SOEP

The German
Socio-Economic
Panel Study

113

SOEP Survey Papers

Series D - Variable Description and Coding

SOEP 2011 – Documentation of Household-related Status and Generated Variables in HGEN for SOEP v28

SOEP Group

DIW SOEP

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

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

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

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

Please cite this paper as follows:

SOEP Group. 2012. SOEP 2011 – Documentation of Household-related Status and Generated Variables in HGEN for SOEP v28. SOEP Survey Papers 113: Series D. Berlin: DIW/SOEP

SOEP GROUP

SOEP 2011 – DOCUMENTATION OF HOUSEHOLD-RELATED STATUS AND GENERATED VARIABLES IN HGEN FOR SOEP V28

Berlin, 2012

General information:

Up to Wave G (1990), "old" households already known to the SOEP were surveyed at the old address with a "green" questionnaire; old households that had moved received a "blue" questionnaire. While a number of questions in the blue questionnaire surveyed information for the first time (e.g., living space in square meters), this information was only asked for again in the green questionnaire in the case of changes. Otherwise, the information collected in the previous year was still valid.

The variables described in the following are in part status variables in this sense: information collected once is carried forward to subsequent years if no address change has taken place since the previous year. This is the case for: CNSTYR$$, CONDIT$$, SIZE$$, ROOM$$, EQPKIT$$, EQPSHW$$, EQPIWC$$, EQPHEA$$, EQPTER$$, EQPBAS$$, EQPGAR$$, EQPWAT$$, EQPTEL$$, EQPALM$$, EQPSOL$$, EQPAIR$$, EQPNRJ$$, EQPLIF$$, MOVEYR$$, ACQUIS$$, SUBSID$$, RSUBS$$ and REDUC$$.

Furthermore, identical information is recorded in the green and blue questionnaire in separate variables (e.g. housing tenure as owner or renter). The corresponding status variables are therefore just a compilation of these more dispersed pieces of information. Since only one common questionnaire has been used since Wave H (1991) rather than the former "blue" and "green" versions, the necessity for the aforementioned status variables disappears but this "user-friendly redundance" is maintained for reasons of consistency.

List of Variables:

CNSTYR$$ (formerly $BAUJ)

```
Var Label   :   CNSTYR$$  "Year house was constructed"
Value Label :   CNSTYR$$  (1)'before 1919'
                          (2)'1919 to 1948'
                          (3)'1949 to 1971'
                          (4)'1972 or later'
        from Wave G (East) or Wave H (West) on
                          (4)'1972 to 1980'
                          (5)'1981 or later'
        from Wave M on  (5)'1981 to 1990'
                          (6)'1991 or later'
        from Wave R on  (6)'1991 to 2000'
                          (7)'2001 or later'

Variable format :       CNSTYR$$ (D010)
$$ - Survey Year :      $$=84..11


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


For more information, contact:     Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>
```

CONDIT$$ (formerly $RENOV)

```
Var Label    :   CONDIT$$ "Condition of building"
Value Label  :   CONDIT$$ (1)"no renovations needed"
                          (2)"partial renovation needed"
                          (3)"major renovation needed"

         from Wave G (East) or Wave H (West) on
                          (1)"in good condition"
                          (2)"partial renovation needed"
                          (3)"major renovation needed"
                          (4)"ready for demolition"
Variable format :        CONDIT$$ (D010)
$$ - Survey Year :       $$=84..11
```

```
Comment:      Respondent's estimation of the condition of the building.
              In the West German sub-samples for Waves B (1985) to Wave G
              (1990), information on CONDIT$$ was collected only for new
              households and for households with a residential move since
              the previous wave (households with "blue" questionnaires).
              For immobile households ("green" questionnaire) information
              collected in previous waves was carried forward.
              The wording in the questionnaire was changed in the first
              wave of the East German sub-sample in 1990 (Wave G) as to
              better capture the extremely rundown condition of some
              residential buildings in East Germany.
              Since Wave H (1991) the wording is identical for the entire
              SOEP-sample in East and West Germany.
```

```
For more information, contact:      Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>
```

SIZE$$ (formerly $WOHNFL)

```
Var Label    :  SIZE$$    "Size of housing unit in square meters"
Variable format :       SIZE$$  (D010)
$$ – Survey Year :      $$=84..11
```

Comment: Up to Wave N (1997) this information is collected only in the first interview with new households, in case a household has moved or with old households which still reside at their old address but whose housing unit size has changed due to renovations or additions (up to Wave G (1990), these households filled out a "green" questionnaire). For old households still residing at their old address the information on the size of the housing unit is carried forward as a status variable in order to provide valid current information. From Wave O (1998) on with the exception of Waves P (1999) and R (2001), the question on the housing unit size was posed to all households.

Up to Wave N (1997), therefore, for old households still residing at their old address the information on SIZE$$ is carried forward from the previous wave or the wave before.

From Wave O (1998) on, this is only done for missing values.

In the following cases the information on SIZE$$ 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.

In these cases, the information on the size of the housing unit was imputed using a regression model with the following independent variables:

- Number of rooms larger than 6 square meters (see variable ROOM$$)
- Size of the household ($HHGR)
- Status of ownership (OWNER$$)
- Year house was constructed (CNSTYR$$)
- Type of building ($WUM1)
- Adequacy of living space in housing unit (SEVAL$$)
- SOEP Subsample (HSAMPLE)
- From Wave G (1990) on: 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 SIZE$$ is imputed, the flag variable FSIZE$$ takes the value "1".

