

SOEP_info

Rocco Paolillo

SOEP (German Socio-Economic Panel): representative household panel survey of the non-institutionalized resident population in Germany (Longhi and Nandi 2014), hosted by German Institute for Economic Research (DIW Berlin) (Goebel et al. 2019). Surveyed yearly around 15,000 households and 30,000 individuals (SOEP 2019a)

Possible strategy below.

1 History

- 1984: West Germany (Longhi and Nandi 2014)
- subsample A: non-ethnic household heads (German origin) subsample B: Turkish, Greek, Yugoslavian, Spanish or Italian household heads (Longhi and Nandi 2014)
- 1990:
 - East German Democratic Republic was added (Longhi and Nandi 2014)
 - subsample C: non-ethnic households (Longhi and Nandi 2014), n. 2,179 after the reunification (Goebel et al. 2019)
- 1994/1995: subsample D: ethnic Germans after Soviet collapse (n. 531) (Goebel et al. 2019)
- 1998: subsample E: refreshment sample (samples added randomly over time) (Longhi and Nandi 2014)
- 2002: subsample G: high-income earners (Longhi and Nandi 2014)
- 2000: innovation sample (blends design A and E) (Longhi and Nandi 2014)
- 2009: innovation sample (SOEP 2019e)
- 2011-2012: refreshment
- 2010: low-income family
- 2011: single-parent family
- 2012: SOEP-IS (Goebel et al. 2019) (deadline Nov. 30, some days to full Dec. 31)
- 2013: migration 1995-2010 sample M1
- 2015: migration 2009-2013 M2
- 2016: refugee of years 2013-2015, M3/M4 including Syria, M5 immigration in 2016

For summary of funding, history and sampling see Goebel et al. (2019).

“Data from scientists for scientists” (Goebel et al. 2019, 351). For studies using SOEP check Siedler et al. (2008)

2 Method Analysis

Useful links:

time-series R package: `mclg` choice regression

3 Procedure

- All people in the household are eligible of personal interview, prospectively and retrospectively on employment, partnership and fertility histories etc. (Longhi and Nandi 2014)



Figure 1: Clusters of topics. Retrieved from SOEP (2019g).

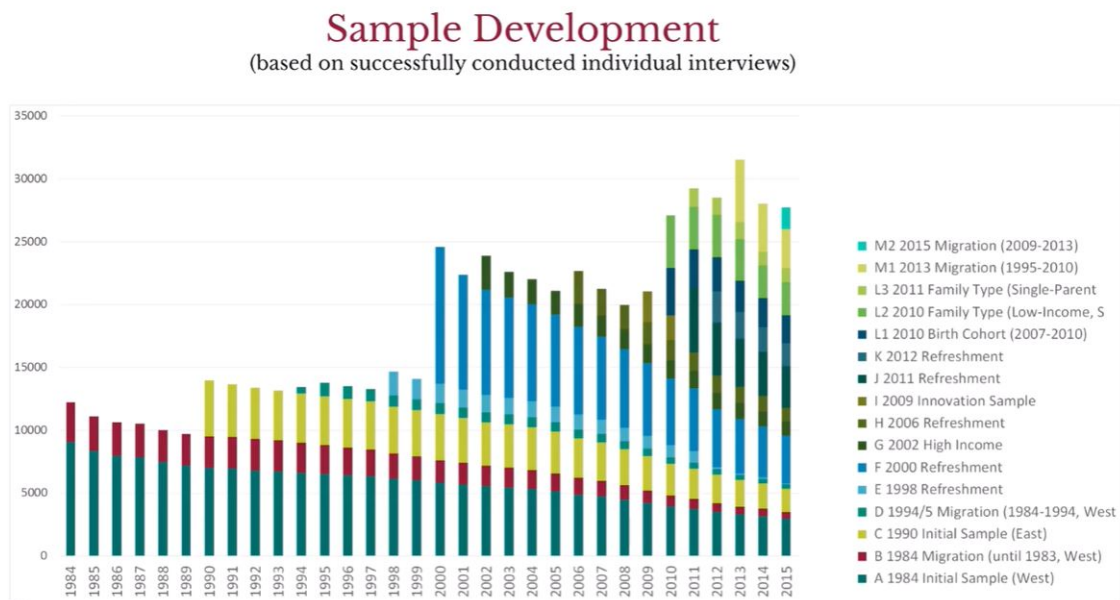


Figure 2: Composition of SOEP over time. Retrieved from SOEP (2019e).

