SOEP The German Socio-Economic Panel study

409

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel study at DIW Berlin 2017

SOEP-Core v32 – Documentation of Household-related Status and Generated Variables in $HGEN

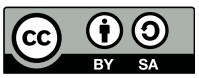
SOEP Group

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, 2017. SOEP-Core v32 – Documentation of Household-related Status and Generated Variables in $HGEN. SOEP Survey Papers 409: 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 v32 – Documentation of Household-related Status and Generated Variables in $HGEN

SOEP Group

2017

The files $HGEN are part of a collection, which is released with doi:10.5684/soep.v32.

Contents

1 General Information

The $$HGEN data provides a set of time-consistent variables generated from the SOEP household questionnaire. It only includes households who participated in the respective year. Up to 1990, "old" households already known to the SOEP were surveyed at the old address with a "green" questionnaire; old households that had moved and new households received a "blue" questionnaire. While a number of questions in the blue questionnaire surveyed information for the first time (e.g., living space in square meters), this information was only asked for again in the green questionnaire in the case of changes. Otherwise, the information collected in the previous year was still valid. The variables described in the following are in part status variables in this sense: information collected once is carried forward to subsequent years if no address change has taken place since the previous year. This is the case for: CNSTYRMIN$$, CNSTYRMAX$$, CONDIT$$ (for years 1985 to 1990), SIZE$$, ROOM$$, EQPKIT$$, EQPSHW$$, EQPIWC$$, EQPHEA$$, EQPTER$$, EQPBAS$$, EQPGAR$$, EQPWAT$$, EQPALM$$, EQPSOL$$, EQPAIR$$, MOVEYR$$, ACQUIS$$, SUBSID$$, RSUBS$$, REDUC$$ and SEVAL$$ (for years 1986 to 1990). Furthermore, identical information is recorded in the green and blue questionnaire in separate variables (e.g. housing tenure as owner or tenant). The corresponding status variables are therefore just a compilation of these more dispersed pieces of information. Since only one common questionnaire has been used since Wave H (1991) rather than the former "blue" and "green" versions, the necessity for the aforementioned status variables disappears but this "user-friendly redundance" is maintained for reasons of consistency.
In addition, we impute missing values of the variables ROOM$$, 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.

2 Sample Information

hhnr – Original Household Number

hhnrakt – Current Wave Number (=BFHHNR)

bfhhnr – Current Wave HH Number

3 General Housing Information

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

1	[1] Owner	6784
2	[2] Main Tenant	8841
3	[3] Sub-Tenant	323
4	[4] Tenant	0
5	[5] Living In A Home	43
-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

Up to 1991, the information for OWNER$$ was collected in separate questionnaires for "old" and first-time respondents, respectively ("blue" and "green" questionnaires). In all waves, codes 1 and 4 are used if the original variable is coded as -1 ("no answer"), but if at least one answer that is specific to owners, respectively to tenants, was given. Code 4 is also used if a change in ownership (from owner to tenant) has taken place, but no original information for OWNER$$ was given. Code 5 ('resident of a home or institutional living facility') has only been assigned by interviewers during fieldwork since 1999. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

moveyr$$ – Year Moved Into Dwelling [generic]

1907		1
1909		1
1915		1
1922		1
1923		1
1924		2
1925		3
1926		2
1927		3
1928		1
1929		5
1930		8
1931		3
1933		5
1934		5
...	(67 rows omitted)	5789
2002		383
2003		423
2004		470
2005		512
2006		545
2007		570
2008		552
2009		647
2010		747
2011		895
2012		1100
2013		1290
2014		1390
2015		523
-1		113