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

FSIZE$$

```
Var Label    :   FSIZE$$   "Size of housing unit imputed"
Variable format :         FSIZE$$   (D010)
Value Label  :   FSIZE$$   (1)"yes"
$$ – Survey Year :        $$=84..11

Comment:         See description of SIZE$$.
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

ROOM$$ (formerly $WOHNR)

```
Var Label    : ROOM$$   "Number of rooms larger than 6 square meters"
Variable format :        ROOM$$   (D010)
$$ - Survey Year :       $$=84..11
```

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

              In the year 1998 (Wave O), the question on the number of
              rooms was posed to all households in order to correct
              mistakes that may have occurred in the carrying forward of
              data or in the process of imputation.

              In case ROOM$$ is imputed, the flag variable FROOM$$ takes
              the value "1".
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

FROOM$$

```
Var Label    : FROOM$$  "Number of rooms imputed"
Variable format :        FROOM$$  (D010)
Value Label  : FROOM$$  (1)"yes"
$$ - Survey Year :       $$=84..11
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

SEVAL$$ (formerly $WGURT)

```
Var Label    :    SEVAL$$    "Adequacy of living space in the housing unit"
Value Label  :    SEVAL$$    (1)"much too small"
                             (2)"a bit too small"
                             (3)"just right"
                             (4)"a bit too large"
                             (5)"much too large"
Variable format :          SEVAL$$   (D010)
$$ - Survey Year :         $$=84..11
```

Comment: Estimate by the respondent (household head). In Waves C
               (1986) to Wave G (1990), information on SEVAL$$ was collected
               for new households or households that have moved (households
               with "blue" questionnaires) and immobile households whose
               sizing unit has changed. In these waves, SEVAL$$ is carried
               forward.

Detailed description:
For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

EQPKIT$$ (formerly $AUS1)

```
Var Label    :   EQPKIT$$  "Dwelling has kitchen"
Value Label  :   EQPKIT$$  (1)"yes"
                           (2)"no"
Variable format :        EQPKIT$$       (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPSHW$$ (formerly $AUS2)

```
Var Label    :   EQPSHW$$  "Dwelling has indoor bath / shower"
Value Label  :   EQPSHW$$  (1)"yes"
                           (2)"no"
Variable format :        EQPSHW$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPIWC$$ (formerly $AUS3)

```
Var Label    :   EQPIWC$$  "Dwelling has indoor toilet"
Value Label  :   EQPIWC$$  (1)"yes"
                           (2)"no"
Variable format :        EQPIWC$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPHEA$$ (formerly $AUS4)

```
Var Label    :   EQPHEA$$  "Dwelling has central / floor heat"
Value Label  :   EQPHEA$$  (1)"yes"
                           (2)"no"
Variable format :        EQPHEA$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPTER$$ (formerly $AUS5)

```
Var Label    :   EQPTER$$  "Dwelling has balcony / terrace"
Value Label  :   EQPTER$$  (1)"yes"
                           (2)"no"
Variable format :        EQPTER$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPBAS$$ (formerly $AUS6)

```
Var Label    :   EQPBAS$$  "Dwelling has basement"
Value Label  :   EQPBAS$$  (1)"yes"
                           (2)"no"
Variable format :        EQPBAS$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPGAR$$ (formerly $AUS7)

```
Var Label    :   EQPGAR$$  "Dwelling has garden"
Value Label  :   EQPGAR$$  (1)"yes"
                           (2)"no"
Variable format :        EQPGAR$$ (D010) / since G:(I5)
$$ - Survey Year :       $$=84..11
```

EQPWAT$$ (formerly $AUS8)

```
Var Label     :   EQPWAT$$  "Dwelling has hot water / boiler"
Value Label   :   EQPWAT$$  (1)"yes"
                            (2)"no"
Variable format :           EQPWAT$$  (I2)
$$ - Survey Year :          90=G (East Germany only), 91..11
```

EQPTEL$$ (formerly $AUS9)

```
Var Label     :   EQPTEL$$  "Dwelling has telephone"
Value Label   :   EQPTEL$$  (1)"yes"
                            (2)"no"
Variable format :           EQPTEL$$  (I2)
$$ - Survey Year :          90=G (East Germany only), 91..11
```

EQPALM$$

```
Var Label     :   EQPALM$$  "Dwelling has alarm device"
Value Label   :   EQPALM$$  (1)"yes"
                            (2)"no"
Variable format :           EQPALM$$  (I2)
$$ - Survey Year :          04..11
```

EQPSOL$$

```
Var Label     :   EQPSOL$$  "Dwelling has solar collector"
Value Label   :   EQPSOL$$  (1)"yes"
                            (2)"no"
Variable format :           EQPSOL$$  (I2)
$$ - Survey Year :          07..11
```

EQPAIR$$

```
Var Label     :   EQPAIR$$  "Dwelling has air conditioning"
Value Label   :   EQPAIR$$  (1)"yes"
                            (2)"no"
Variable format :           EQPAIR$$  (I2)
$$ - Survey Year :          07..11
```

EQPNRJ$$

```
Var Label     :   EQPNRJ$$  "Dwelling has other alternative energy source"
Value Label   :   EQPNRJ$$  (1)"yes"
                            (2)"no"
Variable format :           EQPNRJ$$  (I2)
$$ - Survey Year :          09..11
```

EQPLIF$$

```
Var Label     :   EQPLIF$$  "Dwelling has lift"
Value Label   :   EQPLIF$$  (1)"yes"
                            (2)"no"
Variable format :           EQPLIF$$  (I2)
$$ - Survey Year :          09..11
```

Comment (EQPTEL$$): Information was not collected in waves K (1994), N
(1997), P (1999), Z (2009) and BA (2010). In these waves
and in case of missing information, EQPTEL$$ was updated
using information from the address log $HBRUTTO ($HTEL).

