SOEP The German Socio Economic Panel Study

DIW BERLIN

Documentation PPFAD

_____

Person-related meta-dataset

_____

General information:

This file is designed to support longitudinal analysis linking personal information from various waves. Each person saved in at least one $PBRUTTO file is also in PPFAD. The only sorting key (SORT-ID) is the unchanging person ID number PERSNR. The variables AHHNR, BHHNR, CHHNR, etc. contain the current household number of the particular household in which the person lived and was interviewed at the time of the survey, either as an adult household member or child. The variables ANETTO, BNETTO, CNETTO, etc. give the wave-specific interview status or tell which file contains a person in a particular wave ($P, $KIND, etc.) or whether the person has permanently left SOEP (which leads to that person being entered into the dropout file YPBRUTTO).

The variables $HHNR and $NETTO are coded "–2" for those waves in which a person is not yet or no longer part of the SOEP survey population.

In addition, PPFAD contains basic demographic variables, which have been longitudinally tested to some extent:

- o SEX (sex)
- o GEBJAHR (4-digit birth year)
- o TODJAHR (4-digit year of death, if applicable)
- o IMMIYEAR (year of first immigration to Germany)
- o GERMBORN (born in Germany)
- o CORIGIN (country of origin)
- o MIGBACK (migration background)
- o MIGINFO (information source for MIGBACK)

It is recommended that these tested variables be used for cross-sectional and longitudinal analysis. These variables are adjusted on a wave-by-wave basis in the framework of demographic testing. Since 1993, a number of individuals from different original SOEP households moved into the same household. These individuals share the same current household number for the current wave (HHNRAKT or $HHNR) but have different original household numbers, HHNR. With the variable $CASEMAT, these households are assigned the value of the CASE-ID (HHNR) that also needs to be taken into account. All other households are assigned the value -2.

Further methodological variables (year, four-digit):

- o EINTRITT: year a person joined SOEP, e.g. the year a person was first entered in $PBRUTTO
- o AUSTRITT: year a person ultimately left SOEP (entered in YPBRUTTO)
- o ERSTBEFR: year of a person's first interview
- o LETZTBEF: year of a person's most recent interview

List of variables:

$HHNR

```
Var Label    :   $HHNR    "Original household number $$$$"
$ - Wave     :   $=A..BA
$$$$ - Year  :   $$$$=1984..2010
Var format   :   $HHNR  (I6)
```

```
Comment:        Identifier for the household (HHNRAKT) in which a person
                lived at the time of the interview.
```

```
Detailed description:
For more information, contact:          Peter Krause (Tel. 030-89789-690)
```

SEX

```
Var Label   :   SEX      "Sex"
Value Label :   SEX      (1)"male"
                         (2)"female"
Var format  :   SEX (I1)
```

Comment: Respondent's sex, longitudinally tested.

```
Detailed description:
For more information, contact:     Peter Krause (Tel. 030-89789-690)
```


GEBJAHR

```
Var Label   :   GEBJAHR  "Year of birth (four-digit)"
Var format  :   GEBJAHR  (I4)
```

Comment: Respondent's year of birth, longitudinally tested.

```
Detailed description:
For more information, contact:     Peter Krause (Tel. 030-89789-690)
```

$NETTO

```
Var Label    :   $NETTO    "Current Wave Survey Status"
Value Label  :   $NETTO    see below
$ - Wave     :   $=A..BA
Var format   :   $NETTO (I1)
```

With the year 2006 (wave W, 23rd survey wave) the compilation of data on the survey population has changed fundamentally. Previously, an individual interview was carried out with all household members above the age of 16. As of 2006, the regular individual interviews based on the standard adult questionnaire are introduced one year later when household members reach the age of 18. Seventeen-year-olds instead receive an expanded youth questionnaire in their first year as SOEP respondents. (This applies to the old samples A-G; for the new sample H, distribution of this youth questionnaire will start next year, while this year's 17-year-olds have received the regular individual questionnaire, in line with the old system).

This means that we now have two instruments instead of one to obtain data on respondents: the individual and the youth questionnaire. To ensure a consistent differentiation over time, it will therefore be necessary either to include the youth population of the current year or to increase the age limit for all previous years.

