for each interviewer for each year from 1984 until 2018. The generation of the variable LASTINTD is based upon the variables $ptagin from the $p-files, $jtagin from the $youth-files and upon the following variables from the $h-files: $htagin for the years 1984 until 1988, as well as $hhtagin from 1989 onwards.

ENDINT (year of Last SOEP-Interview) The variable ENDINT provides the year of the last SOEP-interview for each interviewer. It is constant over time. The generation of the variable ENDINTis based upon information from the SOEP fieldwork data that indicates when the interviewer's last year was when he or she worked for the SOEP.

MODBULA (primary interviewer allocation, federal state) The variable MODBULA provides the mode of the German Federal Land in which the household-interviews were conducted. It is given for each interviewer and for each year from 1984 until 2018.

MODGGK (primary interviewer allocation, size of community) The variable MODGGK provides the mode of the population size of the communities where the household-interviews were conducted. It is given for each interviewer and for each year from 1984 until 2018.

2.3.3 Workload and Modes

MEANCONTACTHH (mean number of HH contacts made (gross HH data)) The variable MEANCONTACTHH reports the average number of contacts before interviewers realize the interview (or the household finally refuses to participate). The variable provides the mean per interviewer for each year between 1985 and 2018. It is the mean of the variable $intk from the $hbrutto-file. Note that this mean may underestimate the true value as the maximal category of the number of contacts is "9 or more interviewer contacts".

RESPONSERATE (HH Response Rate according to AAPOR definition rr6) The variable RESPONSERATE gives the (longitudinal) response rate as described by the American Association for Public Opinion Research (AAPOR). This means that RESPONSERATE is defined as the share of the realized interviews in one wave divided by the number of eligible households in interviewers. Eligible households are all responding households from the previous wave, split-offs from these households, and temporal drop-outs from previous waves. We define the resolution of households by death and emigration as ineligibility, and these cases are therefore excluded from the calculation of RESPONSERATE. It is given for each interviewer and for each year from 1985 until 2018. For the generation of the variable RESPONSERATE, we use the variable $hergs from the $hbrutto-files.

AMOUNTINTH (number of household interviews) The variable AMOUNTINTH provides the amount of papi and capi household interviews for each interviewer for each year from 1984 until 2018. For the generation of the variable AMOUNTINTH the number of entries with the respective modes in the $h-files is summed up for every interviewer.

AMOUNTINTP (number of personal interviews) The variable AMOUNTINTP provides the amount of personal interviews conducted in papi or capi mode for each interviewer for each year from 1984 until 2018. For the generation of the variable AMOUNTINTP we use the generated SOEP-variable mode$ (interview method). We count entries with the respective modes in the $p- files.

AMOUNTINTJ (number of personal interviews) The variable AMOUNTINTJ provides the amount of personal interviews conducted in papi or capi mode for each interviewer for each year from 2000 until 2018. For the generation of the variable AMOUNTINTJ we use the generated SOEP-variable mode$ (interview method). We count entries with the respective modes in the $youth- files.

PAPI (Number of paper and pencil interviews of interviewer) The variable PAPI provides the amount of paper and pencil interviews for each interviewer for each year from 1984 until 2018. This includes also self-administered paper-based questionnaires if respondents fill in these questionnaires while the interviewer is present in the household[3]. For the generation of the variable PAPI we use the generated SOEP-variable mode$ (interview method). Informaton is summed up over all $p- $h- and $youth- files.

CAPI (number of interviewer-assisted computer interviews) The variable CAPI provides the amount of personal computer assisted interviews for each interviewer for each year from 1998 until 2018. For the generation of the variable CAPI we use the generated SOEP-variable mode$ (interview method). Informaton is summed up over all $p- $h- and $youth- files.

MAIL (number of mailed paper interviews) The variable MAIL provides the amount of self-administered mailed interviews for each interviewer for each year from 1985 until 2018. For the generation of the variable MAIL we use the generated SOEP-variable mode$ (interview method). Informaton is summed up over all $p- $h- and $youth- files. Telephone contact often precedes the mailed questionnaires. In these cases, we consider these interviews as workload of the respective interviewer in the centralized call center.

