SOEP The German Socio-Economic Panel study

482

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel study at DIW Berlin 2018

SOEP-Core v33.1 – Documentation of Household-related Status and Generated Variables in $HGEN

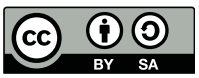
SOEP Group

DIW SOEP

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

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials
The SOEP Survey Papers are available at http://www.diw.de/soepsurveypapers

Please cite this paper as follows:
SOEP Group, 2018. SOEP-Core v33.1 – Documentation of Household-related Status and Generated Variables in $HGEN. SOEP Survey Papers 482: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



ISSN: 2193-5580 (online)
DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany
soeppapers@diw.de

The German Socio Economic Panel study at DIW Berlin

SOEP-Core v33.1 – Documentation of Household-related Status and Generated Variables in $HGEN

SOEP Group

2018

The files $HGEN are part of a collection, which is released with doi:10.5684/soep.v33.1.

Contents

1 General Information

The $$HGEN data provides a set of time-consistent variables generated from the SOEP household questionnaire. It only includes households who participated in the respective year. Up to Wave G, "old" households already known to the SOEP were surveyed at the old address with a "green" questionnaire; old households that had moved and new households received a "blue" questionnaire. While a number of questions in the blue questionnaire surveyed information for the first time (e.g. living space in square meters), this information was only asked for again in the green questionnaire in case of changes. Otherwise, the information collected in the previous wave was still valid. The variables described in the following are in part status variables in this sense: Information collected once is carried forward to subsequent waves if no address change has taken place since the previous wave. This is the case for: CNSTYRMIN$$, CNSTYRMAX$$, CONDIT$$ (for Wave B to G), SIZE$$, ROOM$$, EQPKIT$$, EQPSHW$$, EQPIWC$$, EQPHEA$$, EQPTER$$, EQPBAS$$, EQPGAR$$, EQPWAT$$, EQPALM$$, EQPSOL$$, EQPAIR$$, MOVEYR$$, ACQUIS$$, SUBSID$$, RSUBS$$, REDUC$$ and SEVAL$$ (for Wave C to G). Furthermore, identical information is recorded in the green and blue questionnaire in separate variables (e.g. housing tenure as owner or tenant). The corresponding status variables are therefore just a compilation of these more dispersed pieces of information. Since only one common questionnaire has been used since Wave H rather than the former "blue" and "green" versions, the necessity for the aforementioned status variables disappears. However, this user-friendly redundance is maintained for reasons of consistency.
In addition, we impute missing values of the variables ROOM$$, SIZE$$, RENT$$, HEAT$$, UTIL$$ and HINC$$ to ensure the completeness of the dataset and facilitate its usage. Details of the imputation process can be found in the description of the respective variable.
Since Wave BG, new migration samples (M3 and M4) have been added. These samples provide less household information which explains the high incidence of missing values (=-5). Therefore, in the samples M3 and M4 hgen variables are generally not imputed.

2 Sample Information

hhnr – Original Household Number

hhnrakt – Current Wave HH Number (=BGHHNR)

bghhnr – Current Wave HH Number (=HHNRAKT)

3 General Housing Information

owner$$ – Tenant Or Owner Of Dwelling [generic]

1	[1] Owner	6343
2	[2] Main Tenant	7822
3	[3] Sub-Tenant	292
4	[4] Tenant	0
5	[5] Living In A Home	44
-1	[-1] No Answer	1
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320

-6	[-6] Questionnaire Version With Modified Filter	0

Up to Wave H, the information for OWNER$$ was collected in separate questionnaires for "old" and first-time respondents, respectively ("blue" and "green" questionnaires). In all waves and for each household with the original value of -1 ("no answer"), codes are set to =1 and =4 if at least one owner- or tenant-specific item is given. Code 4 is also used if a change in ownership (from owner to tenant) has taken place but no original information for OWNER$$ was given. Code 5 ('resident of a home or institutional living facility') has only been assigned by interviewers during fieldwork since Wave P. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

moveyr$$ – Year Moved Into Dwelling [generic]

1907		1
1909		1
1915		1
1922		1
1924		2
1925		2
1926		2
1927		3
1928		1
1929		4
1930		6
1931		3
1933		5
1934		4
1935		7
...	(67 rows omitted)	5531
2003		372
2004		406
2005		459
2006		469
2007		496
2008		478
2009		532
2010		622
2011		704
2012		828
2013		964
2014		1238
2015		2149
2016		1209
-1		1322

MOVEYR$$ contains the year of moving into the household of the person who answers the household questionnaire. For old households at their old address data is carried forward for up to two years. For new households in SOEP and for old households that have moved the variable is based on newly collected data. In case the information is missing and an old

household has moved that year or the previous year, MOVEYR$$ is given the value of the year of the respective wave.
The carrying forward of data entails the possibility that the year of moving into the new dwelling may lie before the year of birth of the oldest household member. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

cnstyrmin$$ – Earliest Possible Construction Year of Dwelling [generic]

0		1557
1919		1744
1949		3672
1972		1392
1973		90
1974		107
1975		87
1976		70
1977		36
1978		62
1979		47
1980		88
1981		816
1982		54
1983		43
...	(19 rows omitted)	2265
2003		66
2004		77
2005		72
2006		100
2007		67
2008		60
2009		46
2010		69
2011		91
2012		54
2013		24
2014		71
2015		53
2016		15
-1		4827

CNSTYRMIN$$ provides the lower limit of the time period in which the household's building was constructed. CNSTYRMAX$$ provides the upper limit. E.g. if a household's CNSTYRMIN$$ is 1984 and CNSTYRMAX$$ is 1990, the building was built between 1984 and 1990.
Households new to the SOEP and households who have moved since the last interview are asked for the time period of construction of the building they live in. With this information CNSTYRMIN$$ and CNSTYRMAX$$ are constructed. For old households the variables are carried forward from the last years. Since Wave X, households are also asked for the exact construction year. In these cases, CNSTYRMIN$$ and CNSYTRMAX$$ contain the same

value. If a building was built before 1919, CNSTYRMIN$$ contains a zero and CNSTYRMAX$$ 1919.
In Wave BE, the question about the construction year was not included in the SOEP. All information for CNSTYRMIN$$ and CNSTYRMAX$$ derives from the previous year.
CNSTRMIN$$ and CNSTRMAX$$ replace CNSTYR$$ that has previously been distributed. The latter was a categorial variable grouping about one decade each into one category. Hence, with the last category having a flexible upper limitation and being split up regularily, sorting was ambiguous in some cases. For example, a household responding in Wave L that his house has been built "1981 or later" was not unambiguously assignable to either "1981-1990" or "1990 or later". Using CNSTRYRMIN$$ and CNSTYRMAX$$, changing questions can be far better described. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

cnstyrmax$$ – Latest Possible Construction Year of Dwelling [generic]

