SOEP The German Socio-Economic Panel study

292

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP — The German Socio-Economic Panel study at DIW Berlin 2015

SOEP 2014 – Documentation of Person-related Status and Generated Variables in PGEN for SOEP v31

SOEP Group

DIW SOEP

Running since 1984, the German Socio-Economic Panel study (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:

SOEP Group. 2015. SOEP 2014 – Documentation of Person-related Status and Generated Variables in PGEN for SOEP v31. SOEP Survey Papers 292: Series D. Berlin: DIW Berlin / SOEP

ISSN: 2193-5580 (online)

German Socio-Economic Panel (SOEP) | DIW Berlin
Mohrenstr. 58
10117 Berlin, Germany
Contact: soeppapers@diw.de

SOEP GROUP

SOEP 2014 – DOCUMENTATION OF PERSON-RELATED STATUS AND GENERATED VARIABLES IN PGEN FOR SOEP V31

Berlin, 2015

The German Socio Economic Panel Study at DIW Berlin

SOEP-Core V31—$PGEN

SOEP Group

2015-12-18

This file is part of a collection, which is released with doi:10.5684/soep.v31.

Contents

1 General Information **3**

2 Identifiers **3**
persnr – Never Changing Person ID . . . 3
hhnr – Original Household Number . . . 3
hhnrakt – hhnrakt Current Wave HH Number (=$hhnr) [generic] . . . 3
$hhnr – Current Wave HH Number (=HHNRAKT) [generic] . . . 3

3 Own Nationality **3**
nation$$ – Nationality [generic] . . . 3

4 Familiy Status and Partnership **4**
partz$$ – Partner Indicator [generic] . . . 4
partnr$$ – Partner Person Number [generic] . . . 5
$famstd – Marital Status [generic] . . . 5

5 Wages and Salary **6**
labgro$$ – Current Gross Labor Income in Euro [generic] . . . 6
impgro$$ – Imputation flag for LABGROxx [generic] . . . 7
labnet$$ – Current Net Labor Income in Euro [generic] . . . 8
impnet$$ – Imputation flag for LABNETxx [generic] . . . 9
sndjob$$ – Current Gross Secondary Income In Euro [generic] . . . 9
impsnd$$ – Imputation Flag For SNDJOBxx [generic] . . . 10

6 Current Employment Status **10**
stib$$ – Occupational Position [generic] . . . 10
emplst$$ – Employment Status [generic] . . . 12
lfs$$ – Labor Force Status [generic] . . . 12
erwtyp$$ – Type Of Occupation [generic] . . . 13
jobch$$ – Occupational Change [generic] . . . 14
autono$$ – Autonomy In Occupational Actions [generic] . . . 16

7 Current Occupation **17**
is88$$ – 4 Digit ISCO-88 Occupation Code [generic] . . . 17
isco08$$ – Current Occupational Classification (ISCO-08) [generic] . . . 18
isei$$ – ISEI-Status88 Ganzeboom IS88 [generic] . . . 18
klas$$ – StaBuA 1992 Job Classification [generic] . . . 19
kldb10$$ – Current Occupational Classification (KldB2010) [generic] . . . 20
mps$$ – Magnitude Prestige Scale KLAS [generic] . . . 21
siops$$ – Treiman Standard Int Occ Prestige IS88 [generic] . . . 22
egp$$ – Erikson, Goldthorpe Class Category IS88 [generic] . . . 23
erljob$$ – Working In Occupation Trained For [generic] . . . 24
ausb$$ – Required Training For Job [generic] . . . 24
$erwzeit – Length Of Time With Firm [generic] . . . 25
$tatzeit – Actual Work Time Per Week [generic] . . . 26
$vebzeit – Agreed Upon Work Time Per Week [generic] . . . 27
$uebstd – Overtime Per Week [generic] . . . 28
oeffd$$ – Civil Service [generic] . . . 30
nace$$ – 2 Digit NACE Industry, Sector [generic] . . . 30
betr$$ – Size Of The Company [generic] . . . 31

allbet$$ – Core Category Size Of The Company [generic] 32

8 Last Occupation **33**
jobend$$ – Reasons for occupational change [generic] 33

9 Employment History **34**
expft$$ – Working Experience Full-Time Employment [generic] 34
exppt$$ – Working Experience Part-Time Employment [generic] 35
expue$$ – Unemployment Experience [generic] 36

10 School, Higher and Vocational Education **37**
isced97_$$ – ISCED-1997-Classification [generic] 37
isced11_$$ – ISCED-2011.Classification [generic] 38
casmin$$ – CASMIN Classification [generic] 38
$bilzeit – Amount Of Education Or Training In Years [generic] 39
$psbil – School-Leaving Degree [generic] 40
$pbbil01 – Vocational Degree Received [generic] 40
$pbbil02 – College Degree [generic] 41
$pbbil03 – No Vocational Degree [generic] 41
$psbilo – School-Leaving Degree East Germany [generic] 42
$pbbilo – Vocational Degree Received East Germany [generic] 42
$psbila – School-Leaving Degree Outside Germany [generic] 43
$pbbila – Vocational Degree Outside Germany [generic] 43
field$$ – Field of tertiary education [generic] 44
degree$$ – Type of tertiary degree [generic] 45
traina$$ – Apprenticeship - two-digit occupation KldB92 [generic] 47
trainb$$ – Vocational school - two-digit occupation KldB92 [generic] 49
trainc$$ – Higher vocational school - two-digit occupation KldB92 [generic] . . 50
traind$$ – Civil servant training - two-digit occupation KldB92 [generic] 51
fdt_f$$ – Data source FIELD, DEGREE, TRAIN [generic] 52
bilztch$$ – Change in Education since last survey / last year [generic] 52
bilztev$$ – Change in Education, total observed period [generic] 53

11 Information on the Interview **54**
month$$ – Month Of Interview [generic] 54
mode$$ – Interview Method [generic] 54

1 General Information

The $PGEN-files contain user friendly data on the individual level which are consolidated from different sources. The plausibility is in many respects longitudinally validated, therefore the data here are in most situations superior compared to the data in $P.
The file contains one row for each person (persnr is unique) with a completed personal or youth questionnaire. These are the persons where in PPFAD $netto has the values 10–17 or 19 which is equivalent for values 1 and 5 in $netold.
While frequencies are calculated from the most recent $PGEN the here presented information is basically valid for all $PGEN files. Therefore many variable names depicted here have the generic form $name or name$$ and are flagged with [generic].

2 Identifiers

persnr – Never Changing Person ID

The central individual identifier across time is PERSNR, which is fixed over time (and of course datasets).

hhnr – Original Household Number

The identifier of the household, when it is sampled and selected for interviewing for the first time. The hhnr is attached to all persons living in this household and all new persons inherit this identifier, when they are born or move in a SOEP household. It is fixed no matter how often a person changes the household in the course of time.

hhnrakt – hhnrakt Current Wave HH Number (=$hhnr) [generic]

This identifier groups all persons into households at the time of the most recent wave. [This information can be related to a specific variable and is not necessary generic.]

$hhnr – Current Wave HH Number (=HHNRAKT) [generic]

This identifier groups all persons into households at the time of the most recent wave. [This information can be related to a specific variable and is not necessary generic.]

3 Own Nationality

nation$$ – Nationality [generic]

1	[1] Germany	24858
2	[2] Turkey	576
3	[3] Ex-Yugoslavia	7
4	[4] Greece	218
5	[5] Italy	305
6	[6] Spain	78
7	[7] Ex-GDR (Country Of Origin Only)	0
10	[10] Austria	60
11	[11] France	35
12	[12] Benelux	0

13	[13] Denmark	11
14	[14] Great Britain	30
15	[15] Sweden	5
16	[16] Norway	3
17	[17] Finland	5
...	(156 rows omitted)	1851
174	[174] Madagascar	0
175	[175] Grenada	0
176	[176] Lesotho	0
177	[177] Bhutan	0
178	[178] Rwanda	0
179	[179] Malawi	0
180	[180] Bessarabia	0
222	[222] Eastern Europe	0
333	[333] other unspecified foreign country	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to integrate the information on respondent's nationality for all subsamples. Since some members of Sample B (persons with Turkish, Italian, Spanish, Greek, and Yugoslavian citizenship) received the question items in their own language up to 1995, to carry out an integrated analysis with Sample B, the user must obtain this information from the corresponding $PAUSL files and add it to the individual data. The variable NATION$$ thus offers a variable on nationality for all subsamples. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

4 Familiy Status and Partnership

partz$$ – Partner Indicator [generic]

0	[0] No partner	8976
1	[1] Spouse, registered partner	16244
2	[2] Partner	2606
3	[3] Probably spouse, registered partner	130
4	[4] Probably partner	77
-1	[-1] Unclear	9
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
Partner indicators have the purpose of defining singles, spouses (married and civil union),

and partners (unmarried, no civil union) in SOEP households and thus enabling analyses on the couple level. The variable PARTZ$$ generated in this context reveals whether a person in a SOEP household has a partner in that household, and if so, the type of relationship existing between the partners. Relationships with persons outside the SOEP household are not covered by this variable.
To explain the codes:
Code 0 is assigned to all single persons living in households and those with partners outside the household. Codes 1 to 4 describe relationships. To assign Codes 1 and 2, the partnership has to be definable from the perspective of both partners unanimously. If conflicting information exists between partners, the codes 3 or 4 are assigned. If it is unclear whether an individual has no partner or whether she forms a couple with one other household member, we assign the code -1. Registered partnerships (civil unions) for same-sex couples were introduced in Germany in 2001. Though, registered partnerships are legally not equal to marriage, they are listed in the same category. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Martin Kroh (Tel. +49-30-89789-678)

partnr$$ – Partner Person Number [generic]

Waves: all
Partner indicators have the purpose of defining partnerrelationships in SOEP households and thus to make possible analyses on the couple level. If PARTZ$$ is coded 0, this person has no partner in the household. The variable PARTNR$$ is assigned the missing code of "-2" (=does not apply) for these persons. If PARTZ$$ is coded 1, 2, 3 or 4, a partnership was defined and PARTNR$$ is then assigned the value of the unchanging person ID number (=PERSNR) of the partner. The assignment of the partner ID within households is based on four sources of information: A question in the person-file, that asks (unmarried) respondents to identify their partner in the household (bcp13003 in 2013), the household matrix reported by the head of household at the beginning of the interview (bcstell in 2013), the partnership biography in the lifehistory calendar reported by new respondents (see also, biomars), and self-reports on marital status and life events, such as marriage, move in with partner, separation, etc. In unclear cases, due to temporal non-response for instance, we also consider longitudinal information from previous and prospective waves. Moreover, PARTNR$$ is self-consistent between two individuals. For analyses of partner relationships, this information can be used to link all persons with their respective partners, and all information on both partners can also be stored in a common dataset. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Martin Kroh (Tel. +49-30-89789-678)

$famstd – Marital Status [generic]

1	[1] married, living together	16157
2	[2] married, living separately	172
3	[3] unmarried	7170
4	[4] divorced	2520
5	[5] widowed	1480
6	[6] spouse abroad	0
7	[7] reg. same-sex p., living together	36
8	[8] reg. same-sex p., living separately	6
-1	[-1] Missing	13

