

Data Retrieval from the German Socio-Economic Panel (SOEP, 1984-2010, version 27)

Josef Brüderl
Volker Ludwig

University of Munich
May 2012

Contents

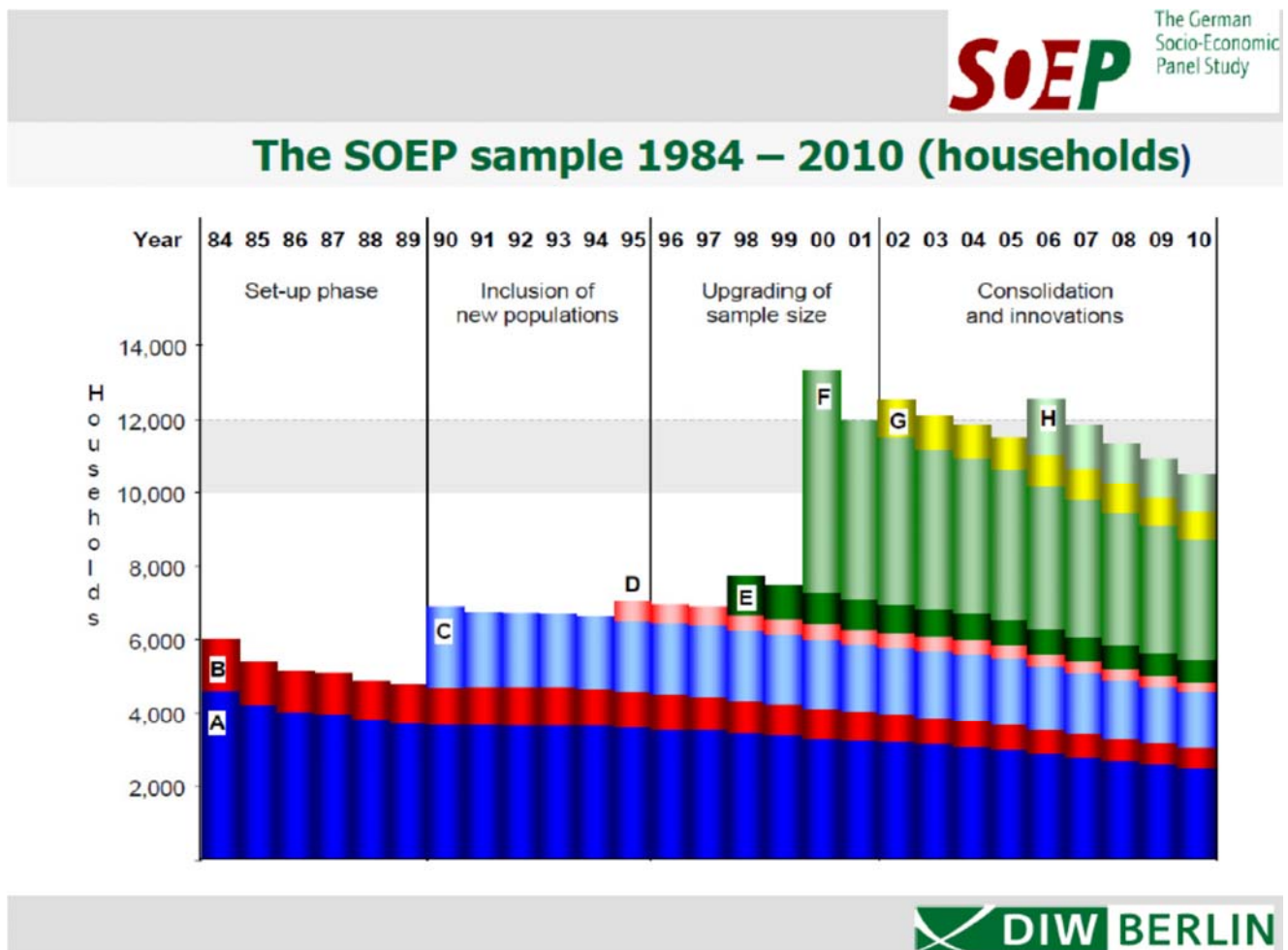
- **SOEP Design**
 - Samples and Questionnaires
- **SOEP Data Structure**
 - Overview of SOEP data sets
- **Data Retrievals**
 - Retrieving a person-year data set
 - \$P, PPFAD, \$PGEN, \$PEQUIV
 - Working with SOEPinfo
 - Hand-made blueprint for a retrieval file
 - Retrieving an episode (spell) data set
 - ARTKALEN
 - Hand-made blueprint for a retrieval file
 - Further Retrievals (only rudimentary)
- **Citing the SOEP**