1918		1557
1948		1744
1971		3672
1972		160
1973		90
1974		107
1975		87
1976		70
1977		36
1978		62
1979		47
1980		1320
1981		59
1982		54
1983		43
...	(19 rows omitted)	2795
2003		87
2004		104
2005		112
2006		152
2007		68
2008		65
2009		47
2010		144
2011		79
2012		57
2013		26
2014		71
2015		57
2016		23
-1		4827

CNSTYRMAX$$ provides the upper limit of the time period in which the household's building was constructed. The generation of CNSTYRMAX$$ is analogous to CNSTYRMIN$$.

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

condit$$ – Condition Of House [generic]

1	[1] In A Good Condition	0
2	[2] Some Renovations	0
3	[3] Full Renovations	0
4	[4] Dilapidated	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Respondent's subjective assessment of the condition of the building. In the West German sub-samples from Wave B to Wave G, the information on CONDIT$$ was collected only for new households and for households with a residential move since the previous year (households with "blue" questionnaires). For households who had not moved ("green" questionnaire), the information collected in previous years was carried forward. The wording in the questionnaire was changed in the first wave of the East German sub-sample (Wave G) to better capture the rundown condition of some residential buildings in East Germany. Since Wave H the wording is identical for the entire SOEP-sample in East and West Germany.
Since Wave BG, CONDIT$$ is not asked anymore and information is not carried forward.
[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

acquis$$ – Means Of Acquiring Dwelling [generic]

1	[1] Bought From Owner	1134
2	[2] Inheritance, Gift	435
3	[3] Bought, Built New	609
4	[4] Got Back From Public Property	0
-1	[-1] No Answer	4166
-2	[-2] Does Not Apply	8158
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

If a household did not provide any information on ACQUIS$$, the information from previous waves is carried forward under the condition that a change of address or ownership status (OWNER$$) has not taken place.
In Wave A to Wave G, ACQUIS$$ was asked only if a household was new to the SOEP or if it had changed its address. From Wave H to Wave R, ACQUIS$$ was also asked if a change in ownership status had taken place in the last year. Since Wave S, ACQUIS$$ is only asked if a change in ownership status has taken place. In consequence, households new to the SOEP without a change in ownership in the last year, do not state any information on ACQUIS$$ anymore. Furthermore, since Wave S, the category „Bought New or Built" is no

longer included in the household questionnaire but is carried forward from last years. Given these changes in the categories and the sample that provides information for ACQUIS$$, we recommend not to compare the shares of the different categories of ACQUIS$$ over time.
„Returned to private ownership“ was surveyed only in Wave I in East Germany and is carried forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

reval$$ – Rent Of Dwelling Vs Comparable Dwellings [generic]

1	[1] Very Inexpensive	0
2	[2] Inexpensive	0
3	[3] Reasonable	0
4	[4] Slightly Expensive	0
5	[5] Too Expensive	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Subjective assessment by respondent (household head). This variable was not surveyed in Wave T and Wave U. The corresponding information from the previous year is not carried forward longitudinally due to the possibility of changes in rent and income, residential moves, and change in the person responding.
Since Wave BF, REVAL$$ is not asked and information is neither carried backward nor forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

seval$$ – Adequacy Of Living Space In Housing Unit [generic]

1	[1] Much Too Small	699
2	[2] A Bit Too Small	2594
3	[3] Just Right	11369
4	[4] A Bit Too Large	1696
5	[5] Much Too Large	245
-1	[-1] No Answer	22
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1197
-6	[-6] Questionnaire Version With Modified Filter	0

Subjective assessment by respondent (household head). From Wave C to Wave G, the information on SEVAL$$ was only collected for new households or households that had moved (households with “blue” questionnaires) and immobile households whose SIZE$$ had changed. In these Waves, SEVAL$$ is carried forward from previous years. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

4 Equipment of the Dwelling

eqpkit$$ – Dwelling Has Kitchen [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Only in Wave H, Wave O, and from Wave V onwards, EQPKIT$$ is asked from every household. In previous years, the variable was only collected from new households and households who had moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPKIT$$ is carried forward from previous years. Additionally, from Wave B on, the information is updated if the household has stated that it modernized its kitchen since January of the previous year. Since Wave BE, EQPKIT$$ is neither asked nor carried forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpshw$$ – Dwelling Has Indoor Bath,Shower [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Only in Wave H, Wave O, and from Wave V onwards, EQPSHW$$ is asked from every household. In previous years, the variable was only collected from new households and households who had moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPSHW$$ is carried forward from the previous years. Additionally, from Wave B onwards, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. Since Wave BE, EQPSHW$$ is not asked anymore. Since Wave BF, the information is not carried forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpiwc$$ – Dwelling Has Indoor Toilet [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0

-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Only in Wave H, Wave O, and from Wave V to Wave Y, EQPIWC$$ is asked from every household. In previous years, the variable was only collected from new households and households who had moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPIWC$$ is carried forward from the previous years. Additionally, from Wave B onwards, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. In a second step, the information of the variable EQPSHW$$ is used to replace missing and inconsistent values of EQPIWC$$. For every household with EQPSHW$$=1, EQPIWC$$ is set to 1. Since Wave Z, EQPWC$$ is not asked. Since Wave BA, the information is neither carried backward nor forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqphea$$ – Dwelling Has Central,Floor Heat [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Only in Wave H, Wave O, and from Wave V onwards, EQPHEA$$ is asked from every household. In previous years, the variable was only collected from new households and households who had moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPHEA$$ is carried forward from the previous years. Additionally, from Wave G onwards the information is updated if the household has stated that it modernized its heating since January of the previous year. Since Wave BF, EQPHEA$$ is not asked and the information is neither carried backward nor forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpter$$ – Dwelling Has Balcony,Terrace [generic]

1	[1] Yes	12909
2	[2] No	3707
-1	[-1] No Answer	7
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1199
-6	[-6] Questionnaire Version With Modified Filter	0

