

DIW Berlin / SOEP (Ed.)

**Research Report**

## SOEP-IS 2020 - HGEN: Household-related status and generated variables

SOEP Survey Papers, No. 1156

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* DIW Berlin / SOEP (Ed.) (2022) : SOEP-IS 2020 - HGEN: Household-related status and generated variables, SOEP Survey Papers, No. 1156, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/263171>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-sa/4.0/>

# 1156<sup>2022</sup>

**SOEP** Survey Papers  
Series D – Variable Descriptions and Coding

## SOEP-IS 2020 – HGEN: Household-Related Status and Generated Variables

SOEP-IS 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. Jan Goebel, DIW Berlin  
Prof. Dr. Stefan Liebig, DIW Berlin and Freie Universität Berlin  
Prof. Dr. David Richter, DIW Berlin and Freie Universität 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-IS Group, 2022. SOEP-IS 2020 – HGEN: Household-Related Status and Generated Variables.  
SOEP Survey Papers 1156: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2022 by SOEP

ISSN: 2193-5580 (online)

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

[soeppapers@diw.de](mailto:soeppapers@diw.de)

# SOEP-IS 2020 – HGEN: Household-Related Status and Generated Variables

SOEP-IS Group  
2022

## Contents

<b>1 Overview</b>	<b>4</b>
<b>2 Variables in file hgen</b>	<b>4</b>
cid – Original Household Number . . . . .	4
hid – Current Wave HH Number . . . . .	4
syear – Survey Year . . . . .	4
hgnrpers – Number Of Persons In HH . . . . .	4
hgnrkid14 – Number Of Children Under 14 In HH . . . . .	5
hgnrkid16 – Number Of Children Under 16 In HH . . . . .	5
hgmoveyr – Year Moved Into Dwelling . . . . .	5
hgcnstyr – Year House Was Built . . . . .	6
hgcondit – Condition Of House . . . . .	7
hgseval – Adequacy Of Living Space In Housing Unit . . . . .	7
hgeqpkit – Dwelling Has Kitchen . . . . .	7
hgeqpshw – Dwelling Has Indoor Bath,Shower . . . . .	8
hgeqpiwc – Dwelling Has Indoor Toilet . . . . .	8
hgeqphea – Dwelling Has Central,Floor Heat . . . . .	8
hgeqpter – Dwelling Has Balcony,Terrace . . . . .	9
hgeqpbas – Dwelling Has Basement . . . . .	9
hgeqpgar – Dwelling Has Garden . . . . .	9
hgeqpwat – Dwelling Has Water Boiler . . . . .	9
hgeqptel – Dwelling Has Telephone . . . . .	10
hgeqpalm – Dwelling Has Alarm System . . . . .	10
hgeqpsol – Dwelling Has Solar System . . . . .	10
hgeqpair – Dwelling Has Air Conditioner . . . . .	11
hgeqplif – Dwelling Has An Elevator . . . . .	11
hgeqpnrj – Dwelling Has Alternative Energy Source . . . . .	11
hgowner – Tenant Or Owner Of Dwelling . . . . .	12
hgacquis – Means Of Acquiring Dwelling . . . . .	12
hgsubsid – Government Subsidizes Housing Payments . . . . .	12
hgsubss – Amount Of Subsidies Last Year . . . . .	13
hgrevall – Rent Of Dwelling Vs Comparable Dwellings . . . . .	13
hgrsubs – Government Subsidized Housing . . . . .	14
hgreduc – Dwelling At A Reduced Prize . . . . .	14
hgsize – Size of Housing Unit in Square Meters . . . . .	14
hgroom – Number Of Rooms Larger than 6 Square Meters . . . . .	16
hgheat – Costs Of Warm Water (gen) . . . . .	16
hgutil – Cost Allocation (gen) . . . . .	17
hgrent – Amount of Rent Minus Heating Costs (EUR) . . . . .	18
hgelectr – Costs of Electricity (gen) . . . . .	19
hgnovent – Does Not Pay Rent . . . . .	20
hgfsiz – Imputation Flag, Size Of Housing . . . . .	20
hgfrorm – Imputation Flag, Number Of Rooms Lager than 6 Square Meters . . . . .	20
hgfhheat – Imputation Flag, Costs Of Heating And Warm Water . . . . .	21
hgffutil – Imputation Flag, Cost Allocation . . . . .	21
hgffrent – Imputation Flag, Amount Of Rent Minus Heating Costs . . . . .	21
hgffelectr – Imputation Flag, Costs of Electricity . . . . .	22
hgtyp1hh – Household Typology, One Digit . . . . .	22
hgtyp2hh – Household Typology, Two Digit . . . . .	23

