SOEP The German Socio-Economic Panel study

583

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP — The German Socio-Economic Panel study at DIW Berlin 2018

SOEP-Core v33.1 – BIOJOB: Detailed Information on First and Last Job

Paul Schmelzer, Tobias Wolfram, and SOEP Group

DIW SOEP

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

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

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

Editors:
Dr. Jan Goebel, DIW Berlin
Prof. Dr. Stefan Liebig, DIW Berlin and Universität Bielefeld
Dr. David Richter, DIW Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Paul Schmelzer, Tobias Wolfram, and SOEP Group. 2018. SOEP-Core v33.1 – BIOJOB: Detailed Information on First and Last Job. SOEP Survey Papers 583: Series D. Berlin: DIW/SOEP

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
© 2018 by SOEP
ISSN: 2193-5580 (online)
DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany
soeppapers@diw.de

SOEP-Core v33.1 – BIOJOB: Detailed Information on First and Last Job

Paul Schmelzer, Tobias Wolfram, and SOEP Group

BIOJOB: Detailed Information on First and Last Job

by Paul Schmelzer and Tobias Wolfram[1]

1 Overview

Biographical data in the SOEP stem from various sources. All information for the waves 1984 to 1995 is compiled in the BIOLELA-file of the SIR-SOEP-database. Since 1996 a standardised version for all samples has been provided, and new biographical data is stored in wave-specific files ($LELA). To have a general phrasing, all biographical files are referred to as LELA-files. (LELA stems from the German 'LEbensLAuf', curriculum vitae.)

The LELA-data relevant for BIOJOB consists of

- the age at and year of entry into the working force given by different theoretical considerations
- the type of occupation at entry (blue/white collar worker, self-employed, civil servant)
- detailed occupational information at entry
- changes of occupation
- intended educational degree or vocational/professional training
- the year of the last employment
- the type of occupation in the last job.

Since 2000 a new questionnaire (in the following referred to as Youth Questionnaire) has been provided for respondents who are 16 or 17 years old. The youth respondents answer the Youth Questionnaire instead of the biographical one. The Youth Questionnaire provides less detailed information about the job biography because respondents usually have not entered the labour market at the age of 16 or 17.

In 2001 members of the F sample became part of the biojob population. They had to answer the biography questionnaire if their year of birth was prior to 1982. Members of the F sample with a birth year in the range from 1982 to 1984 answered the Youth Questionnaire.

Members of sample G (2002) answered the biography questionnaire in 2003, Persons who were born between 1986 and 1987 answered the Youth Questionnaire.

[1] Based on work of Tanja Schmidt, Anita Kottwitz, Daniel Wachtlin, Mathis Schroeder, Thorsten Schneider; and Hansjoerg Haas.

Members of sample H (2006) answered the biography questionnaire first time in 2007 and therefore are part of the BIOJOB population.

Sample I has been moved to the SOEP-Innovation study and, since 2011 (wave 28/BB), are no longer part of the core SOEP population. Members of sample I are still part of of the BIOJOB population until 2010.

Since 2006 respondents who are 16 or 17 years old filled in a youth questionnaire instead of the standard Individual Questionnaire, which provides less detailed information about the current job.

Recently several new projects were integrated into BIOJOB: The IAB-SOEP migration sample M1 (2013) and migration sample M2 (2015) but also the "Familien in Deutschland" (L1-L3) data which was incorporated into the SOEP in 2015 and is for the first time available with the distribution of wave BE. In 2016 IAB-BAMF-SOEP Survey of Refugees (M3, M4) were also integrated in BIOJOB.

The purpose of BIOJOB is to provide a file, that offers the user convenient access to biographical information on past job activities. Up to now all but two variables of BIOJOB are time-invariant. Information on occupational changes and on the age at the most recent change of occupation refer to the date of the respondent's biography interview.

2 Structure and Contents of BIOJOB

BIOJOB consists of generated variables as well as plain questionnaire information. In this section the generated variables are explained and their coding is illustrated.

Concerning different sources of information, the following priority scheme is applied: First the plain information stemming directly from questions on the relevant topic in the latest valid LELA-file is used. In case of inconsistencies, which will be explained later on, the latest valid information stemming from the PBIOSPE file is also used. The PBIOSPE file consists of spell data concerning the retrospective question 'what did you do since the age of 15' in the Biography Questionnaire as well as the question on activities in the last year in the Individual Questionnaire (for detailed information see chapter 3).