MOVEYR$$ contains the year of moving into the household of the person who answers the household questionnaire. For old households at their old address, data is carried forward for up to two waves. For new households in SOEP and for old households that have moved, the variable is based on newly collected data. In case the information is missing and an old household has moved that year or the previous year, MOVEYR$$ 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0		1708
1919		1914
1949		4070
1972		1505
1973		92
1974		120
1975		98
1976		79
1977		37
1978		63
1979		57
1980		98
1981		880
1982		53
1983		46
...	(18 rows omitted)	2406
2002		60
2003		74
2004		84
2005		82
2006		107
2007		73
2008		65
2009		54
2010		74
2011		91
2012		53
2013		22
2014		65
2015		14
-1		1847

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, the CNSTYRMIN$$ and CNSTYRMAX$$ are constructed. For old households, the variables are carried forward from the last years. Since 2007, 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 2014, the question about the construction year was not included in the SOEP. All information for CNSTYRMIN$$ and CNSTYRMAX$$ derives from the past years.
CNSTRMIN$$ and CNSTRMAX $$ replace CNSTYR$$ that has previously been distributed. The latter was a categorial variable grouping about one deceny each into one category. Hence, with the last category having a flexible upper limitation and being split up regularily, sorting was ambiguous in some cases. For example, a household responding in 1995, that his house has been built "1981 or later" was not unambiguously assignable to either "1981-1990" or "1991 or later". Using CNSTRYRMIN$$ and CNSTYRMAX$$, changing questions can be far better described. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1918		1708
1948		1914
1971		4070
1972		181
1973		92
1974		120
1975		98
1976		79
1977		37
1978		63
1979		57
1980		1422
1981		70
1982		53
1983		46
...	(18 rows omitted)	2955
2002		82
2003		96
2004		118
2005		129
2006		166
2007		74
2008		70
2009		55
2010		135
2011		86
2012		57
2013		24
2014		65
2015		22
-1		1847

CNSTYRMAX$$ provides the upper limit of the time period in which the household's building was constructed. CNSTYRMIN$$ provides the lower limit. The generation of CNSTYRMAX$$ is analogous to CNSTYRMIN$$. [This information can be related to a specific variable and is not necessary generic.]

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

condit$$ – Condition Of House [generic]

1	[1] In A Good Condition	10975
2	[2] Some Renovations	4583
3	[3] Full Renovations	398
4	[4] Dilapidated	21
-1	[-1] No Answer	14
-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

Respondent's subjective assessment of the condition of the building. In the West German sub-samples from 1985 to 1990, information on CONDIT$$ was collected only for new households and for households with a residential move since the previous wave (households with "blue" questionnaires). For households who had not moved ("green" questionnaire), information collected in previous waves was carried forward. The wording in the questionnaire was changed in the first wave of the East German sub-sample in 1990 as to better capture the rundown condition of some residential buildings in East Germany. Since 1991 the wording is identical for the entire SOEP-sample in East and West Germany. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

acquis$$ – Means Of Acquiring Dwelling [generic]

1	[1] Bought From Owner	1154
2	[2] Inheritance, Gift	467
3	[3] Bought, Built New	673
4	[4] Got Back From Public Property	0
-1	[-1] No Answer	4490
-2	[-2] Does Not Apply	9207
-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

If a household does not provide information on ACQUIS$$, information from previous years if carried forward under the condition that a change of address or ownership status (OWNER$$) has not taken place.
In 1984 to 1990, ACQUIS$$ was asked only if a household was new to the SOEP or if it had changed its address. In 1991 to 2001, ACQUIS$$ was also asked if a change in ownership status had taken place in the last year. Since 2002, 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 2002, 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 1992 in East Germany, but is also carried forward. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Subjective assessment by respondent (household head). This variable was not surveyed in 2003 and 2004. 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. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	[1] Much Too Small	460
2	[2] A Bit Too Small	2410
3	[3] Just Right	11074
4	[4] A Bit Too Large	1773
5	[5] Much Too Large	260
-1	[-1] No Answer	14
-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