hghmonth – Month Of Interview . . . . .	23
hghmode – Interview Method . . . . .	24
hghinc – Monthly Household Net Income (EUR) . . . . .	24
hgahinc – Adjusted Monthly Household Net Income (EUR) . . . . .	25
hgnuts1 – NUTS-Systematic-1 (Federal State) . . . . .	27
<b>3 Imputation of Household Income</b> . . . . .	<b>27</b>
hgi1hinc – 1. Imputed Monthly Net Household Income (EUR) [1/15] . . . . .	28
hgi2hinc – 2. Imputed Monthly Net Household Income (EUR) [2/15] . . . . .	29
hgi3hinc – 3. Imputed Monthly Net Household Income (EUR) [3/15] . . . . .	30
hgi4hinc – 4. Imputed Monthly Net Household Income (EUR) [4/15] . . . . .	30
hgi5hinc – 5. Imputed Monthly Net Household Income (EUR) [5/15] . . . . .	31
hgi6hinc – 6. Imputed Monthly Net Household Income (EUR) [6/15] . . . . .	32
hgi7hinc – 7. Imputed Monthly Net Household Income (EUR) [7/15] . . . . .	33
hgi8hinc – 8. Imputed Monthly Net Household Income (EUR) [8/15] . . . . .	33
hgi9hinc – 9. Imputed Monthly Net Household Income (EUR) [9/15] . . . . .	34
hgi10hinc – 10. Imputed Monthly Net Household Income (EUR) [10/15] . . . . .	35
hgi11hinc – 11. Imputed Monthly Net Household Income (EUR) [11/15] . . . . .	35
hgi12hinc – 12. Imputed Monthly Net Household Income (EUR) [12/15] . . . . .	36
hgi13hinc – 13. Imputed Monthly Net Household Income (EUR) [13/15] . . . . .	37
hgi14hinc – 14. Imputed Monthly Net Household Income (EUR) [14/15] . . . . .	37
hgi15hinc – 15. Imputed Monthly Net Household Income (EUR) [15/15] . . . . .	38
hgfhinc – Imputation Flag, Monthly Net Household Income (EUR) . . . . .	39

## 1 Overview

Variables in the file hgen documented here are generated mostly from the answers in the household questionnaire. There is one row for each wave (syear) a household (hid) participated in the survey.

## 2 Variables in file hgen

**cid** – Original Household Number

---

**hid** – Current Wave HH Number

---

**syear** – Survey Year

---

1998	373
1999	383
2000	391
2001	409
2002	415
2003	424
2004	433
2005	441
2006	452
2007	451
2008	447
2009	1942
2010	1628
2011	1504
2012	2277
2013	3173
2014	4232
2015	3743
2016	4599
2017	4073
2018	3717
2019	3351
2020	3729

**hgnrpers** – Number Of Persons In HH

---

1	13107
2	16040
3	5768
4	5256
5	1809
6	437
7	126
8	31
9	13
-1 [-1] No Answer	0

-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

#### hgnrkid14 – Number Of Children Under 14 In HH

---

1		4831
2		3187
3		806
4		120
5		27
6		4
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	33612
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

#### hgnrkid16 – Number Of Children Under 16 In HH

---

1		5173
2		3967
3		1127
4		190
5		46
6		12
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	32072
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

#### hgmmoveyr – Year Moved Into Dwelling

---

? (from: soep-is/soep-bip-2020-e/Q4:hwj[])

? When did you move into this dwelling? // Year (from: soep-is/soep-is-2020/Q4:hwj[])

? How long have you lived in your home? // Year (from: soep-is/soep-is-2020-fgz/Q010:hwj[])

1917	1
1921	19
1925	7
1926	13
1928	8
1930	11
1931	7
1932	12



1933	11
1934	25
1935	49
1936	70
1937	45
1938	30
1939	52
... (73 rows omitted)	31759
2013	1432
2014	1201
2015	1011
2016	1003
2017	740
2018	545
2019	350
2020	158
2021	1
-1 [-1] No Answer	948
-2 [-2] Does Not Apply	2054
-3 [-3] Answer Improbable	16
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	1009
-6 [-6] Questionnaire Version With Modified Filtering	0

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, 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.

### hgcnstyr – Year House Was Built

**?** When, approximately, was the building in which your dwelling is located built? (*from: soep-is/soep-is-2020/Q8:hwohn09[]*)

1 [1] Before 1919	4475
2 [2] 1919 To 1948	4893
3 [3] 1949 To 1971	10642
4 [4] 1972 To 1980	5569
5 [5] 1981 To 1990	3576
6 [6] 1991 To 2000	3536
7 [7] 2001 To 2010	1485
8 [8] 2011 Or Later	631
-1 [-1] No Answer	1779
-2 [-2] Does Not Apply	1901
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	4100
-6 [-6] Questionnaire Version With Modified Filtering	0

Classified statement of the year the building in which a household lives at the time of the survey was built.

### hgcondit – Condition Of House

---

1	[1] In A Good Condition	6997
2	[2] Partial Renovation	2422
3	[3] Full Renovations	250
4	[4] Dilapidated	11
-1	[-1] No Answer	13
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	32894
-6	[-6] Questionnaire Version With Modified Filtering	0

Respondent's estimation of the condition of the building.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgseval – Adequacy Of Living Space In Housing Unit

---

1	[1] Much Too Small	132
2	[2] A Bit Too Small	1107
3	[3] Just Right	6302
4	[4] A Bit Too Large	947
5	[5] Much Too Large	148
-1	[-1] No Answer	17
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

Estimate by the respondent (household head).

