

1309²⁰²⁴

SOEP Survey Papers
Series D – Variable Descriptions and Coding

SOEP-Core v38.1 – HGEN: Household-Related Status and Generated Variables

SOEP Group

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

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

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

The SOEP Survey Papers are available at <http://www.diw.de/soepsurveyspapers>

Editors:

Dr. Carina Cornesse, DIW Berlin and University of Bremen
Dr. Jan Goebel, DIW Berlin
Prof. Dr. Cornelia Kristen, University of Bamberg and DIW Berlin
Prof. Dr. Philipp Lersch, DIW Berlin and Humboldt-Universität zu Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin
Prof. Dr. Sabine Zinn, DIW Berlin and Humboldt-Universität zu Berlin

Please cite this paper as follows:

SOEP Group, 2024. SOEP-Core v38.1 – HGEN: Household-Related Status and Generated Variables. SOEP Survey Papers 1309: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
© 2024 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core v38.1 – HGEN: Household-Related Status and Generated Variables

SOEP Group

2024

Contents

1	General Information	4
2	Sample Information	4
	hid – Current Wave HH Number	4
	cid – Original Household Number, Case ID	5
	year – Survey Year	6
3	General Housing Information	6
	hgacquis – Means Of Acquiring Dwelling	6
	hgcstyrmin – Earliest Possible Construction Year of Dwelling	7
	hgcstyrmax – Latest Possible Construction Year of Dwelling	8
	hgcondit – Condition Of House	9
	hgreval – Rent Of Dwelling Vs Comparable Dwellings	9
	hgnuts1 – NUTS-Systematic-1 (Federal State)	10
	hgnuts1_ew – NUTS-Systematic-1 (East-West Version)	11
4	Equipment of the Dwelling	11
	hgroom – Number Of Rooms Larger Than 6 Sq M	11
	hgeqpbas – Dwelling Has Basement	12
	hgeqpfhea – Equipment: Floor Heating	13
	hgeqpfire – Dwelling Has Fireplace or Ceramic Tiled Stove	13
	hgeqpgar – Dwelling Has Garden	13
	hgeqphea – Dwelling Has Central,Floor Heat	14
	hgeqipinsul – Equipment: Thermal Insulation	14
	hgeqpiwc – Dwelling Has Indoor Toilet	15
	hgeqpkit – Dwelling Has Kitchen	15
	hgeqplif – Dwelling Has An Elevator	16
	hgeqpmglass – Equipment: at Least Double Glazing	16
	hgeqpnobar – Equipment: Barrier Free Furnishing	17
	hgeqpnrj – Dwelling Has Alternative Energy Source	17
	hgeqppark – Equipment: Garage/ Parking Space	17
	hgeqpshw – Dwelling Has Indoor Bath,Shower	18
	hgeqpsol – Dwelling Has Solar System	18
	hgeqptel – Dwelling Has Telephone	19
	hgeqpter – Dwelling Has Balcony,Terrace	19
	hgeqpwat – Dwelling Has Water, Bioler	20
	hgeqpair – Dwelling Has Air Conditioner	20
	hgeqpalm – Dwelling Has Alarm System	20
5	Costs of Living, Size, and Rooms (Imputed Variables)	21
	hgrent – Amount Of Rent Minus Heating Costs	21
	hgrentinfo – Pays Rent and/or Utilities	23
	hgnorent – Does Not Pay Rent	24
	hgelectr – Costs of Electricity (gen)	24
	hgelectrinfo – Reason for Missing Electricity Costs	25
	hggas – Costs of Gas (EUR)	26
	hggasinfo – Reason for Missing Gas Costs	26
	hgheat – Costs Of Warm Water, Gen	27
	hgheatinfo – Reason for Missing Heating Costs	28

hgmoveyr – Year Moved Into Dwelling	28
hgutil – Other Monthly Utility Costs	29
hgutilinfo – Reason for Missing Additional Utility Costs	30
hgseval – Adequacy Of Living Space In Housing Unit	30
hgsize – Size Of Housing Unit In Sq M	31
hgf2rent – Detailed Imputation Flag, Rent Minus Heating Costs	32
hgfelectr – Imputation Flag, Costs of Electricity	32
hgfheat – Imputation Flag, Costs Of Heating And Warm Water	33
hgfrent – Imputation Flag, Amount Of Rent Minus Heating Costs	33
hgfroom – Imputation Flag, Number Of Rooms Lager Than 6 Sq M	34
hgfsiz – Imputation Flag, Size Of Housing	34
hgfutil – Imputation Flag, Other Monthly Utility Costs	34
6 Subsidization of Housing Costs	35
hgosubs – Amount Of Subsidies Last Year	35
hgreduc – Dwelling At A Reduced Prize	35
hgrsubs – Government Subsidized Housing	36
hgsubsid – Government Subsidizes Housing Payments	36
7 Typology of Household	37
hgowner – Tenant Or Owner Of Dwelling	37
hgtyp1hh – Household Typology, One Digit	37
hgtyp2hh – Household Typology, Two Digit	38
8 Household Income	39
hghinc – Monthly Household Net Income (EUR)	39
hgi1hinc – 1. Imputed Monthly Net Household Income (EUR) [1/5]	40
hgi2hinc – 2. Imputed Monthly Net Household Income (EUR) [2/5]	42
hgi3hinc – 3. Imputed Monthly Net Household Income (EUR) [3/5]	43
hgi4hinc – 4. Imputed Monthly Net Household Income (EUR) [4/5]	43
hgi5hinc – 5. Imputed Monthly Net Household Income (EUR) [5/5]	44
hgfhinc – Imputation Flag, Monthly Net Household Income	45
9 Time and Method of Interview	45
iyear – Household Related Survey Year	45
hghmonth – Month Of Interview	45
hghmode – Interview Method	46
inputdataset – Source dataset	47

1 General Information

The HGEN file provides a set of time-consistent variables generated from the SOEP household questionnaire. It only includes households who participated in the respective year.

Up to Wave G, “old” households already known to the SOEP were surveyed at the old address with a “green” questionnaire; old households that had moved and new households received a “blue” questionnaire. While a number of questions in the blue questionnaire surveyed information for the first time (e.g. living space in square meters), this information was only asked for again in the green questionnaire in case of changes. Otherwise, the information collected in the previous wave was still valid. The variables described in the following are in part status variables in this sense: Information collected once is carried forward to subsequent waves if no address change has taken place since the previous wave. This is the case for: HGCNSTYRMIN, HGCNSTYRMAX, HGCONDIT (for Wave B to G), HGSIZE, HGRROOM, HGEQPKIT, HGEQPSHW, HGEQPIWC, HGEQPHEA, HGEQPTER, HGEQP-BAS, HGEQPGAR, HGEQPWAT, HGEQPALM, HGEQPSOL, HGEQPAIR, HGMOVEYR, HGACQUIS, HGSUBSID, HGRSUBS, HGREDC and HGSEVAL (for Wave C to G). Furthermore, identical information is recorded in the green and blue questionnaire in separate variables (e.g. housing tenure as owner or tenant). The corresponding status variables are therefore just a compilation of these more dispersed pieces of information. Since only one common questionnaire has been used since Wave H rather than the former “blue” and “green” versions, the necessity for the aforementioned status variables disappears. However, this user-friendly redundancy is maintained for reasons of consistency.