CAWI (number of online interviews) The variable CAWI provides the amount of self-administered online interviews for each year from 2014 until 2018. For the generation of the variable CAWI we use the generated SOEP-variable mode$ (interview method). Informaton is summed up over all $p- $h- and $youth- files.

[3]The information whether the repondent filled in the questionnaire while the interviewer was present is detailed in the variable mode$ for most of the survey years.

LENGTHINTH (average length of personal interviews) The variable LENGTHINTH provides the average length of the household paper and pencil (papi) interviews as well as computer assisted (capi) interviews for each interviewer for each year from 1984 until 2018. The generation of the variable LENGTHINTH is based upon the variables $hdaer1 (for the years 1984 until 1988) and $hhdaer1 (from 1989 onwards) from the $h-files. These variables state the length of the interview in minutes. Their mean is calculated for each interviewer and each year.

LENGTHINTP (average length of personal interviews) The variable LENGTHINTP provides the average length of the individual person paper and pencil (papi) interviews as well as computer assisted (capi) interviews for each interviewer for each year from 1984 until 2018. The generation of the variable LENGTHINTP is based upon the variables $pinda1 (for the years 1984 until 1998) and $pdauer (from 1999 onwards) from the $p-files. These variables state the length of the interview in minutes. Their mean is calculated for each interviewer and each year.

LENGTHINTJ (average length of personal interviews) The variable LENGTHINTJ provides the average length of the paper and pencil (papi) interviews as well as computer assisted (capi) interviews for each interviewer for each year from 2000 until 2018. The generation of the variable LENGTHINTJ is based upon the variables $jdauer1 and $jdauer2 from the $youth-files (from 2000 onwards). These variables state the length of the interview in minutes. Their mean is calculated for each interviewer and each year.

2.4 Variables from Multiple Sources

For some variables there is information from more than one source. This reduces missing data and gives an opportunity for verification. If there is conflicting information, we indicate this by a flag-variable. Then a decision for one, presumed more valid, value of the conflicting information is made.

GENDER The variable GENDER provides the gender of the interviewer. It is constant over time. The generation of the variable GENDER is based upon the variable intsex_IB from the Kantar Public personnel data and the variable isex from the survey reports. If there is conflicting information, we use the self-reported isex for GENDER. The variable GENDERCONFL indicates if there was conflicting information.

BIRTH (year of birth of interviewer) The variable BIRTH provides the year of birth of the interviewer. It is constant over time. The generation of the variable BIRTH is based upon the variable igebjahr_IB from the Kantar Public personnel data and the variable igeb from the survey reports. If there is conflicting information, we use the self-reported igeb for BIRTH. The variable BIRTHCONFL indicates if there was conflicting information.

MARITAL (marital status of interviewer) The variable MARITAL provides the marital status of the interviewer for each interviewer for each year from 2000 until 2018. The generation of the variable MARITAL is based upon the variable ifamst_IB from the Kantar Public personnel data and the variable ifamst from the survey reports. If there is conflicting or ambiguous information, ifamst is taken as MARITAL. The variable MARITALCONFL indicates if there was

conflicting information. Conflicts happen frequently as the option «divorced» is missing in ifamst_IB.

EDUC (education of interviewer) The variable EDUC provides the education of the interviewer for each interviewer for each year from 2000 until 2018. The generation of the variable EDUC is based upon the variable ischul_IB from the Kantar Public personnel data and the variable ischabs from the survey reports. If there is conflicting or ambiguous information, ischabs is taken as EDUC. The variable EDUCCONFL indicates if there was conflicting information.