-2	[-2] Does not apply	0
-3	[-3] Not Valid	488
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: 1984-2012
Marital status is describing the institutional status of marriage at the time of the person interview. Marital status is based on information given by the respective person on his or her current relationship as well as on retrospective information about previous relationships asked in the biography questionnaire. Information on marital status when a child was born (provided in the biography information) is not used here, so contradicting information to BIOBIRTH might still be possible. For those whose partner was identified within the household, marital status is counter-checked with the information given by the partner. Where contradictions can be found, indication of the person information is compiled if reasonable. If no information is available, the indication by position related to head of household is deferred. Remaining contradictions are solved using information on marriage status when a child was born as well as future reports on a given relationship. Marital status is only available for people, who were interviewed. Data on marital status provided here is generated from and hence consistent with the corresponding BIOMARSY data, although value labels differ between $FAMSTD and SPELLTYP. Note that marital status in a specific wave can be different between data distributions due to consistency checks using up-to-date information from following waves. Because of the newly introduced construction process of BIOMARSY, changes might be more than usual. Above, the partner indicator PARTZ$$ supplied in the $PGEN data files as well might not match the information provided in $FAMSTD in its entirety. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Paul Schmelzer (Tel. +49-30-89789-526, pschmelzer@diw.de)

5 Wages and Salary

labgro$$ – Current Gross Labor Income in Euro [generic]

0		106
7		1
12		1
17		1
18		1
26		1
27		1
30		6
35		2
40		9
45		1
50		8
52		1
53		1
55		5
...	(2610 rows omitted)	16834
20000		14

20300	1
21000	1
24221	1
25000	3
25117	1
27000	1
30000	4
34000	1
35000	2
40000	2
42000	1
50000	3
50989	1
-2	11027

Waves: all
The variable LABGRO$$ represents the imputed current gross labor income generated for all SOEP respondents who are employed in a main job in each respective wave. Income details are consistently provided in euros for all waves. Item nonresponse is imputed in a two-stage procedure: first, with the "Row-and-Column" method of Little und Su (1989) using individual longitudinal data as well as cross-sectional trend data (cf. Joachim R. Frick and Markus M. Grabka (2005): Item-Non-Response on Income Questions in Panel surveys: Incidence, Imputation and the Impact on the Income Distribution. Allgemeines Statistisches Archiv (ASTA) 89, 49-61). Alternatively, if no individual longitudinal information is available, we base the imputation on a regression using different Mincer covariates, also taking into account current net labor income. If both types of income information are lacking, first we impute current net labor income and then current gross labor income. Imputed values are flagged (IMPGRO$$).
The original variables coming from the $P-files and are: ap3301, bp4301, cp5201, dp4401, ep4401, fp4501, gp4301, hp5401, ip5401, jp5401, kp6401, lp5301, mp4701, np5401, op4501, pp6001, qp5601, rp5701, sp5801, tp7601, up5901, vp7101, wp5901, xp7301, yp6801, zp7201, bap6101, bbp7301, bcp5901, bdp7701, bep5701. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

impgro$$ – Imputation flag for LABGROxx [generic]

0	[0] Not Imputed	15768
1	[1] Imputed	1247
-1	[-1] No Answer	0
-2	[-2] Does not apply	11027
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
The variable IMPGRO$$ designates imputations of item non-response in the variable LABGRO$$ (current gross labor income). IMPGRO$$ can take the value 0 = "no imputation", 1 = "imputed income statement" and - 2 = "does not apply, not working". [This information

can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

labnet$$ – Current Net Labor Income in Euro [generic]

0		110
7		1
12		1
17		1
18		1
20		1
22		1
24		1
29		2
30		6
35		1
40		8
41		1
45		1
46		1
...	(2330 rows omitted)	16847
11542		1
12000		7
13000		3
14400		1
15000		3
18000		3
19000		1
19442		1
20000		4
20567		1
21000		1
22000		1
23000		1
25000		3
-2		11027

Waves: all
The variable LABNET$$ represents the generated and imputed current net labor income in a main job of all persons in SOEP working in the respective wave. Income details are consistently provided in euros for all waves. The imputation of item nonresponse takes place in a two-stage procedure: first, with the "Row-and- Column" method of Little und Su (1989) using individual longitudinal data as well as cross-sectional trend data (cf. Joachim R. Frick and Markus M. Grabka (2005): Item-Non-Response on Income Questions in Panel surveys: Incidence, Imputation and the Impact on the Income Distribution. Allgemeines Statistisches Archiv (ASTA) 89, 49-61). Alternatively, if no individual longitudinal information is available, we base the imputation on a regression using different Mincer covariates, also taking into account current gross labor income. If both types of income information are lacking, first we impute current gross labor income and then current net labor income.

Imputed values are flagged (IMPNET$$).
The original variables coming from the $P-files and are: ap3302, bp4302, cp5202, dp4402, ep4402, fp4502, gp4302, hp5402, ip5402, jp5402, kp6402, lp5302, mp4702, np5402, op4502, pp6002, qp5602, rp5702, sp5802, tp7602, up5902, vp7102, wp5902, xp7302, yp6802, zp7202, bap6102, bbp7302, bcp5902, bdp7702, bep5702. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

impnet$$ – Imputation flag for LABNETxx [generic]

0	[0] Not Imputed	16018
1	[1] Imputed	997
-1	[-1] No Answer	0
-2	[-2] Does not apply	11027
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
The variable IMPNET$$ designates imputations of item non-response in the variable LABNET$$ (current net labor income). IMPNET$$ can take the value 0 = “no imputation”, 1 = “imputed income statement” and - 2 = “does not apply, not working”. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

sndjob$$ – Current Gross Secondary Income In Euro [generic]

0		199
1		1
12		2
13		1
14		2
15		2
20		21
21		1
22		1
23		1
24		1
25		4
30		18
32		3
33		3
...	(241 rows omitted)	1564
1997		1
2000		6
2400		1
2500		2
2800		1

3200	2
3273	1
3400	1
3850	1
4000	3
4500	1
5000	1
6000	1
25000	1
-2	26195

Waves: all
The variable SNDJOB$$ represents the imputed current gross labor income from second job generated for all SOEP respondents in each respective wave. Income details are consistently provided in euros for all waves. Item nonresponse is imputed in a two-stage procedure: first, with the "Row-and-Column" method of Little und Su (1989) using individual longitudinal data as well as cross-sectional trend data (cf. Joachim R. Frick and Markus M. Grabka (2005): Item-Non-Response on Income Questions in Panel surveys: Incidence, Imputation and the Impact on the Income Distribution. Allgemeines Statistisches Archiv (ASTA) 89, 49–61). Alternatively, if no individual longitudinal information is available, we base the imputation on a regression using a subset of different Mincer covariates. Information about gross income from second job is firstly asked in wave 1995 (L). In 2013 information about income from second jobs was not collected for sub-sample M. We use information from the the subsequent survey year (2013) about income from second job in the previous year (variable BEP2C03). If persons from the sub-sample M stated that he/she had income in the previous year from second jobs, then IMPSND13 was set to -1 and SNDJOB$$ had been imputed. Imputed values are flagged (IMPSND$$).
The original variables coming from the $P-files and are: lp7702, mp5802, np5802, op4902, pp6602, qp6302, rp6602, sp6602, tp8402, up67, vp79, wp70, xp81, yp76, zp78, bap69, bbp79, bcp67, bdp85, bep64. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

impsnd$$ – Imputation Flag For SNDJOBxx [generic]

0	1700
1	147
-2	26195

Waves: all
The variable IMPSND$$ designates imputations of itemnonresponse in the variable SNDJOB$$ (current gross labor income from second job). IMPSND$$ can take the value 0 = "no imputation", 1 = "imputed income statement" and - 2 = "does not apply, not working". [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Markus Grabka (Tel. +49-30-89789-339 / mgrabka@diw.de)

6 Current Employment Status

stib$$ – Occupational Position [generic]

0	[0] Do Not Know	0
10	[10] Not Employed	1893
11	[11] In Training	973
12	[12] Unemployed, Not Employer	1612
13	[13] Pensioner	5645
15	[15] Military, Community Service	37
110	[110] Apprentice	0
120	[120] Apprentice,Trainee Industry Technology	362
130	[130] Apprentice, Trainee Trade And Commerce	212
140	[140] Trainee, Intern	87
150	[150] Aspirant	0
210	[210] Untrained Worker	771
220	[220] Semi-Trained Worker	1385
230	[230] Trained Worker	1433
240	[240] Foreman, Team Leader	174
...	(21 rows omitted)	2924
522	[522] Trained Employee With Simple Tasks	1433
530	[530] Qualified Professional	4217
540	[540] H. Qualified Professional	2309
550	[550] Managerial	242
610	[610] Low-Level Civil Service	24
620	[620] Middle-Level Civil Service	216
630	[630] High-Level Civil Service	490
640	[640] Executive Civil Service	299
999	[999] Employed Without StiB Info	0
-1	[-1] No Answer	281
-2	[-2] Does not apply	1023
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

The variable represents a compilation of all relevant information on current occupational position. It is generated by combining information on “occupational group”, “unemployed (yes/no)”, “military/community service”, “in education (yes/no)”, and “pensioner”. A hierarchical scheme is used to determine which data is given precedence when a variety of divergent information exists (increasing dominance):

value	label
10	not employed
13	pensioner
11	currently in education
15	military / community service
12	registered unemployed
110–150	apprentice
410–440	self-employed
210–250	manual laborer
510–550	employee
610–640	civil service

The categories (150) and (310) to (340) were only assigned to respondents in East Germany in 1990. In STIB$$, non-working persons are only assigned to the category (13) "pensioner" if they are recipients of retirement pension or if they are recipients of widow's pension AND are older than 60 years. Moreover, if there is missing information on pension receipt, additional information from ARTKALEN (retrospective information from the activity calendar for the previous year) is used in the generation process to determine if a person was in retirement or early retirement (Vorruhestand) at the time of the interview.
The code (-2) is assigned to first time respondents aged 16 or 17 who answer the youth questionnaire (since wave W (2006)). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

emplst$$ – Employment Status [generic]

1	[1] Full-Time Employment	10333
2	[2] Regular Part-Time Employment	3939
3	[3] Vocational Training	783
4	[4] Marginal, Irregular Part-Time Employment	2017
5	[5] Not Employed	10932
6	[6] Sheltered workshop	30
-1	[-1] No Answer	5
-2	[-2] Does not apply	3
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is generated from the annual question on current employment status, which has a central filter function in the questionnaire to separate employed people from non-employed people for further questions. It is designed to provide consistent longitudinal data on employment status across all waves.
Since the beginning of the SOEP in the year 1984, a consistent status variable has been used to differentiate among different types of employment status. The category "not employed" comprises non-working individuals, those in military/community service, those on maternity leave, and employed persons in a phased retirement scheme (Altersteilzeit) whose current actual working hours are zero. From 1998 on, the additional category "sheltered workshop" is included for disabled persons in sheltered employment.
EMPLST$$ supplements the variable LFS$$, which differentiates among persons who are not employed. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

lfs$$ – Labor Force Status [generic]

1	[1] Non-Working	2165
2	[2] NW-Age 65 And Older	4590
3	[3] NW-In Education-Training	1253
4	[4] NW-Maternity Leave	585

5	[5] NW-Military-Community Service	16
6	[6] NW-Unemployed	1452
8	[8] NW-But Sometimes Sec. Job	312
9	[9] NW-but work past 7 days	208
10	[10] NW-But Reg. Sec. Job	241
11	[11] Working	16883
12	[12] Working But NW Past 7 Days	337
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is based on the annual question on current employment status, combined with additional information on activities of non-working individuals. The number of values assigned has been based, since the beginning of the SOEP in the year 1984, on a large number of highly differentiated answer categories. It is designed to provide consistent longitudinal data on labor force participation across all waves.
LFS$$ provides a differentiation between "working" (Code 11–12) and "non-working" (Code 1–10), categories which are constant over all waves. Non-employment is subdivided further in order to make it possible to efficiently apply different labor market concepts in studying the data. To calculate this variable, the variables on employment status, age, maternity leave, second jobs, registration at the employment office, participation in paid work during the past 7 days and training status are used. Code (12) was added in 2000.
For respondents who have multiple status codes and different values for this variable, the following hierarchy was used to determine which of the values would play the determining role (increasing dominance):

value	label
11	working
1	non-working without further information
2	non-working, and older than 65
3	non-working, and currently in a training program
6	non-working, and registered unemployed
4	non-working, on maternity leave
5	non-working, in military/community service
9	non-working, but working past 7 days
10	non-working, but regular second job
8	non-working, but occasional second job
12	working, but non-working past 7 days