Comment (EQPKIT$$-EQPTEL$$): In some waves, these variables are only
collected from new households and households who have moved
since the previous interview. For this reason, in case no
address change has taken place the information for EQPKIT$$-
EQPTEL$$ is carried forward from the previous years for up to
two waves. Additionally, from Wave B (1985) on, the
information on EQPKIT$$ to EQPHEA$$ can be updated to some
extent using the information on modernization projects (on
kitchens, bathrooms or modern heating systems) undertaken
since January of the previous year. Please note that this
update cannot be done for EQPTER$$ to EQPTEL$$.

Comment (EQPNRJ$$ and EQPLIF$$): Information was collected in wave Z (2009)
for the first time.

Comment (EQPIWC$$): Collection of this information was discontinued after
wave Y (2008). In waves Z (2009) and BA (2010) data is
carried forward from the previous year. Nevertheless,
information is missing for mobile households and those
responding for the first time (including all households in
Sample I which was started only in 2009).

Detailed description:
For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

MOVEYR$$ (formerly $EINZUG)

```
Var Label        :        MOVEYR$$ "Year moved into dwelling"
Variable format  :        MOVEYR$$ (D010)
$$ – Survey Year :        $$=84..11
```

```
COMMENT:         For old households at their old address, data is carried
                 forward for up to two waves. For new households in SOEP and
                 for old households that have moved, the variable is based on
                 newly collected data. In case the information is missing and
                 an old household has moved that year or the previous year,
                 MOVEYR$$ is given the value of the year of the respective
                 wave.

                 The carrying forward of data entails the possibility that the
                 year of moving into the new dwelling may lie before the year
                 of birth of the oldest household member.

                 Help for "(very) old friends": this was converted to four-
                 digit annual data starting with the 2000 data distribution.
```

```
Detailed description:
For more information, contact:        Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>
```

OWNER$$ (formerly $EIGEN)

```
Var Label    :   OWNER$$   "Tenant or owner of dwelling"
Value Label  :   OWNER$$   (1)"Owner"
                           (2)"Main tenant"
                           (3)"Subtenant"
                           (4)"Tenant"
                           (5)"Resident of a home or inst. living facility"
Variable format :         OWNER$$   (D010)
$$ – Survey Year :         $$=84..11
```

Comment: Up to Wave H (1991), the information given in OWNER$$ was
              collected in separate questionnaires for "old" and first-time
              respondents, respectively ("blue" and "green"
              questionnaires). In all waves, codes 1 and 4 are used if the
              original variable is coded as -1 ("no answer") but if at
              least one answer that is specific to owners, respectively to
              tenants, was given. Code 4 is also used if a change in
              ownership (from owner to tenant) has taken place, but no
              original information for OWNER$$ was given. Code 5 ('resident
              of a home or institutional living facility') has only been
              assigned by interviewers during fieldwork since Wave P
              (1999).

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

ACQUIS$$ (formerly $ERWERB)

Target Population: Owner-occupiers, only

```
Var Label    :   ACQUIS$$ "Means of acquiring dwelling"
Value Label  :   ACQUIS$$ (1) "bought from previous owner"
                          (2) "inheritance, gift"
                          (3) "bought it new"
                          (4) "returned to private ownership
                          "Rückübertragung")"  (only East Germany)
Variable format :        ACQUIS$$ (D010)
$$ – Survey Year :       $$=84..11
```

```
Comment:     Statement by respondent (household head). If no new
             information is provided and a change of address or ownership
             status (OWNER$$) has not taken place, the information of the
             previous year is carried forward. Code(4) was surveyed only
             in Wave I (1992) in East Germany, but is also carried
             forward.
```

```
For more information, contact:      Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>
```

SUBSID$$

Target Population: Owner-occupiers, only

```
Var Label    :   SUBSID$$ "Government-subsidized housing payments"
Value Label  :   SUBSID$$ (1)"yes"
                          (2)"no"
Variable format :       SUBSID$$ (D010)
$$ – Survey Year :      $$=84..99 (West); $$=94..99 (East)
```

Comment: Statement by respondent. SUBSID$$ contains information on
              government subsidies at the time the housing was built or
              bought. From Wave B (1985) to Wave N (1997), this has only
              been asked to new households or in case an old household has
              moved. Information is then carried forward. In Waves O (1998)
              and P (1999), the question was again posed to the whole
              population.

OSUBS$$

Target Population: Owner-occupiers, only

```
Var Label    :   OSUBS$$  "Received government housing subsidies last year"
Value Label  :   OSUBS$$  (1)"yes"
                          (2)"no"
Variable format :       OSUBS$$  (D010)
$$ – Survey Year :      $$=00..11
```

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

              Please note: The old variable $FOERD (available until SOEP
              data release 2008) most likely misrepresented the true
              percentage of households receiving subsidies. Homeowner
              subsidies in Germany have been subject to major revisions and
              fluctuations over time. The corresponding question in SOEP
              was in some years only posed to new households and those that
              have moved, in some years it was not surveyed at all. For
              these reasons, the question for government housing subsidies
              was changed in Wave Q (2000) to cover direct subsidies
              received the previous year. SUBSID$$ and OSUBS$$ replace the
              old variable $FOERD.

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

REVAL$$ (formerly $MURT)

```
Var Label    :   REVAL$$   "Evaluation of rent paid"
Value Label  :   REVAL$$   (1)"very inexpensive"
                           (2)"inexpensive"
                           (3)"reasonable"
                           (4)"slightly expensive"
                           (5)"too expensive"
Variable format :          REVAL$$   (D010)
$$ – Survey Year :         $$=84..02, 05..11
```

Comment: Subjective assessment by respondent (household head). This
                variable was not surveyed in Waves T and U (2003-04). 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.

