1587 2026

SOEP Survey Papers
Series D – Variable Descriptions and Coding

SOEP-IS 2024 – PGEN: Person-Related Status and Generated Variables

SOEP-IS Group

DIW BERLIN

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

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

Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

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

Please cite this paper as follows:
SOEP-IS Group, 2026. SOEP-IS 2024 – PGEN: Person-Related Status and Generated Variables. SOEP Survey Papers 1587: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Anton-Wilhelm-Amo-Straße 58
10117 Berlin
Germany

soeppapers@diw.de

DIW BERLIN The German Socio Economic Panel at DIW Berlin

SOEP-IS 2024 – PGEN: Person-Related Status and Generated Variables

SOEP-IS Group

2026

The file pgen is part of a collection, which is released with doi:10.5684/soep.is.2024.

Contents

1 Identifiers

pid – Never Changing Person ID

-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

hid – Current Wave HH Number

-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

cid – Original Household Number, Case ID

-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

syear – Survey Year

1998	724
1999	750
2000	755
2001	766
2002	780

2003		795
2004		792
2005		799
2006		797
2007		797
2008		794
2009		3226
2010		2745
2011		2506
2012		3696
...	(6 rows omitted)	36585
2019		4983
2020		5339
2021		5339
2022		2507
2023		4760
2024		3774
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

2 Own Nationality and Residential Status

pgnation – 1. Nationality

1	Germany	62411
2	Turkey	794
3	Ex-Yugoslavia	5
4	Greece	190
5	Italy	332
6	Spain	75
7	Ex-GDR (Country Of Origin Only)	3
10	Austria	178
11	France	75
12	Benelux	0
13	Denmark	4
14	Great Britain	62
15	Sweden	6
16	Norway	3
17	Finland	14
...	(176 rows omitted)	1629
194	Sahara	0
195	Ingushetia	0

196	Kosovo	7
222	Eastern Europe	0
333	Other Unspecified Foreign Country	0
999	ethnic minorities (e.g. Yazidi, Roma)	0
-1	No answer	1841
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	16380
-9	Missing due to a terminated interview	0

This variable is designed to integrate the information on respondents' nationality for all subsamples.

pgnationiso3n – 1. Nationality (ISO3N country code list)

4	Afghanistan	2
8	Albania	5
10	Antarctica	0
12	Algeria	0
16	American Samoa	0
20	Andorra	0
24	Angola	0
28	Antigua and Barbuda	0
31	Azerbaijan	2
32	Argentina	1
36	Australia	0
40	Austria	30
44	Bahamas	0
48	Bahrain	0
50	Bangladesh	0
...	(230 rows omitted)	10933
876	Wallis and Futuna	0
882	Samoa	0
887	Yemen	0
894	Zambia	0
900	Kosovo	0
997	stateless	0
-1	No answer	68
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	72968
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0

-9	Missing due to a terminated interview	0

3 Familiy Status and Partnership

pgpartz – Partner Indicator

0	No partner	26136
1	Spouse, registered partner	43985
2	Partner	7571
3	Probably spouse, registered partner	414
4	Probably partner	229
5	not clear	5674
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

Partner indentifier have the purpose of clearly defining spouse (married) and partner (unmarried) relationships in SOEP-IS households and thus enabling analyses on the couple level. The variable PGPARTZ generated in this context reveals whether a person in a SOEP-IS 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 automatically assigned to all persons living in households in which there is clearly no partnership. These include:

(a) one-person households

(b) single-parent households

(c) household head living together with only one parent (or parent-in-law)

Codes 1 to 4 define these relationships. To assign Codes 1 and 2, the partnership has to be clearly definable from the perspective of both partners. This implies agreement between the codes of the variable STELL (= relationship to head of household in PBRUTTO) pointing to a possible partnership (e.g., the combination 0 (=head of household) and 1 (=spouse of household head)), as well as agreement between the codes for family status in that wave (e.g., married couples both have the Code 1 (=married, living together)). In case of ambiguity, the marital history is taken into account as well. If there are inconsistencies between the answers provided by the two persons, or between data on marital status and relationship to head of household, each person is examined individually within his or her household context. If uncertainty remains, the codes 3 or 4 are assigned.

pgpartnr – Partner Person Number

-1	No answer	0
-2	Does not apply	31740
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0
-9	Missing due to a terminated interview	0

Partner indicators have the purpose of clearly defining spouse (married) and partner (unmarried) relationships in SOEP households and thus to make possible analyses on the couple level.
If PGPARTZ is coded 0 or 9, this person has no partner or the partner cannot be identified as such. The variable PGPARTNR is assigned the missing code of "-2" (=does not apply) for these persons.
If PGPARTZ is coded 1, 2, 3 or 4, a partnership was defined and PGPARTNR is then assigned the value of the unchanging person ID number (=PERSNR) of the partner.
For analyses of partner relationships, this information can be used to clearly link all persons with their respective partners, and all information on both partners can also be stored in a common dataset.

pgfamstd – Marital Status In Survey Year

1	Married	44438
2	Married, But Separated	1460
3	Single	18563
4	Divorced	8330
5	[de] Verwitwet / Lebenspartner/-in aus eingetragener gleichgeschlechtlicher Partnerschaft verstorben	5576
6	husband/wife abroad	112
7	Registered Same-Sex Partnership, Living Together	20
8	Registered Same-Sex Partnership, Living Apart	0
-1	No answer	60
-2	Does not apply	0
-3	Implausible value	35
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	76
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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. 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.
Note that the partner indicator PGPARTZ supplied in the PGEN data files as well might not match the information provided in PGFAMSTD in its entirety.

4 Wages and Salary

pglabgro – Current Gross Labor Income in Euro

-1	No answer	3442
-2	Does not apply	34011
-3	Implausible value	48
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable PGLABGRO represents the current gross labor income of all SOEP respondents who are employed in each respective wave. The variable contains both generated and imputed values for Sample E (until 2011) & Sample I (until 2010); since then, the variable contains only generated values for all samples. The imputed values are available in the variables PGI1-PGI5LABGRO. Income details are consistently provided in euros for all waves. Item nonresponse is imputed in a twostage 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 (PGIMPGRO).

pgimpgro – Imputation flag for LABGROxx

0	Not imputed	39465
1	Imputed	5194
-1	No answer	0
-2	Does not apply	34011
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0

-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable PGIMPGRO designates imputations of item-nonresponse in the variable PGLABGRO (current gross labor income).

pglabnet – Current Net Labor Income in Euro

-1	No answer	2599
-2	Does not apply	34011
-3	Implausible value	36
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable PGLABNET represents the current net labor income of all persons in SOEP working in the respective wave. The variable contains both generated and imputed values for Sample E (until 2011) & Sample I (until 2010); since then, the variable contains only generated values for all samples. The imputed values are available in the variables PGI1-PGI5LABNET. 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 (PGIMPNET).

pgimpnet – Imputation flag for LABNETxx

0	Not imputed	40710
1	Imputed	3949
-1	No answer	0
-2	Does not apply	34011
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339

-9	Missing due to a terminated interview	0

The variable PGIMPNET designates imputations of item-nonresponse in the variable PGLABNET (current net labor income).

pgsndjob – Current Gross Secondary Income In Euro

-1	No answer	1206
-2	Does not apply	70636
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable PGSNDJOB represents the imputed current gross labor income from second job generated for all SOEP respondents in each respective wave. Because missing information on PGSNDJOB is not imputed in the SOEP Innovation Sample, PGSNDJOB not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGSNDJOB is coded to "-5" (not contained in questionnaire).

pgimpsnd – Imputation Flag For SNDJOBxx

0	Not imputed	758
1	Imputed	114
-1	No answer	0
-2	Does not apply	16154
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	61644
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable PGIMPSND indicates imputations of item nonresponse in the variable PGSNDJOB (current gross labor income from second job). Because missing information on PGSNDJOB is not imputed in the SOEP Innovation Sample, PGIMPSND not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGIMPSND is coded to "-5" (not contained in questionnaire).

pgi1labgro – 1. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [1/15]

0		308

1		4
2		2
5		1
6		2
8		2
9		1
12		4
13		2
15		1
16		2
19		2
20		3
22		1
23		1
...	(2999 rows omitted)	35882
30000		9
32500		1
34000		1
35000		3
40000		2
45000		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.
Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical

packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi2labgro – 2. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [2/15]

0		318
1		5
2		3
5		2
6		1
8		2
9		1
12		3
13		1
15		1
16		2
19		2
20		3
22		1
23		1
...	(2999 rows omitted)	35869
30000		12
32500		1
34000		1
35000		3
40000		2
45000		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339

-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.
Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi3labgro – 3. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [3/15]

0	312
1	5
2	3
5	1
6	1
8	2
9	1
12	3

13		1
15		1
16		3
19		2
20		4
22		1
23		2
...	(2998 rows omitted)	35876
30000		9
32500		1
34000		1
35000		3
40000		2
45000		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.

· IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi4labgro – 4. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [4/15]

0		319
1		4
2		2
5		1
6		1
8		2
9		1
12		5
13		1
15		1
16		3
19		2
20		3
22		2
23		1
...	(2998 rows omitted)	35865
30000		14
32500		1
34000		1
35000		3
40000		2
45000		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations

are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.
Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi5labgro – 5. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [5/15]

0	320
1	4
2	2
5	1
6	1
8	2
9	1
12	4
13	1
15	1
16	3
19	3
20	4
22	1

23		1
...	(2998 rows omitted)	35868
30000		10
32500		1
34000		1
35000		3
40000		2
45000		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536.

Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi6labgro – 6. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [6/15]

0		109
1		2
2		2
5		1
6		1
8		1
9		1
12		4
13		1
16		4
19		2
20		1
23		2
25		1
31		1
...	(1690 rows omitted)	9533
23000		2
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi7labgro – 7. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [7/15]

0		133
1		2
2		3
5		2
6		1
8		2
9		1

12		6
13		1
16		3
19		2
20		2
23		1
25		1
31		1
...	(1690 rows omitted)	9505
23000		2
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi8labgro – 8. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [8/15]

0		109
1		2
2		3
5		1
6		1
8		1
9		1
12		3
13		1
16		2
19		2
20		1
23		1
25		1
31		1
...	(1690 rows omitted)	9535
23000		2
24000		1
25000		2
30000		4
32500		1
35000		3

-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi9labgro – 9. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [9/15]

0		121
1		2
2		2
5		1
6		1
8		2
9		1
12		3
13		1
16		2
19		3
20		2
23		2
25		1
31		1
...	(1690 rows omitted)	9520
23000		3
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi10labgro – 10. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [10/15]

0		119

1		3
2		3
5		1
6		1
8		1
9		1
12		3
13		2
16		3
19		3
20		1
23		1
25		1
31		1
...	(1690 rows omitted)	9522
23000		2
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi11labgro – 11. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [11/15]

0		118
1		2
2		2
5		1
6		1
8		1
9		1
12		5
13		1
16		2
19		2
20		1
23		1
25		1
31		2
...	(1690 rows omitted)	9525

23000		2
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi12labgro – 12. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [12/15]

0		110
1		2
2		2
5		1
6		1
8		1
9		1
12		4
13		1
16		3
19		2
20		2
23		1
25		1
31		1
...	(1690 rows omitted)	9532
23000		2
24000		1
25000		2
30000		4
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi13labgro – 13. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [13/15]

0		115
1		3
2		2
5		2
6		2
8		1
9		1
12		4
13		1
16		3
19		2
20		1
23		1
25		1
31		2
...	(1690 rows omitted)	9523
23000		2
24000		1
25000		2
30000		5
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi14labgro – 14. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [14/15]

0		113
1		2
2		2
5		2
6		2
8		1
9		1
12		3
13		1
16		2

19		2
20		1
23		1
25		2
31		1
...	(1690 rows omitted)	9529
23000		3
24000		1
25000		2
30000		3
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi15labgro – 15. Imput. Akt. Bruttoerwerbseink.(gen) in Euro [15/15]

0		119
1		2
2		2
5		1
6		2
8		1
9		1
12		3
13		2
16		3
19		4
20		1
23		2
25		2
31		1
...	(1690 rows omitted)	9519
23000		2
24000		1
25000		2
30000		4
32500		1
35000		3
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0

-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi1labnet – 1. Imput. Akt. Nettoerwerbseink.(gen) in Euro [1/15]

0		293
1		4
2		2
4		1
6		1
8		2
9		1
11		1
12		5
13		2
15		2
16		3
18		1
20		2
22		1
...	(2798 rows omitted)	35900
17600		1
18000		2
18500		6
20000		3
21000		1
27796		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood

2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi2labnet – 2. Imput. Akt. Nettoerwerbseink.(gen) in Euro [2/15]

0		302
1		4
2		3
4		1
6		1
8		2
9		1
11		2
12		3
13		2
15		2
16		2
18		1
20		2
22		1
...	(2798 rows omitted)	35889
17600		1
18000		2

18500		9
20000		3
21000		1
27796		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi3labnet – 3. Imput. Akt. Nettoerwerbseink.(gen) in Euro [3/15]

0		307
1		4
2		3
4		1
6		1
8		3
9		1
11		1
12		3
13		2
15		2
16		2
18		1
20		2
22		1
...	(2798 rows omitted)	35887
17600		1
18000		2
18500		6
20000		3
21000		1
27796		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi4labnet – 4. Imput. Akt. Nettoerwerbseink.(gen) in Euro [4/15]

0		297
1		4
2		3
4		1
6		1
8		2
9		1
11		1
12		4
13		2
15		2
16		2
18		2
20		2
22		1
...	(2798 rows omitted)	35891
17600		1
18000		2
18500		11
20000		3
21000		1
27796		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521

-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.
Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several (m) estimates with the within-and between-imputation variance.

- STATA provides a built-in functionality called mi.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi5labnet – 5. Imput. Akt. Nettoerwerbseink.(gen) in Euro [5/15]

0	298
1	4
2	2
4	2
6	1

8		3
9		1
11		1
12		3
13		4
15		3
16		2
18		1
20		2
22		1
...	(2798 rows omitted)	35892
17600		1
18000		2
18500		7
20000		3
21000		1
27796		1
-1	No answer	0
-2	Does not apply	27914
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	14521
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Multiple imputation procedures provide a way to deal with missing values on the variable current net labor income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the $PGEN datasets: the variables pgi1labnet-pgi5labnet. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current net labor income is the gross labor income of the previous year. A complete list of the variables used for modelling is available upon request. The missing observations were assumed to be missing at random. We set the number of imputations m=5 and get 5 multiple imputed values for pglabnet. The number of iterations carried out in each prediction model was specified to be 2000. Sample E&I and the supplementary sample S1 were imputed separately.

Analysing multiply imputed data: For analysing multiple imputed data, one does not necessarily need special methods; however, such tools exist and simplify the use of multiply imputed data. Below is given a short overview of some useful tools for various statistical packages. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared stan-

dard errors of the several (m) estimates with the within-and between-imputation variance.
· STATA provides a built-in functionality called mi.
· Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
· IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.
Royston, P. 2004. Multiple imputation of missing values. Stata Journal 4: 227–241. Royston, P. 2005a. Multiple imputation of missing values: Update. Stata Journal 5: 188–201. Royston, P. 2005b. Multiple imputation of missing values: Update of ice. Stata Journal 5: 527–536. Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 7: 445–464. Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 9: 466–477.

pgi6labnet – 6. Imput. Akt. Nettoerwerbseink.(gen) in Euro [6/15]

0		102
1		2
2		2
4		1
6		1
8		1
9		1
11		1
12		4
13		2
15		1
16		2
18		1
31		1
40		3
...	(1772 rows omitted)	9545
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi7labnet – 7. Imput. Akt. Nettoerwerbseink.(gen) in Euro [7/15]

0		103
1		2
2		2
4		1
6		1
8		1
9		1
11		1
12		3
13		2
15		1
16		1
18		1
31		1
40		5
...	(1771 rows omitted)	9544
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi8labnet – 8. Imput. Akt. Nettoerwerbseink.(gen) in Euro [8/15]

0		103
1		2
2		2
4		1
6		1
8		1
9		1
11		1
12		3
13		3
15		1
16		1

18		1
31		1
40		6
...	(1773 rows omitted)	9542
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi9labnet – 9. Imput. Akt. Nettoerwerbseink.(gen) in Euro [9/15]

0		100
1		2
2		2
4		1
6		1
8		1
9		1
11		4
12		5
13		2
15		1
16		1
18		2
31		2
40		2
...	(1772 rows omitted)	9543
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291

-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi10labnet – 10. Imput. Akt. Nettoerwerbseink.(gen) in Euro [10/15]

0		99
1		2
2		2
4		1
6		1
8		1
9		1
11		1
12		4
13		4
15		2
16		1
18		2
31		1
40		4
...	(1771 rows omitted)	9544
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi11labnet – 11. Imput. Akt. Nettoerwerbseink.(gen) in Euro [11/15]

0		98
1		2
2		3
4		1
6		1
8		2

9		2
11		1
12		4
13		2
15		2
16		1
18		1
31		1
40		3
...	(1772 rows omitted)	9545
16000		2
17500		3
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi12labnet – 12. Imput. Akt. Nettoerwerbseink.(gen) in Euro [12/15]

0		101
1		2
2		2
4		1
6		1
8		1
9		1
11		1
12		4
13		2
15		1
16		1
18		1
31		1
40		3
...	(1771 rows omitted)	9547
16000		2
17500		2
17600		1
18000		1
20000		1

27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi13labnet – 13. Imput. Akt. Nettoerwerbseink.(gen) in Euro [13/15]

0		95
1		3
2		2
4		2
6		1
8		1
9		1
11		1
12		4
13		2
15		2
16		1
18		1
31		1
40		2
...	(1771 rows omitted)	9551
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi14labnet – 14. Imput. Akt. Nettoerwerbseink.(gen) in Euro [14/15]

0		106
1		2
2		3
4		1
6		1
8		1
9		1
11		1
12		4
13		3
15		2
16		1
18		1
31		1
40		2
...	(1771 rows omitted)	9540
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgi15labnet – 15. Imput. Akt. Nettoerwerbseink.(gen) in Euro [15/15]

0		101
1		2
2		2
4		1
6		2
8		1
9		1
11		1
12		4
13		2
15		1
16		1
18		1
31		1
40		3

...	(1771 rows omitted)	9546
16000		2
17500		2
17600		1
18000		1
20000		1
27796		1
-1	No answer	0
-2	Does not apply	6701
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	62291
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

5 Current Employment Status

pgstib – Occupational Position

10	Not employed (NE): without further information	5064
11	NE: In education, further education, apprenticeship	2276
12	NE: registered unemployed	3664
13	NE: Pensioner	22623
15	NE: Military, Community Service	141
100	Apprentice / trainee, no further information	0
110	Apprentice (until 1999), trainee (1990 East)	43
120	Apprentice / trainee in industry or technology (since 2000)	793
130	Apprentice / trainee in trade or commerce (since 2000)	520
140	Student trainee, intern	198
150	Aspirant	0
200	Worker, no further information	0
210	Untrained Worker	1123
220	Semi-Trained Worker	2648
230	Trained Worker	3038
...	(28 rows omitted)	32542
560	Managing partner or comparable employee in own company (since 2019)	26
600	Civil servant, no further information	0
610	Civil servant, lower level	92
620	Civil servant, middle level	736
630	Civil servant, upper level	1338
640	Civil servant, executive level	764
-1	No answer	1040
-2	Does not apply	1
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0

-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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):
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
In PGSTIB, 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.

pgemplst – Employment Status

1	Full-Time Employment	28078
2	Regular Part-Time Employment	10715
3	Vocational Training	1469
4	Marginal, Irregular Part-Time Employment	4216
5	Not Employed	34060
6	Sheltered workshop (1998-2020)	117
7	In short-time work (1991-1995; since 2021)	0
-1	No answer	1
-2	Does not apply	14
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
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.