LFS$$ supplements the variable EMPLST$$, which differentiates among persons who are employed. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

erwtyp$$ – Type Of Occupation [generic]

1	[1] Not Employed, Green	10821
2	[2] Not Employed (First Surveyed) Not Applicable Since 94	0
3	[3] Employed (First Surveyed) Not Applicable Since 94	0
4	[4] Empl. Exc Change	12998
5	[5] Empl. No Info If Change	340
6	[6] Empl. With Change, Also First Time Employment	3769
7	[7] Empl. With Near-Retirement Part-time	111
-1	[-1] No Answer	3
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is generated from the question on whether a respondent has changed jobs since the beginning of the previous year, which is a central filter variable in the questionnaire. In years with a partial survey – 1985, 1986, 1987, 1988, 1990 (West), 1992 (West), 1994, 1996, 1999, 2001, 2003, 2005, 2006, 2008, 2010, 2012, and 2014 – only employed persons who changed jobs and first-time respondents are asked to provide up-to-date information on time-invariant job characteristics. Therefore, in years with a partial survey, for persons without a job change, many of the generated variables related to the job contain information from the previous year.
The variable ERWTYP$$ was originally created to integrate the "blue" (first-time respondent) and "green" (follow-up respondent) questionnaires used up to 1993 to differentiate between employed persons with and without a change of job (as a central filter variable). Since the 1994 wave, there has been only one questionnaire for all respondents - both first-time and follow-up respondents. For this reason, since the 1994 wave, Codes (2) and (3) are no longer assigned. Codes (1), (4), and (5) have been assigned since 1994 to first-time respondents (who would have previously received the blue questionnaire). The variable was recalculated for all waves and is assigned the code (6) only when the respondent started a new job since the last interview or was hired for his or her first job ever. The variable also includes a new code (7) since wave T (2003) for employed persons in a phased retirement scheme (Altersteilzeit) whose current actual working hours are zero.
An alternative variable is JOBCH$$ (see below), which is an improved version of ERWTYP$$, as it is generated in a longitudinally consistent way and contains an additional category for first-time employed persons. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

jobch$$ – Occupational Change [generic]

1	[1] Not Employed	10821
2	[2] Employed No Change	13706
3	[3] Employed No Info If Change	402
4	[4] Employed With Change	2831
5	[5] First Time Employed	281
-1	[-1] No Answer	1
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0

-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable indicates a change of job since the previous interview for respondents with a follow-up interview, whereas for first-time respondents, the information refers to a change of job since the beginning of the previous year.
JOCHCH$$ is generated based on the central filter variable, which indicates whether a respondent has changed jobs since the beginning of the previous year. A job change can be within one firm as well as a change to another firm. Information on the date of job change is then combined with interview month of the previous year's interview to identify whether a new job change has taken place since the previous interview.
Hence, JOBCH$$ is a modified version of the variable ERWTYP$$ which indicates whether a respondent has changed jobs since the beginning of the previous year. Unlike ERWTYP$$, the variable is calculated for all waves, and the codes are assigned independently of the respondent being a first-time or follow-up respondent.
In addition to ERWTYP$$, the variable is also designed to identify respondents who have entered employment for the first time. Up to 1993, first-time respondents did not answer the question about job change. Therefore, for first-time respondents up to 1993, the variable was generated by using the information on the start date with the current employer and the respondent's age at entrance into his/her first job.
In addition to ERWTYP$$, the variable is designed to provide consistent longitudinal information on job changes. The JOBCH$$ variable is generated by correcting the original job change information in various ways:

1. We check whether the job changes stated by a respondent in two consecutives interviews refer to one and the same job change. The date of the job change and the interview month are used to correct double entries.
2. If the respondent indicates a job change with a date before the previous interview but did not state a job change in the previous interview, this is coded as a job change in the current interview.
3. If a respondent indicates no job change and was not employed at the time of the previous interview, this is coded as "no job change" despite the seeming implausibility, since there are possible explanations how this information could be plausible, e.g. if there were short-term employment spells between two interview dates.
4. Respondents can be "first-time employed" only once. If a respondent states being "first-time employed" for a second time, this is coded as "employed, with change".

In 2013 the respondents of the newly introduced migration sample were not asked whether they have changed jobs since the beginning of the previous year, therefore the generation of JOBCH$$ for the migration sample was modified in 2013:

1. Respondents who are not employed were coded (1).
2. Respondents who are still in the same occupation and position and are working for the same employer as they had worked in their first job in Germany were coded (2).
3. Respondents who have entered the firm they are currently working after the 31th of December 2011 were coded (4).
4. If a respondent is in her first vocational training this was coded as (5).
5. Respondents who are employed but for whom no further information could be used were coded (3). In 2014 there was again a uniform questionnaire for all respondents. [This information can be related to a specific variable and is not necessary generic.]

For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

autono$$ – Autonomy In Occupational Actions [generic]

0	[0] Apprentice	947
1	[1] Low Autonomy	2188
2		4144
3		5566
4		3451
5	[5] High Autonomy	650
-1	[-1] No Answer	275
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable gives the occupational autonomy for all employed persons. It offers an alternative to the ISCO-based scales on occupational status (ISEI$$), class (EGP$$), or prestige (SIOPS$$). AUTONO$$ is the simplest variable based on the scales of "occupational position" in terms of its construction, and strongly correlated with the Treiman Prestige Scale (SIOPS$$).
The basis for the "autonomy in occupational activity" scale is the classification of occupational position. Self-employed persons are categorized according to the size of the company (with the exception of farmers, who are all classified within the same category of autonomy, independent of farm size in hectares). Civil servants are differentiated according to the civil service laws defining each kind of activity and the amount of autonomy connected to it. Workers are differentiated according to their vocational training, and thus categorized hierarchically according to the different tasks they can be expected to carry out and the different amounts of responsibility associated with each task. Similarly, salaried employees are classified according to how differentiated their tasks are and how much responsibility is associated with each.
The value "1" is assigned mainly to manual workers with a low level of status and a low level of autonomy. Group 2 encompasses work in production, services demanding a minimal level of specialization, and farm work. Activities that require completion of the middle track of secondary education and entail a limited amount of responsibility are classified in Group 3. Group 4 includes activities carried out either with or without supervision that require a degree from a college of applied sciences or university, but are not very high in prestige. Managers and freelance academics are both placed in Group 5 (highest autonomy). Depending on the number of employees, self-employed are categorized in Group 3, Group 4, or Group 5. [This information can be related to a specific variable and is not necessary generic.]
Detailed description: Hoffmeyer-Zlotnik, Jürgen H.P., and Alfons J. Geis (2003) Berufsklassifikation und Messung des beruflichen Status/ Prestige. In: ZUMA-Nachrichten 52, Jg. 27, Mai 2003. pp. 125-138.
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

7 Current Occupation

Not all employed persons are asked the question about occupation on an annual basis. In years with a partial survey – 1985, 1986, 1987, 1988, 1990 (West), 1992 (West), 1994, 1996, 1999, 2001, 2003, 2005, 2006, 2008, 2010, 2012, and 2014 – only those employed persons who changed jobs and first-time respondents are asked to provide up-to-date information. Therefore, in years with a partial survey or in case of non-response the variables in this section usually contain available information from the previous year (persons with JOBCH$$-category (2) "employed, no change"). For some persons without a job change who updated the information on their current occupation without being asked, up-to-date information is used.

is88$$ – 4 Digit ISCO-88 Occupation Code [generic]

1	[1] Armed Forces	0
2	[2] Officer	0
3	[3] Job Seeker	0
4	[4] Insufficient Info	0
5	[5] Student	0
6	[6] Housewife	0
7	[7] Pensioner	0
8	[8] Do Not Know	0
9	[9] No Answer	0
110	[110] Armed Forces	6
1110	[1110] Legislator	1
1120	[1120] Senior Government Official	0
1130	[1130] Traditional Chief, Head Of Village	0
1140	[1140] Sr. Official Organisation	0
1141	[1141] Sr. Official Political Party Org.	0
...	(440 rows omitted)	3562
9910	[9910] Help In Family Business, Not Agri	0
9920	[9920] Apprentice Without Defined Job Role	0
9930	[9930] Intern, Trainee Without Defined Job Role	0
9941	[9941] School Leaver, Seeking Employment	0
9942	[9942] Other Labourer, Seeking Employment	0
9950	[9950] Skilled Labourer, Craftsman, No Further Details	0
9960	[9960] Home Worker, No Further Details	0
9970	[9970] Foreman, Team Leader	0
9980	[9980] Sonstige Arbeitskraefte o.n.T.	0
-1	[-1] No Answer	64
-2	[-2] Does not apply	24409
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Hartmann and Schütz (2002) provide detailed information on occupational coding. [This information can be related to a specific variable and is not necessary generic.]
Hartmann/Schütz (2002): Die Klassifikation der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984–2001. Infratest Sozialforschung, München.

https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.40132.de/vercodung.pdf
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

isco08$$ – Current Occupational Classification (ISCO-08) [generic]

110	[110] Commissioned armed forces officers	1
210	[210] Non-commissioned armed forces officers	0
310	[310] Armed forces occupations, other ranks	5
1111	[1111] Legislators	2
1112	[1112] Senior government officials	3
1113	[1113] Traditional chiefs and heads of village	0
1114	[1114] Senior officials of special-interest organizations	0
1120	[1120] Managing directors and chief executives	16
1211	[1211] Finance managers	4
1212	[1212] Human resource managers	6
1213	[1213] Policy and planning managers	5
1219	[1219] Business services and administration managers not elsewhere classified	5
1221	[1221] Sales and marketing managers	19
1222	[1222] Advertising and public relations managers	2
1223	[1223] Research and development managers	0
...	(443 rows omitted)	3468
9954	[9954] Independent Consultant	0
9955	[9955] Referee	0
9960	[9960] Social Sector	0
9961	[9961] Medical Sector	0
9962	[9962] Therapist	0
9990	[9990] Secondary General School Degree Profess. Prep.	0
9991	[9991] Intermediate School	0
9992	[9992] Vocational School	0
9993	[9993] Specialized School (Field not Detectable) (9996)	0
-1	[-1] No Answer	97
-2	[-2] Does Not Apply	24409
-3	[-3] Retired	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

isei$$ – ISEI-Status88 Ganzeboom IS88 [generic]

16	193
19	25
20	130
21	8
22	2
23	69
24	7

25		187
26		38
27		13
28		15
29		167
30		217
31		18
32		39
...	(32 rows omitted)	2069
69		222
70		18
71		67
74		8
77		38
78		2
79		4
82		10
83		2
85		15
87		2
88		36
90		1
-1		11
-2		24409

Waves: all
This variable reflects the Standard International Socio-Economic Index of Occupational Status for all employed persons. The ISEI Index was developed in 1992 by Ganzeboom, De Graaf, Treiman, and De Leew based on information about income, education, and occupation. Technically, ISEI was created by scaling the ISCO88 classification. The values for the variable range between 16 and 90. In contrast to the prestige scores of Ganzeboom and Treiman (1996) and Wegener (1988), ISEI is a measure of socio-economic status.
Please also see occupational prestige scores (SIOPS$$, MPS$$) and occupational class (EGP$$). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

klas$$ – StaBuA 1992 Job Classification [generic]