The newly revised $NETTO variable assists retrospectively in both differentiations for the entire survey period. The connection between survey population and survey instrument can be retraced with the help of the variable $NETTO in PPFAD or $HNETTO in HPFAD. As a result of the change in the survey population as well as the expansion of the survey instrument to include detailed information on biographical contexts, the corresponding variable $NETTO in PPFAD has been fundamentally revised and is now provided as a two-digit variable. To ease the transition to the new variable, the old one-digit variable is still provided as well under a different name $NETOLD; the variable $HNETTO in HPFAD is unaffected by this and remains unchanged.

The new „Netto"-variable follows the classification of the old version (available in $NETOLD): The Codes 10-39 describe the population in realized and not realized households if these households were part of the current gross population.

Value	German Label	English Label
10	Befragungsperson mit real. Interview (_P)	Interviewee With Succesful Interview (_P)
12	Personen- und Lebenslauffragebogen	Individual Questionnaire And Person Biography
13	Personen- und Jugendfragebogen	Individual Questionnaire And Youth Biography
14	Personenfragebogen und andere Frabos	Individual Questionnaire And Other Questionnaires
15	Personenfragebogen und Experimente, Test	Individual Questionnaire And Experiments, Test
16	Erstbefr, Personenfragebogen, 17 Jahre	Individual Questionnaire, First Time Surveyed, 17 Years
17	Erstbefragung, Jugendfragebogen, 17 Jahr	Youth Biography First Time Surveyed, 17 Years
18	Personenfrabo und Kind unter 17 Jahre	Individual Questionnaire And Child und 17 Years
19	Personenfrabo ohne Haushaltsinterview	Individual Questionnaire Without Household Interview
20	Kinder in realisierten Haushalten (_KIND	Children in Succesfully Interviewed Households (_Kind)
21	Kinder mit Mutter-Kind-Frabo_I, 0-1 Jahr	Children With Mother-Child Questionnaire_I, 0-1 Year
22	Kinder mit Mutter-Kind-Frabo_II 2-3 Jahr	Children With Mother-Child Questionnaire_II, 2-3 Year
23	Kinder mit Mutter-Kind-Frabo_III 5-6 Jahr	Children With Mother-Child Questionnaire_III, 5-6 Year
30	Personen in Brutto-HHen ohne P.Interview	Persons In Successfully Interviewed HH Without P-Interview
31	Realisierte Nacherhebung (_LUECKE)	Successful Gap Interview (_LUECKE)
32	Realisierter Biografiefragebogen	Successfully Completed Biography Questionnaires
33	Realisierter Jugendfragebogen	Successful Youth Questionnaire
34	Realisierte Tests und Experimente	Successful Tests and Experiments
60	Nur Fragebogen ohne P- und HH-Interview	Only Questionnaire Without Indiv. And HH Interview
61	Nacherhobene Luecke ohne HH-Bezug	Gap Interview without HH reference
62	Nacherhobene Luecke bei Ausfall	Gap Interview with drop out
70	Nur Teilnahme an Tests, Experimenten etc	Only Participation In Tests, Experiments, etc.
80	Person ohne aktuelle Angabe ohne Austrit	Individual Without Any Current Information
81	Vormals Befragte(ERSTBEF) ohne akt.Angab	Prior Interviewee Without Any Current Information
88	Rueckkehrer - (zuvor Wegzug Ausland [91]	Repatriate - (moved abroad before [91])
89	Rueckkehrer - (zuvor Ausfaelle [90])	Repatriate - (was drop out [90])
90	Personenausfaelle (_YPBRUTTO)	Individual Dropouts YPBRUTTO
91	Wegzug ins Ausland	Moved abroad
99	Verstorben	Died

```
For more information, contact:    Peter Krause (Tel. 030-89789-690)
```

$NETOLD

```
Var Label    :   $NETOLD        "Current Wave Survey Status"
Value Label  :   $NETOLD
                      (0)'Person Gap (YPBRUTTO)'
                      (1)'Successful Interview (_P [,_JUGEND])'
                      (2)'Below Survey Age (_KIND)'
                      (3)'Did not participate (_PBRUTTO)'
                      (4)'Missing this wave (_PLUECKE)'
                      (5)'Interviewee without household interview'
$ - Wave     :   $=A..BA
Var format   :   $NETOLD (I1)
```

The previous Survey Status ("Netto") Variable will be continued under the name $NETOLD. The code 1 refers from 2006 on not only to persons with a realized personal questionnaire and includes also persons with a realized youth questionnaire.