SOEP Subsamples

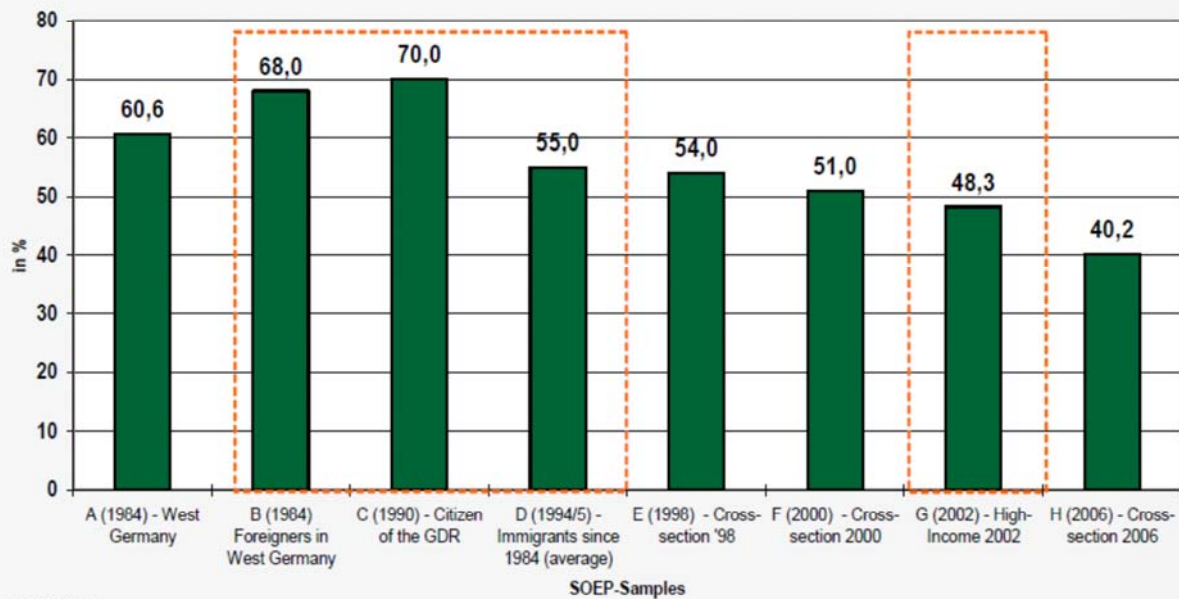
- All SOEP subsamples are multi-step random samples of private hh

Name/ Value	Label	Start Year	House- holds	Persons	Description
A/1	German West	1984	4,528	9,076	Head is either German or other nationality than those in Sample B
B/2	"Foreigner" West	1984	1,393	3,169	Head is either Turkish, Italian, Spanish, Greek or from the former Yugoslavia
C/3	Germans East	1990	2,179	4,453	Head was a citizen of the GDR (expansion of survey territory)
D/4	84-93 Immigrant (West)	1994/1995	522	1,078	At least one household member has moved to Germany after 1989 (expansion of survey population)
E/5	Refreshment 1998	1998	1,056	1,910	Random sample covering all existing subsamples (total population)
F/6	ISOEP 2000	2000	6,043	10,880	InnovationRandom sample covering all existing subsamples (total population)
G/7	High Income	2002	1,224	2,671	Monthly net household income is more than 4.500 Euro (7.500 DM)
H/8	Refreshment 2006	2006	1,506	2,616	Random sample covering all existing subsamples (total population)
I/9	"Incentive"	2009	1,531	2,509	Random sample covering all existing subsamples (total population)