100	[100] Landwirtschaftliche Berufe	0
110	[110] LandwirtInnen, allgemein	4
111	[111] Obst- und Gemuesebauern,-baeuerinnen (ni	0
112	[112] Ackerbauern,-baeuerinnen fuer Spezial-,	0
113	[113] ViehhalterInnen und Gruenlandwirtinnen	1
114	[114] Saat-, PflanzenzuechterInnen, VermehrerI	0
115	[115] PflanzenschuetzerInnen	0
116	[116] LandwirtInnen und GastwirtInnen,Kaufleut	0
118	[118] LandwirtInnen und WinzerInnen	0
120	[120] WinzerInnen, allgemein	0
121	[121] RebenveredlerInnen	0
129	[129] andere WinzerInnen	0

130	[130] Landarbeitskraefte, allgemein	1
131	[131] LandarbeitsaufseherInnen	0
132	[132] LandmaschinenfuehrerInnen	0
...	(2439 rows omitted)	3574
9931	[9931] VorarbeiterInnen, GruppenleiterInnen o.n	1
9940	[9940] Zivildienstleistende ohne naehere Taetig	0
9941	[9941] Zivildienstleistende o.n.T.	0
9950	[9950] Selbstaendige ohne naehere Taetigkeitsan	0
9951	[9951] Selbstaendige o.n.T.	0
9960	[9960] Beratungs-, Planungsfachleute ohne naehe	0
9961	[9961] Beratungs-, Planungsfachleute o.n.T.	2
9970	[9970] Sonstige Arbeitskraefte ohne naehere Tae	0
9971	[9971] Sonstige Arbeitskraefte o.n.T.	39
-1	[-1] No Answer	11
-2	[-2] Does not apply	24409
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Current occupation coded as KldB1992. Hartmann and Schütz (2002) provide detailed information on occupational coding. [This information can be related to a specific variable and is not necessary generic.]
Hartmann/Schütz (2002): Die Klassifikation der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984–2001. Infratest Sozialforschung, München. https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.40132.de/vercodung.pdf
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

kldb10$$ – Current Occupational Classification (KldB2010) [generic]

1203		3
1402		3
11101		6
11102		5
11104		1
11113		1
11194		1
11211		2
11212		1
11232		1
11694		1
11711		2
11712		2
11714		1
12101		13
...	(490 rows omitted)	3458
94214		1
94252		1
94303		1
94342		1

94383	1
94402	1
94404	2
94512	3
94513	2
94532	1
94533	1
94534	1
94704	2
-1	114
-2	24409

[This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

mps$$ – Magnitude Prestige Scale KLAS [generic]

30.1000003814697		7
30.2000007629395		6
30.2999992370605		2
31		45
31.1000003814697		10
31.2000007629395		3
31.5		48
31.7000007629395		15
31.7999992370605		2
32		9
32.0999984741211		43
32.2000007629395		2
32.2999992370605		158
32.4000015258789		36
32.5		8
...	(139 rows omitted)	3029
123.900001525879		4
125.199996948242		2
132.100006103516		19
135.699996948242		15
138.199996948242		2
138.899993896484		2
139.800003051758		5
145.699996948242		18
152.5		32
153.5		1
191.300003051758		36
207.199996948242		6
216		3
-1		65
-2		24409

Waves: all

This variable gives the occupational prestige score developed by Wegener (1988) for all employed persons. Like the SIOPS$$ prestige sore, Wegener's prestige scala measures a person's occupational prestige and was developed especially for use in the Federal Republic of Germany. MPS$$ is assigned based on the German Federal Statistical Office's occupational classification of 1992 (KLAS$$). The procedure has been documented in Frietsch and Wirth (2001).
Please also see occupational prestige scores (SIOPS$$), occupational status (ISEI$$), and occupational class (EGP$$). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

siops$$ – Treiman Standard Int Occ Prestige IS88 [generic]

13		3
15		8
17		1
19		85
20		65
21		287
22		45
23		75
24		6
25		30
26		3
27		2
28		27
29		34
30		70
...	(30 rows omitted)	2666
63		1
64		7
65		21
66		26
67		14
69		5
70		37
71		10
72		10
73		12
75		1
76		1
78		70
-1		11
-2		24409

Waves: all
This variable gives the occupational prestige score index for all employed persons. SIOPS$$ is based on ISCO-88 and was developed by Donald Treiman et al. The scale ranges from 6 to 78. The alogorithm is based on Fritsche and Wirth (2001).
Please also see occupational prestige scores (MPS$$), occupational status (ISEI$$), and

occupational class (EGP$$). [This information can be related to a specific variable and is not necessary generic.]
Frietsch, Rainer/Wirth, Heike (2001): Die Uebertragung der Magnitude-Prestigeskala von Wegener auf die Klassifikation der Berufe. In: ZUMA Nachrichten 48 (Jg.25): 139–165
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

egp$$ – Erikson, Goldthorpe Class Category IS88 [generic]

1	[1] High Service	367
2	[2] Low Service	759
3	[3] Routine NonManual	373
4	[4] routine service-sales	535
5	[5] SEmpl With Empl	597
6	[6] SEmpl No Empl	880
8	[8] Skilled Manual	435
9	[9] Semi - Unskilld Manual	910
10	[10] Farm Labor	52
11	[11] SEmpl Farm	57
15	[15] Not Working - Unemployed	0
18	[18] Not Working - Pensioner	0
-1	[-1] No Answer	11
-2	[-2] Does not apply	23066
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable gives the occupational class for all employed persons. EGP$$ is derived from the Standard International Socio-Economic Index of Occupational Status (ISEI). Technically, the variable was created by scaling the ISCO-88 classification. In addition, it is based on information about income, education and occupation. The EGP Index was documented by Ganzeboom/Treiman in 1996 and revised in 2003.
The values for the variable range between 1 and 11; additional categories are (15) not working - registered unemployed and (18) not working - pensioner.
Non-working persons are only assigned to the category “not working - pensioner” if they are recipients of retirement pension or if they are recipients of widow’s pension AND are older than 60 years. Moreover, if there is missing information on pension receipt, additional information from ARTKALEN (retrospective information from the activity calendar for the previous year) is used in the generation process to determine if a person was in retirement or early retirement (Vorruhestand) at the time of the interview. Hence, the category “not working - pensioner” in the most recent wave will be updated with retrospective information of the following wave. All other non-working persons are assigned to category (-2) “does not apply” as long as they are not registered as unemployed (category 15).
As information about supervisory status is only available from wave X (2007) on, it is not used to generate the corresponding EGP$$ category. Hence, the potential category (7) “Manual workers with supervisory status” is not assigned.
Annual information on the occupational position is used to generate the EGP-categories for the self-employed. In case no information on the number of employees is available,

the EGP$$-categories (5) and (6) contain information on the firm size for self-employed persons.
Based on the new classification developed by Ganzeboom/Treiman (2003), several ISCO values were recoded in EGP$$ as follows:

- ISCO 2470 becomes EGP=1.
- ISCO 2500 becomes EGP=2.
- ISCO 4300, 4400, 4500 become EGP=4.
- ISCO 7900 becomes EGP=7.
- ISCO 9910-9990 become EGP=9.

Please also see occupational status (ISEI$$) and occupational prestige scores (SIOPS$$, MPS$$). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Knut Wenzig (Tel. +49 30 89789 341, kwenzig@diw.de)

erljob$$ – Working In Occupation Trained For [generic]

1	[1] Yes	8808
2	[2] No	5358
3	[3] In Training	1010
4	[4] has No Job Training	909
-1	[-1] No Answer	1136
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to offer annual data on all employed persons, indicating whether they are working in the occupation they were trained for. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

ausb$$ – Required Training For Job [generic]

1	[1] No Training	64
2	[2] Intro. To Job	2565
3	[3] On-The-Job Training	940
4	[4] Courses	562
5	[5] Vocational Training	7961
6	[6] Technical School, Engineering (East) 90-96	0
7	[7] Technical College, University until 1998	0
8	[8] Technical College since 1999	1518
9	[9] University since 1999	2260
-1	[-1] No Answer	1351
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0

-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to provide annual data on required job training for all employed persons. The variable is generated using questions on required formal education and required on-the-job-training which are categorized into seven independent variables with 0/1 coding. Out of these, the highest available level of required training is used for the generation of the status variable.
The answer option "completed technical school" was only used from 1990 to 1993 in the East German version of the questionnaire. Since not all employed people are asked the question about required training every year, the value (6) of the variable AUSB$$ is valid up to 1996.
The code (-2) is assigned to all non-employed persons and also includes persons in occupational training, in occupational retraining programs, and those doing an internship at the time of the survey. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

$erwzeit – Length Of Time With Firm [generic]

0		205
0.100000001490116		264
0.200000002980232		238
0.300000011920929		378
0.400000005960464		230
0.5		220
0.600000023841858		206
0.699999988079071		185
0.800000011920929		371
0.899999976158142		171
1		166
1.10000002384186		175
1.20000004768372		163
1.29999995231628		248
1.39999997615814		168
...	(444 rows omitted)	13474
50.7999992370605		1
51.0999984741211		1
51.2000007629395		2
51.7999992370605		1
52.2999992370605		1
52.7999992370605		1
53		1
53.7999992370605		1
54.5999984741211		1
56.7999992370605		1
61.5999984741211		1
65.3000030517578		1
-1		231

-2	10932
-3	4

Waves: all
The variable $ERWZEIT is designed to offer data on the length of time with the firm at the point in time of the interview for all employed persons. This variable is generated from the respondent's start date with the current employer. In the case of a job change within the firm, the full length of time with the firm is calculated. Hence, the variable describes the length of time with the same firm and not the length of time in the same position.
The variable provides consistent longitudinal information on the length of time with the same employer. Data that show longitudinal inconsistencies are corrected.

1. In case of no job change, the information on the start date with the current employer given in the earliest interview available is treated as dominant and carried forward to the subsequent years.

2. In case of a job change between firms, the information on the start of the current position is used and carried forward to the subsequent years.

3. Up to wave Z (2009), a respondent who starts working again after a period of non-employment is assumed to have returned to the former employer if the indicated start date with the current employer was before the previous interview date. In this case, the start date with the current employer given in the previous interview is treated as dominant. Otherwise, the present information on the start date with the current employer is used and carried forward to the subsequent years. For respondents who are assumed to have returned to their former employer, the full length of time with the firm is calculated. There is no deduction for the time during which the respondent was not employed.

4. Since wave BA (2010), there is a modified answer category in the questionnaire which indicates that a respondent returns to his/her former employer after a period of non-employment. If a respondent indicates to have started working again at a former employer, the present information on the start date with the current employer is used and carried forward to the subsequent years. Unlike before wave BA (2010), the present information is treated as dominant even if the indicated start date with the current employer was before the previous interview date. Hence, the full length of time with the firm is calculated, and there is no deduction for the time during which the respondent was not employed or employed in another firm.

5. The length of time with the firm is also provided for the East German sample since its start in 1990. Due to the massive restructuring of the economy that took place in East Germany after reunification, this variable should be dealt with cautiously in the first transition years.