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgeqpklt – Dwelling Has Kitchen

---

1	[1] Yes	8558
2	[2] No	77
-1	[-1] No Answer	18
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934

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

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpshw – Dwelling Has Indoor Bath,Shower

---

1	[1] Yes	8588
2	[2] No	47
-1	[-1] No Answer	18
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpiwc – Dwelling Has Indoor Toilet

---

1	[1] Yes	5876
2	[2] No	19
-1	[-1] No Answer	2758
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqphea – Dwelling Has Central,Floor Heat

---

1	[1] Yes	8327
2	[2] No	306
-1	[-1] No Answer	20
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

tained in questionnaire).

#### hgeqpter – Dwelling Has Balcony,Terrace

---

1	[1] Yes	7023
2	[2] No	1587
-1	[-1] No Answer	43
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpbas – Dwelling Has Basement

---

1	[1] Yes	8202
2	[2] No	419
-1	[-1] No Answer	32
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpgr – Dwelling Has Garden

---

1	[1] Yes	5598
2	[2] No	3028
-1	[-1] No Answer	27
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpwat – Dwelling Has Water Boiler

---

1	[1] Yes	8600
2	[2] No	28
-1	[-1] No Answer	25
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqptel – Dwelling Has Telephone

---

1	[1] Yes	9327
2	[2] No	66
-1	[-1] No Answer	300
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	32894
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpalm – Dwelling Has Alarm System

---

1	[1] Yes	230
2	[2] No	5605
-1	[-1] No Answer	423
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	36329
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpsol – Dwelling Has Solar System

---

1	[1] Yes	254
2	[2] No	4671
-1	[-1] No Answer	7
-2	[-2] Does Not Apply	0

-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	37655
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpair – Dwelling Has Air Conditioner

---

1	[1] Yes	56
2	[2] No	4869
-1	[-1] No Answer	7
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	37655
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqplif – Dwelling Has An Elevator

---

1	[1] Yes	255
2	[2] No	3765
-1	[-1] No Answer	14
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	38553
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgeqpnrj – Dwelling Has Alternative Energy Source

---

1	[1] Yes	75
2	[2] No	3946
-1	[-1] No Answer	13
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	38553
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgowner – Tenant Or Owner Of Dwelling

**?** The next questions deal with the costs of living in your dwelling. First the question: Are you the main tenant, subletter, or owner? (*from: soep-is/soep-is-2020-fgz/Q022:hm01[]*)

1	[1] Owner	19128
2	[2] Main Tenant	19447
3	[3] Sub-tenant	560
4	[4] Tenant	0
5	[5] Home Occupant	31
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	3421
-6	[-6] Questionnaire Version With Modified Filtering	0

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 HGOWNER was given.

### hgacquis – Means Of Acquiring Dwelling

1	[1] Bought From Owner	841
2	[2] Inheritance, Gift	629
3	[3] Bought, Built New	1120
4	[4] Got back from public property	0
-1	[-1] No Answer	1750
-2	[-2] Does Not Apply	4313
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

Statement by respondent (household head). If no new information is provided and a change of address or ownership status (HGOWNER) has not taken place, the information of the previous year is carried forward.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgsubsid – Government Subsidizes Housing Payments

1	[1] Yes	38
2	[2] No	126

-1	[-1] No Answer	114
-2	[-2] Does Not Apply	478
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	41831
-6	[-6] Questionnaire Version With Modified Filtering	0

Statement by respondent. HGSUBSID contains information on government subsidies at the time the housing was built or bought. Information is carried forward until the year 1999.

#### hgosubs – Amount Of Subsidies Last Year

---

1	[1] Yes	416
2	[2] No	2973
-1	[-1] No Answer	44
-2	[-2] Does Not Apply	4464
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	34690
-6	[-6] Questionnaire Version With Modified Filtering	0

Statement by respondent. HGOSUBS contains information on cash housing subsidies received from the government during the year prior to the interview. Information is not carried forward.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

#### hgreal – Rent Of Dwelling Vs Comparable Dwellings

---

1	[1] Very Inexpensive	267
2	[2] Inexpensive	879
3	[3] Reasonable	1711
4	[4] Slightly Expensive	640
5	[5] Too Expensive	136
-1	[-1] No Answer	62
-2	[-2] Does Not Apply	4101
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	34791
-6	[-6] Questionnaire Version With Modified Filtering	0

Subjective assessment by respondent (household head). 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 is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).