Contents of BIOJOB

Population: All persons with an entry in any LELA-/YOUTH-file up to 2015, even if information on employment is missing.

number of cases: 79,802 *waves*: 1984 - 2016 *samples*: A, B, C, D, E, F, G, H, I, J,K,L,M

variables:

HHNR	original household identifier
PERSNR	unique individual identifier
BIOYEAR	year of biography / youth interview
AGEFJOB	age at first job
AGEINFO	information source AGEFJOB
EINSTIEG_ARTK	Year of first job (different generation process involving ARTKALEN)
EINSTIEG_ARTK_INFO	information source EINSTIEG_ARTK
EINSTIEG_PBIO	Year of first job (different generation process involving PBIOSPE)
EINSTIEG_PBIO_INFO	information source EINSTIEG_ARTK
NOJOB	never worked before the time of the interview
STILLFJ	still employed in first job
OCCFJOB	occupational position first job
FULLTIME	first job was a full-time or part-time job
FJBLUE	first job blue collar worker
FJSELFE	first job self-employed
FJSEFSIZ	number of employees FJSELFE
FJWHITE	first job white collar worker
FJCIVS	first job civil servant
ISCO88	International Standard Classification of Occupation 1988, first job
STBA	classification of career according to the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992, first job
EGP	Erikson and Goldthorpe's Class Category (EGP), first job
ISEI	International Socio-Economic Index of Occupational Status after Ganzeboom (ISEI), first job
MPS	Magnitude Prestige Scale after Wegener, first job
SIOPS	Treiman Standard Int. Occ. Prestige Scale, first job
REQEDUC	required education for first job
CIVILSFJ	first job was in civil service
NACEFJ	NACE branch code first job

OCCMOVE	number of occupational changes
AGEATMV	age at most recent occupational change
INTEDUC1 to INTEDUC4	intended educational degree
CURREMPL	employed at time of biography interview
YEARLAST	year of last employment
SCOPELJ	last job was a full-time or part-time job
CIVILSLJ	last job was in civil service
NACELJ	NACE branch code last job
OCCLJOB	occupational position last job
LJBLUE	last job blue collar worker
LJSELFE	last job self-employed
LJSEFSIZ	number of employees LJSELFE
LJWHITE	last job white collar worker
LJCIVS	last job civil servant

If data are missing, we use the SOEP missing value definition:

-1	no answer / don't know: item nonresponse
-2	does not apply
-3	after intensive checks a given value was found to be implausible and was finally deleted (to be interpreted like -1)

Description of variables

AGEFJOB/AGEINFO

The variable AGEFJOB provides the age at entry into the working force. AGEINFO is a pointer variable indicating the source of the age information.

In the Biography Questionnaire people either have to give information on their age at entry into the working force or have to state that they have never worked before the time of the interview. The latter information is used in the variable NOJOB.

In the Youth Questionnaire people have to answer whether they are currently working in a regular occupation. They are not asked about the age at their first occupation, but since people answering the Youth Questionnaire are normally at the age of 16 or 17, in most cases we can assume that a full-time job at this age is their first regular employment.

Information on the coding procedure of AGEFJOB is provided in the following subsections where (a) to (i) refer to LELA respondents, (j) to (p) to youth respondents respectively.

LELA-respondents

a) For people who are or have ever been employed at the time of answering the biographical questions their age at the time of entry into the working force is taken from the LELA-files.

b) When we observe, that the person has not been in the working force at the time of responding, but starts to work later on, data of the PBIOSPE-file is used. Using the spell information in PBIOSPE, we are able to collect the age at the first job.

c) A replacement of the LELA-data takes place, when respondents state that they have worked before the age of fifteen, but have a spell entry later than the age of fifteen. This rule is not applied when the spell starts at the age of fifteen, since this is the minimum value for spell data in the questionnaires.

d) The same procedure is applied, when people answer, that they have never worked at the time of the interview, but have a spell which starts before the first interview.

e) In some cases the AGEFJOB value is higher than the start of the corresponding working spell in PBIOSPE. In general, the AGEFJOB value is maintained. Only when the value is greater than 27, is it replaced by the PBIOSPE data. (95% of these cases have an AGEFJOB below 27.)

f) If we observe item non response concerning AGEFJOB and NOJOB, but spell information is available, the missing value is replaced by the corresponding PBIOSPE spell data.

g) If even the 'What did you do since you were 15' question had not been answered, there still was a chance to extract similar information out of the PBIOSPE-file by considering the question 'What did you do every month last year'.

h) If we still had no valid information, the value of AGEFJOB was left out of the dataset.

i) Due to the fact that PBIOSPE information are collected only until the end of the year preceding the actual wave, for respondents without first job information from both the biography questionnaire and PBIOSPE we further look for a first job using information from the current wave individual questionnaire.