Both monthly and annual information is used in the variables and rounded off as length of time in years (with months in decimal form). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

$tatzeit – Actual Work Time Per Week [generic]

1	16

1.29999995231628		1
1.5		3
2		49
2.5		5
3		70
3.5		3
4		78
4.19999980926514		1
4.5		4
5		109
5.5		5
6		116
6.30000019073486		1
6.5		2
...	(144 rows omitted)	15978
65.5		1
66		4
67		2
68		4
69		1
70		85
72		12
75		12
76		1
77		1
78		1
80		32
-1		607
-2		10821
-3		17

Waves: all
This variable is designed to offer annual data on actual weekly working hours (including overtime) for all persons employed at the time of the survey (including the self-employed). The data are obtained by asking respondents how many hours they work on average per week. Actual weekly working hours were asked up to 1989 only in full hours, and from 1990 on as a three-digit number (counting the first digit after the decimal point).
For implausible answers (actual weekly working hours of more than 80 per week), we assign the value (-3). The variable is rounded off and gives the number of working hours as a decimal number.
Please also see $VEBZEIT and $UEBSTD. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

$vebzeit – Agreed Upon Work Time Per Week [generic]

1	3
1.5	2
2	27
2.5	1

3		41
3.5		2
4		52
4.19999980926514		1
4.5		5
5		58
5.5		3
5.80000019073486		1
6		82
6.30000019073486		1
6.5		7
...	(154 rows omitted)	13726
49.5		1
50		52
51		1
52		6
53		2
54		7
55		12
56		13
60		15
65		2
70		2
72		1
-1		397
-2		13517
-3		2

Waves: all
This variable is designed to offer annual data on agreed weekly working hours. The variable takes into account only those persons who were in dependent employment (not self-employed) at the time of the survey. Agreed weekly working hours were asked up to 1989 only in full hours, and from 1990 on in three-digit form (counting the first digit after the decimal point).
The value (-2) is assigned to non-employed people, employees without set hours and to self-employed people, including self-employed farmers, freelancers, and other self-employed persons. If persons helping out in family businesses report agreed weekly working hours, we assign a non-missing value.
For implausible answers (agreed weekly working time of more than 80 hours per week) we assign the value (-3). The variable is rounded off and gives the number of working hours as a decimal number.
Please also see $TATZEIT and $UEBSTD. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

$uebstd – Overtime Per Week [generic]

0	7678
0.200000002980232	94
0.5	295

0.699999988079071		254
0.899999976158142		362
1.20000004768372		531
1.39999997615814		321
1.60000002384186		124
1.89999997615814		541
2.09999990463257		74
2.29999995231628		1068
2.59999990463257		43
2.79999995231628		341
3		31
3.29999995231628		74
...	(49 rows omitted)	2939
16.1000003814697		1
16.3999996185303		14
16.7999992370605		1
17.5		6
17.7999992370605		1
18.7000007629395		66
19.6000003814697		2
20.1000003814697		1
20.6000003814697		2
21		16
22.3999996185303		1
23.1000003814697		31
-1		671
-2		12458
-3		1

Waves: all
This variable is designed to offer annual data on overtime per week for all persons in dependent employment at the time of the survey. The data is obtained by asking respondents how many overtime hours they worked in the month before the survey. The number of monthly overtime hours is then converted into weekly overtime by dividing the number given by 4.3. Since $UEBSTD refers to weekly overtime during the last month, the number may deviate from the difference between average actual weekly working hours and the agreed weekly working hours.
In the years 1984, 1985 and 1987, respondents were not asked about number of hours of overtime per week. The variables for 1984 and 1985 were therefore generated using the difference between average actual weekly working hours and agreed weekly working hours. It is not possible to create this kind of variable for the year 1987, since all values here were "missings". Respondents were asked for the number of overtime hours up to 1989 only in full hours, and from 1990 on, as a three-digit number (counting the first digit after the decimal point).
The value (-2) is assigned to non-employed people, employees without set hours and to self-employed people, including self-employed farmers, freelancers, and other self-employed persons. If persons helping out in family businesses report overtime hours, we assign a non-missing value. For implausible answers (agreed-upon weekly working time or actual weekly working time of more than 80 hours per week AND weekly overtime of more than 10 hours we assign the value (-3).
The variable is rounded off and gives the number of overtime hours as a decimal number.

Please also see $VEBZEIT and $TATZEIT. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

oeffd$$ – Civil Service [generic]

1	[1] Yes	3792
2	[2] No	11761
-1	[-1] No Answer	1668
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
Status variable: This variable is designed to provide annual data on employment in the civil service for all employed persons. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

nace$$ – 2 Digit NACE Industry, Sector [generic]

1	[1] Agriculture, Hunting, Related Service Activities	230
2	[2] Forestry, Logging, Related Service activities	27
5	[5] Fishing, Operation Of Fish Hatcheries And Fish Farms	3
10	[10] Mining Of Coal And Lignite; Extraction Of Peat	10
11	[11] Extraction Of Crude Petroleum And Natural Gas	6
12	[12] Mining Of Uranium And Thorium Ores	0
13	[13] Mining Of Metal Ores	0
14	[14] Other Mining And Quarrying	5
15	[15] Manuf Food Products And Beverages	354
16	[16] Manuf Tobacco Products	2
17	[17] Manuf Textiles	59
18	[18] Manuf Wearing Apparel; Dressing And Dyeing Of Fur	23
19	[19] Tanning,Dressing Of Leather; Manuf luggage, Footwear	8
20	[20] Manuf Wood Products, Except Furniture	58
21	[21] Manuf Pulp, Paper And Paper Products	63
...	(40 rows omitted)	13830
91	[91] Activities Of Membership Organizations NEC.	204
92	[92] Recreational, Cultural And Sporting Activities	307
93	[93] Other Service Activities	149
95	[95] Private Households With Employed Persons	90
96	[96] Industry - NEC	91
97	[97] Handcraft, Trade - NEC	63
98	[98] Services - NEC	172
99	[99] Extra-territorial Organizations And Bodies	6
100	[100] Manufacturing - NEC	49
-1	[-1] No Answer	1412

-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to provide annual data on the industry of economic activity for all employed persons according to the Statistical Classification of Economic Activities in the European Community (Nomenclature des statistiques des activités économiques de la Communauté européenne - NACE). Respondents answer the question in their own words regarding the industry in which they are currently working, and this response is entered into a blank in the questionnaire. In order to facilitate international comparability, the European industry standard classification system is used by Infratest Sozialforschung to recode this information. This recoding has been documented in Hartmann/Schütz 2002. The codes in NACE Rev.1 also correspond to ISIC Rev.3 (International Standard Classification of All Economic Activities). With the 2001 data distribution, the sector codes formerly used in the SOEP were completely recoded to the NACE classification. Please note that special codes 96–98 as well as 100 were assigned by Infratest Sozialforschung whenever respondents did not provide a more detailed answer. [This information can be related to a specific variable and is not necessary generic.]
Detailed description: Hartmann/Schütz (2002): Die Klassifikation der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984–2001. Infratest Sozialforschung, München. https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.40132.de/verco
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

betr$$ – Size Of The Company [generic]

1	[1] LT 5	1504
2	[2] GE 5 LT 10	1506
3	[3] GE 11 LT 20	1365
4	[4] Until 90: LT 20	0
5	[5] 91-04: GE 5 LT 20	0
6	[6] GE 20 LT 100	2941
7	[7] GE 100 LT 200	1408
8	[8] Until 98: GE 20 LT 200	0
9	[9] GE 200 LT 2000	3275
10	[10] GE 2000	3924
11	[11] Self-Employed Without Coworkers	851
-1	[-1] No Answer	447
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to offer annual data on company size for all employed persons. Please pay attention to special codes 4, 5, and 8! These codes were necessary due to the differentiation of items for small and medium-sized companies over the years. In the years

1991, 1999, and 2005, respondents were asked about company size in a more detailed form, so when the data were recalculated for the years 1984–90, the original codes (1)–(5) were changed to (4)(8)(9)(10) and (11). For the years 1999 to 2004, Code (8) was differentiated into (6) and (7). From 2005 on, the prior category "5 to 20 employees" (5) has been split into the two categories "5 to 10 employees" (2) and "11 to 20 employees" (3).
In 2012, the questionnaire provides a one-time-only information on the size of the local establishment in addition to the size of the entire company (BETR$$). The enriched questionnaire revealed that in previous interviews, persons have mistakenly provided information on the local establishment size instead of the entire company size, especially if their entire company had 2000 and more employees. Due to the importance of longitudinal consistency, these persons where identified, and their 2012 original value of the entire company size was replaced by their value of the local establishment size. Furthermore, data of persons without a job change were modified if:

- the entire company size varies from 2011 to 2012,
- the local establishment size in 2012 matches the entire company size in 2011.

Those person's values of the entire company in 2012 were replaced by their values of the local establishment size in 2012.
Please also see ALLBET$$ for a broader categorization of the firm size, which is appropriate for analyses that include all sample years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

allbet$$ – Core Category Size Of The Company [generic]

1	[1] LT 20	4375
2	[2] GE 20 LT 200	4349
3	[3] GE 200 LT 2000	3275
4	[4] GE 2000	3924
5	[5] Self-Employed Without Coworkers	851
-1	[-1] No Answer	447
-2	[-2] Does not apply	10821
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
This variable is designed to provide annual data on the core size category of the company for all employed persons. Since respondents were asked about company size in more detailed form in the years 1991, 1999, and 2005 (see also BETR$$), the variable ALLBET$$ contains the lowest common denominator of the variable BETR$$, i.e., the firm size categories available across all SOEP waves. This broader categorization corresponds to the values of variables BETR84 to BETR90 and offers a variable that is consistent across all waves.
In order to maintain longitudinal consistency, there were modifications after a one-time change in the questionnaire in 2012. Please also see BETR$$ for more information. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

8 Last Occupation

jobend$$ – Reasons for occupational change [generic]

1	[1] Terminated by employer	351
2	[2] Temporary contract expired	0
3	[3] Education, training completed	0
4	[4] Own resignation	614
5	[5] Mutual termination	169
6	[6] Employee requested transfer	0
7	[7] Company transferred employee	0
8	[8] Ended self-employment	72
9	[9] Temporary contract expired or education/training completed	362
10	[10] Took early retirement	0
11	[11] Company closed down	108
12	[12] Old-age pension	120
13	[13] Leave of absence/sabbatical (1999-2010)	0
14	[14] Leave, maternity leave and parental leave (1991-1998), since 2011	219
15	[15] Other incl. early retirement, company closed, old-age pension, leave of absence/sabbatical (1985-1986)	0
16	[16] Other incl. company closed, old-age pension,leave of absence/sabbatical (1987-1990)	0
17	[17] Other incl. mutual termination (1991-1998)	0
-1	[-1] No Answer	62
-2	[-2] Does not apply	25965
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
This variable is designed to offer annual data on reasons for an occupational change for all formerly employed persons, persons with a job change or persons on leave. For years 1985–1991 also persons who changed positions in the same company are considered. Only persons with valid dates for an occupational or positional change are included. Likewise to the questionnaire the variable offers data from interview date to interview date not from one year to the following. Respondents are asked about their annual and possibly same occupational change in two consecutive interviews, duplicate answers are therefore considered only once and the older statement is dominant. If a respondent stated a job termination in the current interview which was before the interview date in the previous year but didn't reported this in the previous interview this termination has been counted for the current interview. For years 1985 up to 1998 every given reason was coded as separate variable with variable values "Yes" (1) and "Does not apply" (-2), which resulted in up to 13 different variables. Since 1999 all given reasons have been collected in one single variable with diverse values.
Please pay attention to special codes (15), (16), and (17)! These codes were necessary due to the variety of the given values over the years. In any years respondents were asked about reasons for change with more or less given answers and from years 1985–1998 also the answer "Other" was possible. While all explicit reasons have been recoded to uniform values, the answer "Other" then in some years includes reasons for which in other years was separately asked for: "Other" was coded (15) for years 1985 and 1986, (16) for years