Subjective assessment by respondent (household head). From 1986 to 1990, 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 necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Only in 1991, 1998, and from 2005 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 the previous years. Additionally, from 1985 on, the information is updated if the household has stated that it modernized its kitchen since January of the previous year. Since 2014, EQPKIT$$ is not asked anymore. In v32, information is not carried forward. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Only in 1991, 1998, and from 2005 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 1985 on, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. Since 2014, EQPSHW$$ is not asked anymore. In v32, information is not carried forward. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Only in 1991, 1998, and from 2005 to 2008, 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 1985 on, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. In a second step, 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. Beginning with 2009, data is no more collected. In v32, information is not carried forward. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Only in 1991, 1998, and from 2005 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 1990 on, the information is updated if the household has stated that it modernized its heating since January of the previous year. In 2015, EQPHEA$$ is no longer asked for and information is not carried forward. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	[1] Yes	12886
2	[2] No	3098
-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	0
-6	[-6] Questionnaire Version With Modified Filter	0

Except for 1991, 1998, and 2005 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpbas$$ – Dwelling Has Basement [generic]

1	[1] Yes	14678
2	[2] No	1307
-1	[-1] No Answer	6
-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

Except for 1991, 1998, and 2005 onwards, EQPBAS$$ was only been collected from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place the information for EQPBAS$$ is carried forward from the previous years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpgar$$ – Dwelling Has Garden [generic]

1	[1] Yes	9446
2	[2] No	6534
-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	0
-6	[-6] Questionnaire Version With Modified Filter	0

Except for 1991, 1998, and 2005 onwards, EQPGAR$$ was only been collected from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place the information for EQPGAR$$ is carried forward from the previous years. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Except for 1991, 1998, and 2005 onwards, EQPWAT$$ was only been collected from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place the information for EQPWAT$$ is carried forward from the previous years. Beginning from 2014, EQPWAT$$ is no longer asked for. In v32, information is not carried forward. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqptel$$ – Dwelling Has Telefone [generic]

1	[1] Yes	15954
2	[2] No	36
-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	0
-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 when a household stated that it had a telephone, landline telephone, or cellphone. In addition, information from $HBRUTTO ($HTEL) is used to complete missing information. In the years 1997, 1999, and 2009, the household questionnaire contained no information on the possession of a telephone and $HBRUTTO is the only source of data. For 1994, no information is available. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpalm$$ – Dwelling Has Alarm System [generic]

1	[1] Yes	607
2	[2] No	15358
-1	[-1] No Answer	26
-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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpsol$$ – Dwelling Has Solar System [generic]

1	[1] Yes	1587
2	[2] No	14384
-1	[-1] No Answer	20
-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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpair$$ – Dwelling Has Air Conditioner [generic]

1	[1] Yes	302
2	[2] No	15670
-1	[-1] No Answer	19
-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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqplif$$ – Dwelling Has An Elevator [generic]

1	[1] Yes	1609
2	[2] No	14367
-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	0
-6	[-6] Questionnaire Version With Modified Filter	0

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

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

1	[1] Yes	571
2	[2] No	15392
-1	[-1] No Answer	28
-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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

eqpfhea$$ – Equipment: Floor Heating [generic]

1	[1] Yes	2971
2	[2] No	12874
-1	[-1] No Answer	146
-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

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

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

1	[1] Yes	2208
2	[2] No	13758
-1	[-1] No Answer	25
-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

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

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

1	[1] Yes	14287
2	[2] No	1619
-1	[-1] No Answer	85
-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

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

eqpinsul$$ – Equipment: Thermal Insulation [generic]

1	[1] Yes	8613
2	[2] No	6939
-1	[-1] No Answer	439
-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

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

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

1	[1] Yes	10882
2	[2] No	5036
-1	[-1] No Answer	73
-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

[This information can be related to a specific variable and is not necessary 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]

10		3
11		3
12		13
13		4
14		6
15		14
16		7
17		3
18		8
19		3
20		48
21		4
22		12
23		7
24		15
...	(221 rows omitted)	15802
330		5