YOUTH-respondents

j) For respondents who are regularly employed, information is taken from the Youth Questionnaire; AGEFJOB is coded as year of questioning minus year of birth minus one (only if the respondent does not state that he/she is still in school, etc.).

k) If we additionally observe a spell starting before the respondent answers the Youth Questionnaire, information from PBIOSPE is used if the respondent does not state in the current questionnaire that he/she is still in school, etc.

l) If respondents answer that they have no regular employment but provide an employment spell starting after the time of the first interview, information from $P (for details see m) is taken if available (only if the respondent does not state that he/she is still in school, etc.).

m) For respondents with inconsistent first job information (simultaneous employment and school attendance/apprenticeship, differing job info in Youth Questionnaire and PBIOSPE) the question 'Are you currently engaged in paid employment?' asked in the Individual Questionnaire turned out to be the most reliable source of information. If a respondent states to be full- or part-time employed in a wave subsequent to the youth interview, AGEFJOB info is derived from the latest information of that kind.

n) If people do not answer at least one of the questions 'Do you currently earn money?" and 'Do you earn money as an apprentice, full-time worker or part-time-worker?" but have an employment spell, like in m) the earliest $P information is taken if available (only if the respondent does not state that he/she is still in school, etc.).

o) If information from the Youth and the Individual Questionnaire (including PBIOSPE) are inconsistent concerning AGEFJOB, then the variable is set to missing.

p) Due to the fact that PBIOSPE information are collected only until the end of the year preceding the actual wave, for respondents without first job information from both the Youth Questionnaire and PBIOSPE we further look for a first job using information from the current Individual Questionnaire.

The pointer variable AGEINFO provides the coding information described above. Value labels of AGEINFO indicating the source of information are:

(1) LELA-files (case (a) above)

(2) PBIOSPE if AGEFJOB<15, but spell begin > 15 (c)

(3) PBIOSPE if 'not worked' at interview but later spell begin (b)

(4) PBIOSPE if 'not worked' at interview but earlier spell begin (d)

(5) PBIOSPE if AGEFJOB>27 and earlier spell begin (e)

(6) implausible information therefore set missing (h)

(7) PBIOSPE if 'not worked'-question and AGEFJOB not answered, but 'what done at 15'-question answered (f)

(8) PBIOSPE if 'not worked'-question, AGEFJOB and 'what done at 15'-question not answered, but 'what done last year'-question answered (g)

(9) completely missing

(10) SP if no info from bio interview and PBIOSPE but employment in current Individual Questionnaire (i)

(11) info drawn from Youth Questionnaire(j)

(12) info drawn from PBIOSPE for persons who state in the Youth Questionnaire to be regularly employed and additionally have an employment spell starting earlier (k)

(13) info drawn from $P for persons who state in the Youth Questionnaire not to earn money relating to an employment/job or to earn money but relating to a part-time job or a practical training, and have a subsequent employment spell (l)

(14) info drawn from $P for persons with inconsistent first job information from the Youth Questionnaire or PBIOSPE, but valid employment information from an Individual Questionnaire subsequent to the biography interview (m)

(15) info drawn from $P for persons with item non response in one of the questions 'Do you already earn money from jobs?' or 'Do you earn that money as a trainee, full-time or part-time employee?' and with info in PBIOSPE (n)

(16) completely missing

(17) set to missing because of inconsistent information (o)

(18) info drawn out of UP, the last wave of the SOEP (p)

For more than 50% of the cases with AGEINFO = 3, 7, or 8 (AGEINFO=7 or 8 only if information collected after biography interview) it is possible to extract information from the regular questionnaires.

For respondents with AGEINFO=10 or 11, information referring to the variables OCCFJOB, FJBLUE, FJWHITE, FJSELFE, FJSEFSIZ, FJCIVS, REQEDUC and CIVILSFJ are taken from the Individual Questionnaire (same year as of youth interview). While for respondents having AGEINFO=10 this approach is intuitive, for the persons having AGEINFO=11 we act on the assumption that the job declared in the respective Individual Questionnaire is still the first job of that person. This assumption seems plausible due to the low age of all persons responding to the YOUTH Questionnaire.