**hgsubs** – Government Subsidized Housing

---

1	[1] Yes, With due diligence	212
2	[2] Yes, Exp.Subsidy	177
3	[3] No	3905
-1	[-1] No Answer	19
-2	[-2] Does Not Apply	4340
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

**hgreduc** – Dwelling At A Reduced Prize

---

1	[1] Yes	301
2	[2] No	3773
-1	[-1] No Answer	239
-2	[-2] Does Not Apply	4340
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

Information is carried forward from the previous years for old households residing at their old address; for new households and for old households that have moved, newly collected data is used.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

**hgsize** – Size of Housing Unit in Square Meters

? How large is the total living space in this flat? // sqm (from: *soep-is/soep-is-2020/Q11:hqm*  
[])

? What is the total area of this apartment? // sqm (from: *soep-is/soep-is-2020-fgz/Q019:hqm*  
[])

8	3
9	2
10	2
11	1
12	15
13	4
14	8
15	14
16	13

17	3
18	10
19	1
20	67
21	11
22	24
... (208 rows omitted)	32801
320	3
340	4
350	12
360	2
400	15
540	8
600	2
650	1
754	1
-1 [-1] No Answer	528
-2 [-2] Does Not Apply	5607
-3 [-3] Answer Improbable	4
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	3421
-6 [-6] Questionnaire Version With Modified Filtering	0

In the following cases the information on HGSIZE is still missing:

- The information could not be carried forward (e.g. due to the household moving)
- A new household did not provide information
- The size of the housing unit changed, but no new information was provided
- The given value was found to be implausible.

Missing cases in 2009 and 2010 were imputed using a regression model with the following independent variables:

- Number of rooms larger than 6 square meters (see variable HGROOM)
- Size of the household (HHGR)
- Status of ownership (HGOWNER)
- Year house was constructed (HGCNSTYR)
- Type of building (WUM1)
- Adequacy of living space in housing unit (HGSEVAL)
- SOEP Subsample (HSAMPLE)
- Dummy variable coded “1” for households living in Eastern Germany

The results of these imputations are also carried forward for up to two waves unless new information is provided in later waves.

In case HGFSIZE is imputed, the flag variable HGFSIZE takes the value “1”.

The number of missing cases from 2011 on was reduced by carrying on the information from previous years, unless there was a relocation of the household. Note the possibility of a change in the size of the housing unit even when no relocation happened. We estimate that there was a (undetected) change in size of the housing unit for around 3% of the households that did not relocate.

#### hgroom – Number Of Rooms Larger than 6 Square Meters

---

1	256
2	1221
3	2473
4	2000
5	1277
6	828
7	311
8	180
9	70
10	23
11	7
12	7
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	33934
-6 [-6] Questionnaire Version With Modified Filtering	0

The procedure of carrying forward and imputation is identical to that for HGFSIZE. In the regression model the same covariates are used with the exception of HGROOM which is replaced with HGFSIZE (size of the housing unit in square meters). In case information on HGFSIZE is missing, a first imputation was done excluding HGFSIZE. In case information for HGFSIZE exists, these values were replaced with values obtained from a second imputation including HGFSIZE as a covariate.

In case HGROOM is imputed, the flag variable HGFROOM takes the value “1”.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to “-5” (not contained in questionnaire).

#### hgheat – Costs Of Warm Water (gen)

---

0	1
1	1
5	1
7	1
8	3
9	3

10	7
11	3
12	1
13	1
14	4
15	6
16	5
17	2
18	11
... (198 rows omitted)	4033
270	1
296	1
300	4
350	1
358	1
360	1
375	1
475	1
493	1
-1 [-1] No Answer	5
-2 [-2] Does Not Apply	4553
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	33934
-6 [-6] Questionnaire Version With Modified Filtering	0

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgutil – Cost Allocation (gen)

? What are your other average monthly costs for water, garbage removal, etc. in other words, all additional costs? // euros per month (*from: soep-is/soep-is-2020/Q21:hm10[]*)

? What are your other average monthly costs for water, garbage removal, etc. in other words, all additional costs? (*from: soep-is/soep-is-2020/Q21:hm11[]*)

1	1
2	4
3	2
4	2
5	15
6	14
7	12
8	16
9	6
10	122
11	9
12	36
13	14
14	12

15	136
... (219 rows omitted)	12556
352	1
400	6
414	1
422	1
450	2
460	1
480	2
500	1
1213	1
-1 [-1] No Answer	3619
-2 [-2] Does Not Apply	22312
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	3683
-6 [-6] Questionnaire Version With Modified Filtering	0

HGRENT, HGUTIL and HGHEAT are generated variables on the housing cost of households in rental properties. HGRENT is a measure of the gross rent including utility costs (HGUTIL) and excluding heating costs (HGHEAT).

For further information on HGRENT, HGUTIL and HGHEAT please see the corresponding SOEP HGEN documentation.

#### hgent – Amount of Rent Minus Heating Costs (EUR)

? What is your current monthly rent? // Euro (*from: soep-is/soep-is-2020/Q13:hm04[]*)

? What is your current monthly rent? // Rent excluding heating costs: \_\_\_\_\_ Euro  
(*from: soep-is/soep-is-2020-fgz/Q023:hkmiet[]*)

8	1
15	1
20	1
45	1
46	1
47	1
48	1
50	4
57	1
70	3
71	1
73	1
74	1
75	2
76	1
... (1028 rows omitted)	18791
2350	1
2400	2
2500	7
2573	1
2700	1

2902	1
3000	1
3100	1
3300	1
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	20334
-3 [-3] Answer Improbable	4
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	3421
-6 [-6] Questionnaire Version With Modified Filtering	0

Because detailed information on heating costs is not assessed in the Questionnaire of the SOEP Innovation Sample, in the years 2011 and 2012 HGRENT was generated by simply duplicating respondents' answers on the amount of monthly rent from the H dataset.