PGEMPLST supplements the variable PGLFS, which differentiates among persons who are not employed.

pglfs – Labor Force Status

1	Non-working (NW): without further information	6559
2	NW: age 65 and older	17302
3	NW: and currently in training/education	1714
4	NW: on parental leave (since 1991)	956
5	NW: in military/community service	39
6	NW: and registered unemployed	3099
8	NW: but occasional secondary job (1985-2016)	216
9	NW: but paid work in past 7 days (since 1999)	648
10	NW: but regular secondary job (1985-2016)	3309
11	Working	43557
12	Working, but inactive within past 7 days (since 2000)	1256
13	NW: but paid secondary job (since 2017)	0
-1	No answer	15
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is based on the annual question on current employment status, combined with additional information on activities of non-working individuals. Since the beginning of the SOEP in the year 1984, the number of values assigned has been based on a large number of highly differentiated answer categories. It is designed to provide consistent longitudinal data on labor force participation across all waves.
PGLFS 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):

```
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
```

PGLFS supplements the variable PGEMPLST, which differentiates among persons who are employed.

pgjobch – Occupational Change

1	Not Employed	33769
2	Employed No Change	31757
3	Employed No Info If Change	3164
4	Employed With Change	7908
5	First Time Employed	904
-1	No answer	1168
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
PGJOBCH is generated based on the central filter variable, which indicates whether a respondent has changed jobs since the beginning of the previous year. 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, PGJOBCH is a modified version of the variable PGERWTYP which indicates whether a respondent has changed jobs since the beginning of the previous year. Unlike PGERWTYP, 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 PGERWTYP, the variable is also designed to identify respondents who have entered employment for the first time.

In addition to PGERWTYP, the variable is designed to provide consistent longitudinal information on job changes. The PGJOBCH 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".

pgerwtyp – Type of occupation

1	Not Employed, Green	33773
2	Not Employed (First Surveyed) Not Applicable Since 94	0
3	Employed (First Surveyed) Not Applicable Since 94	0
4	Empl. Exc Change	31352
5	Empl. No Info If Change	3861
6	Empl. With Change, Also First Time Employment	8869
7	Empl. With Near-Retirement Part-time	288
-1	No answer	527
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
An alternative variable is PGJOBCH (see below), which is an improved version of PGERWTYP, as it is generated in a longitudinally consistent way and contains an additional category for first-time employed persons.
Respondents from the supplementary samples are not being asked about the information on job change; hence, in the year when these samples enter the SOEP-IS, the majority of the employed persons fall into the category [5] (Employed, no info if change).

6 Current Occupation

pgisei08 – Last Reached Isei Value (International Socio-Economic Index)

11.56		3
11.7399997711182		1
11.74		36
12.01		2
13.2399997711182		1
13.24		2
13.34		3
13.72		1
13.8699998855591		2
13.87		1
14.21		384
14.210000038147		7
14.39		24
14.5699996948242		2
14.57		20
...	(465 rows omitted)	18166
86.81		3
88.31		12
88.6999969482422		9
88.7		125
88.9599990844727		1
88.96		11
-1	No answer	8828
-2	Does not apply	5733
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	45293
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 ISCO-88 classification. The values for the variable range between 11 and 90. In contrast to the prestige scores of Ganzeboom and Treiman (1996) and Wegener (1988), ISEI is a measure of socio-economic status. Starting in SOEP-IS wave 2016, pgisei08 is derived from ISCO-88 scores that themselves are derived from ISCO-08 scores using the „derivescores“ STATA package (https://github.com/dirtyhawk/stata-derivescores).
This recoding is derived from Ganzeboom‘s SPSS script iskoisei.sps. (http://www.harryganzeboom.nl/isco88/in
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgsiops08, pgmps08) and occupational class (pgegp08).

In the SOEP-IS, pgisei08 was first implemented in wave 2016 and is coded [-5] in all previous waves.
Ganzeboom, Harry B.G.; Treiman, Donald J., "International Stratification and Mobility File: Conversion Tools." Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. <2010-01-12>.

pgmps08 – Last Reached MPS Value (Magnitude-Prestige-Scale, Wegener)

20		41
23.9		39
24.7		19
26.7		39
26.9		194
28.6		3
30		499
30.3		24
31.2		95
31.6		7
31.8		99
31.9		42
32.4		78
34.7		3
35.6		8
...	(164 rows omitted)	16022
160.3		28
160.5		81
170.9		64
173.3		42
179.6		228
186.8		12
-1	No answer	9977
-2	Does not apply	5733
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	45293
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The Magnitude-Prestige-Scale (MPS) is a tool for operationalising the social position of occupations in empirical surveys. In contrast to alternative instruments such as the Standard International Occupational Prestige Scale (SIOPS) or the International Socio-economic Index of Occupational Status (ISEI), the MPS is an instrument that was constructed for analyses on the national level only (Christoph, 2005). Starting in SOEP-IS wave 2016, pgmps08 is derived from ISCO-88 scores that themselves are derived from ISCO-08 scores using the „derivescores“ STATA package (https://github.com/dirtyhawk/stata-derivescores). Not all employed persons are asked the relevant input questions on an annual basis. Only those employed

persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgsiops08), occupational status (pgisei08), and occupational class (pgegp08).
In the SOEP-IS, pgmps08 was first implemented in wave 2016 and is coded [-5] in all previous waves.
Christoph, Bernhard: 2005: Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88, https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207543 – Ganzeboom, Harry B.G.; Treiman, Donald J., "International Stratification and Mobility File: Conversion Tools." Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. <2010-01-12>.

pgsiops08 – last reached SIOPS score (stand. internat. occ. prestige scale, Treiman)

12		1
13		27
15		19
16		38
17		1
20		485
20.03		21
20.0300006866455		2
20.2999992370605		6
20.3		109
20.3899993896484		4
20.39		74
21		29
21.0799999237061		1
21.08		140
...	(272 rows omitted)	17415
76.11		11
76.1100006103516		1
78.01		212
78.0100021362305		16
78.16		191
78.1600036621094		13
-1	No answer	8828
-2	Does not apply	5733
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	45293
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational prestige score index for all employed persons. PGSIOPS is based on ISCO-88 and was developed by Donald Treiman et al. The scale ranges from 6 to 78. Starting in SOEP-IS wave 2016, pgsiops08 is derived from ISCO-88 scores

that themselves are derived from ISCO-08 scores using the „derivescores“ STATA package (https://github.com/dirtyhawk/stata-derivescores). This recoding is derived from Ganzeboom‘s SPSS script iskotrei.sps. (http://www.harryganzeboom.nl/isco88/index.htm). Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgmps08), occupational status (pgisei08), and occupational class (pgegp08).
In the SOEP-IS, pgsiops08 was first implemented in wave 2016 and is coded [-5] in all previous waves.
Ganzeboom, Harry B.G.; Treiman, Donald J., “International Stratification and Mobility File: Conversion Tools.” Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. <2010-01-12>.

pgegp08 – Last Reached EPG Value (Erikson, Goldthorpe, Portocarero)

1	[I] Higher Managerial and Professional Workers	2919
2	[II] Lower Managerial and Professional Workers	5416
3	[IIIa] Routine Clerical Work	3243
4	[IIIb] Routine Service and Sales Work	2089
5	[IVa] Small Self-Employed With Employees	227
6	[IVb] Small Self-Employed Without Employees	452
7	[V] Manual Supervisors	46
8	[VI] Skilled Manual Workers	1693
9	[VIIa] Semi- and Unskilled Manual Workers	2467
10	[VIIb] Agricultural Labour	183
11	[IVc] Self-Employed Farmers	81
-1	No answer	8828
-2	Does not apply	5733
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	45293
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational class for all employed persons. PGEGP 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. Starting in SOEP-IS wave 2016, pgegp08 is derived from ISCO-88 scores that themselves are derived from ISCO-08 scores using the „derivescores“ STATA package (https://github.com/dirtyhawk/stata-derivescores). This version of Goldthorpe‘s class categories (EGP) is derived from Ganzeboom‘s SPSS script iskoegp.sps.
Former versions and waves contained additional categories for unemployed persons (15) and pensioners (18). From wave 2015 on the egp-variable has a more standard shape. Information on unemployment and retirement can be found in PGSTIB (occupational position) and PGLFS (labor force status).