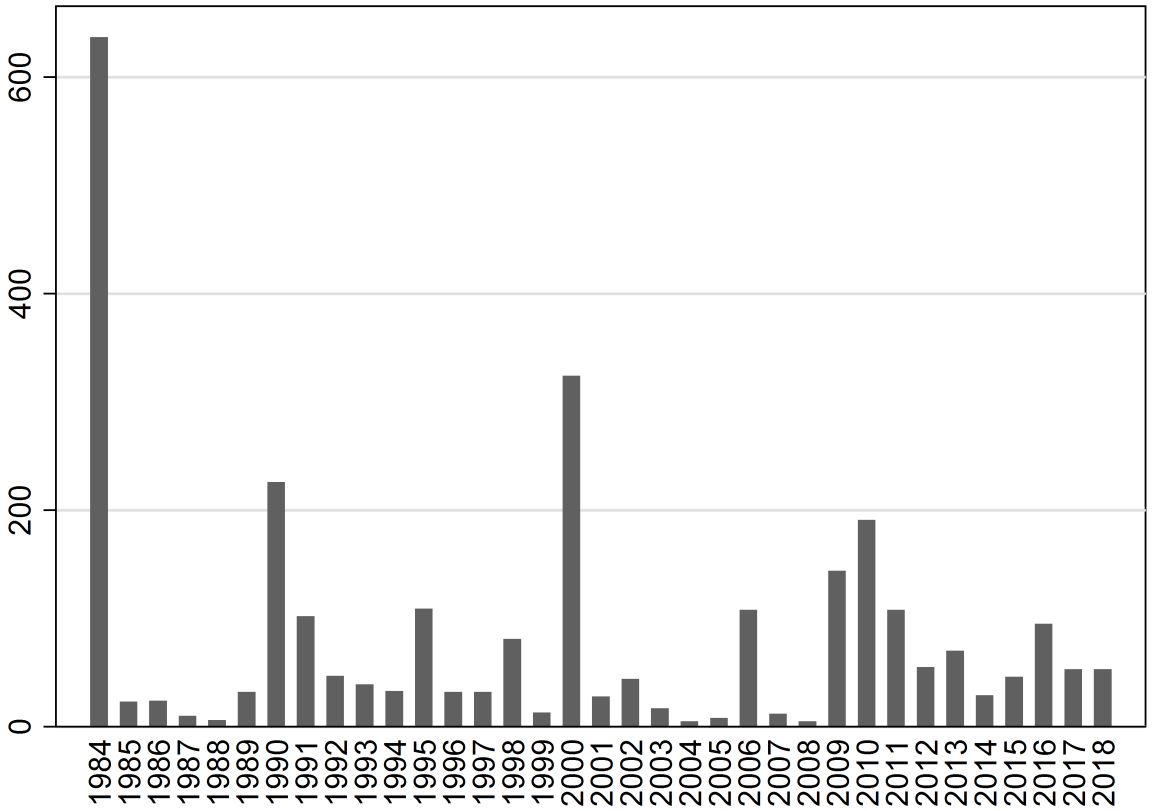
STARTINT (year of First SOEP-Interview) The variable STARTINT provides the year of the first SOEP-interview for each interviewer. It is constant over time. The generation of the variable STARTINT is based upon the variables lidstart from the Kantar Public personnel data, ianf01_ from the survey reports (2006 and 2012) and the first year with activity as an interviewer according to SOEP fieldwork data for each interviewer. If there is conflicting information, information from the SOEP fieldwork data is taken. The variable STARTINTCONFL indicates if there was conflicting information.

3 Descriptive Statistics

INTERVIEWER includes all interviewers who ever conducted at least one interview in the SOEP and a few more interviewers active in SOEP-related studies and pretests who participated in the interviewer surveys of 2006 and 2012. There is information on 2935 interviewers. INTERVIEWER contains information for every year since 1984 (Waves A(1984) - BI(2018)).