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

RSUBS$$ (formerly $SOZIAL)

```
Var Label     :   RSUBS$$   "Government-subsidized rental housing"
Value Label   :   RSUBS$$   (1)"yes"
                            (3)"no"
Variable format  :          RSUBS$$   (D010)
$$ - Survey Year :          $$=84..94 (West); $$=90, 94 (East)


Value Label   :   RSUBS$$   (1)"yes, with subsidy"
                            (2)"yes, with expired subsidy"
                            (3)"no"
Variable format  :          RSUBS$$   (I5)
$$ - Survey Year :          $$=95..11
```

COMMENT: Up to Wave K (1994), information is carried forward from
               previous years for immobile households. The rewording of the
               response categories beginning with Wave L (1995) became
               necessary due to the carrying forward of data: It was
               impossible to identify whether a housing unit had lost its
               subsidization status for any period of time. Thus for
               population estimates, there is a distinct possibility that
               RSUBS$$ produces increasing overestimations of government-
               subsidized housing units up to Wave K (1994). For reasons of
               time series consistency, RSUBS$$ was coded with "3" for "no"
               in Waves A (1984) to K (1994). In Wave L (1995), the code "2"
               was introduced to indicate expired subsidization.

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

REDUC$$ (formerly $BILLIG)

```
Var Label    :   REDUC$$   "Rent-reduced dwelling"
Value Label  :   REDUC$$   (1)"yes"
                           (2)"no"
Variable format :          REDUC$$   (D010)
$ - Wave     :   $=86..02, 08..11 (West); $=90, 92..02, 08..11 (East)
```

COMMENT: 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. In Waves T (2003) to X (2007) this
              information was not collected. It is carried forward from
              Wave S for households who have not moved and whose stated
              amounts of rent vary only slightly. The new information from
              Wave Y is then carried backward for households with the same
              characteristics if REDUC$$ is still missing after carrying
              forward from Wave S.

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

RENT$$ (formerly $MIETEG)

Var Label: RENT$$ **"Amount of monthly rent incl. utilities, excl. heating
            costs, in Euro"**
Variable format : RENT$$ (I5)
$$ - Survey Year : $$=84..11

HEAT$$ (formerly $HEIZG)

Var Label: HEAT$$ **"Amount of monthly costs for heating and hot water in
Euro"**
Variable format : HEAT$$ (I4)
$$ - Survey Year : $$=84..11

UTIL$$

Var Label: UTIL$$ **"Amount of monthly utility costs in Euro"**
Variable format : UTIL$$ (I4)
$$ - Survey Year : $$=91..11

COMMENT: RENT$$, UTIL$$ and HEAT$$ are generated variables on the housing cost
            of households in rental properties. RENT$$ is a measure of the
            gross rent including utility costs (UTIL$$) and excluding
            heating costs (HEAT$$). RENT$$, UTIL$$ and HEAT$$ are converted
            into Euro values for all years, including those prior to 2002.

1.   Motivation

The housing costs of households in rental properties are in general
comprised of
    - net rent, i.e., rent minus all heating costs (basic rent)
    - utility costs (electricity, water, trash removal, etc. but excluding
      any associated heating costs)
    - heating costs (including costs for hot water).
The corresponding questions in the SOEP household questionnaire have
changed significantly over time, particularly from Wave H (1991) up to Wave
N (1997). For this reason it is desirable to generate a variable following
a concept of rent that is comparable over time, i.e. gross rent including
utility costs but excluding heating costs (RENT$$).

2.   Generating time consistent measures of gross rent, utility costs and
      heating costs

2.1 Overview

In Waves A (1984) to G (1990), the amount of rent stated by the households
in SOEP is in principle the desired concept of gross rent, i.e. basic rent
excluding heating costs but including utility costs. In these waves,
however, information on utility costs was not collected. From Wave H (1991)
on, households simply state the amount of rent they pay. Following this
question, it is asked whether heating and utility costs are included in
that amount of rent and what the exact costs for heating and utilities
eventually are (in the latter case only if they are included).

From Wave G (1990) on, it is therefore necessary to deduct heating costs
(if included in the amount of rent stated) or include utility costs (if not
included in the amount of rent stated) in order to obtain the target
measure of gross rent RENT$$. In case of missing information this is done
by an imputation process using cross-sectional regression models (see

below). Using the amount of rent stated and the components UTIL$$ and HEAT$$, the basic rent excluding heating costs but including utility costs is then calculated. Missing values of this basic rent (which in most cases are due to a missing value for the amount of rent stated) are again imputed using a regression model.

2.2 Imputation procedure for heating costs, utility costs and gross rent

1. Heating costs (HEAT$$)

Heating costs are surveyed starting with Wave C (1986), but values for Waves A and B (1984 – 1985) are imputed backwards using the imputation coefficients from Wave C (see below) and the covariates as given in Waves A and B, respectively. These values for Wave A and B are then deflated by 2% per year.

Missing values for heating costs are obtained using a regression model of the following specification:

Dependent variable: Heating costs per square meter of the size of the housing unit
Independent variables:
- EQPHEA$$ ("Dwelling has central, floor heating")
- SIZE$$ ("Size of the housing unit in square meters")
- $HHGR ("Number of persons in household")
- CONDIT$$ ("Condition of house")
- $GGK ("City, district size"); from Wave Q (2000) on: $BIK ("City, district type")
- $HWUM1 ("Type of building")
- HSAMPLE (SOEP Subsample)

This imputation is done for private households only. From Wave G (1990) on, this is done for East and West Germany separately (up until Wave P (1999). From Wave Q on, a joint regression for East and West Germany is done, including a dummy variable coded "1" for "East Germany".