In the YOUTH Questionnaire there is no question on the first job. But we can follow up their professional career by the statements given in the activity calendar in the subsequent waves. This can lead to problems if these youths report student jobs. For that reason we decided to take information from the question "Are you currently engaged in paid employment?" asked in the Individual Questionnaires of subsequent waves as the relevant source of information for this group of respondents. The earliest information of that kind determines the variable AGEFJOB.

Some respondents have very low values with respect to AGEFJOB. Most of these jobs turn out to be low-skilled and starting before 1970. The respective persons are either blue collar workers (mostly unskilled) or self-employed (mostly helping in family business). We think these characteristics suggest that these specifications are valid.

EINSTIEG_ARTK/EINSTIEG_ARTK_INFO

The variable EINSTIEG_ARTK provides the year of available survey information related to the entry into the working force. It is primarily based on information found in the spell dataset ARTKALEN and generally founded on a different conceptualization of job entry than AGEFJOB. EINSTIEG_ARTK_INFO is a pointer variable indicating the source of the information. There are three main reasons why two seemingly redundant variables like AGEFJOB and EINSTIEG_ARTK are both included in BIOJOB.

a) EINSTIEG is based on a more clear and consistent definition of what a first labor market entry is. Here the first labor market entry is conceptualized as the entry in the first job after the completion of (secondary and tertiary) education and apprenticeship. AGEFJOB, though, captures labor market entries at very different stages of the educational and employment biography. One reason for this is that it largely relies on a self-assessment of what a labor market entry is. In the Biography Interview all respondents are asked when they first started to work and this leads to very diverse self-reported labor market entries ranging from the first side-job in high school to the first full-time stable employment matching the own professional field.

b) As described above for people who have never been employed at the time point of the Biography Interview the very first observed labor market entry from the spell-data PBIOSPE is used for the generation of the "age at first job"-variable. This also leads to inconsistent labor market entries since this generating strategy often captures student side-jobs.

c) Additionally, EINSTIEG refers to the earliest yearly measurement after the transition. Although this year is, due to panel structure and interview date, not necessarily to the year of entry, only this strategy allows a clear assignment of covariates from the yearly measurement to the labor market entry.

Among the several plausible concepts and operationalization of labor market entries with SOEP Data (among them AGEFOB), we consequently hold this indicator particularly suited for scholars who want to study the impact of labor market institutions on early career outcomes. By not being focused on first full-time or standard employment, but also regarding shifts into atypical employment as labour market entry, employment biographies spanned from this entry point capture the uncertainties and instabilities associated with the early career phase. At the same time, side-jobs or apprenticeships are explicitly assigned to the educational phase and excluded from the concept of labor market entry. A detailed description of the generation process of EINSTIEG_ARTK is given in the respective documentation file *Introduction to the Variable EINSTIEG.*

EINSTIEG_PBIO/EINSTIEG_PBIO_INFO

The variable EINSTIEG_PBIO provides the year of the entry into the working force. It is primarily based on information found in the spell dataset PBIOSPE and founded on the same concept as EINSTIEG_ARTK. EINSTIEG_PBIO_INFO is a pointer variable indicating the source of the information. The motivation behind another operationalization of job entry is straight-forward: Using the algorithm of EINSTIEG_ARTK only the job market entry-years of less than 7500 respondents in 2015 can be reconstructed. This low number is explained by the fact that to be identified by our algorithm the beginning of the job (and the end of the educational) biography of a SOEP-participant has to be part of the ARTKALEN-dataset. This is only the case if one became part of the survey in late youth or early adolescence and did not leave the sample before a first employment could be observed, so only for a fraction of first jobs as defined by EINSTIEG_ARTK dates can be estimated.

To offer a compromise between the problems of AGEFJOB and the few observations reconstructed by EINSTIEG_ARTK a third variable was created: EINSTIEG_PBIO, which instead of using information from ARTKALEN employs the dataset PBIOSPE, which includes spell-data gathered from the retrospective activity calendar which is part of the biography questionnaire. For understandable practical reasons though this data is just available on a yearly basis and not a monthly one like it is the case with ARTKALEN. The implied loss of granularity induces a potential higher risk of misclassifications compared to EINSTIEG_ARTK while enabling us to reconstruct job entries for a vastly higher amount of respondents, namely almost everyone who ever filled out the biography questionnaire. Still the potential use of EINSTIEG_PBIO compared to EINSTIEG_ARTK is much more restricted as again for most identified first jobs which fall in the time frame before the person became part of the sample and whose information deviates from agefjob there is just no further information available at all. A more detailed description of the generation process of EINSTIEG_PBIO is given in the respective documentation file *Introduction to the Variable EINSTIEG*.