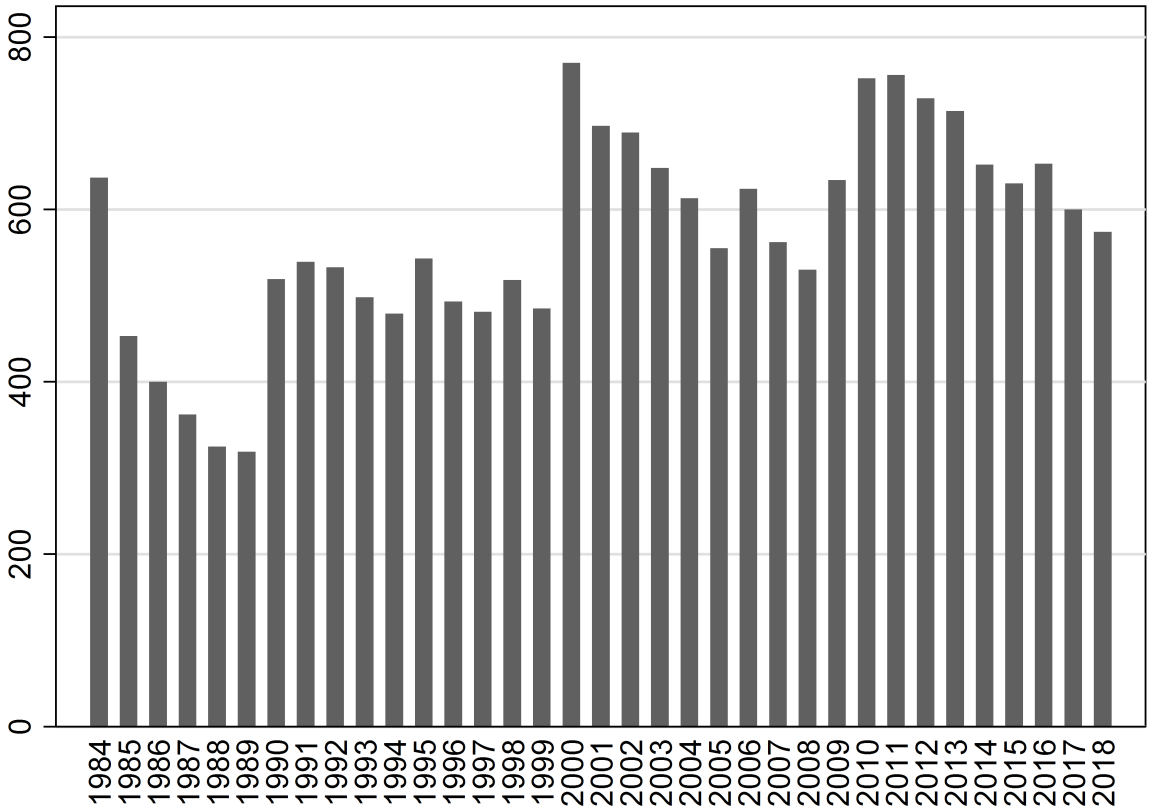
Figure 1 reports the entrants of new interviewers to the SOEP in each wave of the panel.[4] In the first year (1984), more than 600 interviewers began to work for the SOEP. With the enlargement of the survey in East Germany, more than 200 new interviewers entered the panel. Again, with the large refreshment Sample F in the year 2000, more than 350 new interviewers were recruited.

Figure 1: Number of New Interviewers by Year



[4]Only interviewers with non-missing date for first and last SOEP-Interview are included in Figures 1 and 3 and Tables 3.1 to 3.2. In Figure 4 and 5 only field-interviewers are included.

Figure 2: Number of Interviewers by Year



Typically, the number of active interviewers in one wave hovers around 600. As can be seen from Figure 2 the total number of interviewers reached its peak in the year 2000 when almost 800 interviewers were registered in the SOEP. When comparing Figures 1 and 2 one can clearly see the effect of interviewer drop out: In years with few new interviewers, the absolute number of interviewers tends to decrease. There is a recular loss of about 50 interviewers per year. Table 3.1, based on the active interviewers in 2018, reports that around 32 % of them had started to work for the SOEP more than fourteen years ago. 12 % had even joined the SOEP more than 24 years ago.

Table 3.1: Year of Entrants of Active Interviewers until 2018

Year of First Interview	Freq.	Percent	Cum.
1984-1988	35	6.1	6.1
1989-1993	36	6.3	12.4
1994-1998	30	5.2	17.6
1999-2003	81	14.1	31.7
2004-2008	38	6.6	38.3
2009-2013	191	33.3	71.6
2014-2018	163	28.4	100.0
Total	574	100.0	

Figure 3: Median Age and Mean Experience of Interviewers by Wave

Figure 3 shows that the strong commitment of interviewers to the SOEP survey leads to an increasing median age of the interviewers over time. In 1984, the median age is about 42 years and it reaches 67 years in 2017. The general trend of aging interviewers is interrupted in years with a substantial influx of new interviewers as in 2000. At the same time, we observe an increase in interviewer experience, measured in the cumulative number of years in which an interviewer conducted interviews[5]. Starting at one in 1984, the mean experience reaches, on average, 11,7 years of experience in 2018.