Based on the new Netto-Variable the following populations can be identified:

```
Respondents aged 17 and older:             $netto 10-19 [$netold 1,5]
      in realized houdeholds:              $netto 10-18 [$netold 1  ]
      17-years old with Youth-Questionnaire:  $netto   17 [$netold 1  ]

Respondents aged 18 and older:
      (only Person Questionnaire):         $netto 10-16 [$netold 1  ]

Childs in realized households:             $netto 20-29 [$netold 2  ]

Persons in current address-log:            $netto 10-39 [$netold 1-4]
```

```
As in previous years you need the additional variable $HNETTO (HPFAD) if you want
to differentiate between person in realized and non-realized households. However
the variable $POP (provided since several years) includes this differentiation
already and identifies additionally private and institutionalized households.
```

$CASEMAT

```
Var Label    :   $CASEMAT  "Case-match, combined panel households"
Value Label  :   $CASEMAT  (0)"CASE with household details"
$ - Wave     :   $=J..BA
Var format   :   $CASEMAT  (I6)
```

```
Comment:         For the first time in Wave 10 of SOEP, we had three
                 individuals from different original households (CASES) living
                 in one common household. The same was true for two people
                 from the second wave of Sample D1 (immigrants). In these
                 cases, people with identical values for HHNRAKT in one wave
                 have different values for HHNR (CASE-ID).
                 The $CASEMAT variable is assigned a value only for those
                 persons living in the households described above, and
                 contains the HHNR of the other household members.
                 This information is not relevant when linking person and
                 household data based on the current household number HHNRAKT
                 in nonhierarchically structured databanks.
```

Detailed description:

For more information, contact: Peter Krause (Tel. 030-89789-690)

$SAMPREG

```
Var Label    :   $SAMPREG "Current wave sample region $$$$"
Value Label  :   $SAMPREG (1)"West Germany"
                          (2)"East Germany"
$ - Wave     :   $=G..BA
$$$$ - Year  :   $$$$=1990..2010
Var format   :   $SAMPREG (I2)
```

Comment: Independent of the original sample in which a household was
              selected (see the variable PSAMPLE in the file PPFAD or
              HSAMPLE in the file HPFAD, and for "old friends" the former
              variable KSAMPLE), since 1991, households have been
              identified as living in either in the West or the East sample
              region. This information is time-dependent; thus for each
              year (since the base survey for the East conducted in 1990),
              there is the wave-specific information $SAMPREG.
              For analyses differentiating between East and West Germany,
              the current sample region variable $SAMPREG should be used,
              and not the time-independent original sample variable
              PSAMPLE.
              To clarify: In the year 2001, over 360 persons from Sample C
              (PSAMPLE=3) lived in the West sample region, while almost 60
              persons from Sample A (PSAMPLE=1) lived in East Germany.

For more information, contact: Peter Krause (Tel. 030-89789-690)

Documentation on the person-related metafile PPFAD

TODJAHR

```
Var Label    :   TODJAHR  "Year of death, four-digit"
Var format   :   TODJAHR  (I4)
```

TODINFO

```
Var Label    :   TODINFO  "Year of death, information source"
Value Label  :   TODINFO  (1)"from annual survey (pbr_exit)"
                          (2)"Infratest Field Organization Study 1992"
                          (3)"Infratest Field Organization Study 2001"
                          (4)"Infratest Field Organization Study 2007"
                          (5)"Infratest Field Organization Study 2008"
Var format   :   TODINFO  (I2)
```

Comment: For all persons who have been identified as deceased over the
                course of SOEP, the variable TODINFO gives the source of this
                information. The variable TODJAHR contains the four-digit
                year of death or a missing value code:
                (-2): persons, for whom it is unknown whether they are
                    deceased (that is, both persons still living up to that
                    wave, and persons whose exact whereabouts is unknown and
                    have dropped out of SOEP)
                (-1): persons, for whom the fact of death is known, but the
                    year of death is unknown.
                $$$$: factual year of death for persons whose death could be
                    firmly established
                  (1) from the dropout file PBR_EXIT*
                  (2) from the Field Organization Study 1992 (see below).
                  (3) from the Field Organization Study 2001 (see below).
                  (4) from the Field Organization Study 2007 (see below).
                  (5) from the Field Organization Study 2008 (see below).

* The file PBR_EXIT corresponds to the file YPBRUTTO in earlier SOEP
releases.