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 PGEGP-categories (5) and (6) contain information on the firm size for self-employed persons.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational status (pgisei08) and occupational prestige scores (pgsiops08, pgmps08).
In the SOEP-IS, pgegp08 was first implemented in wave 2016 and is coded [-5] in all previous waves.
Harry B.G. Ganzeboom, Paul M. De Graaf, Donald J. Treiman (1992), A standard international socio-economic index of occupational status, in: Social Science Research 21 (1),1-56, https://doi.org/10.1016/0049-089X(92)90017-B. – Ganzeboom, Harry B.G.; Treiman, Donald J., "International Stratification and Mobility File: Conversion Tools." Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. <2010-01-12>.

pgisei88 – Last Reached Isei Value (International Socio-Economic Index)

16		965
19		145
20		454
21		59
22		17
23		809
24		58
25		823
26		374
27		124
28		112
29		1071
30		1503
31		123
32		331
...	(40 rows omitted)	20429
82		58
83		14
85		133
87		44
88		294
90		30
-1	No answer	294
-2	Does not apply	17231
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 ISCO-88 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. It is derived from the ISCO-88 code of the current occupation using the Stata ado iskoisei by John Hendrick which itself is based on Harry Gaanzeboom's SPSS algorithms. Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgsiops88, pgmps92) and occupational class (pgegp88).
In the SOEP-IS, pgisei88 is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only pgisei08 scores will be released.

pgmps92 – Last Reached MPS Value (Magnitude-Prestige-Scale, Wegener)

30		11
30.1		34
30.2		54
30.3		5
31		261
31.1		72
31.2		7
31.5		266
31.7		141
31.8		17
31.9		4
32		24
32.1		277
32.2		14
32.3		808
...	(160 rows omitted)	25210
139.8		45
145.7		149
152.5		170
191.3		294
207.2		41
216		17
-1	No answer	335
-2	Does not apply	17239
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational prestige score developed by Wegener (1988) for all employed persons. Like the PGSIOPS prestige sore, Wegener's prestige scala measures a person's occupational prestige and was developed especially for use in the Federal Republic of Germany. PGMPS is assigned based on the German Federal Statistical Office's occupational classification of 1992 (PGKLAS). The procedure has been documented in Frietsch and Wirth (2001). Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgsiops88), occupational status (pgisei88), and occupational class (pgegp88).
In the SOEP-IS, pgmps92 is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only pgmps08 scores will be released.

pgsiops88 – last reached SIOPS score (stand. internat. occ. prestige scale, Treiman)

13		17
15		57
17		5
18		2
19		474
20		351
21		1129
22		175
23		275
24		51
25		442
26		32
27		22
28		220
29		160
...	(40 rows omitted)	23779
71		58
72		81
73		116
75		19
76		30
78		475
-1	No answer	294
-2	Does not apply	17231
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational prestige score index for all employed persons. PGSIOPS is based on ISCO-88 and was developed by Donald Treiman et al. The scale ranges from 6 to 78.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational prestige scores (pgmps92), occupational status (pgisei88), and occupational class (pgegp88).
In the SOEP-IS, pgsiops88 is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only pgsiops08 scores will be released.
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

pgegp88 – Last Reached EPG Value (Erikson, Goldthorpe, Portocarero)

1	[I] Higher Managerial and Professional Workers	3873
2	[II] Lower Managerial and Professional Workers	6853
3	[IIIa] Routine Clerical Work	3567
4	[IIIb] Routine Service and Sales Work	3818
5	[IVa] Small Self-Employed With Employees	470
6	[IVb] Small Self-Employed Without Employees	762
7	[V] Manual Supervisors	0
8	[VI] Skilled Manual Workers	4070
9	[VIIa] Semi- and Unskilled Manual Workers	3923
10	[VIIb] Agricultural Labour	374
11	[IVc] Self-Employed Farmers	155
-1	No answer	292
-2	Does not apply	17338
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational class for all employed persons. PGEGP 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.
Former versions and waves contained additional categories for unemployed persons (15) and pensioners (18). From wave 2015 on the egp-variable has a more standard shape. Information on unemployment and retirement can be found in PGSTIB (occupational position) and PGLFS (labor force status).
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 PGEGP-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 PGEGP 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.

Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see occupational status (pgisei88) and occupational prestige scores (pgsiops88, pgmps92).
In the SOEP-IS, pgegp88 is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only pgegp08 scores will be released.
John Hendrickx, 2002. "ISKO: Stata module to recode 4 digit ISCO-88 occupational codes," Statistical Software Components S425802, Boston College Department of Economics, revised 20 Oct 2004. `https://ideas.repec.org/c/boc/bocode/s425802.html`
Ganzeboom, Harry B.G.; Treiman, Donald J., "International Stratification and Mobility File: Conversion Tools." Amsterdam: Department of Social Research Methodology, `http://www.harryganzeboom.nl/isco88/`.

pgerljob – Working In Occupation Trained For

1	Yes	4660
2	No	2593
3	In Training	412
4	Has No Job Training	397
-1	No answer	354
-2	Does not apply	6909
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to offer annual data on all employed persons, indicating whether they are working in the occupation they were trained for.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Because detailed information on working in occupation trained for is not assessed in the Questionnaire of the SOEP Innovation Sample, PGERLJOB is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGERLJOB is coded to "-5" (not contained in questionnaire).

pgausb – Required Training For Job

1	No Training	63
2	Vocational Training	1224
3	Technical School, Engineering (East) 90 until 94	482
4	Technical College or University Studies, until 99	232
5	Technical College Studies, since 99	4187
6	University Studies, since 99	0
7	Technical College, University until 1998	82
8	Technical College since 1999	767
9	University since 1999	1009
-1	No answer	370
-2	Does not apply	6909
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 up to 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.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
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.
Because detailed information onrequired training for job is not assessed in the Questionnaire of the SOEP Innovation Sample, PGAUSB is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGAUSB is coded to "-5" (not contained in questionnaire).

pguebstd – Overtime Per Week

0	No Overtime Last Month	17541
-1	No answer	3875
-2	Does not apply	37437
-3	Implausible value	2
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	1644
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339

-9	Missing due to a terminated interview	0

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 PGUEBSTD 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 year 2012, respondents were not asked about the number of hours of overtime per week. PGUEBSTD was therefore generated using the difference between average actual weekly working hours and agreed weekly working hours.
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 down and gives the number of overtime hours as a decimal.
Please also see PGVEBZT and PGTATZT.

pgoeffd – Civil Service

1	Yes	10869
2	No	32905
-1	No answer	1105
-2	Does not apply	33791
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to provide annual data on employment in the civil service for all employed persons.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.

pgnace – Industry Occupation [pbra] (NACE Rev. 1.1, Sector)

1	Crop and animal production, hunting	517
2	Forestry, Logging, Related Service activities	26
5	Fishing, Operation Of Fish Hatcheries And Fish Farms	8
10	Mining Of Coal And Lignite, Extraction Of Peat	89

11	Extraction Of Crude Petroleum And Natural Gas	1
12	Mining Of Uranium And Thorium Ores	0
13	Mining Of Metal Ores	0
14	Other Mining And Quarrying	19
15	Manuf Food Products And Beverages	539
16	Manuf Tobacco Products	9
17	Manuf Textiles	101
18	Manuf Wearing Apparel, Dressing And Dyeing Of Fur	59
19	Tanning,Dressing Of Leather; Manuf luggage, Footwear	10
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	93
21	Manuf Pulp, Paper And Paper Products	81
...	(43 rows omitted)	26048
95	Private Households With Employed Persons	132
96	Undifferentiated goods producing activities of private households for own use	144
97	Undifferentiated services producing activities of private households for own use	128
98	Services - NEC	296
99	Extra-territorial Organizations And Bodies	10
100	Manufacturing - NEC	79
-1	No answer	962
-2	Does not apply	22322
-3	Implausible value	1
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	26996
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
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/vercodung.pdf

pgnace2 – Industry Occupation [pbra] (NACE Rev. 2, Sector)

1	Crop and animal production, hunting and related service activities	69
2	Forestry and logging	7
3	Fishing and aquaculture	2
5	Mining of coal and lignite	5
6	Extraction Of Crude Petroleum And Natural Gas	3
7	Mining Of Metal Ores	0
8	Other Mining And Quarrying	2
9	Mining support service activities	0
10	Manufacture of food products	119
11	Manufacture of beverages	8
12	Manuf Tobacco Products	3
13	Manuf Textiles	8
14	Manuf Wearing Apparel, Dressing And Dyeing Of Fur	17
15	Manufacture of leather and related products	5
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	13
...	(67 rows omitted)	6789
94	Activities of membership organisations	103
95	Repair of computers and personal and household goods	12
96	Other personal service activities	57
97	Private Households With Employed Persons	40
98	Undifferentiated goods- and services-producing activities of private households for own use	0
99	Extra-territorial Organizations And Bodies	2
-1	No answer	5239
-2	Does not apply	8860
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	57307
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

pgbetr – Size of the Company

1	Lt 5	2328
2	Ge 5 Lt 10	2797
3	Ge 11 Lt 20	2672
4	Until 90: Lt 20	0
5	91-04: Ge 5 Lt 20	553
6	Ge 20 Lt 100	6842
7	Ge 100 Lt 200	3321
8	Until 98: Ge 20 Lt 200	229
9	Ge 200 Lt 2000	8352
10	Ge 2000	11730
11	Self-Employed Without Coworkers	0

-1	No answer	1590
-2	Does not apply	38250
-3	Implausible value	6
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
Not all employed persons are asked the question on firm size on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Please also see PGALLBET for a broader categorization of the firm size, which is appropriate for analyses that include all sample years.
Self-employed are not included in this variable. Detailed information about the company size of self-employed is included in the variable PGSTIB.

pgallbet – Core Category Size Of The Company

1	LT 20	8350
2	GE 20 LT 200	10392
3	GE 200 LT 2000	8352
4	GE 2000	11730
5	Self-Employed Without Coworkers	0
-1	No answer	1590
-2	Does not apply	38250
-3	Implausible value	6
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to provide annual data on the core size category of the company for all employed persons.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
Self-employed are not included in this variable. Information about the company size is included in the variable pgstib.