#### hgelectr – Costs of Electricity (gen)

---

3	1
6	1
12	2
15	3
16	3
17	1
18	6
19	3
20	20
21	2
22	7
23	1
24	2
25	27
26	7
... (113 rows omitted)	850
177	1
180	1
183	1
185	1
191	1
196	1
200	1
218	1
279	1
-1 [-1] No Answer	4
-2 [-2] Does Not Apply	1143
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	40495
-6 [-6] Questionnaire Version With Modified Filtering	0

Information was collected in wave BA (2010) for the first time. The procedure of imputation is identical to that for HGHEAT. In the regression model the same covariates are used.

In case HGELECTR is imputed, the flag variable HGFELECTR takes the value "1". In the year 2011, no imputation of HGELECTR was conducted.

This information is not assessed in the Questionnaire of the SOEP Innovation Sample from 2012 on, therefore this variable is not generated for these years and coded to "-5" (not contained in questionnaire).

### hgnorent – Does Not Pay Rent

? What is your current monthly rent? (*from: soep-is/soep-is-2020-fgz/Q023:hm05[]*)

1	[1] Pays No Rent	464
2	[2] Pay No Rent But Utilities	378
3	[3] Pays Neither Rent Nor Utilities	162
-1	[-1] No Answer	108
-2	[-2] Does Not Apply	38054
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	3421
-6	[-6] Questionnaire Version With Modified Filtering	0

Coded as "1" if the household does not pay rent; e.g., if living space is provided by relatives at no cost. Note that in these cases, the information on gross cold rent (HGRENT), heating costs (HGHEAT) and utilities (HGUTIL) is coded "-2" ("not applicable").

Information is not carried forward.

### hgfsiz – Imputation Flag, Size Of Housing

0	[0] Not imputed	8609
1	[1] Imputed	44
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains "-5" (not contained in questionnaire).

### hgffroom – Imputation Flag, Number Of Rooms Larger than 6 Square Meters

0	[0] Not imputed	8632
1	[1] Imputed	21
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0

-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains "-5" (not contained in questionnaire).

#### hgfheat – Imputation Flag, Costs Of Heating And Warm Water

---

0	[0] Not imputed	7464
1	[1] Imputed	1189
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains "-5" (not contained in questionnaire).

#### hgfutil – Imputation Flag, Cost Allocation

---

0	[0] Not imputed	5648
1	[1] Imputed	3005
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

Please note that HGFUTIL is also given the value "1" for households reporting that utility costs are only partly included in the amount of rent stated (see the description of HGUTIL). This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains "-5" (not contained in questionnaire).

#### hgfrent – Imputation Flag, Amount Of Rent Minus Heating Costs

---

0	[0] not imputed	6397
1	[1] Imputed	87
2	[2] imputed cost allocation	1207
3	[3] Imputed Costs Of Heating And Warm Water	962
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0



-5	[-5] Not Contained In Questionnaire	33934
-6	[-6] Questionnaire Version With Modified Filtering	0

HGFGRENT indicates that at least one of the components of HGFGRENT (HGFGUTIL or HGFGHEAT) or HGFGRENT itself is imputed. It is only given positive values if the respective imputed component is part of HGFGRENT.

“1”: The household did not provide information on the amount of HGFGRENT, thus the basic value of HGFGRENT is imputed.

“2”: Utility costs are imputed and had to be added to the amount of rent stated to obtain the target measure of gross HGFGRENT.

“3”: Heating costs are imputed and had to be deducted from the amount of rent stated to obtain the target measure of gross HGFGRENT excluding heating costs.

This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains “-5” (not contained in questionnaire).

#### hgfelectr – Imputation Flag, Costs of Electricity

0	[0] not imputed	1971
1	[1] Imputed	121
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	40495
-6	[-6] Questionnaire Version With Modified Filtering	0

This variable contains information, whether the referring variable has been imputed. If the referring information has not been assessed in the questionnaire, this flag also contains “-5” (not contained in questionnaire).

#### hgtyp1hh – Household Typology, One Digit

1	[1] 1-Pers.-HH	13254
2	[2] Couple without children	13778
3	[3] Single Parent	3587
4	[4] Couple With Children LE 16	6558
5	[5] Couple With Children GT 16	2837
6	[6] Couple With Children LE and GT 16	1698
7	[7] Multiple Generation HH	253
8	[8] Other combination	609
-1	[-1] No Answer	13
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	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.

From 2009 onwards, the data production process switched to a standardized procedure for all waves to ensure longitudinal consistency, resulting in minor changes compared with older distributions.

HGTYP1HH is an aggregation of HGTYP2HH (first column of the two-digit code). Single households are differentiated in HGTYP2HH according to both gender and age.

Due to a slightly modified \$STELL variable in the file PBRUTTO, the generation syntax of HGTYP1HH and HGTYP2HH needed to be updated in the year 2011.