Missing values for HEAT$$ are then replaced with their predicted estimates, adding an error term to avoid a "regression to the mean" effect and multiplied by the household-specific size of the housing unit (see variable SIZE$$). In case the imputed values are negative (in total 57 values in Waves A to X) or implausibly high (i.e. they belong to the highest percentile of the heating costs per square meter; 257 values in Waves A to X), the values are replaced by the median value of utility costs per square meter multiplied by the household-specific size of the housing unit.

2. Utility costs (UTIL$$)

Information on utility costs is surveyed starting with Wave H (1991) for East Germany and starting with Wave K (1994) in West Germany. In all years, it is first asked if utility costs are included in the rent. The amount of utility costs is only asked for if they are – fully or in part – included in the measure of rent.

The missing values for UTIL$$ for Waves H, I and J (1991 – 1993) in West Germany are imputed using the imputation coefficients from Wave K (1994) and the covariates from Waves H, I or J, respectively. These values for Waves H, I and J are then deflated by 2% per year.

The imputation process for missing values is similar to that for missing heating costs:

Dependent variable: Utility costs per square meter of the size of the housing unit
Independent variables:
- EQPHEA$$ ("Dwelling has central, floor heating")
- CONDIT$$ ("Condition of house")
- CNSTYR$$ ("Year house was built")
- $GGK ("City, district size"); from Wave Q (2000) on: $BIK ("City, district type")
- $HWUM1 ("Type of building")
- HSAMPLE (SOEP Subsample)

Just as for HEAT$$, this imputation is done for private households only. From Wave G (1990) on, this is done for East and West Germany separately (up until Wave P (1999). From Wave Q (2000) on, a joint regression model for East and West Germany is estimated, including a dummy variable coded "1" for "East Germany".

Missing values for UTIL$$ are then replaced with their estimates, adding an error term to avoid a "regression to the mean" effect and multiplied by the household-specific size of the housing unit (see variable SIZE$$). In case the imputed values are negative (in total 719 values in Waves H to X) or implausibly high (i.e. they belong to the highest percentile of the utility costs per square meter; 113 values in Waves H to X), the values are replaced by the median value of heating costs per square meter multiplied by the household-specific size of the housing unit.

In case a household states that utility costs are partly included in the amount of rent they pay, the full amount of utility costs is approximated by multiplying the given utility costs with a factor contrasting the mean share of utility costs in rent (when fully included) to the mean share of utility costs when partly included.

3. Gross rent (RENT$$)

The gross rent is obtained by adding utility costs (UTIL$$) to the amount of rent stated if not already fully included in that amount and by deducting the amount of heating costs (HEAT$$) if included in the amount of rent stated.

In case a household did not provide information on the inclusion / exclusion of utility and heating costs in the stated amount of rent, values for the respective filter questions are imputed using a probit regression model with the following independent variables:

- $GGK ("City, district size"); from Wave Q (2000) on: $BIK ("City, district type")
- $HWUM1 ("Type of building")
- Number of years the household has been living in that housing unit (Survey year minus MOVEYR$$)
- Dummy variable coded "1" for households living in Eastern Germany
- HSAMPLE (SOEP Subsample)

In case a household did not provide information the amount of rent (as well as in the few cases where a variable needed for the imputation of utility or heating costs is missing), the gross rent could not be calculated from the given or imputed values. RENT$$ is then imputed using a regression model of the following specification:

Dependent variable: Log of gross rent per square meter of the size of the housing unit

Independent variables:
- SIZE$$: Size of the housing unit in square meters
- Number of years the household has been living in that housing unit (Survey year minus MOVEYR$$)
- CNSTYR$$: Year house was built
- CONDIT$$: Condition of house
- $GGK: City, district size; from Wave Q (2000) on: $BIK: City, district type
- $HWUM1: Type of building
- EQPHEA$$: Dwelling has central, floor heating
- EQPTER$$: Dwelling has balcony, terrace
- EQPWAT$$: Dwelling has hot water, boiler
- HSAMPLE (SOEP Subsample)

From Wave G (1990) on, this regression is done for East and West Germany separately (up until Wave P (1999). From Wave Q (2000) on, a joint regression for East and West Germany is estimated, including a dummy variable coded "1" for "East Germany".

Just as for UTIL$$ and HEAT$$, missing values for RENT$$ are then replaced with the predicted estimates, adding an error term to avoid a "regression to the mean" effect and multiplied by the household-specific size of the housing unit (see variable SIZE$$).

Remarks

- This imputation is done for private households and for tenants living in non-subsidized housing units only (i.e. rent must not be subsidized by the government or reduced by the landlord). Missing values for institutionalized households (e.g. homes for the elderly) cannot be imputed due to the lack of representative data.

- In East Germany, due to the extraordinary low amounts of rent paid, the log of the values of rent entering the imputation process is not taken in Waves G (1990) and H (1991). Also, due to the high prevalence of government subsidies, the population for the imputation process is not restricted as it is done in the West German samples up until Wave K (1994). From Wave L (1995) on, the population in Eastern Germany was restricted to households whose rent is not subsidized by the government.

- In case a household has stated an amount of rent but utility costs had to be imputed and are included in that amount of rent, the imputed amount is capped to the mean share of utility costs in gross rent to avoid amounts of utility costs that are larger than the corresponding amount of rent.

- Similarly, in case a household has stated an amount of rent but heating costs had to be imputed and are stated to be included in that amount of rent, it is assumed that these heating costs are not included in case their imputed value is larger than the amount of rent stated.