Except for Wave H, Wave O, and from Wave V onwards, EQPTER$$ was only collected from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPTER$$ is carried forward from the previous years. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpbas$$ – Dwelling Has Basement [generic]

1	[1] Yes	14978
2	[2] No	1632
-1	[-1] No Answer	7
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1205
-6	[-6] Questionnaire Version With Modified Filter	0

Except for Wave H, Wave O, and from Wave V onwards, EQPBAS$$ has been collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPBAS$$ is carried forward from the previous years. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpgar$$ – Dwelling Has Garden [generic]

1	[1] Yes	9147
2	[2] No	7465
-1	[-1] No Answer	11
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1199
-6	[-6] Questionnaire Version With Modified Filter	0

Except for Wave H, Wave O, and from Wave V onwards, EQPGAR$$ has been collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPGAR$$ is carried forward from the previous years. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpwat$$ – Dwelling Has Water, Bioler [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0

-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Except for Wave H, Wave O, and from Wave V onwards, EQPWAT$$ has been collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for EQPWAT$$ is carried forward from the previous years. Since Wave BE, EQPWAT$$ is not asked and since Wave BF, the information is neither carried backward nor forward. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqptel$$ – Dwelling Has Telefone [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

The question on whether a household has a telephone or not has varied over the years. In some years, households were asked very generally if they had a “telephone“. In other years they were asked more specifically for a “landline telephone“ and a “cellphone“. EQPTEL$$ is set to 1 if a household stated that it had a telephone, landline telephone, or cellphone. Additionally, the information from $HBRUTTO ($HTEL) is used to complete missing information. In Wave N, Wave P, and Wave Z, the household questionnaire did not ask about the possession of a telephone and $HBRUTTO is the only source of data. For Wave K, no information is available.
Since Wave BG, EQPTEL$$ is no longer asked in the household questionnaire (except for the samples M3 and M4 as EQPTEL$$ was asked in the migration household questionaire). For this reason, it is no longer provided since Wave BG for any other samples than M3 and M4. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpalm$$ – Dwelling Has Alarm System [generic]

1	[1] Yes	842
2	[2] No	15757
-1	[-1] No Answer	15
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1208
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpsol$$ – Dwelling Has Solar System [generic]

1	[1] Yes	1442
2	[2] No	13044
-1	[-1] No Answer	16
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpair$$ – Dwelling Has Air Conditioner [generic]

1	[1] Yes	308
2	[2] No	14178
-1	[-1] No Answer	16
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqplif$$ – Dwelling Has An Elevator [generic]

1	[1] Yes	1727
2	[2] No	14879
-1	[-1] No Answer	16
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	1200
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpnrj$$ – Dwelling Has Alternative Energy Source [generic]

1	[1] Yes	539
2	[2] No	13945
-1	[-1] No Answer	18

-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpfhea$$ – Equipment: Floor Heating [generic]

1	[1] Yes	2855
2	[2] No	13663
-1	[-1] No Answer	101
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	1203
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpnobar$$ – Equipment: Barrier Free Furnishing [generic]

1	[1] Yes	2313
2	[2] No	14266
-1	[-1] No Answer	17
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	1226
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpmglass$$ – Equipment: at Least Double Glazing [generic]

1	[1] Yes	13412
2	[2] No	1066
-1	[-1] No Answer	24
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	3320
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpinsul$$ – Equipment: Thermal Insulation [generic]

1	[1] Yes	8682
2	[2] No	5701
-1	[-1] No Answer	119
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	3320
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqppark$$ – Equipment: Garage/ Parking Space [generic]

1	[1] Yes	10483
2	[2] No	6109
-1	[-1] No Answer	19
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	1211
-6	[-6] Version of questionnaire with modified filtering	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

5 Costs of Living, Size, and Rooms (Imputed Variables)

size$$ – Size Of Housing Unit In Sq M [generic]

6	1
8	1
9	3
10	6
11	4
12	22
13	2
14	8
15	27
16	11
17	3
18	15
19	5
20	64
21	6

...	(220 rows omitted)	16182
330		6
340		1
345		1
346		1
350		6
360		6
365		1
370		2
380		1
400		9
425		1
430		1
470		1
-1		1
-5		1424

Up to Wave R (except for Wave O), SIZE$$ was collected only in the first interview with new households. In case a household had moved or whether a household's housing unit size had changed due to renovations or additions (up to Wave G, these households filled out a "green" questionnaire). From Wave S onwards, the question on the size of the housing unit has been posed to all households annually. For households still residing at their old address and neither having moved nor renovated their dwelling, the information on the size of the housing unit is carried forward as a status variable in order to provide valid current information. Analogously, the information is carried back to fill gaps if households missed to state the size of the housing in the first year after moving.
In the case the information on the size of the housing unit is still missing, it is imputed simultaneously with other variables using Stata's chained imputation procedure. Imputed values are not carried forward. For information on the imputation procedure, see the description of RENT$$. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

room$$ – Number Of Rooms Larger Than 6 Sq M [generic]

1	676
2	2558
3	5070
4	3402
5	2226
6	1402
7	698
8	339
9	131
10	65
11	27
12	15
13	4
15	4
16	1
-5	1204

Up to Wave R (except for Wave O), ROOM$$ was collected only in the first interview with new households. In case a household had moved or (from Wave H onwards) whether a household's housing unit size changed due to renovations or additions (up to Wave G, these households filled out a "green" questionnaire). In Wave O, the information had been asked again in order to correct for mistakes that may have occurred in carrying forward or in the process of imputation. From Wave S onwards, the question on the number of rooms has been asked to all households annually.
For old households still residing at their old address and neither having moved nor renovated their dwelling, the information on the number of rooms is carried forward as a status variable in order to provide valid current information. Analogously, the information is carried back to fill gaps if a households missed to state the number of rooms in the first year(s) they moved to a new housing.
In case, the information on the number of rooms is still missing, it is imputed simultaneously with other variables using Stata's chained imputation procedure. Imputed values are not carried forward. For information on the imputation procedure, see the description of RENT$$. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

rent$$ – Amount Of Rent Minus Heating Costs [generic]

0		3
10		1
37		1
39		1
41		1
42		1
45		1
50		2
51		1
60		1
75		1
80		1
81		1
84		1
85		1
...	(892 rows omitted)	7717
2330		1
2355		1
2450		1
2770		1
2800		1
2811		1
2850		1
3235		1
3350		1
3600		1
4038		1
5100		1
-1		5