Legend:

- LE = less or equal
- GT = greater than
- K = child

### hgtyp2hh – Household Typology, Two Digit

---

11	[11] 1-Person HH - Male LE 35 Y.	1171
12	[12] 1-Person HH - Male 35- LT 60 Y.	2359
13	[13] 1-Person HH - Male GE 60 Y.	2261
14	[14] 1-Person HH - Female LE 35 Y.	1010
15	[15] 1-Person HH - Female 35- LT 60 Y.	1786
16	[16] 1-Person HH - Female GE 60 Y.	4667
21	[21] Couple without children	13778
31	[31] Single Parent, 1 child LE 16	1187
32	[32] Single Parent, 2 or more children LE 16	833
33	[33] Single Parent, 1 child GT 16	884
34	[34] Single Parent, 2 or more children GT 16	218
35	[35] Single Parent, 2 children LE and GT 16	297
36	[36] Single Parent, 3 children or more LE and GT 16	168
41	[41] Couple, 1 Child, LE 16	2397
42	[42] Couple, 2 Children, LE 16	2973
...	(2 rows omitted)	3104
52	[52] Couple, 2 Children, GT 16	793
53	[53] Couple, 3 Or More Children, GT 16	128
61	[61] Couple, 2 Children, LE and GT 16	925
62	[62] Couple, 3 Or More Children, LE and GT 16	773
71	[71] 3-generation hh	218
72	[72] 4-generation hh	3
73	[73] GrandParents-GrandChildren-HH	32
81	[81] Other combination without K. LE 16	550
82	[82] Other combination with K. LE 16	59
-1	[-1] No Answer	13
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

### hghmonth – Month Of Interview

---

1	[1] January	2891
2	[2] February	3157
3	[3] March	1835
4	[4] April	1063
5	[5] May	695
6	[6] June	410
7	[7] July	775
8	[8] August	464
9	[9] September	7930
10	[10] October	13690
11	[11] November	6999
12	[12] December	2678
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

The month of participation in the survey is generated using data from the household questionnaire.

#### hghmode – Interview Method

---

100	[100] With Interviewer Assistance	15
110	[110] Oral Interview	1586
120	[120] Written Ques. Interviewer	949
130	[130] Mix Between With/Without Interviewer	0
131	[131] Written Ques. No Interviewer	88
132	[132] Oral And Written	85
133	[133] Proxy	0
134	[134] With Interpreter	0
135	[135] Without Interpreter	0
140	[140] CAPI - Wave O Onwards	38386
200	[200] Telephone Assistance	0
210	[210] Written, By Mail	117
220	[220] Phone interview	814
300	[300] Web-based Interview	547
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

The interview method is generated through data from the household questionnaire.

#### hghinc – Monthly Household Net Income (EUR)

---

 (from: soep-is/soep-bip-2020-e/Q6:hnetto[])

? If you look at the total income of all of the members of your household: what is your monthly household income today? // EUR per month (*from: soep-is/soep-is-2020/Q48:hnetto*)

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Euro per month (*from: soep-is/soep-is-2020-fgz/Q063:hnetto*)

0	3
1	1
100	1
102	1
129	1
130	1
139	1
145	1
150	2
154	1
160	2
171	1
176	1
184	1
200	8
... (2111 rows omitted)	39808
16483	1
17000	3
18000	2
20000	14
21000	1
21250	1
22000	2
25000	1
30000	2
-1 [-1] No Answer	2575
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	151
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	0
-6 [-6] Questionnaire Version With Modified Filtering	0

This variable contains the current monthly net household income asked for in the household questionnaire, always provided in euros. Income is reported by the respondent (head of household).

#### hgahinc – Adjusted Monthly Household Net Income (EUR)

0	3
1	1
138	1
154	2
160	1

205	1
225	1
250	1
261	1
270	1
279	1
281	1
286	1
290	1
300	8
... (1889 rows omitted)	6865
13100	1
15000	5
16211	1
17000	2
20000	1
21000	1
21024	1
28164	1
30000	1
-1 [-1] No Answer	571
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	3
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	35109
-6 [-6] Questionnaire Version With Modified Filtering	0

This variable is based on the current monthly net household income asked for directly in the household questionnaire (“screener”). Since everyone in SOEP over the age of 16 is also interviewed personally, this income can be calculated based on the current individual monthly incomes of all household members. Possible underestimation in “screener” can thus be assessed and corrected. However, in the case of item-nonresponse on the original screener, the procedure is only used for households surveyed completely, without item-non-response on the variables in question.

For personal income, we use monthly net income (from dependent employment and self-employment), extra earnings, pensions, widow’s pensions, unemployment benefits or relief, maintenance payments, early retirement payments, maternity benefits, BaFoeG (state higher education grants), military or civil service pay, compulsory child support, as well as other forms of support from the P files. Civil servants’ pensions income is taxed at the flat rate of 20% and multiple entries on the use of employment office services are corrected for by calculating a median value. We add all the individual incomes of all interviewed household members, also adding to this sum all income from the household context (housing subsidies, child benefits, welfare and home nursing subsidies, social assistance; since 2005 also Unemployment Benefit II (“ALG II”) or Social Benefit).