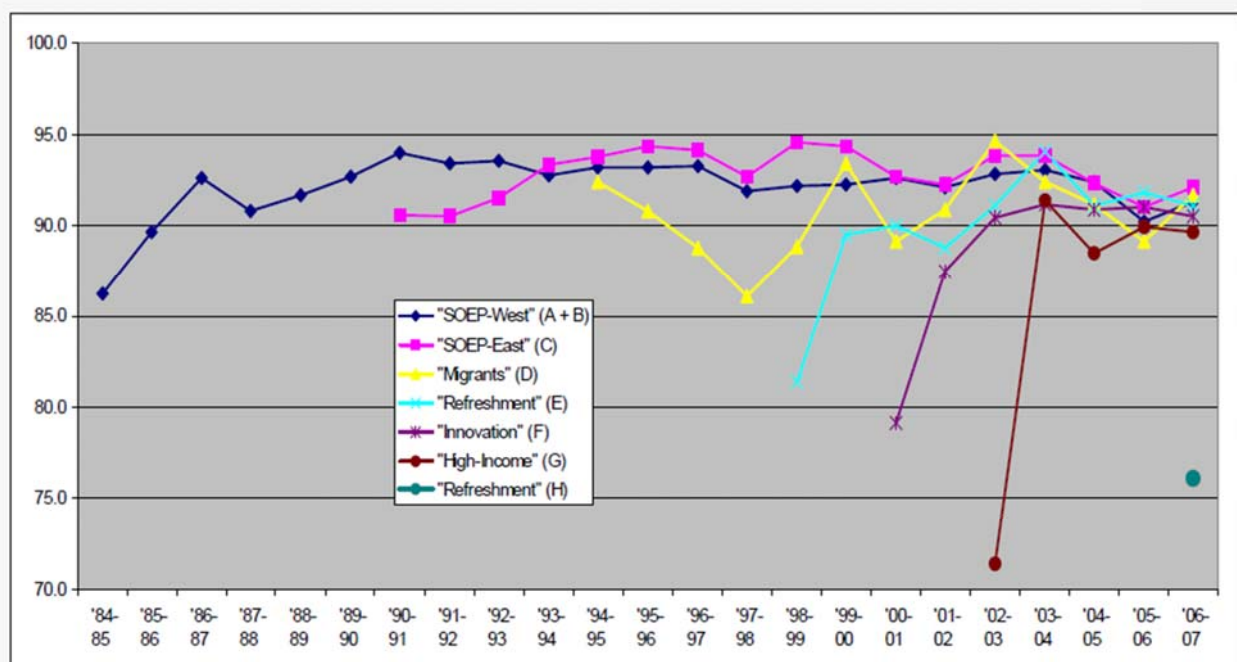
- Sample selection
 - Some authors prefer to delete sample G (high income), because it is not "representative". This argument is not sound, especially if the models used control for income.
 - Some authors prefer to delete persons in institutional hh. All SOEP respondents are originally in private hh. But the SOEP follows them, when moving to an institutional hh. Again I see no sound argument for deleting those respondents.



Longitudinal perspective: Wave 1 response rates



Longitudinal perspective: Wave-to-wave response rates



SOEP Questionnaires

- All persons 18+ in the hh get the “Personenfragebogen”
 - First time respondents in addition “Zusatzfragebogen: Lebenslauf”
 - First time respondents (17 years old) get only the “Jugendfragebogen”
 - Last wave nonrespondents in addition fill out a short quest. “Lücke”
 - In former times there were special questionnaires for foreigners (1984-1994) and East Germans (1990 and 1991)
- The “Haushaltsvorstand” fills out the “Hhfragebogen”
- Since 2003 there are proxy interviews on kids
 - Mothers of newborns (born last year) fill out quest. “Mutter-Kind”
 - Mother/father of kids born 3 years before fill out “Kind 2 - 3”
 - Mother/father of kids born 6 years before fill out “Kind 5 - 6”
 - Mother/father of kids born 8 years before fill out “Kind 7 - 8”
- For dead respondents there is a proxy interview “Exit”
 - More in “SOEP Contents” and “SOEP Fragebögen”

SOEP Data Structure

Longitudinal data in the SOEP comes in two forms

(1) Panel Data

Gathered by prospectively asking respondents the same question in each survey year (person and household questionnaires)

→ Panel data is stored in cross-sectional data files

(2) Event History Data

Asking respondents (retrospectively)

- about occurrence and timing of life events, e.g. birth of children, immigration (biography questionnaires)

→ BIO data sets (single files)

- about information on labor force status, sources of income, marital status (yearly and monthly calendars). These calendars produce episode data (called “spell data” by the SOEP group)

→ SPELL data sets (single files)

Data Sets on Persons

Note: \$ stands for waves a – z, ba (1984-2010)

- **\$P** wave-specific person data
- **\$PGEN** wave-specific generated person data
- **\$PEQUIV** wave-specific CNEF variables (income and more)
- **PPFAD** person meta data
- **PHRF** person weights

- **\$PBRUTTO** wave-specific gross contact protocol
- **\$PKAL** wave-specific monthly calendar
- **\$PAUSL** questions specific for sample B (1984-1994)
- **\$POST** questions specific for East Germans (1990, 1991)
- **\$PLUECKE** information on gaps
- **\$KIND** wave-specific data on kids (<17 from hhquestionnaire)
- **\$PAGE17** questions for youths (since 2006)

More or less detailed descriptions can be found on the SOEP-Homepage
(>FDZ SOEP >Dokumentation >Dokumente >generierte Variablen)

http://www.diw.de/de/diw_02.c.238110.de/generierte_variablen.html

Household Data

- **\$H** wave-specific hh information
- **\$HGEN** wave-specific generated hh information
- **HPFAD** hh meta data
- **HHRF** hh weights
- **\$HBRUTTO** wave-specific gross contact protocol
- **GHOST** wave-specific hh information east (1990)
- **EV** wealth questionnaire (1988)
- **MIHINC** multiple imputed hh income
- **GGKBOU** regional information (only with special user contract)