Detailed description:

53 persons were identified as deceased in the Infratest Field Organization
Study (Follow-up study of drop-outs between 1984 and 1992) carried out from
April – June 1992. The following should be noted regarding the computation
of the variable TODJAHR using this data:

GROUP 1:
For 35 persons, it was possible to determine the exact year of death.

```
  HHNR   PERSNR GEBJ x MM YY y    x: Wave in which person dropped out
  ----------------------------    y: 2 = deceased
   8265    82602 1901 2 10 84 2   YY: Year of death
  25410   254101 1919 2 05 84 2   MM: Month of death
  32042   320402 1910 2 08 84 2
  42501   425002 1905 2 12 84 2
  47449   474401 1920 2 08 84 2
  59285   592802 1928 2 09 84 2
   4464    44601 1908 3 -2 85 2
   5908    59001 1909 2 08 85 2
  42170   421701 1930 2 06 85 2
  35262   352601 1904 2 07 86 2
   5983    59802 1914 2 01 87 2
   7137    71301 1930 5 07 87 2
  24317   243101 1908 3 01 87 2
  26352   263501 1904 2 06 87 2
```

```
35920   359201 1914 2 07 87 2
42501   425001 1905 2 05 87 2
44962   449603 1900 2 09 87 2
22284   222802 1908 4 01 88 2
32034   320301 1902 2 11 88 2
36870   368701 1916 3 05 88 2
39691   396902 1935 3 07 88 2
42358   423501 1903 4 03 88 2
 3107    31001 1919 2 08 89 2
 7064    70601 1939 6 12 89 2
 8346    83402 1908 6 01 89 2
12823   128201 1912 5 01 89 2
29831   298301 1910 7 03 89 2
32042   320401 1912 2 05 89 2
44091   440902 1905 2 04 89 2
 5657    56501 1934 7 05 90 2
 7412    74101 1901 6 07 90 2
 8346    83401 1912 7 11 91 2
31038   310301 1916 2 -2 91 2
32034   320302 1914 2 12 91 2
44091   440903 1906 2 01 92 2
```

GROUP 2:
For 16 persons, the exact year of death is unknown. For them, we used the year in which the
                  person dropped out of SOEP in the place of year of death.

```
HHNR   PERSNR GEBJ x MM YY y ZZ      x: Wave in which person dropped out
-------------------------------       y: 2 = deceased
                                     ZZ: imputed year of death
 4936    49302 1909 2 -2 -2 2 85
 6270    62701 1926 2 -2 -2 2 85
 6270    62702 1928 2 -2 -2 2 85
18279   182703 1909 5 -2 -2 2 88
20184   201802 1911 3 -2 -2 2 86
20770   207701 1927 4 -2 -2 2 87
27928   279201 1928 3 -2 -2 2 86
31178   311701 1907 3 -2 -2 2 86
32778   327701 1904 6 -2 -2 2 89
35793   357902 1897 2 -2 -2 2 85
37559   375501 1923 3 -2 -2 2 86
37800   378001 1908 4 -2 -2 2 87
42773   427702 1915 2 -2 -2 2 85
42803   428001 1908 3 -2 -2 2 86
43044   430401 1919 5 -2 -2 2 88
50539   505301 1929 3 -2 -2 2 86
```

GROUP 3:
With 2 persons, implausible answers were corrected.

```
HHNR   PERSNR GEBJ x MM YY y ZZ      x: Wave in which person dropped out
-------------------------------       y: 2 = deceased
                                     ZZ: Correction of year of death
11452   114502 1894 2 09 98 2 89
42145   421401 1904 3 05 89 2 86
```
****

In the framework of the Infratest Field Organization Study (follow-up study
of drop-outs) of 2001, a total of over 700 persons were identified as
deceased. Among them were several with multiple entries for year of death,
that is, persons who were already identified as deceased in the standard
wave-to-wave follow-up procedure (stored in the file PBR_EXIT) or in the
Infratest Field Organization Study of 1992. A generally very high level of
correspondence was found between the information given in the standard
follow-up procedure and the point of death established ex-post in the
Infratest Field Organization Studies. For ten persons, the year of dropping
out of SOEP was used to impute the missing year of death.
In the third of those follow-up studies which has been conducted in 2007,
another 21 individuals were identified as deceased between 2001 and 2005.
For 18 of those persons a valid year of death could be investigated, for
the remaining three observations for which the exact year of death is
unknown, TODJAHR has been set to the standard missing code "-1".

