Maximilian Trommer | Thorsten Euler | Anne Weber | Henrike Schmidtchen

Documentation of the auxiliary variables for the evaluation of information on gainful employment

Notes on the origin and use of the auxiliary variables in the data set for the 2009 Graduate Panel

fdz.DZHW.
Forschungsdatenzentrum
für Hochschul- und Wissenschaftsforschung

Project team members
Maximilian Trommer
Thorsten Euler
Dr. Anne Weber
Henrike Schmidtchen

Table on Contents

List of Tables/Figures

1 General Hints

In empirical surveys, contradictions in respondents' statements are the rule rather than the exception, even with a carefully designed survey instrument. Case-by-case exclusion due to implausible information is often the best solution but should be minimised due to the effects on inferential statistical conclusions. The correct handling and interpretation of contradictory values always depend on the research questions of the evaluating researchers. Nevertheless, indications of the correct interpretation of values by primary research projects can be useful to use the existing process knowledge and to standardise and simplify the evaluations. In addition, primary research projects can usually draw on extended information that is not published in data sets for reasons of data protection but can support some interpretations.

For the 2009 Graduate Panel, this appears to be particularly useful for the information on the employment of the respondents. Within a wave, the employment history was queried on the one hand via an employment table[1] and on the other hand - in addition to other activities - recorded in a calendar[2] (cf. Data and Methods Report, sections 2.1 and 6.5). In addition, more detailed information on current or most recent employment was requested. Since many employment episodes continue beyond the time of the survey, they are also recorded again in the follow-up wave and intentionally overlap with the information from the previous wave.

This form of recording is helpful to include as much information on employment as possible, but it also makes the data complex and poses certain difficulties. For this reason, several auxiliary variables were generated for the data set of the 2009 graduate panel, for each of which this handout explains the reasoning, the procedure for creating them and tips on how to interpret them correctly. The auxiliary variables are available primarily for the third wave, but also in part for the second wave (cf. Sections 2 and 3).

Even if the critical reflection of our data is never unnecessary for the evaluator, we hope to at least provide some assistance.

---

[1] cf. question 5.4 in wave 1, question 4.5 in wave 2 and page p304-p304o in wave 3
[2] cf. question 4.1 in wave 1, question 1.5 in wave 2 and page p103 in wave 3

2   Reference Number Current/last Digit: bocc240_g1 (wave 2) and cocc240_g1 (wave 3)

2.1   Argumentation

In order to be able to combine the detailed information on the current or last employment with the matching employment episode in the employment table in an uncomplicated way, a so-called reference number was created. This indicates the number of the episode in the employment table which certainly or very probably describes the current or last activity in question. As a rule, this is the only episode still in progress, but occasionally also one of several or a past episode.

2.2   Procedure

First, the reference number was automatically assigned if only one current episode or generally only one episode exists. Then, if available, the last episode to end was also automatically assigned. In addition, an automatic assignment was made in the case of two ongoing episodes, insofar as the working hours and number of hours suggested this.

Finally, a case-based assignment of the episode was made. The reference number was assigned if the assignment was clear or very probable. In case of doubt, no reference number was assigned.

2.3   Hints for Use

None.

3   Adjustment Figure: bocc_241m_g1 (wave 2) and cocc24*m_g1 (wave 3)

3.1   Argumentation

As a rule, at least one, sometimes even more, episodes that were indicated in the employment tableau find a counterpart in episodes of the employment tableau of the previous wave. To facilitate the allocation between the waves, a so-called matching number was formed for each tableau episode in the second or third wave, which reflects the number of the (presumably) corresponding employment episode from the previous wave.

3.2   Procedure

First, an attempt was made to make an automated assignment based on temporal and content-related criteria. In the case of clear combinations (especially exactly one episode with temporal and content-related correspondence) or very high correspondence, the corresponding matching number was assigned. If no matching was possible, the matching number was subsequently completed on a case-by-case basis, as far as possible. If no episode from the previous wave could be assigned clearly enough, the variable takes the value 0.