Cumulative Data Sets

- **(H)PWEALTH** (hh) person wealth (2002, 2007)
- **HEALTH** person health (2002/04/06/08)
- **PFLEGE** persons in need of care (1985-2009)
- **GRIPSTR** grip strength (2006, 2008)
- **COGDJ** results of cognitive tests youth
- **COGNIT06** results of cognitive tests adults (2006)
- **INTVIEW** interviewer data

Biography Data Sets I

- “Real” BIO data sets
 - Here the SOEP group consolidates biographical information from:
 - First time respondents: retrospective biography questionnaire (Lebenslauf)
 - First time respondents (17 years old): youth questionnaire
 - Person questionnaire: what happened last year?
 - BIO format I: timing of important events
 - BIOBIRTH (M): information on kids of female (male) respondents
 - BIOJOB: timing and characteristics of first and last job
 - BIORESID: time moved into home (only from bio quest., not updated)
 - LIFESPELL: (register) information on SOEP dropouts (emigration, death)
 - BIO format II: episode (spell) format
 - ARTKALEN: activity episodes (exact to the month) since year before first time interview until year before last available interview
 - PBIOSPE: activity episodes (year) since age 15 until year before last available interview (ARTKALEN aggregated after entry in SOEP)
 - BIOMARSM: marital (single, married, divorced, widowed) episodes (exact to the month) since month of first time interview
 - BIOMARSY: marital (single, married, divorced, widowed) episodes (exact to the year) since birth
 - EINKALEN: income episodes (1983-1994)
 - SOZKALEN: social assistance episodes (1991-1999)

Biography Data Sets II

- Information from “biography” questionnaires
 - Special Populations
 - BIOPAREN: info on respondents parents either from bio questionnaire (proxy information) or from parents’ personal questionnaires (if in SOEP)
 - BIOIMMIG: person-year dataset for immigrants, migration specific info
 - BIOTWIN: pointers for identified twins
 - Information on childhood
 - BIOSOC: info from bio questionnaire on school and relations to parents
 - Information on special age groups
 - BIOAGE17: time-constant bio info from youth questionnaire (since 2000)
 - \$PAGE17: further info from youth questionnaire (since 2006)
 - BIOAGE01: info from mother child questionnaire (newborns) (since 2003)
 - BIOAGE03 (06): info from child questionnaire (aged 3 or 6) (since 2005)
 - BIOAGE08[A|B]: info from both parent questionnaires (aged 8) (since 2010)

- Frick/Goebel (2011) Biography and Life History Data in the German SOEP. Data Doc. 61.

Systematik Variablennamen

Spalte	Bedeutung
1	Erhebungsjahr (A=1984 ...)
2	Analyseeinheit (P=Person; H=Haushalt)
3-4	Fragennr. im Erhebungsinstrument
5-6	Itemnr. der Frage im Erhebungsinstr.
5 oder 7	Differenzierung spez. Samples (A-Stichprobe B; O=Stichprobe C)
2-8	Erklärender Kurztext (in \$BRUTTO und generierten Files)

- Im SOEP werden drei Arten von fehlenden Werten unterschieden, die wie folgt kodiert sind:
 - 1 keine Angabe, weiß nicht, Item-Non-Response
 - 2 trifft nicht zu
 - 3 Wert nach Plausibilitätsprüfung als unplausibel eingestuft und gelöscht (zu interpretieren wie -1)
- Es gibt im SOEP keine „system missings“ (wie . in Stata).

Retrieving a Person-Year Data Set

- Extracting a long-format person-year panel data set is the main task before doing panel analysis
- In the following we will show how
 - Working with SOEP files
 - “SOEP Retrieval 1 Intro.do”
 - Retrieving a person-year data set via SOEPinfo
 - “SOEP Retrieval 2 SOEPinfo.do”
 - Retrieving a person-year data set via a hand-made program
 - “SOEP Retrieval 3 PersonYears.do”

Working With Person File \$P

- Information from the person questionnaire
 - Stored as cross-sectional files
 - Be careful: not all variables are longitudinally consistent
 - E.g., “welle” changes format, birth year nonsense in 1985, etc.
 - **Therefore, if possible work with generated files!**

	ahhnr	hhnrakt	persnr	welle	ap57	ap3301	ap62z	ap6801
1.	94	94	901	84	2	2000	1951	8
2.	108	108	1001	84	2	-2	1913	9
3.	116	116	1101	84	2	-2	1906	6
4.	124	124	1201	84	1	-2	1911	8
5.	124	124	1202	84	2	-2	1913	8

	bhhnr	hhnrakt	persnr	welle	bp85	bp4301	bp81	bp9301
1.	94	94	901	85	2	2600	951	8
2.	108	108	1001	85	2	-2	913	6
3.	116	116	1101	85	2	-2	906	7
4.	124	124	1201	85	1	-2	911	8
5.	124	124	1202	85	2	-2	913	8