****

Documentation on the person-related metafile PPFAD

When the data from the Infratest Field Organization Study contradicted the
data from PBR_EXIT (help for old friends: YPBRUTTO in earlier releases),
the data from the Field Organization Study was used.

```
*************************************************************
```
For more information, contact: Joachim Frick (Tel. 030-89789-279)

PSAMPLE

```
Var Label    :   PSAMPLE   "Sample member"
Value Label  :   PSAMPLE   (1)'A German (West)'
                           (2)'B Foreigner (West)'
                           (3)'C German (East)'
                           (4)'D 84-93 Immigrant (West)'
                           (5)'E Refreshmemt sample 1998'
                           (6) 'F ISOEP 2000'
                           (7)'G High-income sample 2002'
                           (8)'H Refreshment sample 2006'
                           (9)'I Incentivisation sample 2009'

Var format   :   PSAMPLE   (I2)

Comment:         Unchanging sample information: for example, a person from
                 Sample C would retain this original sample number (the sample
                 in which the household first entered SOEP) even after having
                 later moved to West Germany (see also the comments on the
                 variables $SAMPREG above). Similarly, a person from Sample B
                 who acquired a new nationality would still remain in his or
                 her original sample.
```

Detailed description:

For more information, contact: Peter Krause (Tel. 030-89789-690)

IMMIYEAR

```
Var Label    :   IMMIYEAR "Year of immigration to Germany after 1948"
Var format   :   IMMIYEAR (I4)
```

GERMBORN

```
Var Label    :   GERMBORN "Born in Germany or immigration prior to 1949"
Value Label  :   GERMBORN (1)'born in Germany or immigration prior to 1949'
                          (2)'not born in Germany'
Var format   :   GERMBORN (I2)
```

CORIGIN

```
Var Label    :   CORIGIN   "Country of origin"
Value Label  :   CORIGIN
```

(1) 'Germany'	(2)'Turkey	(3)'Ex-Yugoslavia'	(4) 'Greece'
(5)'Italy'	(6)'Spain'	(7)'Ex-GDR'	(10)'Austria'
(11)'France'	(12)'Benelux'	(13)'Denmark'	(14)'Great Britain'
(15)'Sweden'	(16)'Norway'	(17)'Finland'	(18)'USA'
(19)'Switzerland'	(20)'Chile'	(21)'Rumania'	(22)'Poland'
(23)'Korea'	(24)'Iran'	(25)'Indonesia'	(26)'Hungary'
(27)'Bolivia'	(28)'Portugal'	(29)'Bulgaria'	(30)'Syria'
(31)'Czech Republic '	(32)'Russia'	(33)'Kurdistan'	(34)'Mexico'
(35)'Argentina'	(36)'Cap Verde Is.'	(37)'Benin'	(38)'Philippines'
(39)'Israel'	(40)'Japan'	(41)'Australia'	(42)'India'
(43)'Afghanistan'	(44)'Thailand'	(45)'Jamaica'	(46)'Saudi-Arabia'
(47)'Ethiopia'	(48)'Columbia'	(49)'Ghana'	(50)'Bangladesh'
(51)'Venezuela'	(52)'Tunesia'	(53)'Mauritius'	(54)'Nigeria'
(55)'Canada'	(56)'New Zealand'	(57)'Tanzania'	(58)'Cyprus'
(59)'Cuba'	(60)'Iraq'	(61)'Brasil'	(62)'Monaco'
(63)'Hong Kong'	(64)'Peru'	(65)'Sri Lanka'	(66)'Nepal'
(67)'Morocco'	(68)'China'	(69)'Liechtenstein'	(70)'Iceland'
(71)'Ireland'	(72)'St. Lucia'	(73)'Moldavia'	(74)'Kasakhstan'
(75)'Albania'	(76)'Lebanon'	(77)'Kyrgyzstan'	(78)'Ukraine'
(79)'Algeria'	(80)'Mozambique'	(81)'Egypt'	(82)'Tajikistan'
(83)'Vietnam'	(84)'Somalia'	(85)'Pakistan'	(86)'South Afrika'
(87)'UAE'	(88)'El Salvador'	(89)'Eritrea'	(90)'Jordan'
(91)'Turkmenistan'	(92)'Costa Rica'	(93)'Singapore'	(94)'Burkina Faso'
(95)'Zambia'	(96)'Ecuador'	(97)'Uzbekistan'	(98)'no nationality'
(99)'Puerto Rico'	(100)'Laos'	(101)'Estonia'	(102)'Angola'
(103)'Latvia'	(104)'Malaysia'	(105)'Namibia'	(106)'Montenegro'
(107)'Belize'	(108)'Dominican Republic'	(109)'Nicaragua'	(110)'Kenya'
(111)'Libya'	(112)'Malta'	(113)'Botswana'	(114)'Haiti'
(115)'Trinidad, Tobago'	(116)'Luxembourg'	(117)'Belgium'	(118)'Holland'
(119)'Croatia'	(120)'Bosnia-Herzegovina'	(121)'Macedonia'	(122)'Slovenia'
(123)'Slovakia'	(124)'Paraguay'	(125)'Guinea'	(126)'Kuwait'
(127)'Ivory coast'	(128)'Malaysia'	(129)'Samoa'	(130)'Aserbaijan'
(131)'Seychelles'	(132)'Belarus'	(133)'Uruguay'	(134)'Bahamas'
(135)'Uganda'	(136)'Oman'	(137)'Micronesia'	(138)'Mali'
(139)'Kameroon'	(140)'Kosovo-Albania'	(141)'Georgia'	(142)'Sudan'
(143)'Congo'	(144)'Togo'	(145)'Mongolia'	(146)'Lithuania'
(147)'Chad'	(148)'Armenia'	(149)'Kurdistan'	(150)'Liberia'
(151)'Yemen'	(152)'Palestine'	(153)'Free state of Gdansk'	(154)'Taiwan'
(155)'Turkmenistan'	165"Serbia"	(222)'Eastern Europe'	(333)'other not specified foreign country'