340	3
346	1
350	11
356	1
360	4
365	1
370	1
400	6
420	1
430	1
435	1
450	1
470	1
-1	1

Up to 2001, with an exception in 1998, SIZE$$ was collected only in the first interview with new households, in case a household had moved or with old households which still resided at their old address but whose housing unit size had changed due to renovations or additions (up to 1990, these households filled out a "green" questionnaire). From 2002 onwards the question on the size of the housing unit has been posed to all households annually. For households still residing at their old address and neither having moved nor renovated their dwelling, the information on the size of the housing unit is carried forward as a status variable in order to provide valid current information. Analogously, information is carried back, to fill gaps if households missed to state the size of the housing in the first year after moving. In the case the information on the size of the housing unit is still missing, it is imputed simultaneously with other variables using a chained imputation procedure. Imputed values are not carried forward. For information on the imputation procedure, see the description of RENT$$. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	541
2	2334
3	4800
4	3237
5	2288
6	1463
7	715
8	339
9	163
10	67
11	21
12	13
13	5
14	1
15	3
16	1

Up to 2001 with an exception in 1998 ROOM$$ was collected only in the first interview with new households, in case a household had moved or from 1991 onwards with old households which still resided at their old address but whose housing unit size had changed due to renovations or additions (up to 1990, these households filled out a “green” questionnaire). In 1998, the information had been asked again in order to correct for mistakes that may have occurred in the carrying forward of data or in the process of imputation. From 2002 onwards the question on the number of rooms has been asked to all households annually.
For old households still residing at their old address and neither having moved nor renovated their dwelling, the information on the number of rooms is carried forward as a status variable in order to provide valid current information. Analogously, information is carried back, to fill gaps if households missed to state the number of rooms in the first year(s) they moved to a new housing.
In the case the information on the number of rooms is still missing, it imputed simultaneously with other variables using a chained imputation procedure. Imputed values are not carried forward. For information on the imputation procedure, see the description of RENT$$.
[This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0		9
1		6
2		1
10		1
20		1
22		1
23		1
25		2
28		1
30		1
34		2
35		1
37		1
38		1
39		1
...	(1104 rows omitted)	8738
2427		1
2444		1
2450		1
2465		1
2650		1
2670		1
2811		1
2915		1
3230		1
3252		1
3500		1
4400		1
4920		1
-1		8

-2 7202

RENT$$ is a measure of the gross rent, i.e., it equals the base rent plus utility costs (UTIL$$), but does not include heating (HEAT$$), electricity (ELECTR$$) and additional gas costs (GAS$$). RENT$$ is converted into Euro values for all years, including those prior to 2002. The questions for RENT$$ and utility costs have changed considerably since the first SOEP questionnaire in 1984. In Waves A (1984) to G (1990), the amount of rent stated by the households in SOEP is in principle the desired concept of gross rent, i.e., basic rent excluding heating and electricity costs but including utility costs. In these waves, however, information on utility costs was not collected.
From Wave H (1991) on, households simply state the amount of rent they pay. Following this question, it is asked whether heating and utility costs are included in that amount of rent and what the exact costs for heating and utilities eventually are (in the latter case only if they are included).
In 2014 and 2015, households are asked, in a first step, for their overall monthly housing costs (sum of base rent, heating costs, additional utility costs, electricity, and gas). In a second, step they state the respective expenses. In order to obtain RENT$$, heating, gas, and electricity costs have to be deducted. Due to the 2014 questionnaire design, 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 misunderstand the question. We attribute the drop of rent from 2013 to 2014 and 2015 to the different questionnaire design and do no longer adjust the values as we did in v31. Therefore, the average rent of 2014 is about 10 Euros lower in v32 than it was in v31. Both, in 2014 and 2015, 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.
Missing values of RENT$$ are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address and the dwelling's size has remained the same. The type of rent (heating/utility costs included or not) is also carried forward then. In a second step, if RENT$$ 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 SOEP 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 year, for East and West and for owner 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$$ (1990-2013)
- Dummy: UTIL$$ included in RENT$$ (1991-2013)
- Dummy: UTIL$$ partly included in RENT$$ (1991-2013)
- Dummy: HEAT$$ includes electricity costs (since 2014)
- Dummy: household pays only utility costs (since 2014)
- 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 99-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 for the burn-in period. Furthermore, we only distribute the first imputed value for each observation.
Gross warm rent including electricity (Bruttowarmmiete inkl. Strom) could be computed from this data as follows: RENT$$ + HEAT$$ + ELECTR$$. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