1987–1990 and (17) for years 1991–1998.
For years 1991–1998 and 2011–2012 there is a variable value "Leave, maternity leave and parental leave" whereas for 1999–2010 the given reason covered only "Leave of absence/sabbatical".
Note that codes (2) and (3) for years 1985–1998 have been merged to code (9) since 1999.
In 2013 the respondents of the newly introduced migration sample were not asked about an occupational change, but in 2014 they were. So information on jobend for the migration sample in 2013 were taken from the questionnaire in 2014. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

9 Employment History

expft$$ – Working Experience Full-Time Employment [generic]

0		3562
0.100000001490116		37
0.200000002980232		56
0.300000011920929		124
0.400000005960464		66
0.5		309
0.600000023841858		65
0.699999988079071		44
0.800000011920929		72
0.899999976158142		49
1		486
1.10000002384186		40
1.20000004768372		29
1.29999995231628		76
1.39999997615814		42
...	(483 rows omitted)	21843
50.4000015258789		1
50.7999992370605		1
51		1
51.2000007629395		2
52.5		1
53		3
53.5		1
53.5999984741211		1
54		2
54.5		1
54.7000007629395		1
57		1
61		1
61.2999992370605		1
-1		1124

Waves: all
This variable reflects the total length of full-time employment in the respondent's career up to the point of the interview in a given year. The variable is created by combining monthly information on employment status from the calendar dataset ARTKALEN

(which provides monthly information on activity status since an individual entered the SOEP) and annual information from the biographical dataset PBIOSPE (which provides information on activity status over the individual's life course). EXPFT$$ gives the length of time in years with months in decimal form.
If there is no monthly calendar data available in a given year of a respondent's career, the annual data from PBIOSPE is used for that year. In the most current wave the variable only uses up-to-date information from the newly answered Biography Questionnaires. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 0.5 years is assumed. Persons without annual data (not contained in PBIOSPE) are only assigned a non-missing value for this variable if they joined SOEP by the age of 18 and if there is calendar data on them in ARTKALEN.
Persons whose life course has been observed completely but with no spell of full-time employment are assigned the code (0). The code (-1) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (-3).
Please also see EXPPT$$ and EXPUE$$. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

exppt$$ – Working Experience Part-Time Employment [generic]

0		13092
0.100000001490116		136
0.200000002980232		157
0.300000011920929		302
0.400000005960464		163
0.5		595
0.600000023841858		176
0.699999988079071		140
0.800000011920929		211
0.899999976158142		106
1		1066
1.10000002384186		121
1.20000004768372		114
1.29999995231628		201
1.39999997615814		132
...	(347 rows omitted)	10189
42.4000015258789		1
42.7999992370605		1
42.9000015258789		1
43		1
43.4000015258789		1
43.7000007629395		1
44.2000007629395		1
44.5		2
45		1
45.7999992370605		1
46.0999984741211		2
47		2
47.7000007629395		1

48	1
-1	1124

Waves: all
This variable reflects the total length of part-time employment in the respondent's career up to the point of the interview in a given year. The variable is created by combining monthly information on employment status from the calendar dataset ARTKALEN (which provides monthly information on activity status since an individual entered the SOEP) and annual information from the biographical dataset PBIOSPE (which provides information on activity status over the life course of an individual). EXPPT$$ gives the length of time in years with months in decimal form.
If there is no monthly calendar data available in a given year of a respondent's career, the annual data from PBIOSPE is used for that year. In the most current wave the variable only uses up-to-date information from the newly answered Biography Questionnaires. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 0.5 years is assumed. Persons without annual data (not contained in PBIOSPE) are only assigned a non-missing value for this variable if they joined SOEP by the age of 18 and if there is calendar data on them in ARTKALEN.
Persons whose life course has been observed completely but with no spell of full-time employment are assigned the code (0). The code (-1) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (-3).
Please also see EXPFT$$ and EXPUE$$. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

expue$$ – Unemployment Experience [generic]

0		17261
0.100000001490116		354
0.200000002980232		310
0.300000011920929		441
0.400000005960464		189
0.5		1270
0.600000023841858		193
0.699999988079071		182
0.800000011920929		264
0.899999976158142		151
1		984
1.10000002384186		158
1.20000004768372		133
1.29999995231628		174
1.39999997615814		103
...	(192 rows omitted)	4725
22.3999996185303		2
22.7000007629395		1
23		7
23.2999992370605		1
23.8999996185303		1
24		3

24.2999992370605	3
24.8999996185303	1
26.1000003814697	1
27	2
28	1
29.2999992370605	1
32	1
37	1
-1	1124

Waves: all
This variable reflects the total length of unemployment in the respondent's career up to the point of the interview in a given year. The variable is created by combining monthly information on employment status from the calendar dataset ARTKALEN (which provides monthly information on activity status since an individual entered the SOEP) and annual information from the biographical dataset PBIOSPE (which provides information on activity status over the life course of an individual). EXPUE$$ gives the length of time in years with months in decimal form.
If there is no monthly calendar data available on a given year in a respondent's career, the annual data from PBIOSPE is used for that year. In the most current wave the variable only uses up-to-date information from the newly answered Biography Questionnaires. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 0.5 years is assumed. Persons without annual data (not contained in PBIOSPE) are only assigned a non-missing value for this variable if they joined SOEP by the age of 18 and if there is calendar data on them in ARTKALEN.
Persons whose life course has been observed completely but with no spell of full-time employment are assigned the code (0). The code (-1) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (-3).
Please also see EXPFT$$ and EXPPT$$. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Alexandra Fedorets (Tel. +49-30-89789-321)

10 School, Higher and Vocational Education

isced97_$$ – ISCED-1997-Classification [generic]

0	[0] in school	883
1	[1] inadequately	582
2	[2] general elementary	3731
3	[3] middle vocational	12269
4	[4] vocational + Abi	2052
5	[5] higher vocational	1584
6	[6] higher education	6206
-1	[-1] No Answer	735
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
The educational variable ($ISCED97) classifies all correspondents' educational degrees according to the "International Standard Classification of Education (ISCED)" of 1997 in order to make degrees internationally comparable. The variable is generated retrospectively from 1984 onwards taking into account degrees and diplomas attained in both general schooling and in vocational/university education and indicates the highest degree obtained. E.g., persons who did not indicate secondary school degrees/diplomas but a university degree are placed in the highest ISCED category. Please note that, due to a lack of more detailed information on tertiary degrees in earlier waves – in particular on PhD – we include all tertiary degrees in ISCED category 6. Thus, the ISCED variable provided here is not comparable one-to-one with the ISCED levels as defined by the OECD, since we have included the original ISCED level 5A in our ISCED category 6. OECD (1999): Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries, Paris. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

isced11_$$ – ISCED-2011.Classification [generic]

0	[0] in school	883
1	[1] Primary education	579
2	[2] Lower secondary education	3456
3	[3] Upper secondary education	12648
4	[4] Post-secondary non-tertiary education	2187
5	[5] Short-cycle tertiary education	529
6	[6] Bachelors or equivalent level	4995
7	[7] Masters or equivalent level	1807
8	[8] Doctoral or equivalent level	225
-1	[-1] No Answer	733
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: 2010–
The educational variable ($ISCED11) classifies all correspondents' educational degrees according to the "International Standard Classification of Education (ISCED)" of 2011 in order to make degrees internationally comparable. The variable is generated retrospectively From 2010 onwards taking into account degrees and diplomas attained in both general schooling and in vocational/university education and indicates the highest degree obtained.
[This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

casmin$$ – CASMIN Classification [generic]

0	[0] (0) in school	893
1	[1] (1a) inadequately completed	579
2	[2] (1b) general elementary school	2484

3	[3] (1c) basic vocational qualification	5899
4	[4] (2b) intermediate general qualification	1415
5	[5] (2a) intermediate vocational	6564
6	[6] (2c_gen) general maturity certificate	1072
7	[7] (2c_voc) vocational maturity certificat	1976
8	[8] (3a) lower tertiary education	1782
9	[9] (3b) higher tertiary education	4424
-1	[-1] No Answer	954
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
Another internationally comparable educational variable is $CASMIN where educational degrees/diplomas are classified according to the scheme "Comparative Analysis of Social Mobility in Industrial Nations (CASMIN)". As for $ISCED, the variable is generated for all respondents retroactively from 1984 onwards and indicates the highest degree obtained by the respondent. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

$bilzeit – Amount Of Education Or Training In Years [generic]

7	494
8.5	76
9	2546
10	1481
10.5	4179
11	1460
11.5	4363
12	2646
13	1433
13.5	435
14	483
14.5	693
15	1864
16	848
17	102
18	2982
-1	1064
-2	893

Waves: all
The following statements describe the standard computation for schooling (including years of secondary vocational education). As can be seen, the code is not very differentiated. For example, special schools for health care professions and other kinds of specialized schools are all included in the "technical school" label. However, in Germany, this code is the most commonly used one when earnings functions based on human capital

theory are estimated.
$BILZEIT is now computed for all samples. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$psbil – School-Leaving Degree [generic]

1	[1] Secondary School Degree	6919
2	[2] Intermediate School Degree	7636
3	[3] Technical School Degree	1580
4	[4] Upper Secondary School Degree	5885
5	[5] Other Degree	3434
6	[6] Dropout, No School Degree	631
7	[7] No School Degree Yet	893
-1	[-1] No Answer	1064
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
All respondents in all SOEP subsamples are asked about diplomas/degrees attained for completion of secondary/tertiary education (1984–1993 blue questionnaire; since 1994 biographical questionnaire) the first time they participate in SOEP. First: to generate this variable, the different diploma/degree categories provided for Subsamples B and D (see $PSBILA) as well as C (see $PSBILO) are integrated into the West German diploma/degree categories (Subsample A) and continued on in this form. Second: this data is regularly updated to take into account any changes in highest diploma/degree attained. With the survey of 2000, all educational information was collected again and is reflected in the variables. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$pbbil01 – Vocational Degree Received [generic]

1	[1] Apprenticeship	10499
2	[2] Vocational School	2380
3	[3] Health Care School	205
4	[4] Technical School	1571
5	[5] Civil Service Training	547
6	[6] Other Degree	1303
7	[7] Completed Vocational Training/Education in Germany	273
-1	[-1] No Answer	386
-2	[-2] Does not apply	10878
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

All respondents in all subsamples are asked about vocational degrees attained the first time they participate in SOEP (1984–1993 blue questionnaire; since 1994 biographical questionnaire). To generate the variable, the different vocational degrees for Subsamples B and D (cf. $PBBILA) as well as C (cf. $PBBILO) are integrated into the West German vocational degree categories (Subsample A). The categories that originally each constituted individual variables are combined to make them compatible with the annual question about changes in vocational degrees attained, and this data is updated annually. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$pbbil02 – College Degree [generic]

1	[1] Technical College	1832
2	[2] University, Technical College	3080
3	[3] College Not In Germany	677
4	[4] Engineering, Techncial School (East)	231
5	[5] University (East)	186
6	[6] graduation, state doctorate	198
7	[7] graduation, state doctorate (foreign country, east)	27
8	[8] institution of higher education (youth)	0
9	[9] Dual Studies, University of cooperative education	15
10	[10] Other Colleges	6
-1	[-1] No Answer	386
-2	[-2] Does not apply	21404
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
All respondents in all subsamples are asked about completed college education the first time they participate in SOEP (1984–1993 blue questionnaire; since 1994 biographical questionnaire). To generate the variable, the different degrees/dimplomas for all subsamples are integrated. Category (3) “college abroad” is only defined for persons who completed a foreign-language version of the questionnaire (mainly persons from Samples B and D). Generation of the variable entails combining the categories to make them compatible with the annual question about changes in vocational degrees/diplomas attained. Since 2002, there have been two separate codes (4 and 5) for degrees/diplomas attained in the former GDR. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$pbbil03 – No Vocational Degree [generic]