Comment: Along with SOEP Samples B and D (foreigners and immigrants), Subsamples A, C, E, F, G and H also contain persons who immigrated to Germany from foreign countries. These samples include persons who immigrated to Germany both before and after the selection of the original SOEP samples A and B in the year 1983. The variables IMMIYEAR, GERMBORN and CORIGIN are given for all respondents in all SOEP subsamples (A to H) who were interviewed at least once since 1984, based on information from the files $P, $PAUSL, $PBRUTTO, BIOLELA (integrated biographical data files for Waves A to L) and $LELA (life course information on first-time respondents since Wave M). For persons who were either not yet or no longer SOEP respondents when the immigration information was collected, the variables are set at "-1" (no response) if no adequately reliable decision on country of origin could be made based on the person-related responses. There is one exception to this: underage children, for whom country of origin is approximated based on the child's year of birth and the biological mother's year of immigration. Since the biographical questionnaire was first used with Sample G in the second wave (in the year 2003), the central basis for generating immigration variables is lacking for around 800 persons. For these individuals, the variables IMMIYEAR, GERMBORN and CORIGIN were set at "-3".

Detailed description: **More detailed information on these variables can be found in the documentation on the biographical data**.

For more information, contact: Joachim Frick (Tel. 030-89789-279)

$POP

```
Var Label    :   $POP        "Sample membership $$$$"
Value Label  : $POP          (1)'Private household, German HH-head'
                             (2)'Private household, foreign HH-head'
                             (3)'Institutional household, German HH-head'
                             (4)'Institutional household, foreign HH-head'
                             (5)'Not Complete Private HH, German HH-Head'
                             (6)'Not Complete Private HH, Foreign HH-Head'
                             (7)'Not Complete Institutional. HH, German HH-
                                Head'
                             (8)'Not Complete Institutional. HH, Foreign HH-
                                Head'
$ - Wave     :   $=A..BA
$$$$ - Year  :   $$$$=1984..2010
Var format   :   $POP        (I2)
```

Comment: $POP was derived from $WUM2 (see private household,
              institutional household in $HBRUTTO) as well as $PNAT and
              $STELL (nationality and relationship to head of household in
              $PBRUTTO). Missing values were imputed based on the person's
              history. Thus, the only admissible missing value is –2,
              meaning not applicable.
              This variable is therefore particularly important, as it
              enters into the determination of cross-sectional weights.
              The variable corresponds with $SHPOP in HPFAD. See also the
              description of $NETTO.