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

ELECTR$$

```
Var Label    :   ELECTR$$ "Amount of monthly costs for electricity"
Variable format :        ELECTR$$ (D010)
$$ - Survey Year :        $$=10,11
```

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

              In case ELECTR$$ is imputed, the flag variable FELECTR$$
              takes the value "1".
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

FELECTR$$

```
Var Label    :   FELECTR$$ "Imputation flag for electricity"
Variable format :        FELECTR$$ (D010)
Value Label  :           FELECTR$$ (1)"yes"
$$ - Survey Year :        $$=10,11
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

```
                         FRENT$$
Var Label:                       FRENT$$ "Imputation flag for gross rent"
Variable format :                FRENT$$  (I5)
Value Label     :                FRENT$$  (1)"rent imputed"
                                          (2)"utility costs imputed and added"
                                          (3)"heating costs imputed and deducted"
$$ – Survey Year :               $$=84..11
```

FHEAT$$

```
Var Label:  FHEAT$$ "Imputation flag for heating and hot water costs"
Variable format : FHEAT$$(I5)
Value Label  :    FHEAT$$(1)"yes"
$$ – Survey Year: $$=84..11
```

FUTIL$$

```
Var Label:  FUTIL$$ "Imputation flag for utility costs"
Variable format : FUTIL$$(I5)
Value Label  :    FUTIL$$(1)"yes"
$$ – Survey Year : $$=91..11
```

```
COMMENT:     FRENT$$, FHEAT$$, FUTIL$$ and FELECTR$$ are flag variables to
             indicate whether RENT$$, HEAT$$, UTIL$$ and ELECTR$$ were
             imputed, respectively.

             FRENT$$ indicates that at least one of the components of RENT$$
             (UTIL$$ or HEAT$$) or RENT$$ itself is imputed. It is only
             given positive values if the respective imputed component is
             part of RENT$$:
```

- "1": The household did not provide information on the amount of RENT$$, thus the basic value of RENT$$ 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 RENT$$.

- "3": Heating costs are imputed and had to be deducted from the amount of rent stated to obtain the target measure of gross RENT$$ excluding heating costs.

```
             Please note that FUTIL$$ is also given the value "1" for
             households reporting that utility costs are only partly
             included in the amount of rent stated (see the description of
             UTIL$$).
```

```
For more information, contact:      Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>
```

Diagram 1

Diagram 1: Generating time consistent measures of housing costs in the SOEP
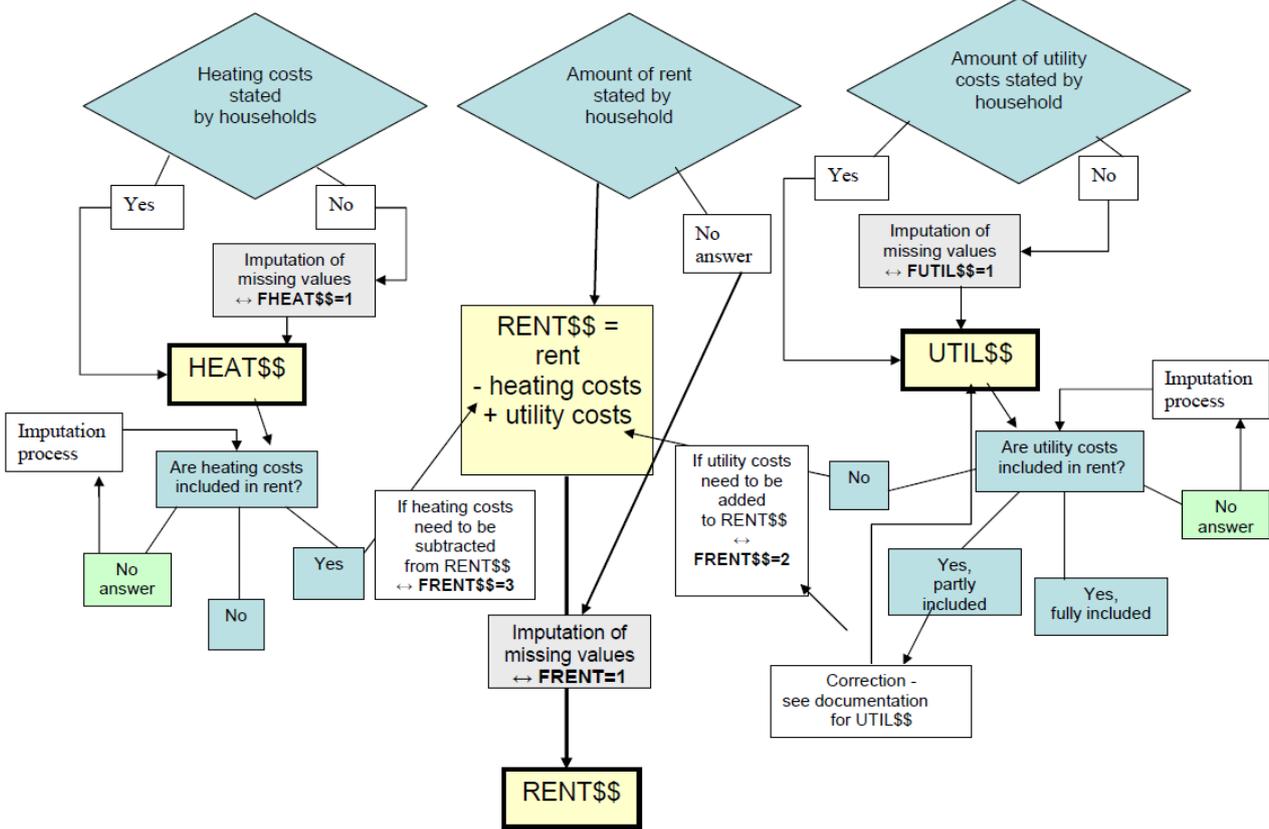
NORENT$$ (formerly $NOMIET)

```
Var Label    :   NORENT$$ "Pays no rent"
Value Label  :   NORENT$$ (1)"pay no rent"
Variable format :        NORENT   (D010)
$$ – Survey Year :       $$=84..11
```

```
Comment:        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 (RENT$$),
                heating costs (HEAT$$) and utilities (UTIL$$) is coded "-2"
                ("not applicable").