In addition, we impute missing values of the variables HGRROOM, HGSIZE, HGRENT, HGHEAT, HGUTIL and HGHINC to ensure the completeness of the dataset and facilitate its usage. Details of the imputation process can be found in the description of the respective variable.

Since Wave BG, new migration samples (M3, M4, M5, and later M6) have been added. These samples provide less household information in some waves which explains the high incidence of missing values (=5). In some waves, missing values in the HGEN variables of the samples M3-M6 are not imputed.

2 Sample Information

hid – Current Wave HH Number

19	6
27	25
35	10
43	3
51	9
60	29
78	3
86	2
94	38
108	4
116	15
124	23
132	14
140	7
159	36

...	(57820 rows omitted)	428131
5557590		1
5557610		1
5557620		1
5557630		1
5557700		1
5557720		1
5557730		1
5557810		1
5557830		1
5557840		1
5557850		1
5557890		1
5557900		1
5557910		1
5557930		1

cid – Original Household Number, Case ID

19		9
27		50
35		10
43		3
51		9
60		92
78		3
86		2
94		38
108		4
116		15
124		23
132		16
140		7
159		36
...	(47689 rows omitted)	428035
5538388		1
5538418		2
5538523		1
5538566		1
5538590		2
5538744		1
5538779		1
5539180		1
5539198		1
5539414		1
5539546		1
5539600		2
5539660		1
5540200		1
5540404		1

syear – Survey Year

1984	5921
1985	5322
1986	5090
1987	5026
1988	4814
1989	4690
1990	6819
1991	6699
1992	6665
1993	6637
1994	6795
1995	6968
1996	6894
1997	6810
1998	7668
... (8 rows omitted)	93072
2007	11689
2008	11058
2009	11889
2010	15414
2011	16819
2012	16508
2013	18093
2014	16037
2015	15991
2016	17775
2017	19701
2018	18682
2019	19030
2020	20183
2021	13611

3 General Housing Information**hgacquis** – Means Of Acquiring Dwelling

1	Bought From Owner	37555
2	Inheritance, Gift	24048
3	Bought, Built New	40747
4	Got Back From Public Property	6
-1	No answer	68503
-2	Does not apply	223636
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	33875
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0

-8 Question not part of the survey program this year 0

Waves: all

If a household did not provide any information on HGACQUIS, the information from previous waves is carried forward under the condition that a change of address or ownership status (HGOWNER) has not taken place.

In Wave A to Wave G, HGACQUIS was asked only if a household was new to the SOEP or if it had changed its address. From Wave H to Wave R, HGACQUIS was also asked if a change in ownership status had taken place in the last year. Since Wave S, HGACQUIS 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 HGACQUIS anymore. Furthermore, since Wave S, the category „Bought New or Built“ is no longer included in the household questionnaire but is carried forward from last years. Given these changes in the categories and the sample that provides information for HGACQUIS, we recommend not to compare the shares of the different categories of HGACQUIS over time.

„Returned to private ownership“ was surveyed only in Wave I in East Germany and is carried forward.

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

hgcnstyrmin – Earliest Possible Construction Year of Dwelling

0	55713
1910	1
1918	7547
1919	52292
1949	117511
1972	54882
1973	1126
1974	1461
1975	1267
1976	932
1977	569
1978	872
1979	751
1980	1229
1981	29948
... (33 rows omitted)	63647
2015	414
2016	395
2017	308
2018	268
2019	183
2020	48
2021	1
-1 No answer	16856
-2 Does not apply	0
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	20149

-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: 1984–2013, 2015–2016

HGCNSTYRMIN provides the lower limit of the time period in which the household's building was constructed. HGCNSTYRMAX provides the upper limit. E.g. if a household's HGCNSTYRMIN is 1984 and HGCNSTYRMAX 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 HGCNSTYRMIN and HGCNSTYRMAX are constructed. For old households the variables are carried forward from the last years. Since Wave X, households are also asked for the exact construction year. In these cases, HGCNSTYRMIN and HGCNSTYRMAX contain the same value. If a building was built before 1919, HGCNSTYRMIN contains a zero and HGCNSTYRMAX 1919.

In Wave BE, the question about the construction year was not included in the SOEP. All information for HGCNSTYRMIN and HGCNSTYRMAX derives from the previous year. HGCNSTRMIN and HGCNSTRMAX replace CNSTYR\$\$ that has previously been distributed in \$HGEN files. The latter was a categorial variable grouping about one decade each into one category. Hence, with the last category having a flexible upper limitation and being split up regularly, sorting was ambiguous in some cases. For example, a household responding in Wave L that his house has been built "1981 or later" was not unambiguously assignable to either "1981-1990" or "1990 or later". Using HGCNSTYRMIN and HGCNSTYRMAX, changing questions can be far better described.

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

hgcnstyrmax – Latest Possible Construction Year of Dwelling

1910	1
1917	6546
1918	49167
1948	59839
1971	117511
1972	2111
1973	1126
1974	1461
1975	1267
1976	932
1977	569
1978	872
1979	751
1980	42001
1981	846
... (33 rows omitted)	104508
2015	430
2016	420
2017	321
2018	303
2019	224

2020		135
2021		24
-1	No answer	16856
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	20149
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: 1984–2013, 2015–2016

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

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

hgcondit – Condition Of House

1	In A Good Condition	224553
2	Some Renovations	96672
3	Full Renovations	12684
4	Dilapidated	499
-1	No answer	1317
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	2693
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	89952

Waves: 1984–2015

Respondent's subjective assessment of the condition of the building. In the West German sub-samples from Wave B to Wave G, the information on HGCONDIT was collected only for new households and for households with a residential move since the previous year (households with "blue" questionnaires). For households who had not moved ("green" questionnaire), the information collected in previous years was carried forward. The wording in the questionnaire was changed in the first wave of the East German sub-sample (Wave G) to better capture the rundown condition of some residential buildings in East Germany. Since Wave H the wording is identical for the entire SOEP-sample in East and West Germany. HGCONDIT was not surveyed between 2016 and 2018 (Waves BG to BI).

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

hgreval – Rent Of Dwelling Vs Comparable Dwellings

1	Very Inexpensive	12957
2	Inexpensive	39152
3	Reasonable	67492

4	Slightly Expensive	27088
5	Too Expensive	5335
-1	No answer	1643
-2	Does not apply	125873
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	148830

Waves: 1984–2002, 2005–2014

Subjective assessment by respondent (household head). This variable was not surveyed in Wave T and Wave U. The corresponding information from the previous year is not carried forward longitudinally due to the possibility of changes in rent and income, residential moves, and change in the person responding.

Since Wave BF, HGREVAL is not asked and information is neither carried backward nor forward.