For more information, contact: Peter Krause (Tel. 030-89789-690)

GEBMONTH

```
Var Label    :   GEBMONTH  "Month of birth"
Value Label  :   GEBMONTH  (1)'January'
                           (2)'February'
                           (3)'March'
                           (4)'April'
                           (5)'May'
                           (6)'June'
                           (7)'July'
                           (8)'August'
                           (9)'September'
                           (10)'October'
                           (11)'November'
                           (12)'December'
Var format   :   GEBMONTH    (I2)
```

GEBMOVAL

```
Var Label    :   GEBMOVAL  "Month of birth, data source"
Value Label  :   GEBMOVAL  (1)'Generated from gebmonth (parents)'
                           (2)'Info from PPFAD'
                           (3)'Info from _KIND'
                           (4)'Info from SP'
                           (5)'Info from _LELA'
                           (6)'Info From bioage01 (mother)'
                           (7)'Info from $PAGE17'

Comment:        GEBMOVAL: Data source on month of birth.

Detailed Description:
```

1. Introduction of variables
From wave T onwards (2003) the data set PPFAD contains – in addition to the year of birth –the month of birth (GEBMONTH). This new variable is accompanied by the supporting variable GEBMOVAL which indicates the data source for the month of birth.
GEBMONTH and GEBMOVAL can take the following characteristics:

- GEBMONTH: Month of birth;
         1 (January) to 12 (December)
- GEBMOVAL: Month of birth - data-source
         1 generated
         2 information as stored in PPFAD
         3 derived from data set $KIND
         4 derived from data set SP
         5 derived from data set $LELA
         6 derived from BIOAGE01
            (mother-child-questionnaire)
            (NEW with Wave W / Surveyyear 2006!)

The month of birth was asked in wave S individual questionnaire (SP). Furthermore, the month of birth was asked in the biography data set, starting with wave T ($lela, file not available with the SOEP data distribution). Additionally the month of birth is recorded for all children within the file $KIND (starting with wave T). While this provides the relevant information for most of the current panel members, the information remains missing for some persons including temporary dropouts or people who exited in a previous wave. For some of them the month of birth could be reconstructed. This reconstruction remains an approximation and might differ from the true month of birth in individual cases.

Documentation on the person-related metafile PPFAD

2. Construction of variables.
The month of birth is constructed in an hierarchical order from the files:
- Generated (basis: $P, $PBRUTTO $KIND)
- $KIND
- SP
- $LELA
- BIOAGE01 (Starting with wave W, 2006)

whereas the latter information overrides the former.
This means the generated information will only be utilized if no further,
questionnaire based information for the month of birth is available.
The generated month of birth could only be constructed for people who were
born while their parents were members of the SOEP. The information was
derived from two sources:
- For newborn children the month of moving into the household was used
  as an approximation of the real month of birth (relevant file
  $PBRUTTO).
- For parents who reported a birth in a certain month, a link to the
  child was established, assigning the month of birth to the child
  (relevant file $P).

Several adjustments and tests of the generated data have been done which
showed that – in the cases in which the generated data was also collected
by SP, $LELA or $KIND – the data generation is almost always congruent with
the collected data and therefore has proven to be reliable.
Frequency examples for the variables "month of birth" and "data source for
month of birth" in the file ppfad (Version 2007 / up to Wave X)

GEBMONAT Month of Birth

Month of Birth	Freq.	Percent	Cum.
[-3] Answer improbable	1	0.00	0.00
[-1] No Answer	26,821	40.14	40.14
[1] January	3,472	5.20	45.34
[2] February	3,294	4.93	50.27
[3] March	3,579	5.36	55.63
[4] April	3,351	5.02	60.64
[5] May	3,475	5.20	65.84
[6] June	3,218	4.82	70.66
[7] July	3,407	5.10	75.76
[8] August	3,311	4.96	80.72
[9] September	3,410	5.10	85.82
[10] October	3,324	4.98	90.80
[11] November	3,018	4.52	95.31
[12] December	3,132	4.69	100.00

GEBMOVAL Month of Birth – Data Source

Month Of Birth, Data Source	Freq.	Percent	Cum.
[-3] Answer improbable	1	0.00	0.00
[-1] No Answer	26,821	40.14	40.14
[1] Generated from gebmonth (parents)	1,485	2.22	42.37
[3] $kind, Info from mother	5,719	8.56	50.93
[4] Info From Sp	21,536	32.23	83.16
[5] Info From $lela	8,078	12.09	95.25
[6] Info From bioage01 (mother)	1,809	2.71	97.96
[7] Info from $PAGE17	1,364	2.04	100.00

© Längsschnittstudie SOEP - DIW Berlin

```
                          Total |     66,813        100.00
```

For more information, contact: Christian Schmitt (Tel. 030-89789-603)

LOC1989

Var Label   : LOC1989   **"Where did you live in 1989?"**
Value Label : LOC1989   (1) "East Germany (GDR) incl. East Berlin"
                          (2) "West Germany (FRG) incl. West Berlin"
                          (3) "abroad (foreign country)"