heat$$ – Amount Of Heating At Hot Water Costs (EUR) [generic]

0		186
1		3
3		2
4		4
5		5
6		6
7		14
8		23
9		28
10		24
11		30
12		47
13		30
14		37
15		37
...	(258 rows omitted)	8697
379		1
380		1

400	2
430	1
450	1
464	1
484	1
487	1
500	2
526	1
540	1
600	1
800	1
-1	24
-2	6779

Heating costs are collected since 1986. Until 2013 only tenants who stated to pay a rent were asked, since 2014 every tenant. Heating costs are reported in Euro for all years.
In 2015, 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 (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 are then added to generate HEAT$$. Due to the changed questionaire, HEAT$$ is lower in 2015 than it was in previous years. In hgen, only HEAT$$ is distributed. All energy cost variables can be found in bfh.
Missing values of HEAT$$ 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$$ for the years 1984 and 1985. For more information on the imputation see the see the description of RENT$$.ate HEAT$$. Due to the changed questionaire, HEAT$$ is lower in 2015 than it was in previous years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

util$$ – Other Monthly Utility Costs [generic]

1	1
2	1
3	9
4	8
5	13
6	10
7	7
8	21

9		13
10		41
11		6
12		16
13		17
14		9
15		52
...	(217 rows omitted)	6137
330		2
340		2
350		2
358		1
379		1
380		1
400		4
430		1
450		3
465		1
479		1
481		1
500		2
600		1
-2		9607

Additional utility costs are collected since 1991 in East German households and since 1993 in West German households. Until 2013 only tenants who stated to pay a rent were asked, since 2014 every tenant. Up to 2013, 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 2015.
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 1984 to 1990 and for West German households from 1991 to 1993. For more information on the imputation see the description of RENT$$.
Utility costs are reported in Euro for all years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	3
1	3
2	2

3		2
4		4
5		5
6		4
7		5
8		3
10		18
11		3
12		8
13		4
14		7
15		23
...	(215 rows omitted)	8911
260		2
262		1
266		1
270		1
273		1
283		1
292		1
300		3
308		2
335		1
341		1
370		1
690		1
-1		20
-2		6949

Costs for electricity have been collected starting in 2010 if tenants pay a rent. From 2014 onwards, they are collected from all tenants. In 2015, total electricity costs have explicitly been asked for, so that costs for heating and warm water have been substracted to generate ELECTR$$ that year. See HEAT$$ for further explanations.
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 2010. 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

gas$$ – Costs of Gas (EUR) [generic]

-5	15991

GAS$$ states the monthly costs for gas not used for heating in Euros. GAS$$ was first collected in 2014. Missing values of GAS$$ are not imputed due to the low number of house-

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

norent$$ – Does Not Pay Rent [generic]

1	[1] Pays No Rent	418
-1	[-1] No Answer	8
-2	[-2] Does Not Apply	15565
-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

NORENT$$ marks tenants who do not a pay a rent. Until 2013 it was a filter question for all housing costs such that all non-rent paying tenants were not asked for HEAT$$, UTIL$$ and ELECTR$$. In 2014 the filter question was rephrased and tenants were asked for their utility costs even though 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	[1] Pays Rent and Utility Costs	8781
2	[2] Pays Utility Costs, but No Rent	253
3	[3] Does Not Pay either Rent or Utilities	165
-1	[-1] No Answer	8
-2	[-2] Does Not Apply	6784
-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