Josef Brüderl, SOEP Data Retrieval, SS 2012

Folie 15

Working With Person File PPFAD

- PPFAD „keeps book“ on all persons who ever have been living in a SOEP household (N=66,813, includes kids!). Variables:
 - Household and person identifiers
 - Technical variables defining the sample population
 - Basic time-constant demographic characteristics
- Important variables:
 - sex, gebjahr, gebmonth
 - \$sampreg (West-East currently), loc1989 (West-East 1989)
 - immiyear, germborn, corigin, migback (migration stuff)
 - erstbefr, letztbefr (year of first and last interview), todjahr (year of death)
 - psample (SOEP subsample)

Working With Person File PPFAD

“Technical” variables used by SOEPinfo to restrict the sample: \$netto and \$pop

. tab znetto [v26]			
Befragungsstatus 2009	Freq.	Percent	Cum.
-----	-----	-----	-----
[-2] trifft nicht zu	17,003	49.83	49.83
[10] Befragungsperson mit real. Intervi	11,476	33.63	83.46
[12] Personenfragebogen und Lebenslauf	97	0.28	83.74
[16] Erstbefr., Personenfragebogen, 17 J	37	0.11	83.85
[17] Erstbefragung, Jugendfragebogen, 1	123	0.36	84.21
[19] Personenfrabo ohne Haushaltsinterv	9	0.03	84.24
[20] Kinder in realisierten Haushalten	2,250	6.59	90.83
[21] Kinder mit Mutter-Kind-Frabo_A, 0-	102	0.30	91.13
[22] Kinder mit Mutter-Kind-Frabo_B 2-3	83	0.24	91.37
[23] Kinder mit Mutter-Kind-Frabo_C 5-6	115	0.34	91.71
[30] Personen in Brutto-HH'en ohne P.Int	2,649	7.76	99.47
[32] Realisierter Biografiefragebogen	1	0.00	99.48
[90] Personenausfaelle (PBR_EXIT)	71	0.21	99.68
[91] Wegzug ins Ausland	28	0.08	99.77
[99] Verstorben	80	0.23	100.00
-----	-----	-----	-----
Total	34,124	100.00	

. tab zpop [v26]			
Populationszugehoerigkeit 2009	Freq.	Percent	Cum.
-----	-----	-----	-----
[-2] trifft nicht zu	17,003	49.83	49.83
[1] Privathaush., deutscher HV	14,456	42.36	92.19
[2] Privathaush., auslaend. HV	902	2.64	94.83
[3] Anstaltshh., deutscher HV	73	0.21	95.05
[4] Anstaltshh., auslaend. HV	5	0.01	95.06
[5] n.real. Privathaush., dt. HV	1,494	4.38	99.44
[6] n.real. Privathaush., ausl. HV	172	0.50	99.94
[7] n.real. Anstaltshh., dt. HV	17	0.05	99.99
[8] n.real. Anstaltshh., ausl. HV	2	0.01	100.00
-----	-----	-----	-----
Total	34,124	100.00	

Josef Brüderl, SOEP Data Retrieval, SS 2012

Folie 17

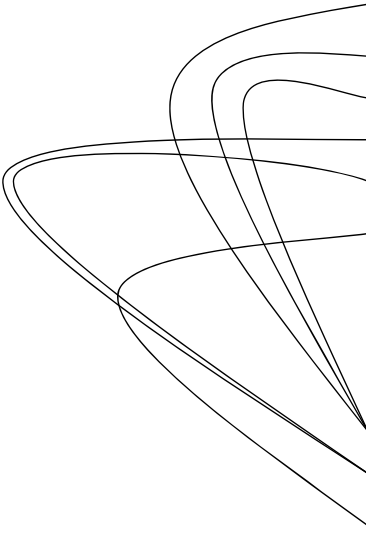
Working With Person File \$PGEN

- Data from person questionnaires
 - Generated variables
- Important variables:
 - Labor force status
 - Characteristics of the firm employed, industry classification
 - Occupational/job classification, ISCO88
 - ISEI, EGP, Prestige Scores
 - Working hours, income
 - Partner (ID), marital status
 - Education, CASMIN
 - Interview month, interview mode

Working With Person File \$PEQUIV

- SOEP data prepared for international comparison (CNEF)
 - Extensive income variables
 - Also household level variables included
- Important variables:
 - Income information of all sorts
 - Consumer price index
 - Household information
 - Residence (Bundesland, East-West)
 - Equivalence scale input
 - Health information
 - Happiness
 - Weights