3.3   Hints for Use

The matching figures are relevant as soon as acquisition episodes are to be evaluated in combination across waves. In this case, according to the matching figure, episodes belonging together from different waves can be considered combined in terms of content. Since the episodes must be similar in terms of content, but by no means identical, the correct approach depends on the research question. For example, an employment relationship may still be temporary at the end of the second wave, but in the corresponding episode of the third wave it may be no longer temporary and the respondent may have indicated this final status. The time of the survey forms a natural censorship here. It would certainly be wrong to set the de-termination at the time of the survey, but it would be equally wrong to evaluate the complete episode across waves as temporary or determined. Accordingly, special care should be taken when merging.

When assigning the matching figures, a certain degree of uncertainty on the part of the respondents was tolerated, e.g. slight deviations in the starting months or professional position. It is up to the users of the data to decide which of the data should be used to harmonise the episodes.

4   Type of Activity: cocc24*o_g1

4.1   Argumentation

In principle, each employment episode in the employment table should also be reflected in the calendar. In contrast to the calendar, however, no activity types are selected by the respondents in the employment tableau, such that these are only partly derived from information on employment. In order to compare tableau episodes and calendar episodes, a variable was created that reflects a type of activity from the calendar on the basis of information from the employment.

4.2   Procedure

First, unique values were assigned from the information provided on occupational status and employment relationship in the employment tableau. The following list shows the unique combination:

Table 1: Information from the employment table

Information from the employment table							
		Type of employment					
	Missing	[1] unlimited	[2] limited	[3] Training / Traineeship	[4] Fee/work contract	[5] Self-employed	[6] Others
Occupational status							
Missing	0. unklar	0. unklar	0. unklar	0. unklar	2. werk	4. selb	0. unklar
[1] Senior employee(s)	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[2] Sci. qualified employee(s) with a middle mgmt. function	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[3] Sci. qualified employee(s) without a mgmt. function	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[4] Qualified employee(s)	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[5] Executive employee(s)	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[6] Self-employed in liberal professions	4. selb	0. unklar	0. unklar	0. unklar	0. unklar	4. selb	4. selb
[7] Self-employed entrepreneur	4. selb	0. unklar	0. unklar	0. unklar	0. unklar	4. selb	4. selb
[8] Self-employed person with fee/work con-	2. werk	0. unklar	0. unklar	0. unklar	2. werk	0. unklar	2. werk

tract							
[9] Civil servant in the higher service	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[10] Civil servant in the upper grade	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[11] Civil servant in the ornary/medium service	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[12] Temporary civil servant (e.g. trainee teacher)	1. ref	0. unklar	0. unklar	1. ref	0. unklar	0. unklar	1. ref
[13] Skilled worker with apprenticeship	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[14] Unskled/skilled worker	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[15] assisting family members	5. nsb	5. nsb	5. nsb	0. unklar	0. unklar	0. unklar	5. nsb
[16] Trainee teacher, recognition trainee	1. ref	0. unklar	0. unklar	1. ref	0. unklar	0. unklar	1. ref

Table 2: Types of activity in calendar

Types of activity in calendar	
ref	Traineeship, recognition internship
werk	Contract for work and labour, fee work
job	Jobbing
selb	Self-employment
nsb	Non-self-employed gainful employment (e.g. as employee or civil servant)
prak	Internship
ausb	Vocational training, retraining, volunteering
trai	Trainee
prof	Junior professorship, tenure-track professorship, junior research group leadership, habilitation

Subsequently, the episodes from the employment table were automatically compared with the calendar and the activity information was added, as far as it was clear. The uniqueness is subject to a very complex comparison: It is checked whether ...

▪ the episode in the calendar completely covers the tableau episode in terms of time,

▪ the episode in the calendar is not covered by another tableau episode with plausible content

▪ and the episode in the calendar at least does not contradict the contents in the tableau.