                Information is not carried forward.
```

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <rsiegers@diw.de>

TYP1HH$$ (formerly $TYPHH1)

```
Var Label    :    TYP1HH$$ "Household typology (1-digit)"
Value Label  :    TYP1HH$$ (1)'1-person HH'
                           (2)'Childless couple'
                           (3)'Single parent'
                           (4)'Couple with children <= 16 yrs.'
                           (5)'Couple with children > 16 yrs.'
                           (6)'Couple with children <= 16 and > 16 yrs'
                           (7)'Multiple generation HH'
                           (8)'Other combination'
Variable format :        TYP1HH$$ (I3)
$$ - Survey Year :       $$=84..11
```

TYP2HH$$ (formerly $TYPHH2)

```
Var Label    :    TYP2HH$$ "Household typology (2-digit)"
Value Label  :    TYP2HH$$ (11)'1-P-HH Man < 35'
                           (12)'1-P-HH Man 35-<60'
                           (13)'1-P-HH Man >=60'
                           (14)'1-P-HH Woman < 35'
                           (15)'1-P-HH Woman 35-<60'
                           (16)'1-P-HH Woman >=60'
                           (21)'Childless couple'
                           (31)'Single parent + 1 child'
                           (32)'Single parent + 2 or more children '
                           (33)'Single parent + 1 EK.'
                           (34)'Single parent + 2 or more EK.'
                           (35)'Single parent + 2 (E)K.'
                           (36)'Single parent + 3 or more(E)K.'
                           (41)'Couple + 1 child'
                           (42)'Couple + 2 children'
                           (43)'Couple + 3 or more children'
                           (51)'Couple + 1 EK.'
                           (52)'Couple + 2 EK.'
                           (53)'Couple + 3 or more EK.'
                           (61)'Couple + 2 (E)K.'
                           (62)'Couple + 3 or more (E)K.'
                           (71)'3-generation HH'
                           (72)'4-generation HH'
                           (73)'Grandparent(s)-grandchild(ren) HH'
                           (81)'Other combination without children'
                           (82)'Other combination + 1 or more children'
Variable format :        TYP2HH$$ (I3)
$$ - Survey Year :       $$=84..11
```

```
COMMENT:      Generated variable created by combining the relationships of
              all persons living in the household to the head of household
              (Variable $STELL in the file $PBRUTTO) at the time of the
              survey.
              With Wave Z (2009) the data production process switched to a
              standardized procedure for all waves to ensure longitudinal
              consistency, resulting in minor changes compared with older
              distributions.
```

TYP1HH$$ is an aggregation of TYP2HH$$ (first column of the two-digit code). Single households are differentiated in TYP2HH$$ according to both gender and age.

<u>Help for old friends</u>: Starting with data distribution 2010 (waves 1984 to 2009) the category "(88) Other combination" has been further differentiated into households with vs. those without children (up to the age of 16).

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

For more information, contact: Rainer Siegers (Tel. 030-89789-239), <<u>rsiegers@diw.de</u>>

HMONTH$$

```
Var Label    :    HMONTH$$ "Month of Interview"
Value Label  :    HMONTH$$ (1)'January'
                           (2)'February'
                           (3)'March'
                           (4)'April'
                           (5)'May'
                           (6)'June'
                           (7)'July'
                           (8)'August'
                           (9)'September'
                           (10)'October'
                           (11)'November'
                           (12)'December'
Variable format :          HMONTH$$ (I1)
$$ - Survey Year :         $$=84..11
```

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

For more information, contact: Jürgen Schupp (Tel. 030-89789-238), <jschupp@diw.de>

HMODE$$