pgerwzt – Length Of Time With Firm

0		430
0.100000001490116		576
0.200000002980232		597
0.300000011920929		932
0.400000005960464		464
0.5		433
0.600000023841858		373
0.699999988079071		341
0.800000011920929		641
0.899999976158142		335
1		430
1.10000002384186		421
1.20000004768372		382
1.29999995231628		628
1.39999997615814		304
...	(506 rows omitted)	76149
-10		80
-11		81
-12		58
-13		18
-14		17
-15		24
-16		26
-17		16
-18		34
-19		30
-20		19
-21		40
-22		46
-23		47
-24		37

The variable PGERWZT 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). If the month was not available a random month is used.

pgtatzt – Actual Weekly Work Time

0		5062
0.400000005960464		1
0.5		3
1		61
1.5		11
2		109
2.5		12
3		126
3.5		11
3.59999990463257		1
4		182
4.30000019073486		1
4.5		15
4.80000019073486		1
5		290
...	(209 rows omitted)	40751
80		25
100		2
120		1
125		1
157		1

168		2
-1	No answer	1532
-2	Does not apply	28723
-3	Implausible value	102
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	1644
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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.
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 PGVEBZT and PGUEBSTD.

pgvebzt – Agreed Upon Weekly Work Time

0		4894
0.600000023841858		1
1		13
1.5		4
2		44
2.5		6
3		77
3.5		8
4		116
4.5		8
5		166
5.5		9
6		184
6.19999980926514		1
6.30000019073486		1
...	(195 rows omitted)	34819
65		6
70		5
72		1
75		5
78		2
80		1
-1	No answer	1870
-2	Does not apply	34286
-3	Implausible value	499
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	1644

-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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. In 2012, the value (-2) was assigned only to non-employed people 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 PGTATZT and PGUEBSTD.

pgis88 – Current Occupational Classification (ISCO-88 Com)

0	Soldiers	0
100	Soldiers	78
1000	Legislators, Senior Officials and Managers	0
1100	Legislators and Senior Government Officials	0
1110	Legislators and Senior Government Officials	1
1140	Senior Officials of Special-Interest Organisations	5
1141	Senior Officials of Political Party Organisations	0
1142	Senior Officials of Employers', Workers' and Other Economic-Interest Organisations	26
1143	Senior Officials of Humanitarian and Other Special-Interest Organisations	0
1200	Corporate Managers	84
1210	Directors and Chief Executives	233
1220	Production and Operations Managers	0
1221	Production and Operations Managers in Agriculture, Hunting, Forestry and Fishing	0
1222	Production and Operations Managers in Manufacturing	85
1223	Production and Operations Managers in Construction	0
...	(471 rows omitted)	23787
9310	Mining and Construction Labourers	0
9311	Mining and Quarrying Labourers	2
9312	Construction and Maintenance Labourers: Roads, Dams and Similar Constructions	3
9313	Building Construction Laborer	39
9320	Manufacturing Laborer	366
9330	Transport Lab., Freight Handler	178
-1	No answer	308
-2	Does not apply	20300
-3	Implausible value	0
-4	Inadmissable multiple response	0

-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Code name (Main group, group):

(1000) Legislators, senior officials, and managers

(1001) Legislators and senior officials

(1002) Corporate managers

(1003) Managers of small enterprises

(1004) Professionals

(1005) Physical, mathematical, and engineering science professionals

(1006) Life science and health professionals

(1007) Teaching professionals

(1008) Other professionals

(1009) Technicians and associate professionals

(1010) Physical and engineering science associate professionals

(1011) Life science and health associate professionals

(1012) Teaching associate professionals

(1013) Other associate professionals

(1014) Clerks

(1015) Office clerks

(1016) Customer services clerks

(1017) Service Workers and shop and market sales workers

(1018) Personal and protective services workers

(1019) Models, salespersons, and demonstrators

(1020) Skilled agricultural and fishery Workers

(1021) Skilled agricultural and fishery workers

(1022) Craft and related trades workers

(1023) Extraction and building trades workers

(1024) Metal, machinery, and related trades workers

(1025) Precision, handicraft, craft printing and related trades workers

(1026) Other craft and related trades workers

(1027) Plant and machine operators and assemblers

(1028) Stationary plant and related operators

(1029) Machine operators and assemblers

(1030) Drivers and mobile plant operators

(1031) Elementary occupations

(1032) Sales and services elementary occupations

(1033) Agricultural, fishery, and related laborers

(1034) Laborers in mining, construction, manufacturing, and transport

This variable is designed to provide annual data on occupational activity for all employed persons according to the International Standard Classification of Occupations ISCO-88. Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire.
ISCO-88 is a strictly four-digit classification, and this variable is therefore coded in four-digit form. In contrast to the previous version of the classification system, ISCO-68, ISCO-88 does not use blanks if there is no adequate information for specific coding, but uses zeros instead. Thus 4000 stands for an unspecified office job; 2300 stands for teachers and 2000 stands for scientists, both without closer specification. There is no conversion key since the two classifications differ significantly. Hartmann and Schütz (2002) provide detailed information on the conducted occupational coding. This result has been slightly modified to fit to the ISCO-88 version for European Union purposes (ISCO-88(COM)).
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
In the SOEP-IS, ISCO-88 (pgis88) is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only ISCO-08 scores will be released (see pgis08).

pgis08 – Current Occupational Classification (ISCO-08)

0	Armed Forces Occupations	0
100	Commissioned Armed Forces Officers	0
110	Commissioned Armed Forces Officers	4
200	Non-commissioned Armed Forces Officers	0
210	Non-commissioned Armed Forces Officers	2
300	Armed Forces Occupations, Other Ranks	0
310	Armed Forces Occupations, Other Ranks	9
1000	Managers	0
1100	Chief Executives, Senior Officials and Legislators	0
1110	Legislators and Senior Officials	0
1111	Legislators	3
1112	Senior Government Officials	25
1113	Traditional Chiefs and Heads of Villages	0
1114	Senior Officials of Special-interest Organizations	10

1120	Managing Directors and Chief Executives	169
...	(569 rows omitted)	18458
9620	Other Elementary Workers	0
9621	Messengers, Package Deliverers and Luggage Porters	67
9622	Odd-job Persons	0
9623	Meter Readers and Vending-machine Collectors	1
9624	Water and Firewood Collectors	0
9629	Elementary Workers Not Elsewhere Classified	29
-1	No answer	8867
-2	Does not apply	12628
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	38398
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to provide annual data on occupational activity for all employed persons according to the International Standard Classification of Occupations ISCO-08 (Version of the ILO). Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire. ISCO-08 is a strictly four-digit classification, and this variable is therefore coded in four-digit form. Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
ISCO-08 was adopted through a resolution of a Tripartite Meeting of Experts on Labour Statistics held in December 2007. This resolution was subsequently endorsed by the Governing Body of the ILO in March 2008.
Sourcelink: http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm
Source of German Labels http://ec.europa.eu/eurostat/ramon/documents/SCL/isco08/SCL_isco08.zip
In the SOEP-IS, ISCO-08 (pgis08) was first implemented in wave 2016 and is coded [-5] in all previous waves.

pgklas92 – Current Occupational Classification (KldB92)

110	Farmers, general	128
111	Fruit and vegetable farmers (non-horticultural)	4
112	Arable farmers (special, permanent crops)	0
113	Livestock farmers and pasture farmers	0
114	Seed, crop producers, propagators (non-horticultural)	0
115	Crop protectors	3
116	Farmers and landlords	0
118	Farmers and wine growers	0
120	Wine growers, general	0
121	Vine propagators	0
129	Other wine growers	0
130	Agricultural workers, general	12
131	Agricultural supervisors	0

132	Agricultural machinery drivers	8
133	Vineyard workers	0
...	(2264 rows omitted)	24695
9921	Homeworkers without further specification	2
9931	Forepersons, group leaders without further specification	15
9941	Persons doing community service without further specification	0
9951	Self-employed persons without further specification	36
9961	Consultancy, planning professionals without further specification	18
9971	Other employees without further specification	124
-1	No answer	159
-2	Does not apply	20291
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	33175
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to provide annual data on job classification for all employed persons according to the classification of the German Federal Statistical Office (StaBuA). Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire. Due to data protection regulations, this information cannot be provided to data users and was therefore completely recoded by Infratest Sozialforschung in the year 2002. This recoding has been documented in Hartmann/Schütz 2002.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
In the SOEP-IS, pgklas92 is released up until 2016 and coded to [-5] afterwards. Starting in 2017, only pgklas10 scores will be released.
The occupational classification of the German Federal Statistical Office differentiates among six main occupational types (see next page):
I KLAS-Codes 0100-0629 Berufe in der Land-, Tier-, Forstwirtschaft und im Gartenbau
II KLAS-Codes 0700-0809 Bergleute, Mineralgewinner
III Fertigungsberufe
IIIa KLAS-Codes 1000-1129 Berufe in der Steinbearbeitung und Baustoffherstellung
IIIb KLAS-Codes 1200-1359 Keramik-, Glasberufe
IIIc KLAS-Codes 1400-1539 Chemie-, Kunststoffberufe
IIId KLAS-Codes 1600-1799 Berufe in der Papierherstellung, -verarbeitung und im Druck
IIIe KLAS-Codes 1800-1859 Berufe in der Holzverarbeitung, Holz- und Flechtwarenherstellung
IIIf KLAS-Codes 1900-2459 Berufe in der Metallerzeugung und –bearbeitung
IIIg KLAS-Codes 2500-3099 Metall-, Maschinenbau- und verwandte Berufe
IIIh KLAS-Codes 3100-3189 Elektroberufe
IIIi KLAS-Codes 3200-3239 MontiererInnen und Metallberufe, a.n.g.
IIIk KLAS-Codes 3300-3619 Textil- und Bekleidungsberufe
IIIl KLAS-Codes 3700-3789 Berufe in der Lederherstellung, Leder- und Fellverarbeitung
IIIm KLAS-Codes 3900-4359 Ernährungsberufe
IIIn KLAS-Codes 4400-4729 Hoch-, Tiefbauberufe

