SOEP The German Socio-Economic Panel study

294

SOEP Survey Papers

Series D – Variable Descriptions and Coding

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

SOEP 2014 – Documentation of Household-related Status and Generated Variables in HGEN for SOEP v31

SOEP Group

DIW SOEP

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

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

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

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

Editors:
Dr. Jan Goebel, DIW Berlin
Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

SOEP Group. 2015. SOEP 2014 – Documentation of Household-related Status and Generated Variables in HGEN for SOEP v31. SOEP Survey Papers 294: Series D. Berlin: DIW Berlin / SOEP

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. © 2015 by SOEP

ISSN: 2193-5580 (online)

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

SOEP GROUP

SOEP 2014 – DOCUMENTATION OF HOUSEHOLD-RELATED STATUS AND GENERATED VARIABLES IN HGEN FOR SOEP V31

Berlin, 2015

The German Socio Economic Panel Study at DIW Berlin

SOEP-Core V31—BEHGEN

SOEP Group

2015-12-18

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

Contents

1 General Information

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 HH Number

hhnrakt – Current Wave HH Number (=BEHHNR)

behhnr – Current Wave HH Number

3 General Housing Information

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

1	[1] Owner	7098
2	[2] Main Tenant	8625
3	[3] Sub-Tenant	271
4	[4] Tenant	2
5	[5] Living In A Home	41
-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		2
1924		3
1925		4
1926		2
1927		3
1928		1
1929		4
1930		8
1931		3
1932		1
1933		5
...	(67 rows omitted)	5959
2001		432
2002		438
2003		480
2004		537
2005		608
2006		633
2007		671
2008		659
2009		745
2010		858
2011		975
2012		1181
2013		1226
2014		486
-1		109

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		1711
1918		1903
1949		4046
1972		1464
1973		89
1974		129
1975		108
1976		83
1977		47
1978		62
1979		57
1980		93
1981		861
1982		57
1983		44
...	(16 rows omitted)	1939
2000		125
2001		381
2002		62
2003		82
2004		90
2005		89
2006		116
2007		72
2008		69
2009		56
2010		68
2011		85
2012		57
2013		13
-5		1979

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. [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]

1917		1711
1948		1903
1971		4046
1972		178
1973		89
1974		129
1975		108
1976		83
1977		47
1978		62
1979		57
1980		1379
1981		74
1982		57
1983		44
...	(16 rows omitted)	2156
2000		695
2001		118
2002		85
2003		105
2004		125
2005		139
2006		182
2007		73
2008		75
2009		59
2010		110
2011		88
2012		66
2013		15
-5		1979

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	10767
2	[2] Some Renovations	4788

3	[3] Full Renovations	424
4	[4] Dilapidated	24
-1	[-1] No Answer	34
-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	1153
2	[2] Inheritance, Gift	486
3	[3] Bought, Built New	723
4	[4] Got Back From Public Property	0
-1	[-1] No Answer	4736
-2	[-2] Does Not Apply	8939
-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	755
2	[2] Inexpensive	2281
3	[3] Reasonable	3833
4	[4] Slightly Expensive	1510
5	[5] Too Expensive	258
-1	[-1] No Answer	166
-2	[-2] Does Not Apply	7234
-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). 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	431
2	[2] A Bit Too Small	2469
3	[3] Just Right	10995
4	[4] A Bit Too Large	1862
5	[5] Much Too Large	254
-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

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	14524
2	[2] No	75
-1	[-1] No Answer	1438

-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

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, but information 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>

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

1	[1] Yes	14525
2	[2] No	75
-1	[-1] No Answer	1437
-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

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, but information 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>

eqpiwc$$ – Dwelling Has Indoor Toilet [generic]

1	[1] Yes	5686
2	[2] No	18
-1	[-1] No Answer	10333
-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

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$$ is carried forward from the previous years. Additionally, from 1985 on, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. Beginning with 2009, data is no more collected, but carried forward from the two previous years for households which have not moved. In a second step, information of the variable EQPSHW$$ is used to replace missing and inconsistent values of EQPIWC$$. For every household with EQPSHW$$=1, EQPIWC$$ is set to 1. [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	15105
2	[2] No	924
-1	[-1] No Answer	8
-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