-2	6750
-5	3320

RENT$$ is a measure of gross rent, i.e. it equal to the base rent plus utility costs (UTIL$$) but does not include heating (HEAT$$), electricity costs (ELECTR$$) and additional gas costs (GAS$$). RENT$$ is converted into Euro values for all waves including those prior to Wave S.
The questions for RENT$$ and utility costs have changed considerably since the first SOEP questionnaire in Wave A. From Waves A to Wave G, the amount of rent stated by the households in SOEP is in principle the desired concept of gross rent, i.e. basic rent excluding heating and electricity costs but including utility costs. However, in these waves the information on utility costs was not collected.
From Wave H onwards, households simply state the amount of rent they pay. Following this question, it is asked whether heating and utility costs are included in that amount of rent and what the exact costs for heating and utilities eventually are (in the latter case only if they are included).
In Wave BE and Wave BF, the questionnaire design was different to all other waves. In a first step, households were asked for their overall monthly housing costs (sum of base rent, heating costs, additional utility costs, electricity and gas). In a second step, they stated the respective expenses. In order to obtain RENT$$, heating, gas, and electricity costs have to be deducted from these overall monthly housing costs. In Wave BE, some households state monthly housing costs that presumably do not include all possible expenses. They might do so because they do not know all their expenses or because they misunderstood the question. We attribute the drop of rent from Wave BD to Wave BE and Wave BF to the different questionnaire design and do no longer adjust the values as we did in data release v31. Therefore, the average rent of Wave BE (data release v31) is about 10 Euros lower in data release v32 and v33 than it was in v31. Both, in Wave BE and Wave BF, implausible values of rent (e.g. if utility costs are above total living costs) have been set to missing to be imputed in the following step.
The questionnaire design in Wave BG was set back to the format of Wave BD.
Missing values of RENT$$ are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address and the dwelling's size has remained the same. The type of rent (heating/utility costs included or not) is then also carried forward. In a second step, if RENT$$ is still missing, values of the two subsequent years are carried backwards in the same manner. In a third step, the remaining missing values of RENT$$ are imputed by Stata's chained imputation procedure. Imputed values are not carried forward.

General Information on the Imputation Procedure Since data release v31, missing values of ROOM$$, SIZE$$, RENT$$, HEAT$$, UTIL$$, and ELECTR$$ are imputed with Stata's chained imputation procedure (mi impute chained). We impute separately for each wave, for East and West and for owners and tenants. The regression specifications vary between these groups but in general they include:

- For owner:
  - ROOM$$, SIZE$$, CONDIT$$, CNSTYRMAX$$
  - Number of persons in household
  - Household net income as stated in questionnaire
  - Type of dwelling (family house, apartment, etc.)
- For tenants:

- ROOM$$, SIZE$$, RENT$$, HEAT$$, UTIL$$, ELECTR$$, CONDIT$$, CNSTYRMAX$$, EQPHEA$$, EQPFHEA$$, EQPTER$$, EQPGAR$$, EQPSOL$$, EQPAIR$$, CNSTYRMAX$$, REVAL$$, REDUC$$
- SIZE$$ squared
- Number of persons in household
- Household net income as stated in questionnaire
- Type of dwelling (family house, apartment, etc., $$wum1)
- Type of tenant
- Dummy: HEAT$$ included in RENT$$ (Wave G to Wave BD)
- Dummy: UTIL$$ included in RENT$$ (Wave H to Wave BD)
- Dummy: UTIL$$ partly included in RENT$$ (Wave H to Wave BD)
- Dummy: HEAT$$ includes electricity costs (since Wave BE)
- Dummy: household pays only utility costs (since Wave BE)
- Household has children younger than 16

In addition, the regression of RENT$$ contains RSUBS$$, a dummy for children under 16, information on the residential area ($$wum3) and BIK regions. The regression of HEAT$$ includes a dummy for children under 16. The regression of UTIL$$ includes EQPLIF$$ and omits RENT$$, CONDIT$$, the type of dwelling and the dummy for whether HEAT$$ is included in RENT$$. The regression of SIZE$$ omits SIZE$$ squared.
RENT$$ is not normalized across households before imputation and may include heating and utility costs or not. Solely partially included utility costs are subtracted. We use dummies for HEAT$$/UTIL$$ included in RENT$$ to account for the different types of rent. Note that these dummies are also imputed if they are missing. If UTIL$$ contains a partial amount, it is set to missing and then also imputed.
Residents of homes (OWNER$$==5) are excluded from the imputation. Values larger than four times the 99th percentile in one of the target variables are also excluded.
For all six target variables, the predictive mean matching imputation method is applied. We use 175 iterations as burn-in period. Furthermore, we only distribute one imputed value for each observation.
Gross warm rent including electricity (Bruttowarmmiete inkl. Strom) could be computed from this data as follows: RENT$$ + HEAT$$ + ELECTR$$.
RENT$$ is set missing (=-5) for all M3 and M4 respondents although the migration household questionnaire surveyed rent information. However, the surveyed information for M3 and M4 is not comparable with the information for the other samples as information on HEAT$$, UTIL$$, ELECTR$$, etc. is not surveyed. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

heat$$ – Costs Of Warm Water, Gen [generic]

2	1
5	1
7	1
9	2
10	9
13	2
14	1
15	5
17	2

18		3
19		2
20		58
21		1
22		3
23		4
...	(201 rows omitted)	7406
312		1
316		1
320		3
343		1
350		3
361		2
375		1
400		5
420		1
450		3
481		1
500		2
-1		164
-2		6813
-5		3320

Heating costs are collected since Wave C. Until Wave BD, only tenants who stated to pay a rent were asked. Since Wave BE, every tenant was asked. Heating costs are reported in Euro for all years.
In Wave BF, HEAT$$ has not been asked for directly. Instead, respondents stated their costs for different energy sources (district heating, gas, liquefied gas, electricity, heating oil, coal, wood, biomass, solar, and geothermal heat) and whether they use them for heating, warm water, cooking, and/or electronic devices respectively (if applicable). To generate HEAT$$, costs for cooking and electricity costs other than heating have been excluded by regressing the usage-dummies and controlling for household-specific (number of people living in the household, type of tenant) and dwelling-specific variables (number of rooms, size of the dwelling, construction year, living environment). The shares for heating and warm water were then added to generate HEAT$$. Due to changes in the questionaire design, firstly, HEAT$$ is lower in Wave BF than it has been in previous and following years and secondly, HEAT$$ from Wave BF is neither carried backward nor forward. In hgen, only HEAT$$ is distributed. All energy cost variables can be found in the dataset bfh.
Missing values of HEAT$$ are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address, the dwelling's size has remained the same and the rent has not changed considerably. In a second step, if HEAT$$ is still missing, values of the two subsequent years are carried backwards in the same manner. In a third step, the remaining missing values of HEAT$$ are imputed by Stata's chained imputation procedure. Imputed values are not carried forward. In contrast to previous versions of the SOEP, we do not impute HEAT$$ for Wave A and Wave B anymore, since data release v32. For more information on the imputation see the see the description of RENT$$. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

util$$ – Other Monthly Utility Costs [generic]