Var format   : LOC1989   (I2)

Comment: After asking this information from all respondents in 2003
             (variable TP121 in file TP), a corresponding question was
             included in the biography questionnaire in Wave U (2004)
             [Question 16, variable UB16 in file ULELA] which will collect
             this time-independent information from all future first-time
             respondents. For all respondents interviewed up to 2003, the
             following information was used as input to generate LOC1989:
             - Information on place and date of last school attendance (see
                variables BSSCHEND and BSSCHWO in file BIOSOC; variables UB38
                and UB3701 in file ULELA)
             - Sample affiliation (variable PSAMPLE in file PPFAD)
             - Year moved to current address (variable VRMOVEIN in file
                BIORESID; variable UB68 in file ULELA)
             - Sample region (variables $SAMPREG in file PPFAD)
             - Year of first immigration to Germany (variable IMMIYEAR in
                file PPFAD)
             In cases of inconsistent information from these various sources,
             the data collected in 2003 via variable TP121 and the
             information from the biography questionnaire collected since
             2004 is considered superior. Persons without any individual
             information and aged less than 18 years in 1989 have been
             assigned parental information, if at all available.

Detailed description: More detailed information on this variable can be found in the
                        documentation on the biographical data in the section on PPFAD.

For more information, contact: Joachim Frick (Tel. 030-89789-279)

MIGBACK

```
Var Label   : MIGBACK    "Migration Background"
Value Label : MIGBACK    (1) "No migration background"
                         (2) "Direct migration background"
                         (3) "Indirect migration background"
                         (4) "Migration background, not differentiated"

Var format  : MIGBACK    (I2)
```

Comment: MIGBACK provides information on migration background. It con-
              siders the nationality of the individual, his/her migration
              history as well as that of his/her parents for respondents born
              in Germany. In order to do so information on country of birth,
              citizenship and its change over time (in case of naturalization
              for example), and if available, parental information (direct
              information or proxy information) are combined.
              Direct/Indirect migration background indicates whether the per-
              son immigrated by him/herself (direct migration background) or
              if she/he is of migrant origin but born in Germany (indirect
              migration background). Respondents assigned the code 4 have a
              migration background as the information on the region of birth
              is missing, it is not possible to know whether they migrated by
              themselves of if they were born in Germany.

              Note that code 1 ("no migration background") also includes per-
              sons of German nationality born in Germany whose parental in-
              formation are not available. Some of them might be migrants'
              offspring. Thus, the MIGBACK variable slightly underestimates
              the number of persons belonging to code 3, i.e. indirect migra-
              tion background.

              Note that in principle, an individual who gains a foreign na-
              tionality will be classified as being of migrant origin even if
              this was not the case before. Thus, the migration background
              variable as applied in SOEP takes a retrospective perspective
              and will be updated every year.

              Ethnic German repatriates ("Spät-/Aussiedler") in this classi-
              fication are coded "2" (direct migration background). Natural-
              ized respondents who were born in Germany are coded "3" (indir-
              ect migration background).

              In order to provide highest possible transparency we include a
              variable informing about the sources used to create the migra-
              tion background variable. This information is provided in MI-
              GINFO.

```
Detailed description:   More detailed information on this variable can be found in the
                        documentation on the biographical data in the section on PPFAD.
```

For more information, contact: Ingrid Tucci (Tel. 030-89789-465)

MIGINFO

```
Var Label    : MIGINFO    "Information Source for MIGBACK"
Value Label  : MIGINFO    (1) "Direct info without parental info"
                          (2) "Proxy info without parental info"
                          (3) "Direct info with parental info"
                          (4) "Proxy info with parental info"

Var format   : MIGINFO    (I2)
```

Comment: The MIGINFO variable informs on the information used to generate MIGBACK. There are two distinctions between used information. Firstly, the personal information. Either it stems from the interviewee or it is reported by someone else. Secondly, there is a further differentiation between persons for which we have parental information and those whose parents were not interviewed.

Direct information stem from the respondent (NATION$$, GERMBORN or variables about citizenship acquisition in $P). A person's migration background is based on proxy information if at least one used information is either reported by the parents (AK07A or EK03A) or by the interviewer ($PNAT out of $PBRUTTO or the respective Infratest information).

Detailed description: More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Ingrid Tucci (Tel. 030-89789-465)