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

hgnuts1 - NUTS-Systematic-1 (Federal State)

1	Baden-Wuerttemberg	51043
2	Bavaria	62247
3	Berlin	18439
4	Brandenburg	16567
5	Bremen	3659
6	Hamburg	7678
7	Hesse	30507
8	Mecklenburg-West Pomerania	9658
9	Lower Saxony	38091
10	North Rhine-Westphalia	88211
11	Rhineland-Palatinate	20574
12	Saarland	3782
13	Saxony	26677
14	Saxony-Anhalt	15659
15	Schleswig-Holstein	13873
16	Thuringia	15779
-1	No answer	5
-2	Does not apply	5921
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

NUTS (“Nomenclature of Statistical Territorial Units”) URL is a hierarchical system for dividing up the economic territory of the European Union. It was introduced by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the

production of regional statistics. NUTS 1 especially contains the major socio-economic regions for analyzing regional Community problems. It subdivides the European Union by now into 97 regions, whereas in Germany there are equivalent to the German Federal States 16 regions. Before Wave Q, Rhineland-Palatinate and Saarland were defined as one region.

For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

hgnuts1_ew – NUTS-Systematic-1 (East-West Version)

21	West-Germany	319665
22	East-Germany	102779
-1	No answer	5
-2	Does not apply	5921
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

4 Equipment of the Dwelling

hgroom – Number Of Rooms Larger Than 6 Sq M

0	14	
1	16469	
2	68728	
2.5	1	
3	129246	
4	88263	
5	56305	
6	34373	
7	16634	
8	8407	
9	3459	
10	1655	
11	559	
12	302	
13	117	
... (10 rows omitted)	211	
35	3	
42	1	
45	3	
51	11	
60	19	
80	20	
82	12	
-1	No answer	3558
-2	Does not apply	0
-3	Not valid	0

-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

Up to Wave R (except for Wave O), HGROOM was collected only in the first interview with new households. In case a household had moved or (from Wave H onwards) whether a household's housing unit size changed due to renovations or additions (up to Wave G, these households filled out a "green" questionnaire). In Wave O, the information had been asked again in order to correct for mistakes that may have occurred in carrying forward or in the process of imputation. From Wave S onwards, the question on the number of rooms has been asked to all households annually.

For old households still residing at their old address and neither having moved nor renovated their dwelling, the information on the number of rooms is carried forward as a status variable in order to provide valid current information. Analogously, the information is carried back to fill gaps if a household missed to state the number of rooms in the first year(s) they moved to a new housing.

In case, the information on the number of rooms is still missing, it is imputed simultaneously with other variables using Stata's chained imputation procedure. Imputed values are not carried forward. For information on the imputation procedure, see the description of HGRENT.

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

hgeqpbas – Dwelling Has Basement

1	Yes	388205
2	No	28794
-1	No answer	11371
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

The variable indicates whether a household has a cellar or storage areas. It was introduced in Wave A (1984). Before Wave V, HGEQPBAS was collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for HGEQPBAS is carried forward from the previous years. Exceptions are Waves H and O, when all households were surveyed. In Wave BJ (2019), HGEQPBAS was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpfhea – Equipment: Floor Heating

1	Yes	40135
2	No	169803
-1	No answer	9795
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	208637

Waves: Since 2009

The variable indicates whether a household has a floor heating (full or partial). It was introduced in Wave Z (2009). In Wave BJ (2019), HGEQPFHEA was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpfire – Dwelling Has Fireplace or Ceramic Tiled Stove

1	Yes	6498
2	No	20068
-1	No Answer	224
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	7004
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	394576

Waves: Since 2020

The variable indicates whether a household has a fireplace or ceramic tiled stove. It was introduced in Wave BK (2020).

hgeqpgar – Dwelling Has Garden

1	Yes	240473
2	No	173724
-1	No answer	14173
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

The variable indicates whether a household has (access to) a yard or garden. It was introduced in Wave A (1984). Before Wave V, HGEQPGAR was collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for HGEQPGAR is carried forward from the previous years. Exceptions are Waves H and O, when all households were surveyed. In Wave BJ (2019), HGEQPGAR was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqphea – Dwelling Has Central,Floor Heat

1	Yes	286359
2	No	26483
-1	No answer	6922
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	79004
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	29602

Waves: 1984-2014

The variable indicates whether a household has a central heating. It was introduced in wave A (1984). Only in Wave H, Wave O, and from Wave V onwards, HGEQPHEA 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 HGEQPHEA is carried forward from the previous years. Additionally, from Wave G onwards the information is updated if the household has stated that it modernized its heating since January of the previous year. In Wave BF (2015), HGEQPHEA was not collected. Since wave BG (2016), HGEQPHEA is only collected from samples M3, M4, and M5.

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

hgeqpinsul – Equipment: Thermal Insulation

1	Yes	62812
2	No	37833
-1	No answer	9771
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14557
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	303397

Waves: Since 2015

The variable indicates whether a household has thermal insulation (e.g., on the facade, roof, basement ceiling). It was introduced in Wave BF (2015). In Wave BJ (2019), HGEQINSUL was not collected, but information is carried forward from previous years if households have not moved.

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

hgeqpiwc – Dwelling Has Indoor Toilet

1	Yes	202210
2	No	4371
-1	No answer	2056
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	219733

Waves: 1984-2008

The variable indicates whether a household has a toilet in the flat. It was introduced in Wave A (1984). Only in Wave H, Wave O, and from Wave V to Wave Y, HGEQPIWC 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 HGEQPIWC is carried forward from the previous years. Additionally, from Wave B onwards, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. In a second step, the information of the variable HGEQPSHW is used to replace missing and inconsistent values of HGEQPIWC. For every household with HGEQPSHW=1, HGEQPIWC is set to 1. Since Wave Z, HGEQPWC is not asked. Since Wave BA, the information is neither carried backward nor forward.

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

hgeqpkit – Dwelling Has Kitchen

1	Yes	282602
2	No	3187
-1	No answer	1571
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	141010

Waves: 1984-2013

The variable indicates whether a household has a kitchen. It was introduced in Wave A (1984). Only in Wave H, Wave O, and from Wave V onwards, HGEQPKIT 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 HGEQPKIT is carried forward from previous years. Additionally, from Wave B on, the information is updated if the household has stated that it modernized its kitchen since January of the previous year. Since Wave BE, HGEQPKIT is neither asked nor carried forward.

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

hgeqplif – Dwelling Has An Elevator

1	Yes	20215
2	No	189844
-1	No answer	9674
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	208637

Waves: Since 2009

The variable indicates whether a household has an elevator of lift inside the building. It was introduced in Wave Z (2009). In Wave BJ (2019), HGEQPLIF was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpmglass – Equipment: at Least Double Glazing

1	Yes	94641
2	No	7104
-1	No answer	8671
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14557
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	303397

Waves: Since 2015

The variable indicates whether a household has windows with at least double glazing. It was introduced in Wave BF (2015). In Wave BJ (2019), HGEQPMGLASS was not collected, but information is carried forward from previous years if households have not moved.

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

hgeqpnobar – Equipment: Barrier Free Furnishing

1	Yes	21782
2	No	127805
-1	No answer	9516
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	269267

Waves: Since 2013

The variable indicates whether a household has home furnishings that are accessible for the elderly or disabled. It was introduced in Wave BD (2013). In Wave BJ (2019), HGEQPNOBAR was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpnrj – Dwelling Has Alternative Energy Source

1	Yes	7083
2	No	188832
-1	No answer	9261
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14557
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	208637

Waves: Since 2009

The variable indicates whether a household has other alternative energy sources besides solar collectors or a solar energy system (e.g., geothermal heat pump). It was introduced in Wave Z (2009). In Wave BJ (2019), HGEQPNRJ was not collected, but information is carried forward from previous years if households have not moved.