IIIo KLAS-Codes 4800-4929 Ausbauberufe, PolsterInnen
IIIp KLAS-Codes 5000-5069 Berufe in der Holz- und Kunststoffverarbeitung
IIIq KLAS-Codes 5100-5149 MalerInnen, LackiererInnen und verwandte Berufe
IIIr KLAS-Codes 5200-5239 WarenprüferInnen, VersandfertigmacherInnen
IIIs KLAS-Codes 5300-5319 HilfsarbeiterInnen ohne nähere Tätigkeitsangabe
IIIt KLAS-Codes 5400-5509 MaschinistInnen und zugehörige Berufe
IV Technische Berufe
IVa KLAS-Codes 6000-6129 IngenieurInnen, ChemikerInnen, PhysikerInnen, MathematikerInnen
IVb KLAS-Codes 6200-6529 TechnikerInnen, Technische Sonderfachkräfte
V Dienstleistungsberufe
Va KLAS-Codes 6600-6899 Warenkaufleute
Vb KLAS-Codes 6900-7069 Dienstleistungskaufleute und zugehörige Berufe
Vc KLAS-Codes 7100-7449 Verkehrsberufe
Vd KLAS-Codes 7500-7899 Organisations-, Verwaltungs-, Büroberufe
Ve KLAS-Codes 7900-8149 Ordnungs- und Sicherheitsberufe
Vf KLAS-Codes 8200-8399 Schriftwerkschaffende, -ordnende und künstlerische Berufe
Vg KLAS-Codes 8400-8599 Gesundheitsdienstberufe
Vh KLAS-Codes 8600-8949 Sozial- und Erziehungsberufe, anderweitig nicht genannte geistes- und sozialwissenschaftliche Berufe
Vi KLAS-Codes 9000-9379 Sonstige Dienstleistungsberufe
VI KLAS-Codes 9700-9979 Sonstige Arbeitskräfte
Because of gaps in the answers provided by respondents, the following "new" codes were created:
9711 - Mithelfende Familienangehörige außerhalb der Landwirtschaft, anderweitig nicht genannt
9811 - Auszubildende mit (noch) nicht feststehendem Ausbildungs-beruf
9821 - Praktikanten/Praktikantinnen, Volontäre/ Volontärinnen mit (noch) nicht feststehendem Beruf
9911 - Facharbeiter/innen, ohne nähere Tätigkeitsangabe
9921 - Heimarbeiter/innen, ohne nähere Tätigkeitsangabe
9931 - Vorarbeiter/innen, Gruppenleiter/innen, ohne nähere Tätigkeitsangabe
9971 - Sonstige Arbeitskräfte, ohne nähere Tätigkeitsangabe

pgklas10 – Current Occupational Classification (KldB2010)

1104	Officer	7
1203	Senior Non-Commissioned Officers and Higher	9
1302	Junior Non-Commissioned Officers	1
1402	Armed Forces Personnel in Other Ranks	49
11101	Occupations in Farming (without Specialisation)-Unskilled/Semiskilled Tasks	15
11102	Occupations in Farming (without Specialisation)-Skilled Tasks	76
11103	Occupations in Farming (without Specialisation)-Complex Tasks	1
11104	Occupations in Farming (without Specialisation)-Highly Complex Tasks	4
11113	Technical Occup. in Farming-Complex Tasks	0
11114	Technical Occup. in Farming-Highly Complex Tasks	0
11123	Agricultural Experts-Complex Tasks	0
11124	Agricultural Experts-High Complex Tasks	0
11132	Technical Laboratory Occup. in Agriculture-Skilled Tasks	1
11133	Technical Laboratory Occup. in Agriculture-Complex Tasks	0

11182	Occupations in Farming (with Specialisation, Not Elsewhere Classified)-Skilled Tasks	0
...	(1265 rows omitted)	19142
94704	Occupations in Museums (without Specialisation)-Highly Complex Tasks	2
94712	Technical Occup. in Museums and Exhibitions-Skilled Tasks	1
94713	Technical Occup. in Museums and Exhibitions-Complex Tasks	0
94714	Technical Occup. in Museums and Exhibitions-Highly Complex Tasks	0
94724	Art Experts-Highly Complex Tasks	0
94794	Managers in Museum	0
-1	No answer	8336
-2	Does not apply	12628
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	38398
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable is designed to provide annual data on job classification for all employed persons according to the classification of the German Federal Statistical Office (StaBuA). Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire.
Not all employed persons are asked the relevant input questions on an annual basis. Only those employed persons who changed jobs and first-time respondents were asked to provide up-to-date information.
In the SOEP-IS, pgklas10 was first implemented in wave 2016 and is coded [-5] in all previous waves.
German Classification of Occupations 2010. https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikationder-Berufe/KldB2010/Arbeitshilfen/EnglischeKldB2010/KldBEnglischl-Nav.html

pgautono – Autonomy in occupational activity

0	Apprentice	1554
1	Low Autonomy	3771
2	[2/5]	9774
3	[3/5]	16722
4	[4/5]	10325
5	High Autonomy	1684
-1	No answer	1071
-2	Does not apply	33769
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable gives the occupational autonomy for all employed persons. It offers an alternative to the ISCO-based scales on occupational status (PGISEI), class (PGEGP), or prestige (PGSIOPS). PGAUTONO is the simplest variable based on the scales of "occupational position" in terms of its construction, and strongly correlated with the Treiman Prestige Scale (PGSIOPS).
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.

7 Employment History

pgexpft – Working Experience Full-Time Employment

-1	No answer	840
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable reflects the total length of full-time employment in the respondent's career. 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). PGEXPFT uses calendar information up to December of the previous year and 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 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).
Because detailed information on the acitivity status of the respondents is not assessed in the Questionnaire of the SOEP Innovation Sample, PGEXPFT is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGEXPFT is coded to "-5" (not contained in questionnaire).
Please also see PGEXPPT and PGEXPUE.

pgexppt – Working Experience Part-Time Employment

0		9031
0.100000001490116		70
0.200000002980232		117
0.300000011920929		112
0.400000005960464		76
0.5		249
0.600000023841858		67
0.699999988079071		67
0.800000011920929		91
0.899999976158142		28
1		436
1.10000002384186		57
1.20000004768372		62
1.29999995231628		97
1.39999997615814		37
...	(279 rows omitted)	3868
40		14
40.2000007629395		1
41.2999992370605		1
42		2
42.2000007629395		1
43		1
-1	No answer	840
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable reflects the total length of part-time employment in the respondent's career. 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). PGEXPPT uses calendar information up to December of the previous year and 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 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).
Because detailed information on the activity status of the respondents is not assessed in the Questionnaire of the SOEP Innovation Sample, PGEXPPT is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGEXPPT is coded to "-5" (not contained in questionnaire).
Please also see PGEXPFT and PGEXPUE.

pgexpue – Unemployment Experience

0		10635
0.100000001490116		149
0.200000002980232		123
0.300000011920929		173
0.400000005960464		83
0.5		531
0.600000023841858		72
0.699999988079071		91
0.800000011920929		128
0.899999976158142		66
1		409
1.10000002384186		103
1.20000004768372		39
1.29999995231628		59
1.39999997615814		102
...	(149 rows omitted)	1715
24		1
24.5		1
25.2999992370605		1
26.3999996185303		2
32		1
33		1
-1	No answer	840
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

This variable reflects the total length of unemployment in the respondent's career. 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). PGEXPUE uses calendar information up to December of the previous year and 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 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).
Because detailed information on the activity status of the respondents is not assessed in the Questionnaire of the SOEP Innovation Sample, PGEXPUE is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGEXPUE is coded to "-5" (not contained in questionnaire).
Please also see PGEXPFT.

8 School, Higher and Vocational Education

pgcasmin – CASMIN Classification

0	(0) In School	878
1	(1a) Inadequately Completed	1225
2	(1b) General Elementary School	5266
3	(1c) Basic Vocational Qualification	18093
4	(2b) Intermediate General Qualification	3186
5	(2a) Intermediate Vocational	20241
6	(2c_gen) General Maturity Certificate	3575
7	(2c_voc) Vocational Maturity Certificate	7023
8	(3a) Lower Tertiary Education	4737
9	(3b) Higher Tertiary Education	12764
-1	No answer	1644
-2	Does not apply	23
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	15
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

As an alternative to PGISCED, a second educational variable is generated (PGCASMIN) that also enables comparison with international educational degrees/diplomas. Based on the modified CASMIN classification scheme (Comparative Analysis of Social Mobility in Industrial

Nations), this variable has been computed retroactively from 1984 on for all respondents. Taken into account are both secondary-level and university/college-level degrees and diplomas. As with PGISCED, the higher-level occupational degrees override the lower-level secondary school degrees.