RENTINFO$$ is a filter question for housing costs. It was introduced in 2014 and replaces NORENT$$. [This information can be related to a specific variable and is not necessary 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	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	15991

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

HEATINFO$$ was introduced in 2014 and indicates why HEAT$$ 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	[1] Utility Costs Unknown	13
3	[3] No Other Additional Utility Costs	2645
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	13333
-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

UTILINFO$$ was introduced in 2014 and indicates why UTIL$$ is missing. If a household stated that it did not know its additional utility costs, we imputed UTIL$$ and set UTILINFO$$ to -2. [This information can be related to a specific variable and is not necessary 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	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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

ELECTRINFO$$ was introduced in 2014 and indicates why ELECTR$$ is missing. If a household stated that it did not know its electricity costs, we imputed ELECTR$$ and set ELECTRINFO$$ to -2. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

GASINFO$$ was introduced in 2014 and indicates why a household does not state additional gas costs that are not included in the heating costs. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	[0] Not imputed	15952
1	[1] Imputed	39
-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 necessary 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	15985
1	[1] Imputed	6
-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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	[0] Not imputed	13776
1	[1] Rent and/or Sub-Aggregate Imputed by PMM	1139
2	[2] Rent and/or Sub-Aggregate from Previous or Subsequent Years	943
3	[3] Rent and/or Sub-Aggregate From Prev./Subs. Years and/or PMM	133
-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$$ have been imputed (=1) by Stata's chained imputation procedure, whether RENT$$ itself or expenses added or subtracted to RENT$$ have been carried for- or backwards (=2) from past/subsequent years, or whether RENT$$ is the result of a mixture of imputed and past/subsequent year's values. The fourth category marks households who very likely did not include all utility costs in their overall housing costs in 2014 (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 have been added or subtracted. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	[0] Not imputed	13709
1	[1] Reported Rent Corrected for Imputed Heat	1081
2	[2] Reported Rent Corrected for Imputed Util	0
3	[3] Reported Rent Corrected for Imputed Electr	330
4	[4] Reported Rent Corrected for Heat and Util	0
5	[5] Reported Rent Corrected for Heat and Electr	712
6	[6] Reported Rent Corrected for Util and Electr	0
7	[7] Reported Rent Corrected for Heat, Util and Electr	0
8	[8] Rent Imp., Sub-Aggregates Not Imputed	0
9	[9] Rent Imp., Sub-Aggregates Not Imputed	44
10	[10] Rent Imp. and Corrected For Imp. Heat	0
11	[11] Rent Imp. and Corrected For Imp. Util	0
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	115
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	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

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

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

0	[0] Not imputed	14083
1	[1] Imputed by PMM	1199
2	[2] Value from Prev./Foll. Years	709
-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 years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	[0] Not imputed	14807
1	[1] Imputed by PMM	926
2	[2] Value from Prev./Foll. Years	258
-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 years. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

0	[0] Not imputed	14834
1	[1] Imputed by PMM	540
2	[2] Value from Prev./Foll. Years	617
-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 years. [This information can be related to a specific variable and is not necessary 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	15991
-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 1985 to 1997, this was only asked to new households or in case an old household had moved. Information is then carried forward. In 1998 and 1999, the question was again posed to the whole population. [This information can be related to a specific variable and is not necessary 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	15991
-6	[-6] Questionnaire Version With Modified Filter	0

Statement by respondent. OSUBS$$ contains information on cash housing subsidies received from the government during the year prior to the interview. 2006 was the last year house owners could apply for the 8-years-long financial support such that since 2015 OSUBS$$ is not asked any longer. Information is not carried forward.
Please note: The old variable $FOERD (available until SOEP data release 2008) is discarded. Homeowner subsidies in Germany have been subject to major revisions and fluctuations over time. The corresponding question in SOEP was in some years only posed to new households and those that have moved, in some years it was not surveyed at all. For these reasons, the question for government housing subsidies was changed in 2000 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 necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

rsubs$$ – Government Subsidized Housing [generic]