For example, a tableau episode with the employment relationship "training/trainee" is set to the activity type "ausb" or "trai" if the corresponding calendar episode had vocational training or trainee position as an activity type and the tableau episode does not already have the activity type "ref" (due to the professional position).

Finally, missing values were added in a case-based plausibility check, if clear or obvious.

4.3   Hints for Use

The variable can be used to facilitate the linking of data from the employment table with data from the calendar and, accordingly, the integrated evaluation of the history data. As a rule, a corresponding calendar episode (matching taet variable) should also belong to an employment episode. However, since this is not always the case - either due to different interpretations of the respondents regarding the selection of episode types in the calendar or an unexpected combination of the contents that form the taet variable - we strongly recommend paying attention to the explanations on the flag variable in the following section 5.

5   Flag Variable for Activity Type Tableau/Calendar: cocc24*p_g1

5.1   Argumentation

In principle, each acquisition episode from the tableau should also be found in the Calender. However, many tableau episodes do not have a matching parallel calendar episode, or even contradictory episodes. This makes interpretation difficult or even impossible. In order to facilitate and standardise the evaluation, a flag variable was inserted per acquisition episode in the tableau, which indicates the type of parallel calendar activity.

5.2   Procedure

To create the flag variable, the following steps were roughly taken, following on from the previously generated auxiliary variable for the type of activity (see section 4):

1. Insofar as the acquisition episode from the tableau is matched by an undoubtedly matching episode in the calendar, the value 0 (no contradiction between the calendar and the tableau) was assigned.

2. Otherwise, an attempt was first made to automatically fill the flag variable as far as possible with values that represent unambiguous (but unsuitable) assignments. As a rule, this was successful if the acquisition episode in the tableau was completely covered by exactly one parallel episode in the calendar.

3. Subsequently, further assignments were made in the case-based plausibility check and episodes were assigned on the basis of the time periods and contents, as far as this made sense. There was certainly a margin of discretion. Thus, in the case of several, unsuitable calendar episodes, the relevant and typical one was usually chosen. For example, parallel parental leave is more relevant for us than further training.

5.3   Hints for Use

| Table 3: Calendar/Tableau |

Tableau								
Calendar	1 ref	2 werk	3 job	4 selb	5 usre	6 prak	7 ausb	8 trai	10 prof

	1	2	3	4	5	6	7	8	9
1 job	1	2	3	4	5	6	7	8	9
2 nsb	10	11	12	13	14	15	16	17	18
3 selb	19	20	21	22	23	24	25	26	27
4 werk	28	29	30	31	32	33	34	35	36
5 ref	37	38	39	40	41	42	43	44	45
6 trai	46	47	48	49	50	51	52	53	54
7 prak	55	56	57	58	59	60	61	62	63
8 stud	64	65	66	67	68	69	70	71	72
9 pro-mo	73	74	75	76	77	78	79	80	81
10 habil	82	83	84	85	86	87	88	89	90
11 ausb	91	92	93	94	95	96	97	98	99
12 wb	100	101	102	103	104	105	106	107	108
13 arlo	109	110	111	112	113	114	115	116	117
14 fami	118	119	120	121	122	123	124	125	126
15 eltz	127	128	129	130	131	132	133	134	135
16 sons	136	137	138	139	140	141	142	143	144

Legend: Correct standard case, cocc24*p_g1 not assigned

It is advisable to consider the flag variable for the type of activity for evaluations of the acquisition episodes with or without calendar. The correct interpretation of the episodes depends on the research questions. The flag variable only provides an indication of the inconsistency of an assignment, but no information on the correct interpretation. Especially with regard to career breaks, the flag variable on the type of activity should never be evaluated alone, but always in combination with the flag variables on parental leave and unemployment (cf. section 6).