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. [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	13106
2	[2] No	2926
-1	[-1] No Answer	5
-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	14943
2	[2] No	1091
-1	[-1] No Answer	3
-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	9944
2	[2] No	6083
-1	[-1] No Answer	10
-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	14514
2	[2] No	84
-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	1439
-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 rea-

son, in case no address change has taken place the information for EQPWAT$$ 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>

eqptel$$ – Dwelling Has Telefone [generic]

1	[1] Yes	15981
2	[2] No	49
-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

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	588
2	[2] No	15431
-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	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	1552
2	[2] No	14472
-1	[-1] No Answer	13
-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	304
2	[2] No	15717
-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	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	1452
2	[2] No	14467
-1	[-1] No Answer	118
-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	605
2	[2] No	15282
-1	[-1] No Answer	150
-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>

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

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

10		3
11		1
12		13
13		3
14		5
15		11
16		6
17		8
18		5
19		3
20		34
21		5
22		9
23		6
24		6
...	(216 rows omitted)	15874
320		5
330		6
340		1
344		1
346		1
350		11
360		5
365		1
390		1
400		6
430		1
450		3
470		1
500		1
540		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	434
2	2177
3	4775
4	3441
5	2349
6	1467
7	757
8	364
9	155
10	68
11	23
12	14
13	6
14	1
15	4
16	1
20	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	5
5	1
15	1
16	1

20		1
23		1
24		1
28		2
30		4
31		1
32		1
34		1
35		2
36		1
37		1
...	(1009 rows omitted)	8508
2350		1
2390		1
2409		1
2416		1
2546		1
2700		1
2720		1
2850		1
2900		1
3114		1
3310		1
5700		1
-1		7
-2		7470
-3		16

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, 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$$, heating, gas, and electricity costs have to be deducted. Due to the 2014 questionnaire design, some households state monthly housing costs that presumably do not include all possible expenses. They might do so because they do not know all their expenses or because they misunderstand the question. We correct the housing costs for some households by comparing the stated housing costs with the last year's rent and expenses. If it is clear that a household has stated e.g. its base rent instead of the housing costs, we do not subtract the heating costs to obtain RENT$$. However, the mean rent in 2014 is around 5% lower than in 2013. We attribute the drop to the different questionnaire design.

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

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

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

10		6
12		2
14		2
15		13
18		8
20		62
21		2
22		3
23		9
24		5
25		47
27		3
28		12
29		2
30		152
...	(224 rows omitted)	8138
350		6
357		1
370		3
380		1
384		1
386		4
396		3
400		4
410		1
440		1
500		2
550		1
600		2
-1		144
-2		7397

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.
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$$.
[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]

0		1
2		1
3		3
4		1
5		20
6		6
7		5
8		10
9		3
10		54
12		11
13		4
14		4
15		72
16		11
...	(199 rows omitted)	7444
279		1
280		3
282		1
285		1
290		2
300		17
302		3
305		1
306		1
320		1
330		1
358		1
400		1
-1		143
-2		8210

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.

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 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]

4		1
5		1
9		2
10		9
11		1
12		3
13		5
14		2
15		18
16		3
17		9
18		10
19		8
20		100
21		19
...	(179 rows omitted)	8220
255		1
260		2
270		2
275		2
280		2
289		1
300		3
315		1
319		2
342		1
345		1
360		1
540		1
-1		143
-2		7463

Costs for electricity have been collected starting in 2010 if tenants pay a rent. From 2014 onwards, they are collected from all tenants.
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>

norent$$ – Does Not Pay Rent [generic]

1	[1] Pays No Rent	372
-1	[-1] No Answer	110
-2	[-2] Does Not Apply	15555
-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	8560
2	[2] Pays Utility Costs, but No Rent	236
3	[3] Does Not Pay either Rent or Utilities	136
-1	[-1] No Answer	7
-2	[-2] Does not apply	7098
-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

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	16
3	[3] No Heating Costs	283
-1	[-1] No Answer	144
-2	[-2] Does not apply	15594
-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

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	16
3	[3] No Other Additional Utility Costs	1097
-1	[-1] No Answer	143
-2	[-2] Does not apply	14781
-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