1	[1] Yes, and it is still subsidized	485
2	[2] Yes, but no Longer Subsidized	233
3	[3] No	8227
-1	[-1] No Answer	262
-2	[-2] Does Not Apply	6784
-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

RSUBS$$ states whether the rent is subsidized by government or not. Up to 1994, the question was asked only to new households and households who had moved since last year. For the remaining households, the information is carried forward from previous years. East German households are asked for the first time in 1993.
In 1995, the second response category was added to indicate expired subsidization. For reasons of time series consistency, RSUBS$$ was coded with "3" for "no" from 1984 to 1994. 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 1994. [This information can be related to a specific variable and is not necessary generic.]
For more information, contact: Carsten Schröder, <cschroeder@diw.de>

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

1	[1] Yes	740
2	[2] No	8443
-1	[-1] No Answer	24
-2	[-2] Does Not Apply	6784
-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

REDUC$$ states whether the flat is offered by the owner at a reduced rent. Information is carried forward from the previous years for old households residing at their old address; for newly surveyed households and for old households that have moved, newly collected data is used. From 2003 to 2007 this information was not collected. It is carried forward from 2002 for households who have not moved and whose stated amount of rent vary only slightly. The new information from 2008 is then carried backward for households with the same characteristics if REDUC$$ is still missing after carrying forward from 2002. [This information can be related to a specific variable and is not necessary 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	3760
2	[2] Couple Without Children	4425
3	[3] Single Parent	1717
4	[4] Couple With Children LE 16	3783
5	[5] Couple With Children GT 16	1065
6	[6] Couple With Children LE And GT 16	846
7	[7] Multiple Generation-HH	141
8	[8] Other Combination	254
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 necessary 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.	444
12	[12] 1-Person HH Male 35- LT 60 Y.	640
13	[13] 1-Person HH Male GE 60 Y.	560
14	[14] 1-Person HH - Female LE 35 Y.	369
15	[15] 1-Person HH Female 35- LT 60 Y.	568
16	[16] 1-Person HH Female GE 60 Y.	1179
21	[21] Couple Without Children	4425
31	[31] Single Parent,1 Child, LE 16	529
32	[32] Single Parent,2 Or More Children, LE 16	336
33	[33] Single Parent,1 Children, GT 16	509
34	[34] Single Parent,2 Or More Children, GT 16	110
35	[35] Single Parent,2 Children, LE and GT 16	136
36	[36] Single Parent,3 Or More Children, LE and GT 16	97
41	[41] Couple, 1 Child, LE 16	1268
42	[42] Couple, 2 Children, LE 16	1651
...	(2 rows omitted)	1568
52	[52] Couple, 2 Children, GT 16	282
53	[53] Couple, 3 Or More Children, GT 16	79
61	[61] Couple, 2 Children, LE and GT 16	328
62	[62] Couple, 3 Or More Children, LE and GT 16	518
71	[71] 3-Generation-HH	125
72	[72] 4-Generation-HH	0
73	[73] GrandParents-GrandChildren-HH	16
81	[81] Other Combination Without K. LE 16	254
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 2009 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 2010 (waves 1984 to 2009) 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 necessary generic.]