0	1
3	1
4	3
5	6
6	3
7	1
8	4
9	4
10	45
11	3
12	7
13	2
14	2
15	42
17	7
...	(193 rows omitted) 6657
303	2
320	2
323	1
325	1
337	1
350	2
359	1
400	1
420	1
431	1
450	1
500	1
-1	169
-2	7530
-5	3320

Additional utility costs have been collected since Wave H in East German households and since Wave J in West German households. Until Wave BD only tenants who stated to pay a rent were asked, since Wave BE every tenant. Up to Wave BD, households had to state whether the additional utility costs were fully, partly or not included in the stated rent. Households with fully or partially included utility costs then had to state the included amount. In consequence, for all households with partially and no included utility costs, the total amount of UTIL$$ is unknown.
UTIL$$ only includes tenants' utility costs even though these have been asked from owners in the exactly same question in Wave BF.
The considerable share of missing values of UTIL$$ are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address, the dwelling's size has remained the same and the rent has not changed considerably. In a second step, if UTIL$$ is still missing, values of the two subsequent years are carried backwards in the same manner. In a third step, the remaining missing values of UTIL$$ are imputed by Stata's chained imputation procedure. Imputed values are not carried forward. We do not impute UTIL$$ from Wave A to Wave

G and for West German households from Wave H to Wave J. For more information on the imputation see the description of RENT$$.
Utility costs are reported in Euro. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

electr$$ – Costs of Electricity (gen) [generic]

3		1
5		1
9		1
10		13
11		3
12		3
13		3
14		3
15		28
16		3
17		5
18		8
19		7
20		116
21		15
...	(167 rows omitted)	7592
230		4
239		1
240		3
250		7
260		2
275		2
280		2
297		1
300		5
320		1
328		1
400		1
-1		162
-2		6508
-5		3320

Costs for electricity (excluding arising expenses for heating if applicable) have been collected starting from Wave BA if tenants pay a rent. From Wave BE onwards, they have been collected from all tenants. In Wave BF, total electricity costs were explicitly asked for so that costs for heating and warm water were substracted to generate ELECTR$$ that year. Furthermore, due to changes in the questionaire design in Wave BF, ELECTR$$ is neither carried backward nor forward. See HEAT$$ for further explanations on housing costs in Wave BF.
Missing values of ELECTR$$ are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address, the dwelling's size has remained the same and the rent has not changed considerably. In a second step, if ELECTR$$ is still missing, values of the two subsequent years

are carried backwards in the same manner. In a third step, the remaining missing values of ELECTR$$ are imputed by Stata's chained imputation procedure. Imputed values are not carried forward. We do not impute ELECTR$$ for years before Wave BA. For more information on the imputation procedure, see the see the description of RENT$$. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

gas$$ – Costs of Gas (EUR) [generic]

-5 17822

GAS$$ states the monthly costs for gas not used for heating in Euros. GAS$$ was only collected in Wave BA. Missing values of GAS$$ are not imputed due to the low number of households which state gas costs. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

norent$$ – Does Not Pay Rent [generic]

1	[1] Pays No Rent	407
-1	[-1] No Answer	4
-2	[-2] Does Not Apply	14090
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3321
-6	[-6] Questionnaire Version With Modified Filter	0

NORENT$$ marks tenants who do not a pay rent. Until Wave BD, it was a filter question for all housing costs such that non-rent paying tenants were not asked for HEAT$$, UTIL$$ and ELECTR$$. In Wave BE, the filter question was rephrased and tenants were asked for their utility costs even if they stated not to pay a rent. To maintain consistency over time, we set NORENT$$ to 1 for all tenants who only pay utility costs. RENTINFO$$ contains the information of the new filter question. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

rentinfo$$ – Pays Rent and/or Utilities [generic]

1	[1] Pays Rent	7678
2	[2] Pays Utility Costs, but No Rent	254
3	[3] Does Not Pay either Rent or Utilities	153
-1	[-1] No Answer	74
-2	[-2] Does Not Apply	6343
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

RENTINFO$$ is a filter question for housing costs. It was introduced in Wave BE and replaces NORENT$$. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

heatinfo$$ – Reason for Missing Heating Costs [generic]

1	[1] Heating Costs Unknown	0
3	[3] No Heating Costs	470
-1	[-1] No Answer	14032
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

HEATINFO$$ was introduced in Wave BE and indicates why HEAT$$ is missing. If a household stated that it did not know its heating costs, we imputed HEAT$$ and set HEATINFO$$ to -2. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

utilinfo$$ – Reason for Missing Additional Utility Costs [generic]

1	[1] Utility Costs Unknown	0
3	[3] No Other Additional Utility Costs	1439
-1	[-1] No Answer	13063
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

UTILINFO$$ was introduced in Wave BE and indicates why UTIL$$ is missing. If a household stated that it did not know its additional utility costs, we impute UTIL$$ and set UTILINFO$$ to -2. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

electrinfo$$ – Reason for Missing Electricty Costs [generic]

1	[1] Electricity Costs Unknown	0
2	[2] Electricity Costs Included in Heating Costs	0
3	[3] No Electricity Costs	165
-1	[-1] No Answer	14337
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

ELECTRINFO$$ was introduced in Wave BE and indicates why ELECTR$$ is missing. If a household stated that it did not know its electricity costs, we impute ELECTR$$ and set ELECTRINFO$$ to -2. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

gasinfo$$ – Reason for Missing Gas Costs [generic]

1	[1] Gas Costs Unknown	0
2	[2] Electricity Costs Included in Heating Costs	0
3	[3] No Gas Costs	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

GASINFO$$ has only been asked in Wave BE and indicates why a household did not state additional gas costs that were not included in the heating costs. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

fsize$$ – Imputation Flag, Size Of Housing [generic]