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	2
2	[2] Electricity Costs Included in Heating Costs	282
3	[3] No Electricity Costs	81
-1	[-1] No Answer	143
-2	[-2] Does not apply	15529
-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

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	704
2	[2] Gas Costs Included in Heating Costs	1551
3	[3] No Gas Costs	5256
-1	[-1] No Answer	531
-2	[-2] Does not apply	7995
-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

GASINFO$$ was introduced in 2014 and indicates why a household does not pay 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	16013
1	[1] Imputed	24
-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	16027
1	[1] Imputed	10
-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	13245
1	[1] Rent and/or Sub-Aggregate Imputed by PMM	1220
2	[2] Rent and/or Sub-Aggregate from Previous or Subsequent Years	929
3	[3] Rent and/or Sub-Aggregate From Prev./Subs. Years and/or PMM	120
4	[4] Rent Corrected for Longitudinal Consistency Before Imputation	523
-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

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 ex-

penses 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	13695
1	[1] Reported Rent Corrected for Imputed Heat	1697
2	[2] Reported Rent Corrected for Imputed Util	0
3	[3] Reported Rent Corrected for Imputed Electr	75
4	[4] Reported Rent Corrected for Heat and Util	0
5	[5] Reported Rent Corrected for Heat and Electr	467
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. and Corrected for Imp. Heat	0
10	[10] Rent Imp. and Corrected for Imp. Util	0
11	[11] Rent Imp. and Corrected for Imp. Electr	1
12	[12] Rent Imp. and Corrected for Imp. Heat and Util	0
13	[13] Rent Imp. and Corrected for Imp. Heat and Electr	102
14	[14] Rent Imp. and Corrected for Imp. Util and Electr	0
15	[15] Rent Imp. and Corrected for Imp. Heat, Util and Electr	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	13771
1	[1] Imputed by PMM	1300
2	[2] Heating Costs From Previous or Subsequent Years	966
-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	13055
1	[1] Imputed by PMM	2383
2	[2] Other Costs From Previous or Subsequent Years	599
-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	15392
1	[1] Imputed by PMM	346
2	[2] Electricity Costs from Previous or Subsequent Years	299
-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	16037
-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	125
2	[2] No	6925
-1	[-1] No Answer	48
-2	[-2] Does Not Apply	8939
-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

Statement by respondent. OSUBS$$ contains information on cash housing subsidies received from the government during the year prior to the interview. Information is not carried forward.
Please note: The old variable $FOERD (available until SOEP data release 2008) is discarded. Homeowner subsidies in Germany have been subject to major revisions and fluctuations over time. The corresponding question in SOEP was in some years only posed to new households and those that have moved, in some years it was not surveyed at all. For these reasons, the question for government housing subsidies was changed in 2000 to cover direct subsidies received the previous year. SUBSID$$ 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, With Due Diligence	464
2	[2] Yes, With Run Out Due Diligence	263
3	[3] No	8179
-1	[-1] No Answer	33
-2	[-2] Does Not Apply	7098
-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	711
2	[2] No	8210
-1	[-1] No Answer	18
-2	[-2] Does Not Apply	7098
-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	3689
2	[2] Couple Without Children	4404
3	[3] Single Parent	1819
4	[4] Couple With Children LE 16	3783
5	[5] Couple With Children GT 16	1091
6	[6] Couple With Children LE And GT 16	862
7	[7] Multiple Generation-HH	127
8	[8] Other Combination	262
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.	389
12	[12] 1-Person HH Male 35- LT 60 Y.	637
13	[13] 1-Person HH Male GE 60 Y.	554
14	[14] 1-Person HH - Female LE 35 Y.	339
15	[15] 1-Person HH Female 35- LT 60 Y.	576
16	[16] 1-Person HH Female GE 60 Y.	1194
21	[21] Couple Without Children	4404
31	[31] Single Parent,1 Child, LE 16	610
32	[32] Single Parent,2 Or More Children, LE 16	375
33	[33] Single Parent,1 Children, GT 16	492
34	[34] Single Parent,2 Or More Children, GT 16	113
35	[35] Single Parent,2 Children, LE and GT 16	115
36	[36] Single Parent,3 Or More Children, LE and GT 16	114
41	[41] Couple, 1 Child, LE 16	1181
42	[42] Couple, 2 Children, LE 16	1680
...	(2 rows omitted)	1642
52	[52] Couple, 2 Children, GT 16	298
53	[53] Couple, 3 Or More Children, GT 16	73
61	[61] Couple, 2 Children, LE and GT 16	348
62	[62] Couple, 3 Or More Children, LE and GT 16	514
71	[71] 3-Generation-HH	113
72	[72] 4-Generation-HH	1
73	[73] GrandParents-GrandChildren-HH	13
81	[81] Other Combination Without K. LE 16	262
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. 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]