```
Var Label    :    HMODE$$   "Interview method"
Value Label  :    HMODE$$   (100)"with interview assistance"
                            (110)"oral interview"
                            (120)"Written questionnaire (without interviewer
                                 assistance)"
                            (130)"Mix between with/without interviewer
                                 assistance"
                            (131)"Written questionnaire (with interviewer
                                 assistance)"
                            (132)"Oral and written"
                            (133)"Proxy"
                            (134)"Third person present"
                            (135)"No third person present"
                            (140)"CAPI – Wave O onwards"
                            (200)"telephone assistance"
                            (210)"written, by mail"
                            (220)"telephone assistance"
Variable format :        HMODE$$   (I2)
$$ – Survey Year :        $$=84..11
```

Comment: The interview method is generated through data from the
              household questionnaire. Missing information is filled in
              with data from the corresponding $HBRUTTO files.

For more information, contact: Jürgen Schupp (Tel. 030-89789-238), <jschupp@diw.de>

HINC$$

Var Label   : HINC$$ **"Monthly household net income in euros"**

Variable format : HINC$$ (I4)
$$ – Survey Year : $$=84..11

Comment: 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>

AHINC$$

Var Label   : AHINC$$ **"Adjusted monthly net household income (euros)"**

Variable format : AHINC$$ (I4)
$$ - Survey Year : $$=84..11

Comment: 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. An overview of the different components (excluding net
monthly income, which is available each year) is provided in the
table on the next page, in which "x" indicates that the particular
component was taken into account in that particular year.

Given that SOEP has only asked for all income components on the
individual level since Wave L (1995) also for the current margin,
we had to use the respondent's statement of previous year's income
for the prior period. In the procedure up to 1994, it is thus
necessary that all household members took part in SOEP for at
least two consecutive years. Each income component is only counted
when the month of the interview was also one of the months of
income receipt stated in the following year. To generate income
and month of receipt, we used the calendar ($PKAL).

While no exact monthly calendar was used in the transitional year
1995, the month of receipt was only partially determined using the
employment status calendar (if possible). For income components
that are not captured by the employment status calendar (e.g.,
support from individuals outside the household), the condition was
more than six months of support.

Year	Additional earnings	Old-age pensions	Widow's pensions	Company pensions **	Private pensions**	Unemployment benefits	Unemployment relief	Maintenance payments while in higher education	Transitional payments ***	Early retirement payments ****	Maternity benefits	Bafög (state higher education grants)	Military or civil service pay	Compulsory child support	Support payments
08	-	x	x	-	-	x	-	x	-	x	x	x	x	x	x
07	-	x	x	-	-	x	-	x	-	x	x	x	x	x	x
06	-	x	x	-	-	x	-	x	-	x	x	x	x	x	x
05	-	x	x	-	-	x	-	x	-	x	x	x	x	x	x
04	-	x	x	-	-	x	x	x	-	x	x	x	x	x	x
03	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x
02	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x
01	x	x	x	-	-	x	x	x	x	x	x	x	x	x	x
00	x	x	x	-	-	x	x	x	x	x	x	x	x	-	x
99	x	x	x	-	-	x	x	x	x	x	x	x	x	-	x
98	x	x	x	-	-	x	x	x	x	-	x	x	x	-	x
97	x	x	x	-	-	x	x	x	x	-	x	x	x	-	x
96	x	x	x	-	-	x	x	x	x	-	x	x	x	-	x
95	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
94	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
93	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
92	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
91	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
90	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
89	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
88	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
87	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
86	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
85	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x
84	x	x	x	-	-	x	x	x	-	-	x	x	-	-	x

*   up to 2001, old-age pensions, from 2001-2004 asked individually, and from 2004 on, old-age pensions again
**   only asked for in 2002 and 2003; since 2004 subsumed under old-age pensions.
*** from 2002 on combined with maintenance payments
**** 1996, 1997 and 1998 combined under transitional payments

For more information, contact: Jan Goebel (Tel. +49-30-89789-377), <jgoebel@diw.de>
                               Peter Krause (Tel. +49-30-89789-690), <pkrause@diw.de>

NUTS1$$

```
Var Label          : NUTS1$$ "NUTS-Systematic-1 (Federal State)"

Value Label        : NUTS1$$ (1)'[DE100] Baden-Wuerttemberg'
                             (2)'[DE200] Bavaria'
                             (3)'[DE300] Berlin'
                             (5)'[DE500] Bremen'
                             (6)'[DE600] Hamburg'
                             (7)'[DE700] Hesse'
                             (9)'[DE900] Lower Saxony'
                             (10)'[DEA00] North Rhine-Westphalia'
                             (11)'[DEB00] Rhineland-Palatinate'
                             (15)'[DEF00] Schleswig-Holstein'
          from Wave G (East) on
                             (4)'[DE400] Brandenburg'
                             (8)'[DE800] Mecklenburg-Western Pomera'
                             (13)'[DED00] Saxony'
                             (14)'[DEE00] Saxony Anhalt'
                             (16)'[DEG00] Thuringia'
          from Wave Q on
                             (12)'[DEC00] Saarland'

Variable format  : NUTS1$$ (I4)
$$ – Survey Year : $$=84..90, 91, 92..11
```

Comment: NUTS ("Nomenclature of Statistical Territorial Units") is a
          hierarchical system for dividing up the economic territory of the
          European Union. It was introduced by Eurostat more than 30 years
          ago in order to provide a single uniform breakdown of territorial
          units for the production of regional statistics. NUTS 1 especially
          contains the major socio-economic regions for analyzing regional
          Community problems. It subdivides the European Union by now into
          97 regions, whereas in Germany there are equivalent to the German
          Federal States 16 regions. Before the year 2000 (wave Q)
          Rhineland-Palatinate and Saarland were defined as one region.

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

I_HINC$$

Var Label   : I_HINC$$ **"Multiply imputed monthly net household income (euros)"**

Variable format : I1HINC$$ I2HINC$$ I3HINC$$ I4HINC$$ I5HINC$$ (I4)
$$ – Survey Year : $$=84..11

Comment: Multiple imputation procedures provide a way to deal with missing values on the variable Current Monthly Net Household Income by using information about components and determinants of the household income and replacing item-nonresponse with multiply imputed data. The first five imputations are available within the $HGEN datasets: the variables I1HINC$$-I5HINC$$.

The imputations were calculated using the program ICE (Imputation by Chained Equations) of STATA which was written by Patrick Royston (vgl. Royston 2004, 2005a, 2005b) and which is based on the program MICE in S-Plus and R. The missing observations are assumed to be missing at random. We set the number of imputations m=10 and get 10 multiple imputed values for I_HINC$$. For a discussion on the choice of m, see Rubin (1987) and Royston (2004).

The dataset MIHINC contains the complete imputation results and is separately available. To be compatible with methods for analysing multiply imputed data, MIHINC is constructed in the so called stacked or MIM Dataset Format. It contains the following variables: HHNRAKT, SVYYEAR, MJ, MI, IHINC and IMPFLAG. For every survey household in all survey years (1995-2007) 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 250. 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 assistence, Social assistence, 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 non-response)

- Cross-sectional weights

Analysing multiply imputed data

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

- STATA provides various a builtint 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.

References

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.

FHINC$$

```
Var Label    :  FHINC$$       "Imputation Flag"

Variable format : FHINC $$   (I1)
$ - Wave     :  $= L..Z
$$ - Jahr    :  $$=95..11


Comment: FHINC$$ is a dummy variable indicating whether an observation was
         missing on HINC$$ and was therefore imputed or not.

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