0	[0] Not imputed	16376
1	[1] Imputed	1446
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FSIZE$$ indicates whether or not SIZE$$ has been imputed (=1) by Stata's chained imputation procedure. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

froom$$ – Imputation Flag, Number Of Rooms Lager Than 6 Sq M [generic]

0	[0] Not imputed	16615
1	[1] Imputed	1207
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FROOM$$ indicates whether or not ROOM$$ has been imputed (=1) by Stata's chained imputation procedure. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

frent$$ – Imputation Flag, Amount Of Rent Minus Heating Costs [generic]

0	[0] Not imputed	15448
1	[1] Rent and/or Sub-Aggregate Imputed by PMM	2111
2	[2] Rent and/or Sub-Aggregate from Previous or Subsequent Years	152
3	[3] Rent and/or Sub-Aggregate From Prev./Subs. Years and/or PMM	111
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

FRENT$$ indicates whether RENT$$ or expenses added or subtracted to RENT$$ has been imputed (=1) by Stata's chained imputation procedure, whether RENT$$ itself or expenses added or subtracted to RENT$$ was carried for- or backwards (=2) from past/subsequent waves or whether RENT$$ is the result of a mixture of imputed and past/subsequent waves' values. The fourth category marks households who very likely did not include all utility costs in their overall housing costs in Wave BE (the question from which RENT$$ is generated). We identified these households by their longitudinal information on housing costs and did not subtract HEAT$$ and/or ELECTR$$ to obtain their RENT$$. F2RENT$$ contains more detailed information about which imputed expenses were added or subtracted. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

f2rent$$ – Detailed Imputation Flag, Rent Minus Heating Costs [generic]

0	[0] Not imputed	14886
1	[1] Reported Rent Corrected for Imputed Heat	516
2	[2] Reported Rent Corrected for Imputed Util	1491
3	[3] Reported Rent Corrected for Imputed Electr	55
4	[4] Reported Rent Corrected for Heat and Util	60
5	[5] Reported Rent Corrected for Heat and Electr	124
6	[6] Reported Rent Corrected for Util and Electr	36
7	[7] Reported Rent Corrected for Heat, Util and Electr	10
8	[8] Rent Imp., Sub-Aggregates Not Imputed	514
9	[9] Rent Imp., Sub-Aggregates Not Imputed	38
10	[10] Rent Imp. and Corrected For Imp. Heat	1
11	[11] Rent Imp. and Corrected For Imp. Util	72
12	[12] Rent Imp. and Corrected For Imp. Electr	0
13	[13] Rent Imp. and Corrected For Imp. Heat and Util	0
14	[14] Rent Imp. and Corrected For Imp. Heat and Electr	0
15	[15] Rent Imp. and Corrected For Imp. Util and Electr	0
16	[16] Rent Imp. and Corrected For Imp. Heat, Util and Electr	0

17	[17] Rent Imp. & Imp. Heating Costs /Levies â€" Filter Question	19
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

F2RENT$$ indicates whether or not RENT$$ has been imputed (=1) by Stata's chained imputation procedure. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

fheat$$ – Imputation Flag, Costs Of Heating And Warm Water [generic]

0	[0] Not imputed	16914
1	[1] Imputed by PMM	700
2	[2] Electricity Costs from Previous or Subsequent Years	208
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FHEAT$$ indicates whether HEAT$$ has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

futil$$ – Imputation Flag, Other Monthly Utility Costs [generic]

0	[0] Not imputed	15113
1	[1] Imputed by PMM	2375
2	[2] Electricity Costs from Previous or Subsequent Years	334
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FUTIL$$ indicates whether UTIL$$ value has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

felectr$$ – Imputation Flag, Costs of Electricity [generic]

0	[0] Not imputed	17499
1	[1] Imputed by PMM	230
2	[2] Electricity Costs from Previous or Subsequent Years	93
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FELECTR$$ indicates whether ELECTR$$ value has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

6 Subsidization of Housing Costs

subsid$$ – Government Subsidizes Housing Payments [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Statement by respondent. SUBSID$$ contains information on government subsidies at the time the housing was built or bought. From Wave B to Wave N, this was only asked to new households or in case an old household had moved. Information is then carried forward. In Wave O and Wave P, the question was again posed to the whole population. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

osubs$$ – Amount Of Subsidies Last Year [generic]

1	[1] Yes	0
2	[2] No	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	17822
-6	[-6] Questionnaire Version With Modified Filter	0

Statement by respondent. OSUBS$$ contains information on cash housing subsidies received by owners from the government during the year prior to the interview. Wave W was the last wave house owners could apply for the 8-years-long financial support such that since Wave BF, OSUBS$$ is not asked any longer. Information is not carried forward.

Please note: The old variable $FOERD (available until SOEP data release v25) is discarded. Homeowner subsidies in Germany have been subject to major revisions and fluctuations over time. The corresponding question in SOEP was in some waves only posed to new households and those that have moved, in some waves it was not surveyed at all. For these reasons, the question for government housing subsidies was changed in Wave Q to cover direct subsidies received the previous year. SUBSID$$ and OSUBS$$ replace the old variable $FOERD. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

rsubs$$ – Government Subsidized Housing [generic]

1	[1] Yes, With Due Diligence	534
2	[2] Yes, With Run Out Due Diligence	13
3	[3] No	7542
-1	[-1] No Answer	70
-2	[-2] Does Not Apply	6343
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

RSUBS$$ states whether the rent is subsidized by government or not. Up to Wave K, the question was asked only to new households and households who had moved since last year. For the remaining households, the information is carried forward from previous waves. East German households were asked for the first time in Wave J.
In Wave L, the second response category was added to indicate expired subsidization. For reasons of time series consistency, RSUBS$$ was coded with "3" for "no" from Wave A to Wave K. The rewording of the response categories became necessary due to the carrying forward of data: It was impossible to identify whether a housing unit had lost its subsidization status for any period of time. Thus, for population estimates there is a distinct possibility that RSUBS$$ produces increasing overestimations of government-subsidized housing units up to Wave K. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

reduc$$ – Dwelling At A Reduced Prize [generic]

1	[1] Yes	618
2	[2] No	7527
-1	[-1] No Answer	14
-2	[-2] Does Not Apply	6343
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	3320
-6	[-6] Questionnaire Version With Modified Filter	0