150		1
158		1
160		1
200		1
205		1
250		2
261		1
291		1
300		9
301		1
340		2
350		4
356		1
359		1
365		1
...	(1639 rows omitted)	15296
19000		1
20000		4
21000		1
22000		1
23000		1
24518		1
25000		3
25500		1
28000		1
29000		1
35000		1
36000		1
65000		1
-1		585
-3		110

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]

150		3
158		1
160		1
200		1
205		1
250		2
261		1
291		1
300		9
301		1
340		3
350		4
356		1
359		1
365		1
...	(1637 rows omitted)	15985
18000		2
18500		1
19000		1
20000		4
21000		1
22000		1
23000		1
24518		1
25000		3
25500		1
28000		1
29000		1
35000		1
36000		1
65000		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 house-

hold 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]

150		1
158		1
160		1
200		1
205		1
250		4
261		1
291		1
300		9
301		1
340		3
350		5
356		1
359		1
365		1
...	(1637 rows omitted)	15982
18000		2
18500		1
19000		1
20000		4
21000		2
22000		1
23000		1
24518		1
25000		4
25500		1
28000		1
29000		1
35000		1
36000		1
65000		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]

150		1
158		1
160		1
200		1
205		1
250		2
261		1
291		1
300		10
301		1
340		2
350		4
356		1
359		1
365		1
...	(1637 rows omitted)	15987
18000		2
18500		1
19000		1
20000		4
21000		1
22000		1
23000		1
24518		1
25000		3
25500		1
28000		1
29000		1
35000		1
36000		1
65000		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]

150		1
158		1
160		1
200		1
205		1
250		2
261		1
291		1
300		9
301		1
340		3
350		4
356		1
359		1

365		1
...	(1637 rows omitted)	15986
18000		2
18500		1
19000		1
20000		5
21000		1
22000		1
23000		1
24518		1
25000		3
25500		1
28000		1
29000		1
35000		1
36000		1
65000		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]

150		1
158		1
160		2
200		1
205		1
250		2
261		1
291		1
300		9
301		2
340		2
350		4
356		1
359		1
365		1
...	(1637 rows omitted)	15985
18000		3
18500		1
19000		1
20000		4
21000		1
22000		1
23000		1
24518		1
25000		3
25500		1
28000		1
29000		1

35000	1
36000	1
65000	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	15342
1	[1] Imputed	695
-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 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	1849
2	[2] Bavaria	2587
3	[3] Berlin	664
4	[4] Brandenburg	630
5	[5] Bremen	128
6	[6] Hamburg	276
7	[7] Hesse	1132
8	[8] Mecklenburg-Western Pomerania	358
9	[9] Lower Saxony	1484
10	[10] North Rhine-Westphalia	3271
11	[11] Rhineland-Palatinate	787
12	[12] Saarland	145
13	[13] Saxony	964
14	[14] Saxony-Anhalt	586
15	[15] Schleswig-Holstein	575
16	[16] Thuringia	601
-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	23
2	[2] February	4438
3	[3] March	3299
4	[4] April	1737
5	[5] May	2007
6	[6] June	1604
7	[7] July	1267
8	[8] August	491
9	[9] September	561
10	[10] October	597
11	[11] November	13
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	16
110	[110] Oral Interview	1165
120	[120] Written Ques. Interviewer	1842
130	[130] Mixed Type	0
131	[131] Written Ques. No Interviewer	100
132	[132] Oral And Written	161
133	[133] Proxy	0
134	[134] Third Person Present	0
135	[135] No Third Person Present	0

140	[140] CAPI - Wave O Onwards	9961
150		1122
200	[200] Telephone Assistance	0
210	[210] Written, By Mail	1670
220	[220] Phone Interview	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 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>