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

hgeqppark – Equipment: Garage/ Parking Space

1	Yes	74869
2	No	40861
-1	No answer	9243
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0

-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	303397

Waves: Since 2015

The variable indicates whether a household has a garage or parking space. It was introduced in Wave BF (2015). In Wave BJ (2019), HGEQPPARK was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpswh – Dwelling Has Indoor Bath,Shower

1	Yes	280620
2	No	4860
-1	No answer	1880
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	141010

Waves: 1984-2013

The variable indicates whether a household has a bath or shower in the flat. It was introduced in Wave A (1984). Only in Wave H, Wave O, and from Wave V onwards, HGEQPSHW 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 HGEQPSHW is carried forward from the previous years. Additionally, from Wave B onwards, the information is updated if the household has stated that it modernized its bathroom since January of the previous year. Since Wave BE, HGEQPSHW is not asked anymore. Since Wave BF, the information is not carried forward.

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

hgeqpsol – Dwelling Has Solar System

1	Yes	20235
2	No	198338
-1	No answer	9350
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14557
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	185890

Waves: Since 2007

The variable indicates whether a household has a solar collector or solar energy system. It was introduced in Wave X (2007). In Wave BJ (2019), HGEQPSOL was not collected, but information is carried forward from previous years if households have not moved.

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

hgeqptel – Dwelling Has Telephone

1	Yes	265719
2	No	6768
-1	No answer	4603
-2	Does not apply	4640
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	7083
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	139557

Waves: Since 1990 (east) or 1991 (west) to 2008, 2010–2015

The variable indicates whether a household has a telephone. It was introduced in Wave G (1990). The question 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“. HGEQPTEL is set to 1 if a household stated that it had a telephone, landline telephone, or cellphone. Additionally, the information from HBRUTTO (HTEL) is used to complete missing information. In Wave N, Wave P, and Wave Z, the household questionnaire did not ask about the possession of a telephone and HBRUTTO is the only source of data. For Wave K, no information is available.

Since Wave BG, HGEQPTEL is no longer asked in the household questionnaire (except for the samples M3 and M4 as HGEQPTEL was asked in the migration household questionnaire). For this reason, it is no longer provided since Wave BG for any other samples than M3 and M4.

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

hgeqpter – Dwelling Has Balcony,Terrace

1	Yes	311344
2	No	102946
-1	No answer	14080
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

The variable indicates whether a household has a balcony, terrace, or porch. It was introduced in Wave A (1984). Before Wave V, HGEQPTER was collected only from new households and households who have moved since the previous interview. For this reason, in case no

address change has taken place, the information for HGEQPTER is carried forward from the previous years. Exceptions are Waves H and O, when all households were surveyed. In Wave BJ (2019), HGEQPTER was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

hgeqpwat – Dwelling Has Water, Bioler

1	Yes	246907
2	No	3103
-1	No answer	1847
-2	Does not apply	4640
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	171873

Waves: Since 1990 (east) or 1991 (west) to 2013

The variable indicates whether a household has warm water or a boiler. It was introduced in Wave G (1990). Except for Wave H, Wave O, and from Wave V onwards, HGEQPWAT has been collected only from new households and households who have moved since the previous interview. For this reason, in case no address change has taken place, the information for HGEQPWAT is carried forward from the previous years. Since Wave BE, HGEQPWAT is not asked and since Wave BF, the information is neither carried backward nor forward.

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

hgeqpair – Dwelling Has Air Conditioner

1	Yes	4690
2	No	213822
-1	No answer	9411
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14557
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	185890

Waves: Since 2007

The variable indicates whether a household has an air conditioner (controlled ventilation system). It was introduced in Wave X (2007). In Wave BJ (2019), HGEQPAIR was not collected, but information is carried forward from previous years if households have not moved.

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

hgeqpalm – Dwelling Has Alarm System

1	Yes	15911
2	No	241418
-1	No answer	20886
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	150155

Waves: Since 2004

The variable indicates whether a household has an alarm system. It was introduced in Wave U (2004). In Wave BJ (2019), HGEQPALM was only collected from samples M3, M4, and M5, but information is carried forward for other samples if households have not moved.

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

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

hgent – Amount Of Rent Minus Heating Costs

0	33
1	6
2	4
3	7
4	5
5	21
6	15
7	12
8	18
9	27
10	61
11	32
12	62
13	68
14	61
... (1889 rows omitted)	224938
5040	1
5148	1
5700	1
5915	1
6000	2
6700	2
7318	1
-1 No answer	310
-2 Does not apply	189863
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	12818
-6 Questionnaire Version With Modified Filter	0

-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGRENT is a measure of gross rent, i.e. it equal to the base rent plus utility costs (HGUTIL) but does not include heating (HGHEAT), electricity costs (HGELECTR) and additional gas costs (HGGAS). HGRENT is converted into Euro values for all waves including those prior to Wave S.

The questions for HGRENT and utility costs have changed considerably since the first SOEP questionnaire in Wave A. From Waves A to Wave G, the amount of rent stated by the households in SOEP is in principle the desired concept of gross rent, i.e. basic rent excluding heating and electricity costs but including utility costs. However, in these waves the information on utility costs was not collected.

From Wave H onwards, households simply state the amount of rent they pay. Following this question, it is asked whether heating and utility costs are included in that amount of rent and what the exact costs for heating and utilities eventually are (in the latter case only if they are included).

In Wave BE and Wave BF, the questionnaire design was different to all other waves. In a first step, households were asked for their overall monthly housing costs (sum of base rent, heating costs, additional utility costs, electricity and gas). In a second step, they stated the respective expenses. In order to obtain RENT\$\$, heating, gas, and electricity costs have to be deducted from these overall monthly housing costs. In Wave BE, some households state monthly housing costs that presumably do not include all possible expenses. They might do so because they do not know all their expenses or because they misunderstood the question. We attribute the drop of rent from Wave BD to Wave BE and Wave BF to the different questionnaire design and do no longer adjust the values as we did in data release v31. Therefore, the average rent of Wave BE (data release v31) is about 10 Euros lower in data release v32 and v33 than it was in v31. Both, in Wave BE and Wave BF, implausible values of rent (e.g. if utility costs are above total living costs) have been set to missing to be imputed in the following step.

The questionnaire design in Wave BG was set back to the format of Wave BD.

Missing values of HGRENT are treated as follows: In a first step, past values of the two last years are carried forward and adjusted for inflation given that the household still lives at the same address and the dwelling's size has remained the same. The type of rent (heating/utility costs included or not) is then also carried forward. In a second step, if HGRENT 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 HGRENT are imputed by Stata's chained imputation procedure. Imputed values are not carried forward.