pgfield – Field of tertiary education

1	General Language and Cultural Studies	0
2	Evangelical Theology, -Religious Studies	3
3	Catholic theology, -religious studies	0
4	Philosophy	0
5	History	2
6	Library Science, Documentation	12
7	General and Comparative Literature and Linguistics	4
8	Classical Philology, Modern Greek	0
9	German Studies (German, Germanic languages excluding English studies)	12
10	English Studies, American Studies	12
11	Romance Studies	0
12	Slavic Studies, Baltic Studies, Finno-Ugric Studies	0
13	Non-European Language and Cultural Studies	0
14	Cultural Studies in the narrow sense.	2
15	Psychology	5
...	(43 rows omitted)	563
83	Outside the study area structure	0
87		0
98	Not categorizable	57
99		0
151		0
996		0
-1	No answer	70
-2	Does not apply	14582
-3	Implausible value	1
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 PGBBIL02. While the latter variable records if a person holds a degree PGFIELD contains more detailed information on the type of the degree. The data of the generated variable PGFIELD 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 PGFIELD. 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 PGFIELD we combine these two types of information. Each year the variable contains the most recently collected information.
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 PGFIELD 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.
Because detailed information on the field of tertiary education is not assessed in the Questionnaire of the SOEP Innovation Sample, PGFIELD is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGFIELD is coded to "-5" (not contained in questionnaire).
Statistisches Bundesamt (2009): Bildung und Kultur. Studierende an Hochschulen, Fachserie 11, Reihe 4.1, Wiesbaden: 446ff, Übersicht 1: "Fächergruppen, Studienbereiche und Studienfächer".

pgdegree – Type of tertiary degree

1	Multi-subject Bachelor's with teaching option	0
2	Master's degree	0
3	Licentiate	0
4	Ecclesiastical examination	0
5	Certified translator	0
6	Promotion	0
7	Promotion without a required final examination	0
8	State examination / 1st state examination	0
9	State examination (single-phase training)	0
10	Faculty examination	0
11	Diploma (U)	12
12	Diploma (U) - Interpreter	164
13	Diploma (U) - Translator	6
14	Diploma I (U-GH)	2
15	LA Bachelor's Primary Schools	13
...	(77 rows omitted)	327
94	Certificate of completion	0
95	Other qualification in Germany	0
96	Final exam abroad	0

97	No final exam possible or sought.	0
98	LA Bachelor's in Primary and Secondary Schools	98
99	LA Master Primary and Secondary Schools	0
-1	No answer	120
-2	Does not apply	14582
-3	Implausible value	1
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 PGBBIL02. While the latter variable records if a person holds a degree PGDEGREE contains more detailed information on the type of the degree. The data of the generated variable PGDEGREE 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 PGDEGREE. 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 PGDEGREE 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.
Each year the variable contains the most recently collected information. 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). Since 2010 the data were coded according to the classification presented here. In the variable PGDEGREE 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 of 2009 are stored in the respective variables in the dataset P.
Because detailed information on the type of tertiary degree is not assessed in the Questionnaire of the SOEP Innovation Sample, PGDEGREE is not generated for Sample E (since 2012),

I (since 2011) and the supplementary samples (since 2012). For this purpose, PGDEGREE is coded to "-5" (not contained in questionnaire).
Statistiches Bundesamt (2009): Bildung und Kultur. Studierende an Hochschulen, Fachserie 11, Reihe 4.1, Wiesbaden: 449ff, Übersicht 2: „Prüfungsgruppen und Abschlussprüfungen".

pgtraina – Apprenticeship - two-digit occupation KldB92

1	Agricultural Occupations (Crops)	21
2	Agricultural Occupations (Livestock)	15
3	Administrative/Advisory/Technical Specialist In Agriculture	1
5	Horticultural Occupations	19
7	Mineworkers	3
8	Mineral Exploitation and Processing	1
13	Glass Manufacturing Occupations	1
14	Chemical Industry Occupations	4
15	Plastics Manufacturing Occupations	1
16	Paper Manufacturing and Processing	3
17	Printing Occupations	14
18	Wood and Woodworking, Wickerwork Occupations	1
22	Metal Processing Occupations (Chip Forming)	28
23	Occupations in the Metal Surface Treatment and Finishing Industry	1
25	Metal and Plant Construction Occupations	37
...	(2332 rows omitted)	1537
9921	Homeworkers without further specification	0
9931	Forepersons, group leaders without further specification	0
9941	Persons doing community service without further specification	0
9951	Self-employed persons without further specification	0
9961	Consultancy, planning professionals without further specification	0
9971	Other employees without further specification	0
-1	No answer	13
-2	Does not apply	13625
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable is designed to provide information on the occupation of vocational training which adds details to the information recorded in the variable PGBBIL01. In addition to the variable PGTRAINA we provide the variables PGTRAINB, PGTRAINC and PGTRAIND. All these variables record the occupation of vocational training. The difference is that PGTRAINA contains information on vocational training within the German dual system which combines firm-based and school-based training (apprenticeship). PGTRAINB is designed to provide information on the occupation of full-time school based vocational training. PGTRAINC contains information on level vocational training (e.g., Meister, Techniker). PGTRAIND is designed to provide information on the occupation of civil servant training ("Beamtenausbildung"). We describe in brief detail the construction of the variable PGTRAINA. PGTRAINB, PGTRAINC and PGTRAIND are constructed in an analogous manner.

The data of the generated variable PGTRAINA 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 PGTRAINA. 2. Biography questionnaire: 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 PGTRAINA we combine these two types of information. Each year the variable contains the most recently collected information.
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.
Because detailed information on the specific fields of occupation is not assessed in the Questionnaire of the SOEP Innovation Sample, PGTRAINA is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGTRAINA is coded to "-5" (not contained in questionnaire).
Hartmann/Schütz (2002): Die Klassifikation der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984–2001. Infratest Sozialforschung, München.

pgtrainb – Vocational school - two-digit occupation KldB92

1	Agricultural Occupations (Crops)	4
6	Forestry and Hunting Occupations	1
17	Printing Occupations	1
25	Metal and Plant Construction Occupations	5
26	Sheet Metal Manufacturing Occupations	11
27	Mechanical Engineering and Maintenance Occupations	2
28	Automotive and Aircraft Manufacturing and Maintenance Occupations	1
30	Precision Engineering and Related Occupations	7
41	Chefs	3
44	Occupations in Structural Engineering	1
48	Construction Finishing Occupations	1
50	Occupations in Woodworking and Polymer Processing	1
62	Technicians, not otherwise mentioned	2
66	Salespeople	4
67	Wholesale and Retail Salespeople, Purchasing and Sales Staff	13
...	(2293 rows omitted)	207
9921	Homeworkers without further specification	0
9931	Forepersons, group leaders without further specification	0
9941	Persons doing community service without further specification	0
9951	Self-employed persons without further specification	0
9961	Consultancy, planning professionals without further specification	0
9971	Other employees without further specification	0

-1	No answer	1
-2	Does not apply	14334
-3	Implausible value	2
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	64069
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

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 PGTRAINA for more details on the construction and the values of the variable.
Because detailed information on the specific fields of occupation is not assessed in the Questionnaire of the SOEP Innovation Sample, PGTRAINB is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGTRAINB is coded to "-5" (not contained in questionnaire).

pgtrainc – Higher vocational school - two-digit occupation KldB92

1	Agricultural Occupations (Crops)	3
2	Agricultural Occupations (Livestock)	2
3	Administrative/Advisory/Technical Specialist In Agriculture	3
5	Horticultural Occupations	1
6	Forestry and Hunting Occupations	1
17	Printing Occupations	1
26	Sheet Metal Manufacturing Occupations	1
27	Mechanical Engineering and Maintenance Occupations	1
28	Automotive and Aircraft Manufacturing and Maintenance Occupations	8
30	Precision Engineering and Related Occupations	1
31	Electrical Occupations	8
35	Occupations in Textile Processing	2
39	Occupations in Baking, Confectionery, and Candy Production	1
41	Chefs	1
48	Construction Finishing Occupations	2
...	(2301 rows omitted)	150
9921	Homeworkers without further specification	0
9931	Forepersons, group leaders without further specification	0
9941	Persons doing community service without further specification	0
9951	Self-employed persons without further specification	0
9961	Consultancy, planning professionals without further specification	0
9971	Other employees without further specification	0
-1	No answer	0
-2	Does not apply	14415
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	64069
-6	Version of questionnaire with modified filtering	0

-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable is designed to provide information on the occupation of higher level vocational training (e.g., Meister, Techniker). See the description of variable PGTRAINA for more details on the construction and the values of the variable.
Because detailed information on the specific fields of occupation is not assessed in the Questionnaire of the SOEP Innovation Sample, PGTRAINC is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGTRAINC is coded to "-5" (not contained in questionnaire).

pgtraind – Civil servant training - two-digit occupation KldB92

6	Forestry and Hunting Occupations	2
60	Engineers, not mentioned mentioned	4
62	Technicians, not otherwise mentioned	1
71	Surface Transport Occupations	2
76	Members of Parliament, Administrative Staff	15
78	Office Occupations, Commercial Staff, not otherwise mentioned	27
80	Security Occupations, not otherwise mentioned	37
81	Occupations in Law and Law Enforcement	10
82	Journalism, Translation, Library Science, and Similar Occupations	1
110	Farmers, general	0
111	Fruit and vegetable farmers (non-horticultural)	0
112	Arable farmers (special, permanent crops)	0
113	Livestock farmers and pasture farmers	0
114	Seed, crop producers, propagators (non-horticultural)	0
115	Crop protectors	0
...	(2273 rows omitted)	0
9921	Homeworkers without further specification	0
9931	Forepersons, group leaders without further specification	0
9941	Persons doing community service without further specification	0
9951	Self-employed persons without further specification	0
9961	Consultancy, planning professionals without further specification	0
9971	Other employees without further specification	0
-1	No answer	0
-2	Does not apply	11449
-3	Implausible value	2
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	67120
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The variable is designed to provide information on the occupation of civil servant training ("Beamtenausbildung"). See the description of variable PGTRAINA for more details on the construction and the values of the variable.