When no answer was provided on net household income, we use the net household income calculated as described above, under the condition that all household members gave valid answers. If the net household income generated in this way is higher than the household income stated in the questionnaire, we correct the value upwards. When no answer was given for the different components of income, we set the value of the particular component at zero. HGAHINC was not generated from 2011 on.

**hgnuts1** – NUTS-Systematic-1 (Federal State)

1	[1] [DE100] Baden-Württemberg	4231
2	[2] [DE200] Bavaria	6231
3	[3] [DE300] Berlin	1642
4	[4] [DE400] Brandenburg	1237
5	[5] [DE500] Bremen	379
6	[6] [DE600] Hamburg	720
7	[7] [DE700] Hesse	2684
8	[8] [DE800] Mecklenburg-Western Pomerania	1047
9	[9] [DE900] Lower Saxony	4545
10	[10] [DEA00] North Rhine-Westphalia	11465
11	[11] [DEB00] Rhineland-Palatinate	1663
12	[12] [DEC00] Saarland	445
13	[13] [DED00] Saxony	2283
14	[14] [DEE00] Saxony-Anhalt	1359
15	[15] [DEF00] Schleswig-Holstein	1394
16	[16] [DEG00] Thuringia	1262
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	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.

### 3 Imputation of Household Income

#### Procedure

Multiple imputation procedures provide a way to deal with missing values on the variable HGHINC current monthly net household income in Euros by using information about determinants of the household income and replacing item-nonresponse with multiply imputed data. Five imputations are available within the \$HGEN datasets: the variables hgi1hinc-hgi5hinc. The imputations were calculated using the method of chained equations predictive mean matching in STATA. The procedures were written by Patrick Royston (see Royston 2004, 2005a, 2005b, 2007, 2009) and Ian White (see White, Daniel and Royston 2010; White, Royston and Wood 2011). Predicted mean matching means that for each missing observation on income, the particular non-missing observation is found whose prediction on observed data is closest. This closest observation is used to impute the missing value. The most important variable for modelling the current monthly net household income is monthly net household income of the previous year. A complete list of the variables used for modelling is available upon request.

The missing observations were assumed to be missing at random. We set the number of imputations  $m=5$  and get 5 multiple imputed values for HGHINC. The number of iterations carried out in each prediction model was specified to be 2000. Samples E&I&S1 and the supplementary sample S2 were imputed separately in 2014.

#### Analysing multiple imputed data

For analysing multiple imputed data, you do not necessarily need special methods, however such tools exist and simplify the use of multiply imputed data. Below a short overview of some useful tools for various statistical packages is given. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by combining the average of the squared standard errors of the several ( $m$ ) estimates with the within-and between-imputation variance. STATA provides a builtin functionality called `mi`. Within SAS, PROC MIANALYZE combines the results of analyses on the data sets. IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module `regress` to perform multiple imputation analysis.

Royston, P. 2004. Multiple imputation of missing values. *Stata Journal* 4: 227–241.

Royston, P. 2005a. Multiple imputation of missing values: Update. *Stata Journal* 5: 188–201.

Royston, P. 2005b. Multiple imputation of missing values: Update of ice. *Stata Journal* 5: 527–536.

Royston, P. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. *Stata Journal* 7: 445–464.

Royston, P. 2009. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *Stata Journal* 9: 466–477.

White, I. R., R. Daniel, and P. Royston. 2010. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical data. *Computational Statistics & Data Analysis* 54: 2267–2275.

White, I. R., P. Royston, and A. M. Wood. 2011. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 30: 377–399.

#### hgi1hinc – 1. Imputed Monthly Net Household Income (EUR) [1/15]

? If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (from: *soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	3
1	1
100	2
102	1
129	1
130	1
139	1
145	1
150	3
154	1
160	2
171	1
176	1
184	1

200	12
... (2112 rows omitted)	41487
16483	1
17000	3
18000	2
20000	15
21000	1
21250	1
22000	2
25000	1
30000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	1040
-6 [-6] Questionnaire Version With Modified Filtering	0

### hgi2hinc – 2. Imputed Monthly Net Household Income (EUR) [2/15]

**?** If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (*from: soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	4
1	1
100	2
102	1
129	1
130	1
139	1
145	1
150	2
154	1
160	2
171	1
176	1
184	2
200	8
... (2121 rows omitted)	41491
16483	1
17000	3
18000	2
20000	14
21000	1
21250	1
22000	2
25000	1
30000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0



-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	1040
-6	[-6] Questionnaire Version With Modified Filtering	0

### hgi3hinc – 3. Imputed Monthly Net Household Income (EUR) [3/15]

**?** If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (*from: soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	3	
1	1	
100	1	
102	1	
129	1	
130	1	
139	1	
145	1	
150	2	
154	1	
160	2	
171	1	
176	1	
184	1	
200	8	
... (2123 rows omitted)	41494	
16483	1	
17000	3	
18000	2	
20000	14	
21000	1	
21250	1	
22000	2	
25000	1	
30000	2	
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	1040
-6	[-6] Questionnaire Version With Modified Filtering	0