- *active panel*: set of respondents in a particular year, considering refresher samples to keep the sample representative of German population (demographic inflows) (SOEP 2019e)
- Reason to leave:
 - death, moving abroad
 - attritors: definitive final refusal
- *active panel*:
 - information on household: asked the household head (Longhi and Nandi 2014; SOEP 2019e, 2019f)
 - * individual questionnaire - pl dataset
 - * household questionnaire - hl dataset
 - proxy information:
 - * children below 10 years
 - * person deceased for their last period
 - personal interview:
 - * age 11+ : interviewed themselves if parent consent
 - * adults : interviewed themselves
- People who participated in the first wave and descendants are asked to participate as long as they live in Germany. (Longhi and Nandi 2014)
- yearly, some topics even more than one year. Not all topic every year (SOEP 2019g)
- only random probability sample: at national level sample points sampled by federal state and municipality level, then stratified by locals and migrants. Use of weights see (Goebel et al. 2019) (see weights in Longhi and Nandi (2014))

Collaboration IAB-SOEP-MIG for migration-specific sample and in-depth knowledge of work assimilation, but 2013 and 2015 (Goebel et al. 2019)

4 Structure

Family of studies (SOEP 2019b):

- CORE
- Innovation sample SOEP-IS: from “Innovation” and “Innovation and Incentivisation”, accepting ideas from people outside DIW, with an exclusive usage of data for one year
- Related Studies SOEP-RS: other ventures that can be used for comparison (e.g. BASE II for elder people). FiD: Evaluation of family policies in Germany, part of SOEP from v31 (Goebel et al. 2019).

Levels:

- individual level: members of family +16 years, life-course event approach from “cradle grave”
- household level
- regional data with local information: privacy, contract and specific institution

Dataset:

- pl: individual questionnaire on civil participation, lifestyle and work. . .
- hl: household info, can contain info on education/activities of young people born before 2002
- biol: first time: present life up to present day
- bioparent: first time: info on their parents

- bioagel (proxy questions under 10 asked to parents):
 - A first after childbirth for mother, related to nurturing etc.
 - B after the child is 2 to 3 years for mother, ability of child etc.
 - C child 5 to 6 years for mother, ability of child etc.
 - D child 7 to 8 both parents parenting goals, child aspiration and achievement
 - E child 9 to 10 by mother, education and leisure time of child
- biopupil, children born on 2002 or later:
 - 11-12 first when aged 11: friends and social network
 - 13-14 early youth, idem + what allowed to do alone and political interests
- jugendl: until 17 y/o, introduced in 2001: youth questionnaire, personality, social networks, side jobs etc.
older than 17 y/o: as for all adults
- plueckel (adults + 17 yo): for the years they did not take part
- abroad: 2008-2010 and 2013-2014: Germans who moved abroad, while they were abroad. Only these years
- vpl: on the person that passed away

For general info/questions: **data_documentation** and **paneldata.org** (for variable) and **companion_soep**

Check for visiting grants <http://www.inclusivegrowth.eu/visiting-grants>

Data Structure and Naming Conventions (SOEP 2019c):

- name of datasets: combination of unit, type, format, questionnaires or module
 - Unit
 - * h*: household level
 - * p*: individual level
 - Type of data
 - * tracking: survey units, good to identify the research population. *brutto: who did not provide interview
 - * original: unprocess info from interview
 - * generated: *gen, improved quality, pequiv: armonized for international comparison.
 - * biography: generated dataset from biological perspective (life course): bio*. All at individual level
 - * survey: household
- Dataset Format (SOEP 2019c):
 - Wide Format: ppath, phrf, hpath, hhfr. One row per survey unit
 - Long Format: biogrpahical dataset, *l, person-year combination (see Field, Miles, and Field (2012))
 - Spell format: spells for event-history analysis, marriage and couple

SOEPlong format provides the info pooled and harmonized (e.g. income prior to 2001 in Euros)

- Data Name:
 - Long format and not generated: DVTxxxx see Fig.3
 - * D: dataset identifier: p:person, h:household, l:biography, j:youth
 - * V: variable identifier: l: numeric, a:alphanumeric
 - * T: topic identifier