REDUC$$ states whether the flat is offered by the owner at a reduced rent. Information is carried forward from previous waves for old households residing at their old address. For newly surveyed households and for old households that have moved, newly collected data is used. From Wave T to Wave X, this information was not collected. It is carried forward

from Wave S for households who have not moved and whose stated amount of rent vary only slightly. The new information from Wave Y is then carried backward for households with the same characteristics if REDUC$$ is still missing after carrying forward from Wave S. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

7 Typology of Household

typ1hh$$ – Household Typology, One Digit [generic]

1	[1] 1-Pers.-HH	4549
2	[2] Couple Without Children	4226
3	[3] Single Parent	1844
4	[4] Couple With Children LE 16	4519
5	[5] Couple With Children GT 16	1030
6	[6] Couple With Children LE And GT 16	1029
7	[7] Multiple Generation-HH	158
8	[8] Other Combination	467
9	[9] No Answer	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

typ2hh$$ – Household Typology, Two Digit [generic]

11	[11] 1-Person HH Male LE 35 Y.	1154
12	[12] 1-Person HH Male 35- LT 60 Y.	727
13	[13] 1-Person HH Male GE 60 Y.	560
14	[14] 1-Person HH - Female LE 35 Y.	376
15	[15] 1-Person HH Female 35- LT 60 Y.	576
16	[16] 1-Person HH Female GE 60 Y.	1156
21	[21] Couple Without Children	4226
31	[31] Single Parent,1 Child, LE 16	551
32	[32] Single Parent,2 Or More Children, LE 16	380
33	[33] Single Parent,1 Children, GT 16	523
34	[34] Single Parent,2 Or More Children, GT 16	117
35	[35] Single Parent,2 Children, LE and GT 16	123
36	[36] Single Parent,3 Or More Children, LE and GT 16	150
41	[41] Couple, 1 Child, LE 16	1281
42	[42] Couple, 2 Children, LE 16	1810
...	(2 rows omitted)	2092
52	[52] Couple, 2 Children, GT 16	276
53	[53] Couple, 3 Or More Children, GT 16	90

61	[61] Couple, 2 Children, LE and GT 16	325
62	[62] Couple, 3 Or More Children, LE and GT 16	704
71	[71] 3-Generation-HH	141
72	[72] 4-Generation-HH	0
73	[73] GrandParents-GrandChildren-HH	17
81	[81] Other Combination Without K. LE 16	467
82	[82] Other Combination With K. LE 16	0
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

Generated variable created by combining the relationships of all persons living in the household to the head of household (Variable $STELL in the file $PBRUTTO) at the time of the survey. With data release v26 the data production process switched to a standardized procedure for all Waves to ensure longitudinal consistency, resulting in minor changes compared with older distributions. TYP1HH$$ is an aggregation of TYP2HH$$ (first column of the two-digit code). Single households are differentiated in TYP2HH$$ according to both gender and age. The following abbreviations are used in the labels: LE=less or equal, LT=less than, GE=greater or equal, GT=greater than.
Help for old friends: Starting with data distribution v26 (Wave A to Wave Z), the category "(88) Other combination" has been further differentiated into households with vs. those without children (up to the age of 16).
Legend:

- K = children up to the age of 16
- EK = adult children age 17 and older
- (E)K = children both below and above age 16
- 1-P-HH = one-person households.

[This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

8 Household Income

hinc$$ – Monthly Household Net Income (EUR) [generic]

0	13
50	2
77	1
84	1
100	2
105	1
109	1
110	3
112	1
120	1
124	1

130		6
135		11
136		1
140		2
...	(1817 rows omitted)	16907
18000		1
19000		1
20000		6
20800		1
22000		1
23000		1
25000		2
28200		1
30000		2
32000		2
40000		2
40200		1
45000		1
-1		768
-3		78

This variable contains the current monthly net household income asked for in the household questionnaire, always provided in Euros which was introduced in January 2002 (1 Euro = 1.95583 DM). Income is reported by the respondent (head of household). [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

i1hinc$$ – 1. Imputed Monthly Net Household Income (EUR) [1/5] [generic]

0		14
50		2
77		1
84		1
100		2
105		1
109		1
110		3
112		1
120		1
124		1
130		6
135		12
136		1
140		2
...	(1815 rows omitted)	17745
16000		4
17000		2
18000		1
19000		1
20000		6

20800	1
22000	1
23000	1
25000	2
28200	1
30000	2
32000	2
40000	2
40200	1
45000	1

Multiple imputation procedures provide a way to deal with missing values on the variable Current Monthly Net Household Income by using information about components and determinants of the household income and replacing item-nonresponse with multiply imputed data. The first five imputations are available within the $HGEN datasets: the variables I1HINC$$-I5HINC$$.
The imputations were calculated using multiple imputations by chained equations. Up to Wave BB the program ICE of Stata which was written by Patrick Royston (see Royston 2004, 2005a, 2005b) and which is based on the program MICE in S-Plus and R was used. Since Wave BC the Stata command mi impute is used. The missing observations are assumed to be missing at random. We set the number of imputations m=10 and get 10 multiple imputed values for I_HINC$$. For a discussion on the choice of m, see Rubin (Wave D) and Royston (Wave U).
The dataset MIHINC contains the complete imputation results and is separately available. To be compatible with methods for analyzing multiply imputed data, MIHINC is constructed in the so called stacked or MIM Dataset Format. It contains the following variables: HHNRAKT, SVYYEAR, MJ, MI, IHINC and IMPFLAG. For all household from Wave L to Wave BD, there are ten imputed values for the current household income. MJ identifies the individual dataset to which each observation belongs while MI identifies the observations within each individual dataset. To distinguish between the original data containing missing values and the imputed values, the dummy variable IMPFLAG is added. In the $HGEN files five of these imputed incomes are stored in the conventional wide format.
The number of iterations carried out in each prediction model was specified to be 500. For East and West Germany, imputations were done separately. Furthermore, the option for predicted mean matching was chosen which 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.
Most important variables for modelling the current household net income consist in the household net income of the previous year, in basic information about the household and changes in its composition as well as all relevant income components received.