Merging a Household Data File



vhhnr	persnr	welle	sex	alter
1	1	2005	f	32
1	2	2005	m	34
1	3	2005	f	9
2	4	2005	m	62
2	5	2005	f	60
3	6	2005	f	81

vhhnr	welle	hheink
1	2005	3000
2	2005	2000
3	2005	1000

```
use persdata, clear
merge m:1 vhhnr using hhdata
tab _merge
drop _merge
```

Working With Household Data

- Household questionnaire (\$H)
 - Note that until the 90ies housing questions were posed only for first time households. Since then, they are posed each year.
 - Therefore, it is difficult to work with the original data in \$H
- Generated variables (\$HGEN)
 - Continuous information on housing conditions
 - The SOEP group terms this “status variables”:
The status is carried forward year for year, if there was no change
 - Household typology
 - Household income (use corrected version: ahinc\$\$)
 - Also multiple imputed since 1995! See also MIHINC.DTA
 - Alternative information on hhinc can be found in PEQUIV. It is unclear, which var better to use (see “SOEP Retrieval 3 PersonYears.do”)
- Household meta data (HPFAD)
 - Usually not needed

Working With Weights

- Due to different sampling probabilities and non-response the SOEP is not an equal probability sample
 - Therefore, it is suggested that you use weights (at least for your bivariate analyses)
 - You can pull person weights from PHRF or \$PEQUIV (the latter is more convenient when using a long-format data set)
 - Cross-sectional weights: \$phrf or w11105\$\$
 - These are post-stratification weights
 - To obtain a “representative sample” in a cross-sectional sense
 - Longitudinal weights: \$pbleib or w11103\$\$ (1 / P(staying))
 - Longitudinal weights: aphrf * bpbleib * cpbleib * ...
 - Stata can not use longitudinal weights
 - Therefore, it is not clear what the worth of this exercise should be
 - Therefore, most users do not use (longitudinal) weights in the SOEP

Retrieving a person-year dataset: SOEPinfo

- SOEPinfo: <http://panel.gsoep.de/soepinfo2010/>
 - Should always be the first address when retrieving data
 - Find variable names and data sets where variables are stored
 - Varlists can be copied to do-file (paste and copy)
 - Direct link from variables to questionnaires
 - Produces do-file for retrieving the data (the “official” way to retrieve)
 - Disadvantage: wide format needs to be transformed to long format
- Example: person-year data set with “happiness”
 - Search for “Zufriedenheit”
 - Put all variables of “Allg. Lebenszufriedenheit” in the basket
 - Basket action 1: Frequencies
 - Check for longitudinal consistency!!!
 - Basket action 2: Questionnaires
 - Basket action 3: Generators – Stata
 - “SOEP Retrieval 2 SOEPinfo.do”
 - Note that SOEPinfo automatically pulls \$hhnr, hhnr, persnr, sex, gebjahr, weights

Josef Brüderl, SOEP Data Retrieval, SS 2012

Folie 23

Retrieving a person-year dataset: SOEPinfo

- Reshaping to long format (from “SOEP Retrieval 2 SOEPinfo.do”)
- This is for v26!!

```
rename persnr id //id is convention
replace gp109=gp6401e if gp6401e<. //copy East-values (1990)

local year=1984 //renaming to "happy*"
foreach var in ap6801 bp9301 cp9601 dp9801 ep89 fp108 gp109 hp10901 ip10901 ///
               jp10901 kp10401 lp10401 mp11001 np11701 op12301 pp13501 ///
               qp14301 rp13501 sp13501 tp14201 up14501 vp154 wp142 xp149 ///
               yp15501 zp15701 {
    rename `var' happy`year'
    local year=`year'+1
}
local year=1984 //renaming to "hhnrakt*"
foreach wave in a b c d e f g h i j k l m n o p q r s t u v w x y z{
    rename `wave'hhnr hhnrakt`year'
    local year=`year'+1
}
drop apop anetto aphrf prgroup gp6401e
foreach wave in b c d e f g h i j k l m n o p q r s t u v w x y z{
    drop `wave'pop `wave'netto `wave'pbleib `wave'phrf
}
//drop unnecessary variables to save computation time

reshape long happy hhnrakt, i(id) j(year) //reshape to long format
drop if happy==. //drop empty person-years
```

Josef Brüderl, SOEP Data Retrieval, SS 2012

Folie 24

Retrieving a person-year dataset: hand-made blueprint

- A standard retrieval of a long-format person-year panel data set should include the following variables
 - id unchangeable person identifier (from \$PGEN)
 - year year (self generated) (interv. year: “iyear” in SOEPlong)
 - hhnr household number of origin household (from \$PGEN)
 - hhnrakt number of current household (from \$PGEN)
 - month interview month (from \$PGEN)
 - psample subsample indicator (from PPFAD)
 - sex gender (from PPFAD)
 - gebjahr birth year (from PPFAD)
- A hand-made blueprint (by J. Brüderl and V. Ludwig)
 - “SOEP Retrieval 3 PersonYears.do”
 - A master file containing all person-years from \$PGEN is created
 - Additional variables are pulled from PPFAD, \$PEQUIV, \$P, \$HGEN, ... as needed