[5]This measure ist not always equal to the number of years since the start of working as an interviewer, as interviewers might pause in some years.

Table 3.2: Interviewer Attributes

Year	2000	2006	2012	2016
number of observations	770	624	729	653
male	452	367	404	369
female	307	252	293	277
secondary school degree	155	98	105	58
intermediate school degree	285	219	248	163
technical school degree		54	49	
upper secondary school deg.	99	199	163	63
graduate degree	179	17	63	138
other degree		12	8	
unwed	71	57	65	58
divorced		63	54	
widowed	111	55	56	66
married	574	429	433	305

Table 3.2 gives an overview of some interviewer attributes in 2000, 2006, 2012 and 2016. Please note that due to more detailed response categories, the information on marital status and education is more detailed in years of the interviewer survey, namely 2006 and 2012. On average, interviewers working in the SOEP have about average education and marital status compared to the German population, though the number of male interviewers exceeds the number of females.

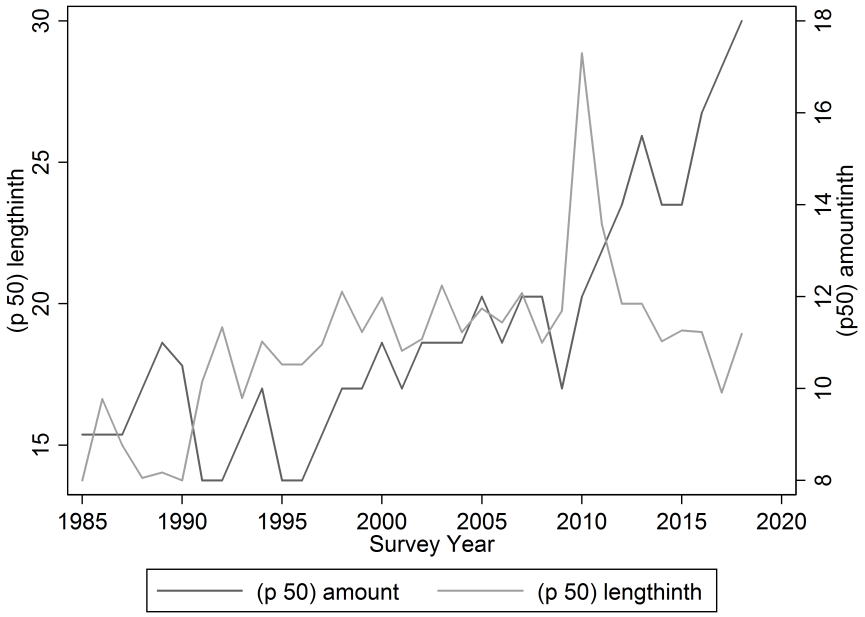


Figure 4: Workload household-level

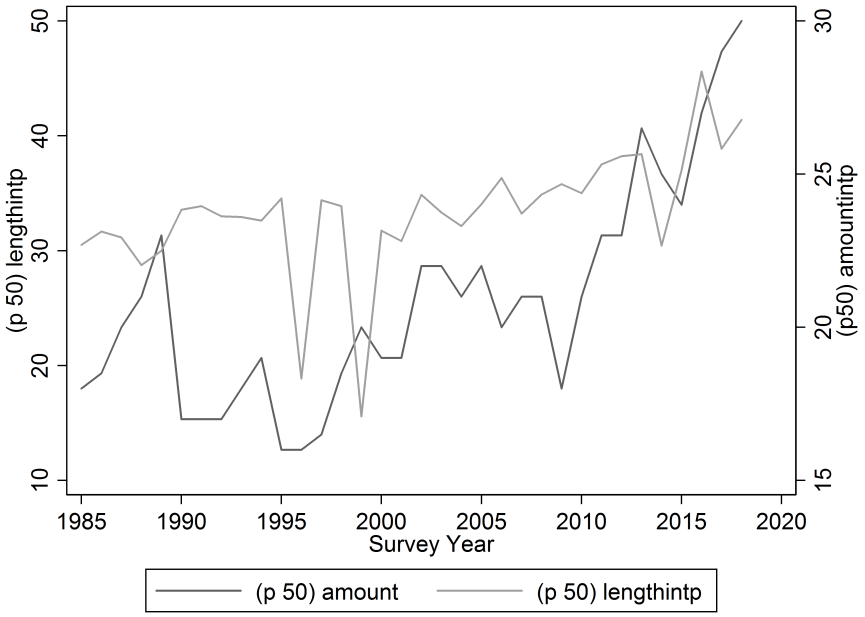


Figure 5: Workload personal-level

Figure 4 shows the median number of household-interviews per interviewer and the median of the mean interview length in minutes for each year between 1985 and 2018. The number of interviews increases clearly over the years and has almost doubled since 1984. The median length of the interviews conducted by interviewers is relatively stable and ranges typically between around 15 and 20 minutes. In 2010 however, an extreme increase in length of interviews can be observed.

Figure 5 shows the same variables only on the person-level. Just as in the previous graphic, you can observe a massive increase of numbers of interviews while the median of the mean interview length is, with a few exceptions, mostly between 30 and 40 minutes.

4 Matching Person and Interviewer Data

The interviewer identifier `INTID` that is recorded in both the `INTERVIEWER` file and all person and household files of the SOEP allows to directly link information of interviewers with information on interviewees.[6] The STATA code on the next page illustrates how to match the *"political party supported"* of respondents with the *"political party supported"* of their interviewers in 2016 (Wave BG).