1	[1] No Vocation Degree	5280
2	[2] Apprenticeship	1028
3	[3] currently: studying	1041
-1	[-1] No Answer	386
-2	[-2] Does not apply	20307

-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
In connection with the question about vocational degrees ($PBBIL01 and $PBBIL02), all first-time respondents to all subsamples are explicitly asked whether they (still) do not possess a vocational degree. In the subsequent years, this data is carried forward or updated. The variable has the Missing Value Code -2 (does not apply) if one of the other two variables on vocational degree has a positive value. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$psbilo – School-Leaving Degree East Germany [generic]

1	[1] 8th Grade Completed	1074
2	[2] 10th Grade Completed	2237
3	[3] College Entrance Exam	817
4	[4] Other Degree	36
5	[5] Dropout, No School Degree	25
6	[6] No School Degree Yet	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	23853
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
As a supplement to the variable $PSBIL the highest secondary school degree/diploma in East Germany is provided as a separate variable and updated if necessary for 1991. Since 1992, secondary degrees/diplomas are asked only in the West German version. New SOEP respondents are also asked about secondary degrees/diplomas obtained in the former GDR; and for old respondents, the same codes are carried forward. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$pbbilo – Vocational Degree Received East Germany [generic]

1	[1] Vocational Training	15
2	[2] Master Craftsman	0
3	[3] Engineering, Technical Degree	0
4	[4] Other Degree	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	28027
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
To supplement the variable $PBBIL01 the highest secondary school degree/diploma in East Germany is provided as a separate variable and updated if necessary for 1991. Since 1992 only the West German version has been used for new vocational degrees. For new SOEP respondents, vocational degrees attained in the former GDR are asked as well; for old respondents, the same codes are carried forward. From 2002 on, the questionnaire was expanded and revised, but this led to an operationalization involving more assumptions on the vocational degrees attained in the GDR; (from 2002 on, Code 3 is also listed as the additional category Code 4 in the integrated variables $PBBIL03 if this degree has not been replaced by a more recently attained, higher-level university or college degree). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$psbila – School-Leaving Degree Outside Germany [generic]

1	[1] School, No Degree	316
2	[2] School, With Degree	1527
3	[3] Vocational Extension School	1769
4	[4] School-Leaving Degree[sbil] Outside Germany	1
-1	[-1] No Answer	3
-2	[-2] Does not apply	24426
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
As a supplement to the $PSBIL, this variable provides annually updated data on the highest secondary school degree/diploma attained abroad. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

$pbbila – Vocational Degree Outside Germany [generic]

1	[1] On-The-Job Training	102
2	[2] Vocational Training	487
3	[3] Vocational School	621
4	[4] College	966
5	[5] Other	93
6	[6] Vocational Degree[bbil01] Acquired Abroad	8
7	[7] College Education[bbil02] Acquired Abroad	15
8	[8] Completed Vocational Training/Education Other Country	4
9	[9] graduation, state doctorate (foreign country)	27
-1	[-1] No Answer	0
-2	[-2] Does not apply	25719
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
As a supplement to the variable $PBBIL01, this variable gives (and updates) the highest-level vocational degree attained abroad. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

field$$ – Field of tertiary education [generic]

1	[1] Applied Linguistics and Cultural Studies	36
2	[2] Protestant Theology	55
3	[3] Catholic Theology	1
4	[4] Philosophy	15
5	[5] History	30
6	[6] Library Science, Archival Studies, Journalism	13
7	[7] Literary Studies, Linguistics	49
8	[8] Classical Philology, Modern Greek	6
9	[9] German Philology	102
10	[10] English Studies	38
11	[11] Roman Studies	20
12	[12] Slavonic Studies	9
13	[13] Non-European Languages and Cultural Studies	11
14	[14] Cultural Studies	9
15	[15] Psychology	63
...	(37 rows omitted)	3268
68	[68] Construction Engineering	123
69	[69] Civil Engineering	12
74	[74] Art, Aesthetics	26
75	[75] Fine Arts	7
76	[76] Design	45
77	[77] Performance, Film and Television, Theater	15
78	[78] Music, Musicology	50
83	[83] Outside the structure of the university system	16
98	[98] Not categorizable	110
-1	[-1] No Answer	364
-2	[-2] Does Not Apply	23549
-3	[-3] Not Valid	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
The variable is designed to provide information on the field of education of tertiary degrees which adds details to the information recorded in the variable $PBBIL02. While the latter variable records if a person holds a degree FIELD$$ contains more detailed information on the type of the degree. The data of the generated variable FIELD$$ stem from two sources:

1. Person questionnaire: Each year since 1985 respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type

and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables FIELD$$.

2. Biography questionnaire: Since 2001 similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained degrees but to degrees obtained during the time before being part of the SOEP sample.

In the variable FIELD$$ we combine these two types of information. However, since the retrospective information was not collected before 2001 the variable covers until 2000 only persons for whom we have prospectively observed the end of study. This explains why the number of valid observations is rather small in these years. Information on the data source is stored in the variable FDT_F$$.
Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed a first degree in sociology in 1987 and a second degree in economics in 1991. For this person the variables FIELD$$ would be filled as follows:

year	value
1984–1986	-2 does not apply
1987–1990	26 political/social science
1991–today	30 economics

If you want to take into account that a person holds two degrees you have to combine the information from all available years. However, only a minority of the population holds more than one tertiary degree. In very few cases we encounter the problem that a respondent provides information on two different degrees in one survey year. This only happens in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information. Furthermore, there are cases who report an applied university degree and a university degree in the biography questionnaire. In these cases, the variable contains information on the university degree only. The variable is coded according to the classification on fields of education („Fächergruppen“) provided by the Statistisches Bundesamt (2009). Until 2009 data from the person questionnaire were coded using an earlier version of this classification (1982). In the variable FIELD$$ we recoded the original values. As the newer version is more precise this could be done with hardly any loss of information. Some categories are collapsed. Category 3 is coded as 2 (no distinction between catholic and protestant theology), 14 as 13, 17 as 16, 24 as 23, 25 as 26 and 48 as 49. The original values of the data collected from the person questionnaire up to 2009 are stored in the respective variables in the dataset $P.
Please note that for respondents from the newly introduced migration sample a valid value to FIELD$$ was only assigned if the tertiary degree was attained in Germany. [This information can be related to a specific variable and is not necessary generic.]
Statistisches Bundesamt (2009): Bildung und Kultur. Studierende an Hochschulen, Fachserie 11, Reihe 4.1, Wiesbaden: 446ff, Übersicht 1: "Fächergruppen, Studienbereiche und Studienfächer".
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

degree$$ – Type of tertiary degree [generic]

11	[11] Masters Degree	118
12	[12] Diplom (University)	896
13	[13] Bachelor	108
14	[14] Master	82
15	[15] 1st State Examination	117
16	[16] Other state examination	130
21	[21] Diplom (at technical college, technical college for administration)	630
22	[22] Bachelor (at technical college, technical college for administration)	77
23	[23] Master (at technical college, technical college for administration)	18
30		1
31	[31] Teacher training,BA,MA at elementary, lower secondary schools/primary level	118
32	[32] Teacher training,BA,MA at 2ndary level 1/elementary schools/primary level	8
33	[33] Teacher training,BA,MA at intermediate scndry schools/scndry level I	50
34	[34] Teacher training, BA, MA at secondary level II and I	3
35	[35] Teacher training,BA,MA at academic 2ndry schools,2ndry levl 2,genrl school	65
36	[36] Teacher training, BA, MA at special needs schools	31
37	[37] Teacher training, BA, MA at vocational schools	13
38	[38] Teacher training, other	316
41	[41] Academic degree in the arts	17
42	[42] Doctorate	61
43	[43] Post-doctoral dissertation (Habilitation)	3
44	[44] Other Degree	120
98	[98] Not categorizable	97
-1	[-1] No Answer	1084
-2	[-2] Does Not Apply	23879
-3	[-3] Not Valid	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985

The variable is designed to provide information on the type of tertiary degree (e.g., Diploma, Bachelor, Master) which adds details to the information recorded in the variable $PBBIL02. While the latter variable records if a persons holds a degree, DEGREE$$ contains more detailed information on the type of the degree. The data of the generated variable DEGREE$$ stem from two sources:

1. Person questionnaire: Each year since 1985 respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables DEGREE$$.
2. Biography questionnaire: Since 2001 similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained degrees but to degrees obtained during the time before being part of the SOEP sample.

In the variable DEGREE$$ we combine these two types of information. However, since the retrospective information was not collected before 2001 the variable covers until 2000

only persons for whom we have prospectively observed the end of study. This explains why the number of valid observations is rather small in these years. Information on the data source is stored in the variable FDT_F$$.
Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed first an applied university diploma in 1987 and a university diploma in 1991. For this person the variables DEGREE$$ would be filled as follows:

year	value
1984–1986	-2 does not apply
1987–1990	21 diploma (applied university)
1991–today	12 diploma (university)

If you want to take into account that a person holds two degrees you have to combine the information from all available years. However, only a minority of the population holds more than one tertiary degree. In very few cases we encounter the problem that a respondent provides information on two different degrees in one survey year. This only happens in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information. Furthermore, there are cases who report an applied university degree and a university degree in the biography questionnaire. In these cases, the variables contain information on the university degree only.
The variable is coded according to a slightly collapsed version of the classification on types of tertiary degrees ("Prüfungsgruppen und Abschlussprüfungen") provided by the Statistisches Bundesamt (2009). Until 2009, data from the person questionnaire were coded using an earlier version of this classification (1982) which was slightly revised in 2009 (inclusion of Bachelor and Master degrees). Since 2010 the data were coded according to the classification presented here. In the variable DEGREE$$ we recoded the original values from years 2009 and earlier. As the newer version is more precise this could be done with hardly any loss of information. Some categories are collapsed. Category 16 was mostly likely coded as 15 in earlier years, 34 as 35 and 43 as 44. The original values of the data collected from the person questionnaire up to 2009 are stored in the respective variables in the dataset $P.
Please note that for respondents from the newly introduced migration sample a valid value to DEGREE$$ was only assigned if the tertiary degree was attained in Germany. [This information can be related to a specific variable and is not necessary generic.]
Statistiches Bundesamt (2009): Bildung und Kultur. Studierende an Hochschulen, Fachserie 11, Reihe 4.1, Wiesbaden: 449ff, Übersicht 2: „Prüfungsgruppen und Abschlussprüfungen".
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

traina$$ – Apprenticeship - two-digit occupation KldB92 [generic]

1	[1] Agricultural Occupations (Crops)	121
2	[2] Agricultural Occupations (Livestock)	41
3	[3] Administrative/Advisory/Technical Specialist In Agriculture	1
5	[5] Horticultural Occupations	160
6	[6] Forestry and Hunting Occupations	17
7	[7] Mineworkers	20

8	[8] Mineral Exploitation and Processing	2
10	[10] Stonemasons	10
11	[11] Manufacturers of Construction Materials	4
12	[12] Ceramicists	10
13	[13] Glass Manufacturing Occupations	20
14	[14] Chemical Industry Occupations	36
15	[15] Plastics Manufacturing Occupations	19
16	[16] Paper Manufacturing and Processing	9
17	[17] Printing Occupations	97
...	(65 rows omitted)	8904
89	[89] Pastoral Occupations	1
90	[90] Personal Care Occupations	252
91	[91] Occupations in Hotels and Hospitality	168
92	[92] Occupations in Domestic and Nutritional Science	123
93	[93] Cleaning and Waste Management Occupations	32
96	[96] Others	27
97	[97] Family members providing assistance,not in agriculture,not otherw. mntnd	0
98	[98] Workers, (still) without specific occupation	0
99	[99] Workers, responsibilities not specified	10
-1	[-1] No Answer	43
-2	[-2] Does not apply	17915
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
The variable is designed to provide information on the occupation of vocational training which adds details to the information recorded in the variable $PBBIL01. In addition to the variable TRAINA$$, we provide the variables TRAINB$$, TRAINC$$ and TRAIND$$. All these variables record the occupation of vocational training. The difference is that TRAINA$$ contains information on vocational training within the German dual system which combines firm-based and school-based training (apprenticeship). TRAINB$$ is designed to provide information on the occupation of full-time school based vocational training. TRAINC$$ contains information on level vocational training (e.g., Meister, Techniker). TRAIND$$ is designed to provide information on the occupation of civil servant training ("Beamtenausbildung"). We describe in brief detail the construction of the variable TRAINA$$. TRAINB$$, TRAINC$$ and TRAIND$$ are constructed in an analogous manner.
The data of the generated variable TRAINA$$ stem from two sources:

1. Person questionnaire: Each year since 1985 respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables TRAINA$$.

2. Biography questionnaire: Since 2001 similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained vocational qualifications but to qualifications obtained during the time before being part of the SOEP sample.