The Logic of the Retrieval

apgen.dta			
	id	year	var1
1.	1	1984	2
2.	2	1984	3
3.	3	1984	1

bpngen.dta			
	id	year	var2
1.	1	1985	6
2.	2	1985	7
3.	3	1985	5

1. Prepare \$PGEN by renaming the important vars
2. Append the resulting working files
3. Save as master.dta

```
use apgen.dta, clear
ren var1 X
save awork.dta, replace
use bpngen.dta, clear
ren var2 X
append using awork.dta
save master.dta
```

4. Make a second long-file from \$PEQUIV with var Z
5. Merge the resulting working file and the master

```
use apequiv.dta, clear
ren var7 Z
save awork.dta, replace
use bpequiv.dta
ren var13 Z
append using awork.dta
merge 1:1 id year using master.dta
```

id	year	X
1	1984	2
1	1985	6
2	1984	3
2	1985	7
3	1984	1
3	1985	5

id	year	X	Z
1	1984	2	3
1	1985	6	4
2	1984	3	5
2	1985	7	5
3	1984	1	6
3	1985	5	3

Retrieving an episode (spell) data set

- Episodes generated from the yearly activity calendar
 - Raw data can be found in \$PKAL and \$PLUECKE
 - Compiled to episodes in ARTKALEN
 - Unfortunately there seems to be no good documentation (2 pages in DTC)

74. And now please think about the entire previous year, in other words about 1993: We have made a sort of calendar. On the left, we have written things that could have happened last year. Please go through the entire list and check each month, in which, for example, you were employed or unemployed, etc. Please make sure you answer for each month.	
1993	J F M A M J J A S O N D
full-time employment, job creation measure	- - - - -
Short-time work	- - - - -
part-time or occasionally employed	- - - - -
vocational training, education, retraining	- - - - -
registered unemployed	- - - - -
retired, early retirement	- - - - -
maternity leave	- - - - -
in school/college	- - - - -
military/civilian service	- - - - -
housewife/househusband	- - - - -
other, namely	- - - - -

Working with ARTKALEN (V26)

Episodes of person 901 from Artkalen:

(variable "spellnr" deleted)

hhnr	persnr	spelltyp	begin	end	zensor
94	901	[1] Voll erwerbstätig	[1] 1983 Jan	[156] 1995 Dez	[4] Links zensiert
94	901	[3] Teilzeit-geringfügig	[157] 1996 Jan	[180] 1997 Dez	[1] unzensiert
94	901	[1] Voll erwerbstätig	[181] 1998 Jan	[296] 2007 Aug	[1] unzensiert
94	901	[12] sonstiges	[297] 2007 Sep	[299] 2007 Nov	[1] unzensiert
94	901	[1] Voll erwerbstätig	[300] 2007 Dez	[312] 2008 Dez	[2] Rechts zensiert

Art des Ereignisses	Freq.
[1] Voll erwerbstätig	51,057
[2] Kurzarbeit	1,741
[3] Teilzeit-geringfügig	23,544
[4] Betriebl. Ausbildung	14,438
[5] Arbeitslos gemeldet	23,272
[6] in Rente-Ruhestand	14,967
[7] Mutterschaft-Freistellung	5,383
[8] Schule-Hochschule	16,347
[9] Wehr-Zivildienst	2,086
[10] Hausfrau-mann	23,281
[11] Nebenberufl. Tätigkeit	514
[12] sonstiges	8,474
[13] in betriebl. Erstausbildung	3,519
[14] in Fortbildung, Umschulung	4,584
[15] Minijob (bis 400 Euro)	5,404
[99] Luecke	4,639
Total	203,250

Zensurierungsvariable	Freq.
[-2] trifft nicht zu	4,639
[1] unzensiert	102,735
[2] Rechts zensiert	31,461
[3] Rechts(KA) zensiert	2,230
[4] Links zensiert	28,823
[5] Links und Rechts zensiert	26,212
[6] Links und Rechts(KA)zensiert	2,141
[7] Links(KA) zensiert	2,796
[8] Links(KA) und Rechts zensiert	1,598
[9] Links(KA) und Rechts(KA)-zensiert	615
Total	203,250

Note: January 1983 is month 1!