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

8 Household Income

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

0		1
150		1
188		1
200		1
220		2
250		2
251		1
270		1
275		1
280		1
300		5
303		1
320		2
321		1
330		1
...	(1628 rows omitted)	15146
15000		9
15200		1
15500		2
16000		4
17000		2
18000		1
19000		1
20000		6
23900		1
24000		1
24700		1
25000		5
60000		1
-1		705
-3		83

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 necessary 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		1
150		1
188		1
200		1
220		2
250		2
251		1
270		1
275		2
280		1
300		5
303		1
320		2
321		1
330		1
...	(1627 rows omitted)	15926
14500		1
15000		11
15200		1
15500		2
16000		5
17000		2
18000		1
19000		1
20000		7
23900		1
24000		1
24700		1
25000		5
60000		2
-1		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 28 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 29 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 (1987) and Royston (2004).
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 every survey household in all survey years (1995-20013) 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 a 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 necessary 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		1
150		1
188		1
200		1
220		3
250		2
251		1
270		1
275		1
280		1
300		5
303		1
320		2
321		1
330		1
...	(1627 rows omitted)	15929
14500		1
15000		9
15200		1
15500		2
16000		4
17000		2
18000		2
19000		1
20000		6
23900		1
24000		1
24700		1
25000		6
60000		1
-1		1

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

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

0		1
150		1
188		1
200		1
220		3
250		2
251		1
270		1
275		1
280		1
300		5
303		1
320		2
321		1
330		1
...	(1627 rows omitted)	15929
14500		1
15000		9
15200		1
15500		2
16000		5
17000		2
18000		1
19000		1
20000		6
23900		1
24000		1
24700		1
25000		5
60000		2
-1		1

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

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

0	1
150	1
188	2
200	1
220	2
250	2
251	1
270	1
275	1
280	1
300	5
303	1

320		2
321		1
330		1
...	(1627 rows omitted)	15929
14500		1
15000		9
15200		1
15500		2
16000		4
17000		2
18000		2
19000		1
20000		6
23900		1
24000		1
24700		1
25000		5
60000		2
-1		1

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

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

0		1
150		1
188		1
200		1
220		2
250		3
251		1
270		1
275		1
280		1
300		5
303		1
320		2
321		1
330		1
...	(1627 rows omitted)	15929
14500		1
15000		10
15200		1
15500		2
16000		4
17000		2
18000		1
19000		1
20000		6
23900		1

24000	1
24700	1
25000	6
60000	1
-1	1

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

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

0	[0] Not imputed	15203
1	[1] Imputed	787
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	1
-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 necessary 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	1854
2	[2] Bavaria	2628
3	[3] Berlin	659
4	[4] Brandenburg	617
5	[5] Bremen	118
6	[6] Hamburg	289
7	[7] Hesse	1148
8	[8] Mecklenburg-West Pomerania	355
9	[9] Lower Saxony	1510
10	[10] North Rhine-Westphalia	3258
11	[11] Rhineland-Palatinate	783
12	[12] Saarland	149
13	[13] Saxony	932
14	[14] Saxony-Anhalt	549
15	[15] Schleswig-Holstein	555
16	[16] Thuringia	587
-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 the year 2000 (wave Q) Rhineland- Palatinate and Saarland were defined as one region. [This information can be related to a specific variable and is not necessary 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	102
2	[2] February	3252
3	[3] March	3686
4	[4] April	1846
5	[5] May	1264
6	[6] June	1684
7	[7] July	1573
8	[8] August	1005
9	[9] September	818
10	[10] October	596
11	[11] November	165
12	[12] December	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

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 necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690), <pkrause@diw.de>

hmode$$ – Interview Method [generic]

100	[100] With Interviewer Assistance	26
110	[110] Oral Interview	961
120	[120] Written Ques. Interviewer	1678
130	[130] Mixed Type	0
131	[131] Written Ques. No Interviewer	108
132	[132] Oral And Written	179

133	[133] Proxy	0
134	[134] Third Person Present	0
135	[135] No Third Person Present	0
140	[140] CAPI - Wave O Onwards	10468
150	[150] CAWI - Wave BE (2014) Onwards	992
200	[200] Telephone Assistance	0
210	[210] Written, By Mail	1578
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 necessary generic.]
For more information, contact: Peter Krause (Tel. +49-30-89789-690), <pkrause@diw.de>