6 Flag Variable on Career Breaks Parental Leave and Unemployment: cocc24*q_g1 and cocc24*r_g1

6.1 Argumentation

In most cases, the parental leave and unemployment periods indicated in the calendar result in parallel interruptions in employment in the employment table. However, this is not always the case, because in some cases parallel employment was also indicated. Since there are clear indications that the respondents indicated their respective employment constellations parallel to parental leave and unemployment phases very differently, there are certain ambiguities that entail a certain scope for interpretation.

For example, only full-time employment during parental leave and unemployment is not possible in principle.

On the one hand, the indication of continuous full-time employment during parental leave could be interpreted as a continuing contractual relationship, but a dormant employment relationship, so that the gainful employment is de facto not exercised. On the other hand, it would also be possible that the working hours are reduced within the framework of the existing full-time job and a part-time job is exercised parallel to parental leave (up to 30 h/week, so-called parental part-time). For unemployed persons, a continuation of the contractual relationship is rather unlikely, but for seasonal workers it is theoretically conceivable.

Even more difficult is the interpretation of stated part-time employment that continues during phases of parental leave or unemployment. Thus, for parental leave, analogous to full-time employment above, this could be interpreted as a representation of dormant employment, but equally as part-time parental leave. Reduced part-time employment parallel to unemployment (up to 15 hours per week) would also be possible.

Without further information from the respondents, it is thus not automatically apparent whether a stated employment episode parallel to parental leave represents dormant employment, a reduction in working hours or part-time employment.

Since the sensible handling of the problem described is strongly dependent on the research question being worked on, the introduction of flag variables that assign a code with the underlying data constellation to each episode in the employment tableau serves as a solution to this.

6.2   Procedure

The procedure for parental leave and unemployment is almost identical, differing only in the number of hours possible in parallel and the subsequent interpretation.

First, without exception, all episodes of employment of the period covered in the third wave were separated in the tableau into before, during and after a possible interruption.[3] In the case of several interruptions, they were also separated several times.

Since changes within a month are also possible, they were only separated and flagged if there was an overlap of more than one month, unless the interruption lasted only one month in total (see chart).

Figure 1: Interruption

		1	2	3	4	5	6	7	8	9	10	11	12	Splits
	Tableau: Acquisition episode		0									0		
Interruption	Calendar: Interruption													0
	Calendar: Acquisition Episode													
	Tableau: Acquisition episode		0			1				0				
Simply Enclosed	Calendar: Interruption													2
	Calendar: Acquisition Episode													
	Tableau: Acquisition episode		0				11*				0			
Double Enclosed	Calendar: Interruption													2
	Calendar: Acquisition Episode													
	Tableau: Acquisition episode		0		11*		1		11*		0			
Overlapping	Calendar: Interruption													4
	Calendar: Acquisition Episode													
	Tableau: Acquisition episode		0		11*		0							
Enclosed month	Calendar: Interruption													2
	Calendar: Acquisition Episode													

* = Codes 11, 21, 31 or 41 Depending on the hours of the acquisition episode in the tableau (<=30 / >30 / . / 0)

The specified activity characteristics such as occupational position, working hours, location, etc. remained unchanged. The respondents' original episodes remain identifiable in a separate variable (see section 7).

Finally, the episodes were subjected to a verification scheme based on of the parallel activity in the calendar and the information on working hours and the corresponding values were stored coded in separate variables for parental leave and for unemployment (see Charts & Tables).

---

[3] It follows that the number of episodes in the data set was expanded to up to 16 episodes, although only a maximum of 15 episodes could be indicated in the questionnaire.