NOJOB

The underlying question for the variable NOJOB is ‘I have never been employed up to this date’. This variable has the label ‘never been employed until the date of the interview” (1).

If NOJOB has a missing value, in general there should exist AGEFJOB information, for special cases, see above. Due to the lack of a comparable question in the Youth Questionnaire, respondents of this questionnaire are given the value (1) as long as no consistent AGEFJOB information is available.

STILLFJ

This variable is based on the question ‘Are you still employed in the same job and at the same place?’. It applies only to LELA respondents who do not state ‘I have never been gainfully employed’ and whose biography interview was after 2000.

Value labels:

(1) Yes
(2) No

FULLTIME

The FULLTIME-variable is used to indicate, whether the first job of a person was a full-time or a part-time job. The value labels are

(0) part-time job or marginal employment

(1) full-time job.

This variable is generated out of the file PBIOSPE for all respondents. For persons with first job information stemming from the Biography Questionnaires, FULLTIME possibly does not refer to the declared first job if PBIOSPE does not contain the respective job spell (i.e. due to item non response or incomplete answering of the activity biography within the Biography Questionnaire).

OCCFJOB

The variable OCCFJOB provides information on the occupational position at the first job. Due to different versions of the questionnaires in the GSOEP´s different samples we face some difficulties. Table 1 gives an overview.

Table 1: Number of Possible Values for Occupational Classifications in the First Job

	Farmers (not self-employed)	Blue Collar Workers	Self-employed	White Collar Workers	Civil Servants
Sample A, B (84-95)	-	5	5	5	4
Sample C (90-95)	4	5	5	4	4
Sample D (94/95)	4	5	5	4	4
Sample A,B,C,D (96)	-	3	4	3	4
Sample A,B,C,D (97-99), E (99)	-	3	4	4	4
Sample A,B,C,D,E (00)	-	3	6	4	4
Sample A,B,C,D,E,F (01)	-	3	10	4	4
Sample A,B,C,D, E,F (02)	-	5	10	6	4
Sample A,B,C,D, E,F,G(06),H(06), I(10),J(11),K(12)	-	5	10	6	4

Facing these differences we decided to standardise the occupational classification. Only four types of occupational status were taken into account: blue collar workers, white collar workers, civil servants, and self-employed. The group ‘Farmers’ is included in the blue collar worker group.

The potential value labels for OCCFJOB are:

(1) blue collar worker
(2) self-employed
(3) white collar worker
(4) civil servant

Further details are provided by the variables FJBLUE (for blue collar workers), FJSELFE (self-employed), FJWHITE (white collar workers), and FJCIVS (civil servants). Table 2 shows the number of possible values.

Table 2: Number of Possible Values for the subcategories of the variable OCCFJOB

	FJBLUE	FJSELFE	FJWHITE	FJCIVS
Sample A,B,C,D, E,F,G,H,I,J,K(12)	9	4	7	4

Due to the fact that the PBIOSPE-file is used for the coding of AGEFJOB in certain cases (see above) there is less information on OCCFJOB than on AGEFJOB.

FJBLUE

The FJBLUE variable provides detailed information on the first occupational status if the person was a blue collar worker. Certain value labels are only given for certain samples, because of the already mentioned differences in the questionnaires.

The following value labels are assigned:

(10) un- and semiskilled farmers (sample C/D)
(11) unskilled worker
(12) semiskilled worker
(20) skilled worker
(30) farmers (sample C/D) being foreman or master craftsman
(31) foreman (sample A/B)
(32) foreman (sample C/D)
(40) master craftsman
(41) farmers (sample C/D) in middle and higher management

FJSELFE/FJSEFSIZ

The FJSELFE variable provides detailed information on the first occupational status if the person was self-employed. FJSEFSIZ gives the number of employees in the respondent's firm. Again there are differences due to the different versions of questionnaires.