In the variable TRAINA$$ we combine these two types of information. However, since the retrospective information was not collected before 2001 the variable covers until 2000 only persons for whom we have prospectively observed the end of study. This explains why the number of valid observations is rather small in these years. Information on the data source is stored in the variable FDT_F$$.
Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed a first vocational qualification as an electrician in 1987 and a second qualification as a car mechanic in 1991. For this person the variables TRAINA$$ would be filled as follows:

year	value
1984–1986	-2 does not apply
1987–1990	31 electrical occupation
1991–today	28 automotive/flight industry occupation

If you want to take into account that a person holds two vocational qualifications you have to combine the information from all available years. In few cases we encounter the problem that a respondent provides information on two different apprenticeships in one survey year. This only happens once, namely in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information.
The variable is coded according to the classification of occupations at two-digit level („Berufsgruppen“) provided by the Statistisches Bundesamt (1992). Other SOEP occupation variables are coded at four-digit level. The reason why the variable TRAINA$$ is provided at two-digit level only is that until 2009 the data from the two different sources were coded according two different classifications which could be combined at a higher level of aggregation only. The person questionnaire data were coded according to the classification of occupations provided by the Bundesanstalt für Arbeit (1988, four-digit level) while the biography data use the classification provided by the Statistisches Bundesamt (1992, four-digit level). Since 2010 both types of data are coded according to the latter classification. The four-digit version of the biography data can be provided upon request. The original values of the data collected from the person questionnaire up to 2009 are stored in the respective variables in the dataset $P.
Please note that for respondents from the newly introduced migration sample a valid value to TRAINA$$ was only assigned if the vocational training was completed in Germany. [This information can be related to a specific variable and is not necessary generic.]
Hartmann/Schütz (2002): Die Klassifikation der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984–2001. Infratest Sozialforschung, München.
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

trainb$$ – Vocational school - two-digit occupation KldB92 [generic]

1	[1] Agricultural Occupations (Crops)	2
2	[2] Agricultural Occupations (Livestock)	3
3	[3] Administrative/Advisory/Technical Specialist In Agriculture	1
5	[5] Horticultural Occupations	3
6	[6] Forestry and Hunting Occupations	0
7	[7] Mineworkers	0

8	[8] Mineral Exploitation and Processing	0
10	[10] Stonemasons	1
11	[11] Manufacturers of Construction Materials	0
12	[12] Ceramicists	0
13	[13] Glass Manufacturing Occupations	1
14	[14] Chemical Industry Occupations	0
15	[15] Plastics Manufacturing Occupations	0
16	[16] Paper Manufacturing and Processing	2
17	[17] Printing Occupations	10
...	(65 rows omitted)	1595
89	[89] Pastoral Occupations	4
90	[90] Personal Care Occupations	21
91	[91] Occupations in Hotels and Hospitality	17
92	[92] Occupations in Domestic and Nutritional Science	70
93	[93] Cleaning and Waste Management Occupations	2
96	[96] Others	48
97	[97] Family members providing assistance,not in agriculture,not otherw. mntnd	0
98	[98] Workers, (still) without specific occupation	0
99	[99] Workers, responsibilities not specified	3
-1	[-1] No Answer	30
-2	[-2] Does not apply	26229
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
The variable is designed to provide information on the occupation of full-time school based vocational training (e.g., Berufsfachschule, Schule des Gesundheitswesens, Handelsschule). See the description of variable TRAINA$$ for more details on the construction and the values of the variable. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

trainc$$ – Higher vocational school - two-digit occupation KldB92 [generic]

1	[1] Agricultural Occupations (Crops)	26
2	[2] Agricultural Occupations (Livestock)	4
3	[3] Administrative/Advisory/Technical Specialist In Agriculture	11
5	[5] Horticultural Occupations	11
6	[6] Forestry and Hunting Occupations	3
7	[7] Mineworkers	0
8	[8] Mineral Exploitation and Processing	0
10	[10] Stonemasons	3
11	[11] Manufacturers of Construction Materials	0
12	[12] Ceramicists	2
13	[13] Glass Manufacturing Occupations	0
14	[14] Chemical Industry Occupations	3
15	[15] Plastics Manufacturing Occupations	0
16	[16] Paper Manufacturing and Processing	0

17	[17] Printing Occupations	3
...	(65 rows omitted)	955
89	[89] Pastoral Occupations	4
90	[90] Personal Care Occupations	22
91	[91] Occupations in Hotels and Hospitality	9
92	[92] Occupations in Domestic and Nutritional Science	20
93	[93] Cleaning and Waste Management Occupations	6
96	[96] Others	46
97	[97] Family members providing assistance,not in agriculture,not otherw. mntnd	0
98	[98] Workers, (still) without specific occupation	0
99	[99] Workers, responsibilities not specified	1
-1	[-1] No Answer	17
-2	[-2] Does not apply	26896
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
The variable is designed to provide information on the occupation of higher level vocational training (e.g., Meister, Techniker). See the description of variable TRAINA$$ for more details on the construction and the values of the variable. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

traind$$ – Civil servant training - two-digit occupation KldB92 [generic]

1	[1] Agricultural Occupations (Crops)	0
2	[2] Agricultural Occupations (Livestock)	0
3	[3] Administrative/Advisory/Technical Specialist In Agriculture	0
5	[5] Horticultural Occupations	0
6	[6] Forestry and Hunting Occupations	2
7	[7] Mineworkers	0
8	[8] Mineral Exploitation and Processing	0
10	[10] Stonemasons	0
11	[11] Manufacturers of Construction Materials	0
12	[12] Ceramicists	0
13	[13] Glass Manufacturing Occupations	0
14	[14] Chemical Industry Occupations	0
15	[15] Plastics Manufacturing Occupations	0
16	[16] Paper Manufacturing and Processing	0
17	[17] Printing Occupations	0
...	(65 rows omitted)	353
89	[89] Pastoral Occupations	0
90	[90] Personal Care Occupations	0
91	[91] Occupations in Hotels and Hospitality	0
92	[92] Occupations in Domestic and Nutritional Science	0
93	[93] Cleaning and Waste Management Occupations	0
96	[96] Others	23
97	[97] Family members providing assistance,not in agriculture,not otherw. mntnd	0

98	[98] Workers, (still) without specific occupation	0
99	[99] Workers, responsibilities not specified	1
-1	[-1] No Answer	7
-2	[-2] Does not apply	27656
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
The variable is designed to provide information on the occupation of civil servant training ("Beamtenausbildung"). See the description of variable TRAINA$$ for more details on the construction and the values of the variable. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

fdt_f$$ – Data source FIELD, DEGREE, TRAIN [generic]

1	[1] Individual Questionnaire	2948
2	[2] Gap Questionnaire (temporary drop-outs)	15
3	[3] Biographical Questionnaire	13110
4	[4] Various Sources	42
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	11927
-3	[-3] Not Valid	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
This is a flag variable which provides information on the data sources used for the construction of the variables FIELD$$, DEGREE$$, TRAINA$$, TRAINB$$, TRAINC$$ and TRAIND$$ (see the description of the respective variables for details). [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

bilztch$$ – Change in Education since last survey / last year [generic]

0	[0] Consistent educational information since last survey	23789
1	[1] Inconsistent educational information since last survey	34
2	[2] Inconsistent educational information since last year	843
-1	[-1] No Answer	0
-2	[-2] Does not apply	3376
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
This is a flag variable which identifies observations with inconsistent changes in the in-

formation on highest educational qualification compared to the previous observation or year.
Questions on highest educational attainment have been asked in the first survey and were only updated in subsequent years if the respondent reported a change. In the year 2000, every single SOEP participant was asked his highest level of educational attainment which produced a number of inconsistencies between the most recent information from 2000 and the generated information from previous years. These inconsistencies include both higher and lower educational attainment and are not just due to repeating the question about educational attainment in 2000. They also occur more generally, although to a lower degree, in the second survey wave of new samples when respondents to individual and life history questionnaires are asked to state their educational attainment. In both situations, respondents are not only asked annual questions about any changes in educational attainment since the previous year, but are also asked to state their highest level of educational attainment.
In our view there is no means of unequivocally correcting for these inconsistencies. The flag variable helps researchers to identify observations with inconsistent answers to educational questions in the cross-sectional perspective. Researchers need to decide how to deal with these on a case-bycase basis depending on the research question at hand.
So far, we have not found evidence that respondents with a change in the year 2000 differed systematically from other respondents. One possible approach would be to exclude these individuals from the analysis when sample size allows. Alternatively, one could apply the information collected in 2000 to the prior years in which no changes were recorded between two years and test whether the results differ from those obtained when these individuals are left out. Since 2011, a Beta version of BIOEDU has also been made available, containing new data on consistent longitudinally tested educational transitions. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

bilztev$$ – Change in Education, total observed period [generic]

0	[0] Consistent educational information	21733
1	[1] Inconsistent educational decline	945
2	[2] Inconsistent educational increase	2326
3	[3] Inconsistent educational decline and increase	175
-1	[-1] No Answer	0
-2	[-2] Does not apply	2863
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: -1985
This flag variable identifies observations with at least one inconsistent change in the information given on individual highest educational qualification over the whole observation period. See the description of variable BILZTCH$$ for more details on the sources of these inconsistencies. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Charlotte Bartels (Tel. +49-30-89789-346)

11 Information on the Interview

month$$ – Month Of Interview [generic]

1	[1] January	34
2	[2] February	7237
3	[3] March	5737
4	[4] April	3067
5	[5] May	3554
6	[6] June	2960
7	[7] July	2403
8	[8] August	921
9	[9] September	1009
10	[10] October	1098
11	[11] November	22
12	[12] December	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
Month of interview is generated using the answers to the individual questionnaire. Missing answers are filled in using data from the $hbrutto files. Interviews that took place in December and before the 20th of that month were recoded -3. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)

mode$$ – Interview Method [generic]

100	[100] With Interviewer Assistance	0
110	[110] Oral Interview	1797
120	[120] Written Ques. Interviewer	3665
130	[130] Mixed Type	0
131	[131] Written Ques. No Interviewer	307
132	[132] Oral And Written	344
133	[133] Proxy	0
134	[134] Third Person Present	0
135	[135] No Third Person Present	0
140	[140] CAPI - since 1998 (O)	17158
150	[150] CAWI â€“ since 2014 (BE)	2096
200	[200] Telephone Assistance	0
210	[210] Written, By Mail	2674
220	[220] Phone Interview	1
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0

-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
The interview method is generated via the answers to the questions in the individual questionnaire. Missing answers are filled in from the $pbrutto files. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690)