Retrieving an episode dataset: hand-made blueprint

- A standard retrieval
 - Pull episodes from ARTKALEN
 - Pull info on persons from PPFAD
 - Perform episode splitting for discrete-time EHA (episode-month)
 - Now merge time-varying info from \$PGEN, \$HGEN, BIO*
 - ARTKALEN is monthly panel, \$* is yearly panel: How to merge?
 - Simplest way: merge by “id” and “year”
 - Or: merge by “id” and “intmonth” (and interpolate somehow)
- A hand-made blueprint (by J. Brüderl and V. Ludwig)
 - “SOEP Retrieval 4 Episodes.do”
 - Example: retrieve unemployment episodes

Working With BIOBIRTH (M)

- BIOBIRTH for women, BIOBIRTM for men
 - # of kids
 - For each kid (1-15): birth year and month, sex, persnr

Example: 4 observations in file biobirth.dta (only women)

id: person identifier
bioyear: year retrospective birth biography was administered
bioage: age when retrospective birth biography was administered
sumkids: number of kids
biokids: number of kids from retrospective birth biography
newkids: number of kids since retrospective birth biography (from prospective quest.)
kidgeb01-03: birth year of 1st-3rd child

hhnr	id	bioyear	bioage	sumkids	biokids	newkids	kidgeb01	kidgeb02	kidgeb03
272	2702	1989	29	2	1	1	1989	1990	-2
434	4301	1985	71	4	4	0	1935	1936	1938
523	5202	1985	29	1	1	0	1982	-2	-2
523	5203	1999	17	0	0	0	-2	-2	-2


Further Possible Retrievals

- **Extracting child-years from \$KIND**
 - From the hh questionnaire some basic information on kids is known
 - Whether (1) kid gets child care, and (2) which school it visits
 - It is possible to construct child care and educational career
 - A cautionary note:
 - Kids are linked over years by first-name and birth year
 - Unfortunately this process is not documented
 - It is, however, highly likely that many erroneous links occur. This can be seen in highly instable educational careers
 - Since V27 (2011) educational careers are already constructed in BIOEDUC (beta version)
- **SOEPlong**
 - Since V26 (2010) a beta-version of SOEPlong available
 - Advantage: already long format
 - Disadvantage: huge data sets, still many errors

Citing the SOEP

- **Acknowledgement**
 - The data used in this publication were made available by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW Berlin), Berlin.
- **Referencing the data in the text**
 - “We use data from the Socio-Economic Panel years 1984-2010 (SOEP version 27). The SOEP is described in detail in Wagner/Frick/Schupp (2007).”
- **Entry under “References”: reference paper citation**
 - Gert G. Wagner, Joachim R. Frick, and Jürgen Schupp (2007): The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. In: Schmollers Jahrbuch (Journal of Applied Social Science Studies), Vol. 127, No. 1, pp. 139-169.
- **Entry under “References”: data citation**
 - Socio-Economic Panel (SOEP) (2011) Data for years 1984–2010. doi:10.5684/soep.v27

SOEP | [Über uns](#) | [FDZ SOEP](#) | [News and Events](#) | [Publikationen](#) | [Presse](#)



[Einstieg für](#) [Quicklinks](#)

Nachrichten 1 von 3

Allgemein
→ [SOEP Nutzerumfrage gestartet - Teilnahme bis 31. Mai möglich](#)

Allgemein
→ [SOEP auf Facebook](#)

Allgemein
→ [SOEPnewsletter 96 veröffentlicht](#)

Allgemein
→ [FID-Daten für die Forschung verfügbar](#)

Allgemein
→ [Thomas Siedler erhält Ruf nach Hamburg](#)

→ [Zur Nachrichtenseite](#)

Aktuelle SOEPpapers 1 von 3

SOEPpapers
→ [Income Comparisons and Non-Cognitive Skills](#)
Santi Budria, Ada Ferrer-i-Carbonell

SOEPpapers
→ [The Prospects of the Baby Boomers](#)
Christian Westermeier, Anika Rasner, Markus M. Grabka

SOEPpapers
→ [Familienbiographische Verläufe im Kohortenvergleich](#)
Joachim R. Frick, Markus M. Grabka, Anika Rasner, Marian Schmitt, Morten Schuth, Christian Westermeier

SOEPpapers
→ [Taxing Childcare: Effects on Family Labor Supply and Children](#)
Christina Gathmann, Björn Sass

Veranstaltungen

SOEPcampus@Universität Mannheim
18.-20. Juni 2012,
Mannheim/Deutschland
→ [Homepage](#)

SOEP2012 - 10th International German Socio-Economic Panel User Conference
28.-29. Juni 2012, Berlin/Germany
→ [Homepage](#)

SOEPcampus im Rahmen von datalab2012@Uni-Bielefeld
24.-28. September 2012,
Bielefeld/Deutschland
→ [Homepage](#)

→ [alle Veranstaltungen](#)