The following value labels are assigned:

(10) independent farmer
(20) free lances, self employed academics
(30) other self employed workers
(40) helping within family business

FJSEFSIZ has the following value labels:

(10) number of employees ≤ 9 (all subsamples (see exceptions for samples C/D), up until wave M)
(11) no co-workers (all subsamples, from wave R on)
(12) number of co-workers 1-9 (all subsamples, from wave N on)
(20) number of employees > 9 (all subsamples (see exceptions for samples C/D))
(30) number of employees ≤ 10 (sample C (waves I to L) / D (waves K to L), only if info drawn from biography questionnaire)
(40) number of employees > 10 (sample C (waves I to L) / D (waves K to L) , only if info drawn from biography questionnaire)

FJWHITE

FJWHITE gives detailed information on persons, who were first employed as white collar workers. The subvalues of unskilled labour without degree (21), or with degree (22) are, due to uncomparable values in the LELA-files, only drawn from the $P-Files. (Beginning with BIOJOB 2004).

Potential value labels:

(10) industrial foreman
(20) employee / unskilled labour
(21) same as (20), but without degree
(22) same as (20), but with degree
(30) employee / skilled labour
(40) employee / professional labour
(50) employee / managerial labour

FJCIVS

FJCIVS provides detailed information on first employment as a public servant.

The following value labels occur:

(10) low level civil servant
(20) middle level civil servant
(30) high level civil servant
(40) executive civil servant

ISCO88, STBA EGP, ISEI, MPS, SIOPS

These variables – job classifications and different prestige scores – concerning in each case the first job but are not generated within this file and therefore they are not described within this documentation.

REQEDUC

REQEDUC provides information about the required education for the first job. This information has been asked in the Biography Questionnaire for the first time in the year 2001, but comparable information are gathered by the Individual Questionnaire in all waves.

For all respondents having their first job subsequent to their biography interview, information is drawn out of the generated file $PGEN. Neither respective variables in $P nor those in

$PGEN provide full information for all waves. In both data sources no differentiation is made between vocational college degree and university degree. As $PGEN info is equally coded in all waves, it is preferred to $P info.

Potential value labels:

(10) no training
(20) completed vocational training
(30) vocational college or university degree
(31) vocational college degree
(32) university degree

CIVILSFJ

CIVILSFJ indicates if the first job was assigned to the civil service or not. This information has been asked in the 2001 Biography Questionnaire for the first time

For respondents having their first job subsequent to their biography interview, information is drawn out of the generated file $PGEN where this information is provided since the first wave in 1984.

The following value labels occur:

(1) Yes
(2) No

NACEFJ

NACEFJ provides information about the industrial sector of the first job according to the branch classification NACE. This variable is not generated within this file. The description of its value labels is therefore not part of this documentation.

OCCMOVE

The variable OCCMOVE is based on the question ‘Did you change your occupation and if you did, more than once?’. Information stems from the year of the biography interview. For respondents of the Youth Questionnaire as well as persons having their first job after the biography interview no information is available.

Labels of **OCCMOVE:**

(1) never changed occupation
(2) changed once
(3) changed more than once

AGEATMV

This variable is based on the question ‘If you changed your occupation, how old were you at the most recent change?’. Information stems from the year of the biography interview. For respondents of the Youth Questionnaire as well as persons having their first job after the biography interview no information is available.

CURREMPL

This variable is based on the question ‘ Are you gainfully employed at the current time?’. The question applies only to LELA respondents who do not state ‘I have never been gainfully employed’ or ‘Still employed in the first job’. This question has been asked in 1994 for the first time.

Value labels:

(1) Yes
(2) No

YEARLAST

This variable is based on the question ‘When was the last time you were gainfully employed?’. The question applies only to LELA respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

SCOPELJ

SCOPELJ indicates if the last job was a full time or part time job.

Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time. For youth respondents no information is available.

Value labels:

(1) full-time employed
(2) part-time employment
(3) marginal / irregular employment

CIVILSLJ

CIVILSLJ indicates if the last job was assigned to the civil service or not.

Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time. For youth respondents no information is available.

The following value labels occur:

(1) Yes
(2) No

NACELJ

NACELJ provides information about the industrial sector of the last job according to the branch classification NACE. The respective question applies only to respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

This variable is not generated within this file. The description of its value labels is therefore not part of this documentation.