General Information on the Imputation Procedure Since data release v31, missing values of HGROOM, HGSIZE, HGRENT, HGHEAT, HGUTIL, and HGELECTR are imputed with Stata's chained imputation procedure (`mi impute chained`). We impute separately for each wave, for East and West and for owners and tenants. The regression specifications vary between these groups but in general they include:

- For owner:
 - HGROOM, HGSIZE, HGCONDIT, HGCNSTYRMAX
 - Number of persons in household
 - Household net income as stated in questionnaire
 - Type of dwelling (family house, apartment, etc.)

- For tenants:
 - HGROOM, HGSIZE, HGRENT, HGHEAT, HGUTIL, HGELECTR, HGCONDIT, HGCNSTYRMAX, HGEQPHEA, HGEQPFHEA, HGEQPTER, HGEQPGAR, HGEQPSOL, HGEQPAIR, HGCNSTYRMAX, HGREVAL, HGREDDUC
 - HGSIZE 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: HGHEAT included in HGRENT (Wave G to Wave BD)
 - Dummy: HGUTIL included in HGRENT (Wave H to Wave BD)
 - Dummy: HGUTIL partly included in HGRENT (Wave H to Wave BD)
 - Dummy: HGHEAT includes electricity costs (since Wave BE)
 - Dummy: household pays only utility costs (since Wave BE)
 - Household has children younger than 16

In addition, the regression of HGRENT contains HGRSUBS, a dummy for children under 16, information on the residential area (\$\$wum3) and BIK regions. The regression of HGHEAT includes a dummy for children under 16. The regression of HGUTIL includes HGEQPLIF and omits HGRENT, HGCONDIT, the type of dwelling and the dummy for whether HGHEAT is included in HGRENT. The regression of HGSIZE omits HGSIZE squared.

HGRENT 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 HGHEAT/HGUTIL included in HGRENT to account for the different types of rent. Note that these dummies are also imputed if they are missing. If HGUTIL contains a partial amount, it is set to missing and then also imputed.

Residents of homes (HGOWNER==5) are excluded from the imputation. Values larger than four times the 99th percentile in one of the target variables are also excluded.

For all six target variables, the predictive mean matching imputation method is applied. We use 175 iterations as burn-in period. Furthermore, we only distribute one imputed value for each observation.

Gross warm rent including electricity (Bruttowarmmiete inkl. Strom) could be computed from this data as follows: HGRENT + HGHEAT + HGELECTR.

HGRENT is set missing (= -5) for all M3 and M4 respondents although the migration household questionnaire surveyed rent information. However, the surveyed information for M3 and M4 is not comparable with the information for the other samples as information on HGHEAT, HGUTIL, HGELECTR, etc. is not surveyed.

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

hgrentinfo – Pays Rent and/or Utilities

1	Pays Rent	69543
2	Pays Utility Costs, but No Rent	2217
3	Does Not Pay either Rent or Utilities	1311
-1	No answer	217
-2	Does not apply	54904
-3	Not valid	0
-4	Forbidden Multiple Response	0

-5	Not Included In Questionnaire Version	12818
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	287360

Waves: Since 2014

HGRENTINFO is a filter question for housing costs. It was introduced in Wave BE and replaces NORENT\$\$ in \$HGEN files.

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

hgnorent – Does Not Pay Rent

1	Pays No Rent	11764
-1	No answer	302
-2	Does not apply	403478
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	12826
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: 1984–2013

HGNORENT marks tenants who do not pay rent. Until Wave BD, it was a filter question for all housing costs such that non-rent paying tenants were not asked for HGHEAT, HGUTIL and HGELECTR. In Wave BE, the filter question was rephrased and tenants were asked for their utility costs even if they stated not to pay a rent. To maintain consistency over time, we set HGNORENT to 1 for all tenants who only pay utility costs. RENTINFO\$\$ contains the information of the new filter question.

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

hgelectr – Costs of Electricity (gen)

0	5
1	10
2	2
3	5
4	11
5	20
6	9
7	7
8	13
9	10
10	158
11	23
12	48
13	41
14	64
... (306 rows omitted)	104126

530		1
540		1
580		1
690		1
828		1
840		1
950		1
-1	No answer	2123
-2	Does not apply	88344
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	12818
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	220526

Waves: Since 2010

Costs for electricity (excluding arising expenses for heating if applicable) have been collected starting from Wave BA if tenants pay a rent. From Wave BE onwards, they have been collected from all tenants. In Wave BF, total electricity costs were explicitly asked for so that costs for heating and warm water were subtracted to generate HGELECTR that year. Furthermore, due to changes in the questionnaire design in Wave BF, HGELECTR is neither carried backward nor forward. See HGHEAT for further explanations on housing costs in Wave BF. Missing values of HGELECTR 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 HGELECTR 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 HGELECTR are imputed by Stata's chained imputation procedure. Imputed values are not carried forward. We do not impute HGELECTR for years before Wave BA. For more information on the imputation procedure, see the description of HGRENT.

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

hgelectrinfo – Reason for Missing Electricity Costs

1	Electricity Costs Unknown	2
2	Electricity Costs Included in Heating Costs	282
3	No Electricity Costs	2653
-1	No answer	237
-2	Does not apply	109027
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	12818
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	303351

HGELECTRINFO was introduced in Wave BE and indicates why HGELECTR is missing. If a household stated that it did not know its electricity costs, we impute HGELECTR and set HGELECTRINFO to -2.

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

hggas – Costs of Gas (EUR)

0		40
1		1
3		1
4		1
5		9
6		6
7		10
8		13
9		12
10		15
11		6
12		7
13		7
14		3
15		16
...	(115 rows omitted)	783
210		1
213		1
220		1
234		1
250		5
300		3
550		1
-1	No answer	9694
-2	Does not apply	5400
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	287360
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	124973

HGGAS states the monthly costs for gas not used for heating in Euros. HGGAS was only collected in 2014 (wave BE). Missing values of HGGAS are not imputed due to the low number of households which state gas costs.

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

hggasinfo – Reason for Missing Gas Costs

1	Gas Costs Unknown	1051
2	Electricity Costs Included in Heating Costs	1545
3	No Gas Costs	5392
-1	No answer	8
-2	Does not apply	8041
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	287360

-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	124973

HGGASINFO has only been asked in Wave BE and indicates why a household did not state additional gas costs that were not included in the heating costs.

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

hgheat – Costs Of Warm Water, Gen

0		457
1		17
2		25
3		17
4		24
5		156
6		76
6.5		1
7		186
8		102
9		102
10		764
11		139
12		423
13		196
...	(415 rows omitted)	213775
858		1
860		1
880		1
900		3
930		1
999		1
-1	No answer	8026
-2	Does not apply	179814
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	24061
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0
-285		1

Waves: Since 1986

HGHEAT contains the heating costs of tenants. It is collected since Wave C. Until Wave BD, only tenants who stated to pay a rent were asked. Since Wave BE, every tenant was asked. Heating costs are reported in Euro for all years.

In Wave BF, HGHEAT 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 HGHEAT,

costs for cooking and electricity costs other than heating have been excluded by regressing the usage-dummies and controlling for household-specific (number of people living in the household, type of tenant) and dwelling-specific variables (number of rooms, size of the dwelling, construction year, living environment). The shares for heating and warm water were then added to generate HGHEAT. Due to changes in the questionnaire design, firstly, HGHEAT is lower in Wave BF than it has been in previous and following years and secondly, HGHEAT from Wave BF is neither carried backward nor forward. In file HGEN, only HGHEAT is distributed. All energy cost variables can be found in the dataset BFH.