Conventions for naming variables

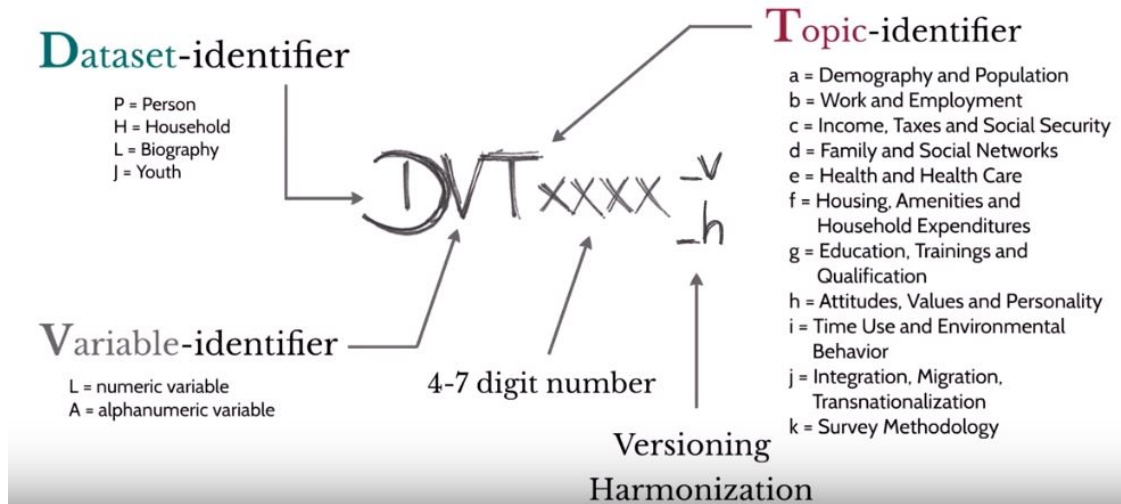


Figure 3: Structure of variable name. Retrieved from SOEP (2019c).

- * xxx: digit
- * `_v/_h`: versioning or harmed
- Generated variables: intuitive names. See Fig.4 for missing values
- Specific:
 - lela: retrospective biography information
- SOEP-Core:
 - files on development of sample
 - generate data
 - originally surveyed data
 - biography prior to the study (biol?)

Data Identifiers and Pointers (SOEP 2019d)

- Identifiers: identify unit surveys (households and individuals) so to let you connect them for nested analysis (individual ↔ household):
 - HID: household, CID: first house where they lived
 - PID: person
 - All dataset individual level have both HID + PID + syear for panel data
 - INTID: interviewer
 - VPID: deceased person (whose info the participant give)
- Pointers: combine information on different persons, link the PID to other people in the dataset:
 - PARID ppathl: partner in the year
 - KIDNRxx biobirth: identify the kids, points the mother to the children, assuming biological mother and that partner of mother can change (and father is one partner of the mother)
 - underage children: K_PMUM kidl child → mother, K_PMUMP kidl child → partner of the mother
 - adults (+ 17 yo): participant ID → parent: bioparen MNR:mother, FNR:partner of mother

Conventions for missing values

- 1 no answer / don't know
- 2 does not apply
- 3 implausible value
- 4 inadmissible multiple response
- 5 not included in this version of the questionnaire
- 6 version of questionnaire with modified filtering
- 8 question not part of the survey in the respective year

Figure 4: Coding for missing values. Retrieved from SOEP (2019c).

- to siblings: SIBPNRxx biosib
- to twins: PNR-twin/trip/quad

You will always merge datasets on the shape:

$MainDataset(ID - ydep.) + AuxiliaryDataset(Pointer - xind.) = MergeddatasetbyID - Pointer$

5 Strategy

Possible strategies will be to identify attitude variables PID of house holder to link with household level HID. To check if at HID there are info on neighborhood of HID (reported by house holder), or check regional data we have. CID in next steps to trace change in residential choice through generations. Comparing PID level between generations (Identifier for parents) to study change in preferences/values and CID level - HID level (original household - current household) for change of neighborhood through generations

5.1 Variables

companion_soep, annually: all years, otherwise it is reported

Neighborhood: Home Amenities > neighborhood amenities,

Individual: Attitudes, Values, Personality > political orientation

References

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. Sage publications.

Goebel, Jan, Markus M Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp. 2019. "The German Socio-Economic Panel (Soep)." *Jahrbücher Für Nationalökonomie Und Statistik* 239 (2): 345–60.

Longhi, Simonetta, and Alita Nandi. 2014. *A Practical Guide to Using Panel Data*. Sage.

Siedler, Thomas, Jürgen Schupp, C Katharina Spiess, and Gert G Wagner. 2008. “The German Socio-Economic Panel as Reference Data Set.”

SOEP. 2019a. “SOEPcompanion.” <http://companion.soep.de/>.

———. 2019b. “SOEPtutorials: Basic Intro.” <https://www.youtube.com/watch?v=XBMtaA9vOf4&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy>.

———. 2019c. “SOEPtutorials: Data Structure and Naming Conventions.” <https://www.youtube.com/watch?v=3aNzO217QNY&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy&index=5>.

———. 2019d. “SOEPtutorials: Identifiers and Pointers.” <https://www.youtube.com/watch?v=w3XNSk3GGW0&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy&index=6>.

———. 2019e. “SOEPtutorials: Panel Composition and Sample Development.” <https://www.youtube.com/watch?v=6V6NtoJYKTA&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy&index=3>.

———. 2019f. “SOEPtutorials: Questionnaires.” <https://www.youtube.com/watch?v=9jvk6UkmHYo&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy&index=4>.

———. 2019g. “SOEPtutorials: Topics of the SOEP Study.” <https://www.youtube.com/watch?v=vm-YZ0R17hQ&list=PL6Z-U6OzQTNPfGep9oS3eMUyLfQ1OlMBy&index=2>.