In the first line the unique ID for individuals "persnr", the survey year "syear", the unique ID for interviewers "intid" and a variable for party preference "bgp14501" are loaded into Stata from the person dataset of the wave BG (2016). Next, the `intid` is used to merge the variable ipol3, the interviewer's party preference. In line four, the value labels for the newly merged variable are taken from the existing one for the purpose of easier readability. In the last step, a cross tabulation for the values 1-8 of the two variables is displayed (assuming both labels are identical as well).

The cross-tabulation of party identification in respondents and interviewers indicates a statistically significant association (Cramer's V = .0993825). Please note that this association may not necessarily reflect causal influence attributable to the social interaction between interviewer and respondent. It may also result from regional differences in party affiliation and the local interviewer allocation in the face-to-face fieldwork. In Table 4.1, interviewers might be counted several times.

[6] Please note that the `INTERVIEWER` file also contains a second identifier, `hhnr`. This second identifier does not refer to any household of SOEP respondents, but serves as an internal case identifier of the SOEP data bank at DIW Berlin.

```
use persnr syear intid bgp14501 using "$DATA/bgp.dta", clear

merge m:1 intid syear using "$OUT/interviewer.dta", keepus(ipol3) keep(match) nogen
label values ipol3 bgp14501
tab ipol3 bgp14501 if bgp14501<9 & bgp14501>0 &  ipol3<9 &  ipol3>0, V
```

Table 4.1: **Party Identification of Respondent and Interviewer in 2016**

Favored party (interviewer)	Parteipraeferenz															
	SPD		CDU/CSU		FDP		Greens		The Left		Right-Wing Parties		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
SPD	370	30	499	26	35	29	211	32	66	22	5	17	18	40	1204	28
CDU/CSU	496	41	823	43	31	26	237	35	100	33	12	40	10	22	1709	40
FDP	72	6	127	7	15	13	53	8	12	4	0	0	6	13	285	7
Greens	170	14	256	14	29	24	115	17	29	10	1	3	8	18	608	14
The Left	77	6	158	8	8	7	34	5	68	23	8	27	2	4	355	8
Right-Wing Parties	24	2	23	1	1	1	18	3	23	8	3	10	0	0	92	2
Other	7	1	7	0	0	0	0	0	1	0	1	3	1	2	17	0
Total	1216	100	1893	100	119	100	668	100	299	100	30	100	45	100	4270	100

Cramer's V = .0993825

5 Appendix

5.1 Useful Links

For further information please use SOEP Companion and Paneldata.org.

5.2 Questionnaires

The German questionnaires to the different interviewer surveys can be accessed under the following links:

2006
2012
2016