Because detailed information on the specific fields of occupation is not assessed in the Questionnaire of the SOEP Innovation Sample, PGTRAIND is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGTRAIND is coded to "-5" (not contained in questionnaire).

pgfdt_f – Data source FIELD, DEGREE, TRAIN

1	Individual Questionnaire	915
2	Gap Questionnaire (temporary drop-outs)	0
3	Biographical Questionnaire	1778
4	Various Sources	26
-1	No answer	0
-2	Does not apply	12606
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Because detailed information on PGFIELD, PGDEGREE and PGTRAIN is not assessed in the Questionnaire of the SOEP Innovation Sample, PGFDT_F is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGFDT_F is coded to "-5" (not contained in questionnaire).

pgbilztch – Change in Education since last survey / last year

0	Inconsistent educational information since last year	10331
1	Inconsistent educational information since last survey	322
2	Inconsistent educational information since last year	441
-1	No answer	0
-2	Does not apply	3507
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	64069
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Because detailed information on PGFIELD, PGDEGREE and PGTRAIN is not assessed in the Questionnaire of the SOEP Innovation Sample, PGBILZTCH is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGBILZTCH is coded to "-5" (not contained in questionnaire).

pgbilztev – Change in Education, total observed period

0	Consistent educational information	9346
1	Inconsistent educational decline	2015
2	Inconsistent educational increase	2553
3	Inconsistent educational decline and increase	56
-1	No answer	0
-2	Does not apply	1355
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	63345
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Because detailed information on PGFIELD, PGDEGREE and PGTRAIN is not assessed in the Questionnaire of the SOEP Innovation Sample, PGBILZTEV is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGBILZTEV is coded to "-5" (not contained in questionnaire).

pgsbil – Diplomas/degrees from secondary/tertiary

1	Sec. Gen. School Leaving Certificate	21429
2	Intermediate School Degree	23033
3	Leaving Certificate From Voc High School	5209
4	College Entrance Exam	18983
5	Other	6021
6	Dropout, No School Certificate	1310
7	Currently In School	892
-1	No answer	1793
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

All respondents in all SOEP subsamples are asked about diplomas/degrees attained for completion of secondary/tertiary education the first time they participate in SOEP.
As the SOEP Innovation Sample does not include a youth questionnaire, since 2012 information usually coming from the youth questionnaire was not included in the generation of PGSBIL.

pgbbil01 – Vocational degree attained

1	Apprenticeship	34304
2	Vocational School	7366
3	Health Care School	200
4	Technical School	4772
5	Civil Servant Training	2052
6	Other Degree	1099
7	Completed Vocational Training/Education in Germany	797
-1	No answer	1676
-2	Does not apply	26404
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

All respondents in all subsamples are asked about vocational degrees attained the first time they participate in SOEP. 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.
Since the SOEP Innovation Sample does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGBBIL01.

pgbbil02 – Completed college education

1	Fachhochschule	4762
2	University, Technical College	9151
3	College Not In Germany	155
4	Engineering, Techncial School (East)	1194
5	University (East)	1063
6	graduation, state doctorate	1163
-1	No answer	1676
-2	Does not apply	59506
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

All respondents in all subsamples are asked about completed college education the first time they participate in SOEP. To generate the variable, the different degrees/diplomas for all subsamples are integrated. Category (3) “college abroad” is only defined for persons who

completed a foreign-language version of the questionnaire. Generation of the variable entails combining the categories to make them compatible with the annual question about changes in vocational degrees/diplomas attained.
Since the SOEP Innovation Sample does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGBBIL02.

pgbbil03 – No vocational degree

1	No Vocation Degree	10184
2	Apprenticeship	4012
3	Studies	3183
-1	No answer	1676
-2	Does not apply	59615
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

In connection with the question about vocational degrees (PGBBIL01 and PGBBIL02), 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.
Since the SOEP Innovation Sample does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGBBIL03.

pgsbila – Secondary school degrees/diplomas abroad

1	Mandatory schooling not completed	247
2	Mandatory schooling completed	1914
3	Higher-level secondary school	3621
4	Secondary school completed abroad	0
-1	No answer	44
-2	Does not apply	72844
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

As a supplement to the PGSBIL, this variable provides annually updated data on the highest secondary school degree/diploma attained abroad.
Since the SOEP Innovation Sample does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGSBILA.

pgbbila – Occupational Training in abroad

1	On-The-Job Training	61
2	Vocational Training	126
3	Vocational School	100
4	College	139
5	Other Training	60
6	Vocational Degree[bbil01] Acquired Abroad	1
7	College Degree Completed Abroad	0
8	Completed Vocational Training/Education Other Country	0
-1	No answer	0
-2	Does not apply	14114
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	64069
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

As a supplement to the variable PGBBIL01, this variable gives (and updates) the highest-level vocational degree attained abroad.
Because detailed information on occupational training abroad is not assessed in the Questionnaire of the SOEP Innovation Sample, PGBBILA is not generated for Sample E (since 2012), I (since 2011) and the supplementary samples (since 2012). For this purpose, PGBBILA is coded to "-5" (not contained in questionnaire).

pgsbilo – Secondary school degree/diploma - East Germany

1	8th Gr. Completed	2919
2	10th Grade Completed	6081
3	Abitur, EOS	2427
4	Other	139
5	Dropout, No School Certificate	51
6	Currently In School	0
-1	No answer	13
-2	Does not apply	67036
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	4
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0

-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

As a supplement to the variable PGSBIL the highest secondary school degree/diploma in East Germany is provided as a separate variable.
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.
As SOEP-IS does not include a youth questionnaire, since 2012 information usually coming from the youth questionnaire was not included in the generation of PGSBILO.

pgbbilo – Vocational degree attained - East Germany

1	Vocational Training	6131
2	Master Craftsman	790
3	Engineering, Technical Degree	1335
4	Other Degree	124
-1	No answer	0
-2	Does not apply	69565
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	725
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

To supplement the variable PGBBIL01 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 PGBBIL03 if this degree has not been replaced by a more recently attained, higher-level university or college degree).
Since the SOEP-IS does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGBBILO.

pgbilzt – Amount of education or training (in years)

7	1043
8.5	153
9	5435
10	3653
10.5	14927
11	2605
11.5	13926

12		7126
13		4393
13.5		1540
14		1682
14.5		2417
15		4914
16		2213
17		296
18		9656
-1	No answer	1795
-2	Does not apply	896
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The following statements describe the standard computation for amount of education or training (in years).
years of education =years of schooling + years of occupational training
Schooling:

- `no degree <= 7 years`
- `lower school degree = 9 years`
- `intermediary school = 10 years`
- `degree for a professional coll. = 12 years`
- `high school degree = 13 years`
- `other = 10 years`

additional occupational training (includes universities):

- `apprenticeship = 1.5 years`
- `technical schools (incl. health) = 2 years`
- `civil servants apprenticeship = 1.5 years`
- `higher technical college = 3 years`
- `university degree = 5 years`

Note that for a high school degree 13 years of education are being taken into account, despite the changes of reducing high school by one year in most German federal states in the period 2001-2007.
Furthermore, the introduction of a Bachelor/Master System in the German higher education system in the early 2000´s is not yet reflected in the calculation of years in (higher) education. Hence, 5 years of university education is taken into account although the respondents could have finished in 3 years with the Bachelor's degree.

pgisced – Highest degree/diploma attained, ISCED-1997

0	(0) in school	863
1	(1) inadequately	1254
2	(2) general elemantary	9230
3	(3) middle vocational	37328
4	(4) vocational + Abi	6265
5	(5) higher vocational	4792
6	(6) higher education	17501
-1	No answer	1409
-2	Does not apply	1
-3	Implausible value	0
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	27
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

To make the educational degrees and diplomas attained in different countries comparable, for all respondents an educational variable (PGISCED) is generated retroactively using the international classification scheme ISCED-1997 (International Standard Classification of Education). It creates the highest degree/diploma attained, taking into account degrees and diplomas attained in both general schooling and in vocational and university education. Here the higher-level vocational and university override lower-level school diplomas. Persons who, for example, have no values for the variables on secondary school degrees/diplomas but state that they have 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 particular on doctoral degrees – we include all tertiary degrees in our 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. See below for more details.
Since the SOEP-IS does not include a youth questionnaire, the information usually coming from the youth questionnaire was not included in the generation of PGBBILA. Furthermore, since the year 2012, input information from PGBBILA is not being used as PGBBILA itself is not being generated.

9 Information on the Interview

pgmode – Interview Method

100	With Interviewer Assistance	30
110	Oral Interview	2511
120	Written Ques. No Interviewer	2038
130	Mixed Type	0
131	Written Ques. Interviewer	306
132	Oral And Written	202
133	Proxy	2

134	Third Person Present	0
135	No Third Person Present	0
140	CAPI - Since 1998 (O)	63441
150	CAWI Since 2014 (BE)	7338
160	CATI since 2021	946
170	CASI since 2021	8
180	CALVI	384
200	Telephone Assistance	0
210	Written, By Mail	181
220	Telephone Interview	0
230	CAPI by phone	120
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	1
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	1162
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

The interview method is generated via the answers to the questions in the individual questionnaire. Missing answers are filled in from the PBRUTTO files.

pgmonth – Month of interview

1	January	5083
2	February	5705
3	March	3461
4	April	1992
5	May	1587
6	June	2218
7	July	2764
8	August	2152
9	September	12906
10	October	22474
11	November	13074
12	December	5253
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	1
-4	Inadmissable multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	5339
-9	Missing due to a terminated interview	0

Month of interview is generated using the answers to the individual questionnaire. Missing answers are filled in using data from the HBRUTTO files.