Missing values of HGHEAT 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 HGHEAT 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 HGHEAT 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 HGHEAT for Wave A and Wave B anymore, since data release v32. For more information on the imputation see the description of HGRENT.

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

hgheatinfo – Reason for Missing Heating Costs

1	Heating Costs Unknown	19
3	No Heating Costs	5594
-1	No answer	255
-2	Does not apply	106333
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	12818
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	303351

Waves: Since 2014

HGHEATINFO was introduced in Wave BE and indicates why HGHEAT is missing. If a household stated that it did not know its heating costs, we imputed HGHEAT and set HGHEATINFO to -2.

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

hgmoveyr – Year Moved Into Dwelling

1900	1
1901	34
1902	11
1903	7
1904	18
1905	4
1906	24
1907	96
1908	8

1909	47
1910	19
1911	18
1912	69
1913	22
1914	33
... (101 rows omitted)	393042
2016	9776
2017	7687
2018	5945
2019	4510
2020	2243
2021	1334
2022	1
-1 No answer	0
-2 Does not apply	3421
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Waves: all

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

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

hgutil - Other Monthly Utility Costs

0	5
1	254
2	469
3	165
4	128
5	647
6	183
7	297
8	311
9	139
10	2373
11	178
12	778
13	223

14		186
...	(377 rows omitted)	176205
831		1
850		1
860		1
880		1
900		5
930		1
1400		1
-1	No answer	7651
-2	Does not apply	173535
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	26950
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	37682

hgutilinfo – Reason for Missing Additional Utility Costs

1	Utility Costs Unknown	30
3	No Other Additional Utility Costs	14092
-1	No answer	249
-2	Does not apply	113821
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	12818
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	287360

hgseval – Adequacy Of Living Space In Housing Unit

1	Much Too Small	12059
2	A Bit Too Small	63820
3	Just Right	298539
4	A Bit Too Large	42618
5	Much Too Large	5594
-1	No answer	5740
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

Subjective assessment by respondent (household head). From Wave C to Wave G, the information on HGSEVAL was only collected for new households or households that had moved

(households with “blue” questionnaires) and immobile households whose HGSIZE had changed. In these Waves, HGSEVAL is carried forward from previous years.

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

hgsizex - Size Of Housing Unit In Sq M

6	16
7	1
8	16
9	35
10	142
11	55
12	358
13	49
14	158
15	401
16	361
17	99
18	309
19	89
20	1229
... (321 rows omitted)	421085
698	1
700	1
720	1
754	1
800	1
838	2
938	1
-1 No answer	3959
-2 Does not apply	0
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Waves: all

Up to Wave R (except for Wave O), HGSIZE was collected only in the first interview with new households. In case a household had moved or whether a household's housing unit size had changed due to renovations or additions (up to Wave G, these households filled out a “green” questionnaire). From Wave S onwards, the question on the size of the housing unit has been posed to all households annually. For households still residing at their old address and neither having moved nor renovated their dwelling, the information on the size of the housing unit is carried forward as a status variable in order to provide valid current information. Analogously, the information is carried back to fill gaps if households missed to state the size of the housing in the first year after moving.

In the case the information on the size of the housing unit is still missing, it is imputed simultaneously with other variables using Stata's chained imputation procedure. Imputed

values are not carried forward. For information on the imputation procedure, see the description of HGRENT.

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

hgf2rent – Detailed Imputation Flag, Rent Minus Heating Costs

0	Not imputed	348705
1	Reported Rent Corrected for Imputed Heat	31286
2	Reported Rent Corrected for Imputed Util	35586
3	Reported Rent Corrected for Imputed Electr	628
4	Reported Rent Corrected for Heat and Util	3661
5	Reported Rent Corrected for Heat and Electr	2174
6	Reported Rent Corrected for Util and Electr	24
7	Reported Rent Corrected for Heat, Util and Electr	15
8	Reported Rent, at least one Sub-Aggregate Imputed	3932
9	Rent Imp., Sub-Aggregates Not Imputed	804
10	Rent Imp. and Corrected For Imp. Heat	588
11	Rent Imp. and Corrected For Imp. Util	436
12	Rent Imp. and Corrected For Imp. Electr	11
13	Rent Imp. and Corrected For Imp. Heat and Util	158
14	Rent Imp. and Corrected For Imp. Heat and Electr	269
15	Rent Imp. and Corrected For Imp. Util and Electr	25
16	Rent Imp. and Corrected For Imp. Heat, Util and Electr	10
17	Rent Imp. & at least one Imp. Sub-Aggregates	58
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGF2RENT indicates whether or not HGRENT has been imputed (=1) by Stata's chained imputation procedure.

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

hgfelectr – Imputation Flag, Costs of Electricity

0	Not imputed	200166
1	Imputed by PMM	4438
2	Electricity Costs from Previous or Subsequent Years	3240
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0

-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	220526

Waves: Since 2010

HGFELECTR indicates whether HGELECTR value has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves.

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

hgfheat – Imputation Flag, Costs Of Heating And Warm Water

0	Not imputed	370508
1	Imputed by PMM	25497
2	Electricity Costs from Previous or Subsequent Years	21122
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	11243
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: Since 1986

HGFHEAT indicates whether HGHEAT has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves.

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

hgfrent – Imputation Flag, Amount Of Rent Minus Heating Costs

0	Not imputed	348903
1	Rent and/or Sub-Aggregate Imputed by PMM	60903
2	Rent and/or Sub-Aggregate from Previous or Subsequent Years	16004
3	Rent and/or Sub-Aggregate From Prev./Subs. Years and/or PMM	2560
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGFRENT indicates whether HGRENT or expenses added or subtracted to HGRENT has been imputed (=1) by Stata's chained imputation procedure, whether HGRENT itself or expenses added or subtracted to HGRENT was carried for- or backwards (=2) from past/subsequent waves or whether HGRENT is the result of a mixture of imputed and past/subsequent waves' values. The fourth category marks households who very likely did not include all utility costs

in their overall housing costs in Wave BE (the question from which HGRENT is generated). We identified these households by their longitudinal information on housing costs and did not subtract HGHEAT and/or HGELECTR to obtain their HGRENT. HGF2RENT contains more detailed information about which imputed expenses were added or subtracted.

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

hgffroom - Imputation Flag, Number Of Rooms Lager Than 6 Sq M

0	Not imputed	424444
1	Imputed	3926
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGFFROOM indicates whether or not HGROOM has been imputed (=1) by Stata's chained imputation procedure.

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

hgffsize - Imputation Flag, Size Of Housing

0	Not imputed	425447
1	Imputed	2923
-1	No answer	0
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGFFSIZE indicates whether or not HGFSIZE has been imputed (=1) by Stata's chained imputation procedure.

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

hgffutil - Imputation Flag, Other Monthly Utility Costs

0	Not imputed	276534
1	Imputed by PMM	83761
2	Electricity Costs from Previous or Subsequent Years	16261
-1	No answer	0
-2	Does not apply	0

-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	14132
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	37682

Waves: Since 1991

HGFUTIL indicates whether HGUTIL value has been imputed (=1) by Stata's chained imputation procedure or whether it has been carried for- or backwards (=2) from past/subsequent waves.