OCCLJOB

The variable OCCLJOB provides information on the occupational position at the last job. The respective question applies only to respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

Due to different versions of the questionnaires in the GSOEP´s different samples we face some difficulties. Table 3 gives an overview:

Table 3: Number of Possible Values for Occupational Classifications in the Last Job

	Farmers (not self-employed)	Blue Collar Workers	Self-employed	White Collar Workers	Civil Servants
Sample A,B (94/95)	-	5	5	5	4
Sample C,D (94/95)	4	5	5	4	4
Sample A,B,C,D (96-99), E (99)	-	5	5	6	4
Sample A,B,C,D,E (00)	-	5	6	6	4
Sample A,B,C,D, E,F (01/02)	-	5	10	6	4
Sample A,B,C,D, E,F,G(06),H(06),I(10), J(11)	-	5	10	6	4

Facing these differences we decided to standardise the occupational classification. Only four types of occupational status were taken into account: blue collar workers, white collar workers, civil servants, and self-employed. The group 'Farmers' is included in the blue collar worker group.

The potential value labels for OCCLJOB are:

(1) blue collar worker
(2) self-employed
(3) white collar worker
(4) civil servant

Further details are provided by the variables LJBLUE (for blue collar workers), LJSELFE (self-employed), LJWHITE (white collar workers), and LJCIVS (civil servants). Table 4 shows the number of possible values.

Table 4: Number of possible values for the subcategories of the variable OCCLJOB

	LJBLUE	LJSELFE	LJWHITE	LJCIVS
Sample A,B,C,D, E,F,G,H,I,J(84-11),K(12)	9	4	7	4

LJBLUE

The LJBLUE variable provides detailed information on the last occupational status if the person was a blue collar worker. Certain value labels are only given for certain samples, because of already mentioned differences in the questionnaires.

The following value labels are assigned:

(10) un- and semiskilled farmers (sample C/D)
(11) unskilled worker
(12) semiskilled worker
(20) skilled worker
(30) farmers (sample C/D) being foreman or master craftsman
(31) foreman (sample A/B)
(32) foreman (sample C/D)
(40) master craftsman
(41) farmers (sample C/D) in middle and higher management

LJSELFE/LJSEFSIZ

The LJSELFE variable provides detailed information on the last occupational status if the person was self-employed. LJSEFSIZ gives the number of employees in the respondent's firm. Again there are differences due to different versions of questionnaires.

The following value labels are assigned:

(10) independent farmer
(20) free lances, self employed academics
(30) other self employed workers
(40) helping within family business

LJSEFSIZ has the following value labels:

(10) number of employees ≤ 9 (all subsamples (see exceptions for samples C/D), until wave M)
(11) number of co-workers = 0 (all subsamples, from wave N on)
(12) number of co-workers 1-9 (all subsamples, from wave N on)
(20) number of employees > 9 (all subsamples (see exceptions for samples C/D))
(30) number of employees ≤ 10 (sample C (waves I to L) / D (waves K to L), only if info drawn from biography questionnaire)
(40) number of employees > 10 (sample C (waves I to L) / D (waves K to L) , only if info drawn from biography questionnaire)

LJWHITE

LJWHITE gives detailed information on persons, who were last employed as white collar workers. The values (21) and (22) are drawn from the BIOLELA-File and from the $P-files.

Potential value labels:

(10) industrial foreman
(20) employee / unskilled labour
(21) same as (20), but without degree
(22) same as (20), but with degree
(30) employee / skilled labour
(40) employee / professional labour
(50) employee / managerial labour

LJCIVS

LJCIVS provides detailed information on last employment as a public servant.

The following value labels occur:

(10) low level civil servant
(20) middle level civil servant
(30) high level civil servant
(40) executive civil servant

INTEDUC1 to INTEDUC4

The variables INTEDUC1, INTEDUC2, INTEDUC3, and INTEDUC4 provide information on the educational degree or the vocational/professional training a respondent intends to complete in the future, asked at the time of the biography interview. We create these four variables since multiple answers are explicitly allowed in the questionnaire. The intended education is stored with respect to the hierarchy given by the questionnaire, i.e., the highest degree is placed in INTEDUC1. For example, a person intending to finish an apprenticeship (1) and university (7) would have INTEDUC1 = 7 and INTEDUC2 = 1. Since this question has been asked for the first time in 1996, we do observe a large number of missing values for INTEDUC1 to INTEDUC4.

(1) apprenticeship
(2) full-time vocational school
(3) technical school
(4) education as a civil servant
(5) accredited professional school
(6) technical or professional college
(7) university

General remark:

Some persons answered more than once the Biography Questionnaire (but this occurs very rarely). The data-set BIOJOB contains only information from one Biography Questionnaire, in most cases the earlier one.