### hgi4hinc – 4. Imputed Monthly Net Household Income (EUR) [4/15]

**?** If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (*from: soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	4
1	1

100	1
102	1
129	1
130	1
139	1
145	1
150	2
154	1
160	3
171	1
176	1
184	1
200	9
... (2126 rows omitted)	41490
16483	1
17000	3
18000	2
20000	14
21000	1
21250	2
22000	2
25000	1
30000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	1040
-6 [-6] Questionnaire Version With Modified Filtering	0

#### hgi5hinc – 5. Imputed Monthly Net Household Income (EUR) [5/15]

**?** If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (*from: soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	3
1	1
96	1
100	1
102	1
129	1
130	1
139	1
145	1
150	3
154	1
160	2
171	1
176	1
184	1

...	(2114 rows omitted)	41500
16483		1
17000		3
18000		2
20000		14
21000		1
21250		1
22000		2
25000		1
30000		2
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	1040
-6	[-6] Questionnaire Version With Modified Filtering	0

#### hgi6hinc – 6. Imputed Monthly Net Household Income (EUR) [6/15]

---

130		1
139		1
150		1
154		1
250		1
260		1
300		4
350		1
370		1
395		1
397		1
400		3
420		3
432		3
442		1
...	(425 rows omitted)	3677
10000		11
10983		1
11000		2
12000		3
12500		1
14000		2
15000		5
16000		1
20000		2
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	38858
-6	[-6] Questionnaire Version With Modified Filtering	0

**hgi7hinc** – 7. Imputed Monthly Net Household Income (EUR) [7/15]

---

130	1
139	1
150	1
154	1
250	1
260	1
300	4
350	1
370	1
395	1
397	1
400	3
420	2
432	3
442	1
... (425 rows omitted)	3678
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgi8hinc** – 8. Imputed Monthly Net Household Income (EUR) [8/15]

---

130	1
139	1
150	1
154	1
250	1
260	1
300	5
350	2
370	1
395	1
397	1
400	3
420	2
432	3

442	1
... (425 rows omitted)	3675
10000	12
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

#### hgi9hinc – 9. Imputed Monthly Net Household Income (EUR) [9/15]

---

130	1
139	1
150	1
154	1
250	1
260	1
300	6
350	1
370	1
395	1
397	1
400	4
420	2
432	3
442	1
... (425 rows omitted)	3675
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgi10hinc** – 10. Imputed Monthly Net Household Income (EUR) [10/15]

---

130	1
139	1
150	1
154	1
250	2
260	2
300	6
350	1
370	1
395	1
397	1
400	4
420	2
432	3
442	1
... (425 rows omitted)	3672
10000	11
10983	1
11000	2
12000	4
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgi11hinc** – 11. Imputed Monthly Net Household Income (EUR) [11/15]

---

130	1
139	1
150	2
154	1
250	1
260	1
300	5
350	1
370	1
395	1
397	1
400	3
420	2
432	3

442	1
... (425 rows omitted)	3676
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

### hgi12hinc – 12. Imputed Monthly Net Household Income (EUR) [12/15]

---

130	1
139	1
150	2
154	1
250	1
260	1
300	4
350	1
370	1
395	1
397	1
400	3
420	2
432	3
442	1
... (425 rows omitted)	3677
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgi13hinc** – 13. Imputed Monthly Net Household Income (EUR) [13/15]

---

130	1
139	1
150	1
154	1
250	1
260	1
300	5
350	1
370	1
395	1
397	1
400	3
420	2
432	4
442	1
... (425 rows omitted)	3676
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgi14hinc** – 14. Imputed Monthly Net Household Income (EUR) [14/15]

---

130	1
139	1
150	1
154	1
250	1
260	1
300	4
350	1
370	1
395	1
397	1
400	3
420	2
432	3



442	1
... (425 rows omitted)	3678
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

#### hgi15hinc – 15. Imputed Monthly Net Household Income (EUR) [15/15]

---

130	1
139	1
150	1
154	1
250	2
260	1
300	6
350	1
370	1
395	1
397	1
400	3
420	2
432	4
442	1
... (425 rows omitted)	3674
10000	11
10983	1
11000	2
12000	3
12500	1
14000	2
15000	5
16000	1
20000	2
-1 [-1] No Answer	0
-2 [-2] Does Not Apply	0
-3 [-3] Answer Improbable	0
-4 [-4] Inadmissible Multiple Answer	0
-5 [-5] Not Contained In Questionnaire	38858
-6 [-6] Questionnaire Version With Modified Filtering	0

**hgfhinc** – Imputation Flag, Monthly Net Household Income (EUR)

**?** If you take a look at the total income of all members of the household: How much is the monthly household net income today? // EUR per month (*from: soep-is/soep-is-2019/Q062:hnetto[28084]*)

0	[0] Not imputed	39861
1	[1] Imputed	2726
-1	[-1] No Answer	0
-2	[-2] Does Not Apply	0
-3	[-3] Answer Improbable	0
-4	[-4] Inadmissible Multiple Answer	0
-5	[-5] Not Contained In Questionnaire	0
-6	[-6] Questionnaire Version With Modified Filtering	0

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