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

6 Subsidization of Housing Costs

hgosubs – Amount Of Subsidies Last Year

1	Yes	13048
2	No	80190
-1	No answer	1225
-2	Does not apply	108728
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	225179

Waves: 2000–2014

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

Please note: The old variable \$FOERD (available until SOEP data release v25 in \$HGEN) is discarded. Homeowner subsidies in Germany have been subject to major revisions and fluctuations over time. The corresponding question in SOEP was in some waves only posed to new households and those that have moved, in some waves it was not surveyed at all. For these reasons, the question for government housing subsidies was changed in Wave Q to cover direct subsidies received the previous year. HGSUBSID and HGOSUBS replace the old variable \$FOERD.

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

hgreduc – Dwelling At A Reduced Prize

1	Yes	18562
2	No	199056
-1	No answer	20960
-2	Does not apply	173757
-3	Not valid	0

-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	4792
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	11243

Waves: 1986–2002 (west), 1990 & 1992–2002 (east), since 2008 (both)

HGREduc states whether the flat is offered by the owner at a reduced rent. Information is carried forward from previous waves for old households residing at their old address. For newly surveyed households and for old households that have moved, newly collected data is used. From Wave T to Wave X, this information was not collected. It is carried forward from Wave S for households who have not moved and whose stated amount of rent vary only slightly. The new information from Wave Y is then carried backward for households with the same characteristics if HGREduc is still missing after carrying forward from Wave S.

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

hgrsubs – Government Subsidized Housing

1	Yes, With Due Diligence	21824
2	Yes, With Run Out Due Diligence	4896
3	No	210421
-1	No answer	5763
-2	Does not apply	175919
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	9547
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: Since 1984 (west), 1990 & since 1994 (east)

HGRSUBS states whether the rent is subsidized by government. Up to Wave K (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 waves. East German households were asked for the first time in Wave J (1993).

The second response category, which indicates expired subsidization, was added in Wave L (1995). In Wave BG (2016), the category was removed. In Germany, rent subsidization is typically set up as a system of subsidized loans to real estate investors and an obligation to offer below market rents. The obligation expires after one to several decades. In principle, there is the possibility that households say to live in a subsidized flat, although subsidization has expired.

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

hgsubsid – Government Subsidizes Housing Payments

1	Yes	5274
2	No	8767
-1	No answer	11769
-2	Does not apply	66348

-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	21422
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	314790

Waves: 1984–1999 (west), 1994–1999 (east)

Statement by respondent. HGSUBSID contains information on government subsidies at the time the housing was built or bought. From Wave B to Wave N, this was only asked to new households or in case an old household had moved. Information is then carried forward. In Wave O and Wave P, the question was again posed to the whole population.

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

7 Typology of Household

hgowner – Tenant Or Owner Of Dwelling

1	Owner	177627
2	Main Tenant	232584
3	Sub-Tenant	10234
4	Tenant	109
5	Living In A Home or Shared Accommodation	2987
-1	No answer	37
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	4792
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

Up to Wave H, the information for HGOWNER was collected in separate questionnaires for “old” and first-time respondents, respectively (“blue” and “green” questionnaires). In all waves and for each household with the original value of -1 (“no answer”), codes are set to =1 and =4 if at least one owner- or tenant-specific item is given. Code 4 is also used if a change in ownership (from owner to tenant) has taken place but no original information for HGOWNER was given. Code 5 (‘resident of a home or institutional living facility’) has only been assigned by interviewers during fieldwork since Wave P.

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

hgtyp1hh – Household Typology, One Digit

1	1-Pers.-HH	104876
2	Couple Without Children	119355
3	Single Parent	35151
4	Couple With Children LE 16	95570
5	Couple With Children GT 16	37483

6	Couple With Children LE And GT 16	23513
7	Multiple Generation-HH	4938
8	Other Combination	7354
-1	No answer	130
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

HGTYP1HH is an aggregation of HGTYP2HH. For further information on HGTYP1HH, please refer to to HGTYP2HH.

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

hgtyp2hh – Household Typology, Two Digit

11	1-Person HH LE 35 Y.	27054
12	1-Person HH 35- LT 60 Y.	32003
13	1-Person HH GE 60 Y.	45819
21	Couple Without Children	119355
31	Single Parent,1 Child, LE 16	10913
32	Single Parent,2 Or More Children, LE 16	6817
33	Single Parent,1 Children, GT 16	10532
34	Single Parent,2 Or More Children, GT 16	2592
35	Single Parent,2 Children, LE and GT 16	2369
36	Single Parent,3 Or More Children, LE and GT 16	1928
41	Couple, 1 Child, LE 16	34812
42	Couple, 2 Children, LE 16	40656
43	Couple, 3 Or More Children, LE 16	20102
51	Couple, 1 Child, GT 16	24736
52	Couple, 2 Children, GT 16	10706
53	Couple, 3 Or More Children, GT 16	2041
61	Couple, 2 Children, LE and GT 16	10810
62	Couple, 3 Or More Children, LE and GT 16	12703
71	3-Generation-HH	4435
72	4-Generation-HH	33
73	GrandParents-GrandChildren-HH	470
81	Other Combination Without K. LE 16	7351
82	Other Combination With K. LE 16	3
-1	No answer	130
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

Generated variable created by combining the relationships of all persons living in the household to the head of household (Variable STELL_H in the file PBRUTTO) at the time of the survey. With data release v26 the data production process switched to a standardized procedure for all Waves to ensure longitudinal consistency, resulting in minor changes compared with older distributions. HGTY1HH is an aggregation of HGTY2HH and corresponds to the first digit of the two-digit code of HGTY2HH. Single households are differentiated in HGTY2HH according to age. The following abbreviations are used in the labels: LE=less or equal, LT=less than, GE=greater or equal, GT=greater than, u.m. = and more, 1-P-HH/1-Pers.-HH = one-person household, K. = children, Alleinerz. = single parent, J. = years, HH = Haushalt, sonst. = sonstige.

Help for old friends: Starting with data distribution v26, the category "(88) Other combination" has been further differentiated into (81) households without vs. (82) those with children (up to the age of 16). Starting with distribution v36, the differentiation by gender of single households has been removed.

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

8 Household Income

hghinc – Monthly Household Net Income (EUR)

0	86
1	12
8	1
10	2
12	1
20	1
26	5
28	1
31	2
35	1
40	1
49	1
50	4
51	5
52	1
... (5877 rows omitted)	405635
400000	1
600800	1
999999	1
1700000	1
2000000	1
-1 No answer	20729
-2 Does not apply	301
-3 Not valid	1574
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0

-8	Question not part of the survey program this year	0
-400		1
-3000		1

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).

For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

hgi1hinc - 1. Imputed Monthly Net Household Income (EUR) [1/5]

0		60
1		13
10		3
12		1
20		1
28		1
35		1
40		1
50		4
51		1
70		2
72		1
77		1
78		1
80		1
...	(5765 rows omitted)	363780
250000		1
300000		2
400000		1
600800		1
999999		1
1700000		1
2000000		1
-1	No answer	9
-2	Does not apply	64481
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

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 dataset: the variables HGI1HINC-HGI5HINC.