3 Steps of Coding

1. Creating a dataset using the data concerning all aspects of the job biography (working force entry, position, etc.) drawn from BIOLELA, MLELA, NLELA, OLELA, PLELA, QLELA, RLELA, SLELA, TLELA, ULELA, VLELA, WLELA, XLELA, YLELA, ZLELA, BALELA, BBLELA, BCLELA (internal DIW files with biographical information up to wave BB), QJUGEND, RJUGEND, SJUGEND, TJUGEND, UJUGEND, VJUGEND, WJUGEND, XJUGEND, YJUGEND, ZJUGEND, BAJUGEND, BBJUGEND, BCJUGEND (internal DIW youth biography files), QP, RP, SP, TP, UP, VP, WP, XP, YP, ZP, BAP, BBP, BCP (needed for consistency checks with respect to the youth biography files).
2. Using the PBIOSPE-data to retrieve spell information during the first occupation.
3. Using PPFAD for personal data (year of birth, sex, sample).
4. Using several files containing generated information about job classification (ISCO), prestige scores and industry sector classification (NACE) concerning the first job.
5. Combining all data concerning the employment biography into a new data file BIOJOB, where priority is set as mentioned above.
6. Coding of AGEFJOB. (for details, see above)
7. Setting the pointer variable AGEINFO indicating the source of the information of AGEFJOB. (for details, see above)
8. Excluding one value for respondents, who stated to have two occupational positions in their first job. Exclusion based on consistency checks.
9. Assignment of the variable OCCFJOB, with respect to the different versions of the questionnaire. Possible value labels: FJBLUE, FJSELFE, FJWHITE, FJCIVS.
10. Definition and assignment of new value-labels for the sub-category FJBLUE, nine labels possible, for details see above.
11. Definition and assignment of new value-labels for the sub-category FJSELFE, four labels possible, for details see above.
12. Definition of the variable FJSEFSIZ, indicating the numbers of employees.
13. Definition and assignment of new value-labels for the sub-category FWHITE, seven labels possible, for details see above.
14. Definition and assignment of new value-labels for the sub-category FJCIVS, four labels possible, for details see above.
15. Coding of the variables REQEDUC and CIVILSFJ.

16. Coding of the variables INTEDUC1 to INTEDUC4.
17. Computing the age at the most recent change of occupation if necessary.
18. Check of consistency: Does information about the age at the most recent change of occupation make sense? If inconsistencies appear, the value is set to a missing value.
19. Assignment of value labels for the variables specifying the last job:
20. Definition and assignment of value labels of the variable CURREMPL indicating if a respondent is gainfully employed at the time of the biography interview.
21. Specification of the year of last employment (YEARLAST).
22. Coding of the variables SCOPELJ and CIVILSLJ.
23. Excluding one value for respondents, who stated to have two occupational positions in their last job. Exclusion based on consistency checks.
24. Assignment of the variable OCCLJOB, with respect to the different versions of the questionnaire. Possible value labels: LJBLUE, LJSELFE, LJWHITE, LJCIVS.
25. Definition and assignment of new value-labels for the sub-category LJBLUE, nine labels possible, for details see above.
26. Definition and assignment of new value-labels for the sub-category LJSELFE, four labels possible, for details see above.
27. Definition of the variable LJSEFSIZ, indicating the numbers of employees.
28. Definition and assignment of new value-labels for the sub-category LJWHITE, seven labels possible, for details see above.
29. Definition and assignment of new value-labels for the sub-category LJCIVS, four labels possible, for details see above.
30. Collecting of job information for people with AGEINFO = 3, 7 or 8, if possible.
31. Collecting of job information for people with AGEINFO = 12, 14 or 16, if possible.
32. Coding of the variable FULLTIME.
33. Definition of missing values for all variables.
34. Hand-editing of inconsistencies between different variables.
35. Final listing
36. Definition and assignment of new value-labels for the sub-category LJCIVS, four labels possible, for details see above.
37. Collecting of job information for people with AGEINFO = 3, 7 or 8, if possible.
38. Collecting of job information for people with AGEINFO = 12, 14 or 16, if possible.

39. Coding of the variable FULLTIME.

40. Coding of the EINSTIEG variables

41. Definition of missing values for all variables.

42. Hand-editing of inconsistencies between different variables.

43. Final listing