The complete list of the variables used for modelling

- Description of household:
  - size, number of children, sample
  - head of household: not German, age, sex
  - changes in household composition between years: births, deaths, persons entering or leaving the household or being temporarily absent
- Financial Situation:
  - Monthly household income previous year

  - Income from employment
  - Pensions
  - Sum of personal incomes (e.g. Support from the "Arbeitsamt", Maternity benefit, Alimony, etc.)
  - Household related incomes (e.g. Child allowance, Housing assistance, Social assistance, Unemployment benefit, Assets, etc.)
  - Fraction of persons greater than 16 in household who refused answering a component of income (0-1)
- Number of persons not attended survey (PUNR, partial unit nonresponse)
- Cross-sectional weights

Analyzing multiply imputed data For analyzing multiple imputed data, you do not necessarily need special methods, however such tools exits 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 various built-in functionality called mi.
- Within SAS, the MIANALYZE procedure combines the results of the analyses of imputations and generates valid statistical inferences: `http://support.sas.com/rnd/app/stat/procedures/mianalyze.html`
- 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.

[This information can be related to a specific variable and is not necessarily generic.]
Royston, Patrick (2004): Multiple imputation of missing values. In: Stata Journal 4(3): 227-241. Royston, Patrick (2005a): Multiple imputation of missing values: update. In: Stata Journal 5(2): 188-201. Royston, Patrick (2005b): Multiple imputation of missing values: Update of ice. In: Stata Journal 5(4): 527-536. Rubin, D.B. (1987): Multiple imputation for non-response in surveys. New York.
For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

i2hinc$$ – 2. Imputed Monthly Net Household Income (EUR) [2/5] [generic]

0	13
50	2
77	1
84	1
100	2
105	1
109	1
110	3
112	1
120	1
124	1
130	7

135		11
136		2
140		2
...	(1815 rows omitted)	17745
16000		4
17000		2
18000		1
19000		1
20000		6
20800		1
22000		1
23000		1
25000		2
28200		1
30000		2
32000		2
40000		2
40200		1
45000		1

[This information can be related to a specific variable and is not necessarily generic.]

i3hinc$$ – 3. Imputed Monthly Net Household Income (EUR) [3/5] [generic]

0		15
50		2
77		1
84		1
100		2
105		1
109		1
110		3
112		1
120		1
124		1
130		7
135		12
136		1
140		2
...	(1815 rows omitted)	17743
16000		4
17000		2
18000		1
19000		1
20000		6
20800		1
22000		1
23000		1
25000		2
28200		1

30000		2
32000		2
40000		2
40200		1
45000		1

[This information can be related to a specific variable and is not necessarily generic.]

i4hinc$$ – 4. Imputed Monthly Net Household Income (EUR) [4/5] [generic]

0		13
50		3
77		1
84		1
100		2
105		1
109		1
110		3
112		1
120		1
124		1
130		7
135		12
136		1
140		2
...	(1815 rows omitted)	17744
16000		4
17000		2
18000		1
19000		1
20000		6
20800		1
22000		1
23000		1
25000		2
28200		1
30000		2
32000		2
40000		2
40200		1
45000		1

[This information can be related to a specific variable and is not necessarily generic.]

i5hinc$$ – 5. Imputed Monthly Net Household Income (EUR) [5/5] [generic]

0		14
50		2
77		1
84		1

100		2
105		1
109		2
110		3
112		1
120		1
124		1
130		7
135		12
136		1
140		2
...	(1815 rows omitted)	17743
16000		4
17000		2
18000		1
19000		1
20000		6
20800		1
22000		1
23000		1
25000		2
28200		1
30000		2
32000		2
40000		2
40200		1
45000		1

[This information can be related to a specific variable and is not necessarily generic.]

fhinc$$ – Imputation Flag, Monthly Net Household Income [generic]

0	[0] Not imputed	16976
1	[1] Imputed	846
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

FHINC$$ is a dummy variable indicating whether an observation was missing on HINC$$ and was therefore imputed or not. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

nuts1$$ – NUTS-Systematic-1 (Federal State) [generic]

1	[1] Baden-Wuerttemberg	2111
2	[2] Bavaria	2875

3	[3] Berlin	807
4	[4] Brandenburg	758
5	[5] Bremen	134
6	[6] Hamburg	338
7	[7] Hesse	1227
8	[8] Mecklenburg-West Pomerania	394
9	[9] Lower Saxony	1633
10	[10] North Rhine-Westphalia	3552
11	[11] Rhineland-Palatinate	845
12	[12] Saarland	232
13	[13] Saxony	1021
14	[14] Saxony-Anhalt	589
15	[15] Schleswig-Holstein	673
16	[16] Thuringia	633
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

NUTS ("Nomenclature of Statistical Territorial Units") is a hierarchical system for dividing up the economic territory of the European Union. It was introduced by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics. NUTS 1 especially contains the major socio-economic regions for analyzing regional Community problems. It subdivides the European Union by now into 97 regions, whereas in Germany there are equivalent to the German Federal States 16 regions. Before Wave Q, Rhineland-Palatinate and Saarland were defined as one region. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

9 Time and Method of Interview

hmonth$$ – Month Of Interview [generic]

1	[1] January	125
2	[2] February	4570
3	[3] March	3029
4	[4] April	2261
5	[5] May	1717
6	[6] June	749
7	[7] July	702
8	[8] August	1569
9	[9] September	1162
10	[10] October	557
11	[11] November	831
12	[12] December	550
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0

-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

The month of participation in the survey is generated using data from the household questionnaire. Missing information is filled in using data from the corresponding $HBRUTTO files. Interviews that took place in the month of December, and prior to the 20th of that month, were recoded to –3. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690), <pkrause@diw.de>

hmode$$ – Interview Method [generic]

100	[100] With Interviewer Assistance	23
110	[110] Oral Interview	833
120	[120] Written Ques. Interviewer	1621
130	[130] Mixed Type	0
131	[131] Written Ques. No Interviewer	84
132	[132] Oral And Written	132
133	[133] Proxy	1
134	[134] Third Person Present	0
135	[135] No Third Person Present	0
140	[140] CAPI - Wave O Onwards	12870
150	[150] CAWI - Wave BE (2014) Onwards	767
200	[200] Telephone Assistance	0
210	[210] Written, By Mail	1490
220	[220] Phone Interview	1
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Not Valid	0
-4	[-4] Forbidden Multiple Response	0
-5	[-5] Not Included In Questionnaire Version	0
-6	[-6] Questionnaire Version With Modified Filter	0

The interview method is generated through data from the household questionnaire. Missing information is filled in with data from the corresponding $HBRUTTO files. [This information can be related to a specific variable and is not necessarily generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690), <pkrause@diw.de>