The imputations were calculated using multiple imputations by chained equations. Up to Wave BB the program ICE of Stata which was written by Patrick Royston (see Royston 2004,

2005a, 2005b) and which is based on the program MICE in S-Plus and R was used. Since Wave BC the Stata command `mi impute` is used. The missing observations are assumed to be missing at random. We set the number of imputations $m=10$ and get 10 multiple imputed values for `IHINC$$`. For a discussion on the choice of m , see Rubin (Wave D) and Royston (Wave U).

The dataset MIHINC contains the complete imputation results and is separately available. To be compatible with methods for analyzing multiply imputed data, MIHINC is constructed in the so called stacked or MIM Dataset Format. It contains the following variables: `HHNRAKT`, `SVYYEAR`, `MJ`, `MI`, `IHINC` and `IMPFLAG`. For all household from Wave L to Wave BD, there are ten imputed values for the current household income. `MJ` identifies the individual dataset to which each observation belongs while `MI` identifies the observations within each individual dataset. To distinguish between the original data containing missing values and the imputed values, the dummy variable `IMPFLAG` is added. In the `$HGEN` files five of these imputed incomes are stored in the conventional wide format.

The number of iterations carried out in each prediction model was specified to be 500. For East and West Germany, imputations were done separately. Furthermore, the option for predicted mean matching was chosen which means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value.

Most important variables for modelling the current household net income consist in the household net income of the previous year, in basic information about the household and changes in its composition as well as all relevant income components received.

####The complete list of the variables used for modelling

- Description of household:
 - size, number of children, sample
 - head of household: not German, age, sex
 - changes in household composition between years: births, deaths, persons entering or leaving the household or being temporarily absent
- Financial Situation:
 - Monthly household income previous year
 - Income from employment
 - Pensions
 - Sum of personal incomes (e.g. Support from the “Arbeitsamt”, Maternity benefit, Alimony, etc.)
 - Household related incomes (e.g. Child allowance, Housing assistance, Social assistance, Unemployment benefit, Assets, etc.)
 - Fraction of persons greater than 16 in household who refused answering a component of income (0-1)
- Number of persons not attended survey (PUNR, partial unit nonresponse)
- Cross-sectional weights

####Analyzing multiply imputed data

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

combining the average of the squared standard errors of the several (m) estimates with the within- and between-imputation variance.

- Stata provides various built-in functionality called mi.
- Within SAS, the MIANALYZE procedure combines the results of the analyses of imputations and generates valid statistical inferences: <http://support.sas.com/rnd/app/stat/procedures/mianalyze.html>
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

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>

hgi2hinc – 2. Imputed Monthly Net Household Income (EUR) [2/5]

0	60
1	13
10	2
12	1
20	1
28	1
35	1
40	1
50	4
51	1
70	2
72	1
77	1
78	2
80	1
... (5765 rows omitted)	363780
250000	1
300000	2
400000	1
600800	1
999999	1
1700000	1
2000000	1
-1 No answer	9
-2 Does not apply	64481
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0

-8 Question not part of the survey program this year 0

hgi3hinc – 3. Imputed Monthly Net Household Income (EUR) [3/5]

0	60
1	12
10	2
12	1
20	1
28	1
35	1
40	1
50	5
51	1
70	2
72	1
77	1
78	1
80	1
... (5765 rows omitted)	363781
250000	1
300000	2
400000	1
600800	1
999999	1
1700000	1
2000000	1
-1 No answer	9
-2 Does not apply	64481
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

hgi4hinc – 4. Imputed Monthly Net Household Income (EUR) [4/5]

0	60
1	15
10	2
12	2
20	1
28	2
35	1
40	1
50	5
51	1
70	2

72	1
77	1
78	1
80	1
... (5765 rows omitted)	363775
250000	2
300000	2
400000	1
600800	1
999999	1
1700000	1
2000000	1
-1 No answer	9
-2 Does not apply	64481
-3 Not valid	0
-4 Forbidden Multiple Response	0
-5 Not Included In Questionnaire Version	0
-6 Questionnaire Version With Modified Filter	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

hgi5hinc – 5. Imputed Monthly Net Household Income (EUR) [5/5]

0	59
1	13
10	2
12	1
20	1
28	1
35	1
40	1
50	4
51	1
70	2
72	1
77	1
78	1
80	1
... (5765 rows omitted)	363782
250000	1
300000	2
400000	1
600800	1
999999	1
1700000	1
2000000	1
-1 No answer	9
-2 Does not apply	64481
-3 Not valid	0
-4 Forbidden Multiple Response	0

-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

hghinc – Imputation Flag, Monthly Net Household Income

0	Not imputed	408551
1	Imputed	19809
-1	No answer	3
-2	Does not apply	7
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

HGFHINC is a dummy variable indicating whether an observation was missing on HGHINC and was therefore imputed or not.

For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>

9 Time and Method of Interview

iyear – Household Related Survey Year

2017	18771	
2018	19033	
2019	19454	
2020	19255	
2021	13882	
2022	812	
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	337163

hghmonth – Month Of Interview

1	January	25204
2	February	101251
3	March	102222
4	April	60476
5	May	38723
6	June	29401
7	July	22002
8	August	16322
9	September	12914
10	October	9132
11	November	6382

12	December	4303
-1	No answer	37
-2	Does not apply	0
-3	Not valid	1
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

The month of participation in the survey is generated using data from the household questionnaire. Missing information is filled in using data from the HBRUTTOL file. Interviews that took place in the month of December, and prior to the 20th of that month, were recoded to -3.

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

hghmode – Interview Method

100	With Interviewer Assistance	4460
110	Oral Interview	101706
120	Written Ques. Interviewer	64992
131	Written Ques. No Interviewer	6176
132	Oral And Written	7396
133	Proxy	12
134	Third Person Present	1334
135	No Third Person Present	9175
140	CAPI - Since 1998 (O)	171280
150	CAWI Since 2014 (BE)	6301
160	CATI since 2021	484
170	CASI since 2021	0
210	Written, By Mail	44093
220	Telephone Interview	3166
230	CAPI by Phone	7784
-1	No answer	11
-2	Does not apply	0
-3	Not valid	0
-4	Forbidden Multiple Response	0
-5	Not Included In Questionnaire Version	0
-6	Questionnaire Version With Modified Filter	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

Waves: all

The interview method is generated through data from the household questionnaire. Missing information is filled in with data from the corresponding HBRUTTO file.

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

inputdataset – Source dataset

ahgen	5921
bahgen	15414
bbhgen	16819
bchgen	16508
bdhgen	18093
behgen	16037
bfhgen	15991
bghgen	17775
bhgen	5322
bhhgen	19701
bihgen	18682
bjhgen	19030
bkhgen	20183
blhgen	13611
chgen	5090
... (8 rows omitted)	48145
lhgen	6968
mhgen	6894
nhgen	6810
ohgen	7668
phgen	7388
qhgen	13249
rhgen	11947
shgen	12692
thgen	12061
uhgen	11796
vhgen	11440
whgen	12499
xhgen	11689
yhgen	11058
zhgen	11889