Figure 2 Parental Leave

Flag ariable	Tableau	Calendar		Interpretation
	Employment	Parental leave	Employment	
0	yes	no	yes	Valid acquisition episode, normal case
1	yes	yes	no	Presumably interruption due to exclusive parental leave, dormant employment relationship with the ongoing contract; interpretation depending on the research question
11	yes (cocc24*h_g1 <=30h.)	yes	yes	Theoretically valid employment episode (parental part-time work), but fraught with uncertainty; interpretation depending on the research question
21	yes (cocc24*h_g1 >30h.)	yes	yes	No valid employment episode, as parental part-time work is only possible up to 30 hours per week. 30 hours per week possible; interpretation depending on the research question
31	yes (cocc24*h_g1 ==.)	yes	yes	Uncertain assessment
41	yes (cocc24*h_g1 ==0)	yes	yes	Ok, exclusive parental leave

Figure 3 Unemployment

Flag ariable	Tableau	Calendar		Interpretation
	Employment	Unemployment	Employment	
0	yes	no	yes	Valid acquisition episode, normal case
1	yes	yes	no	Tableau episode unlikely
11	yes (cocc24*h_g1 <=15h.)	yes	yes	Unemployment & valid, parallel marginal employment
21	yes (cocc24*h_g1_g1 >15h.)	yes	yes	No valid employment episode, as marginal employment is only possible up to 15 hours per week
31	yes (cocc24*h_g1_g1 ==.)	yes	yes	Assessment uncertain
41	yes (cocc24*h_g1_g1 ==0)	yes	yes	Ok, exclusive unemployment

As far as possible, this was done automatically (especially for clear calendar episodes with clear interruptions). In many complex cases, however, a case-based assignment had to be made.

Each acquisition episode thus has a unique value on the flag variables.

6.3   Hints for Use

The two flag variables provide information on the constellation of the tableau and calendar data. If acquisition episodes are to be evaluated independently of acquisition interruptions (e.g. contractual situation), both flag variables can usually be ignored. However, if acquisition interruptions are relevant or even at the core of interest, we strongly recommend using the flag variables.

In this case, it is also advisable to additionally consider the flag variable on type of activity (cf. section 5), which may overlap with the flag variables on parental leave and unemployment.

6.4   Hints for Use of the Parental Leave Flag Variable

The correct interpretation of the acquisition episodes always depends on the research question, but as a rule, we recommend evaluating episodes with the flag value

- 0 as a valid acquisition episode
- 1 to be evaluated as an acquisition interruption
- 11

- To be evaluated as a valid employment episode, if the episode originally starts during parental leave or during parental leave.
- Not to be evaluated if the core of interest is working time or parental leave.
- 21 to be excluded from the evaluations.
- 31 to be excluded from the evaluations
- 41 to be evaluated as a break in employment

6.5   Hints for Use of the Unemployment Flag Variable

The interpretation of employment episodes parallel to unemployment is very similar, but two major differences should be noted:

1. Parallel to unemployment, an activity with a maximum of 15 hours per week is possible.

2. Unemployment, contrary to parental leave, is usually not a dormant employment relationship, an ongoing employment contract (with more than 15h) is not possible in parallel.

Accordingly, we recommend episodes with the flag value

- 0 as a valid episode of employment.
- 1 to be evaluated as an acquisition interruption
- 11 to be evaluated as a valid acquisition episode
- 21 to be excluded from the evaluations
- 31 to be excluded from the evaluations
- 41 to be evaluated as an acquisition interruption

7   Original Episode before Separation due to Interruptions in Employment: cocc24*n_g1

7.1   Argumentation

In the course of data processing, the employment episodes mentioned in the employment table were separated into episodes before, during and after the employment interruptions based on the information collected in parallel in the calendar (e.g. due to parental leave) (cf. section 6). In order to be able to trace at any time which episodes were only separated in the course of data processing and which were already submitted separately by the respondent, a variable was created for each episode that contains the original episode number.

7.2   Procedure

The variable was already created before the episodes were separated and always retained in the event of a separation.

7.3   Hint for Use

The variable contains values from 1-15 (maximum number of recorded episodes in the employment table of the third wave). If several episodes of a case contain the same value, this was originally stated by the respondent as one episode and separated by us.

As a rule, the episodes were sorted chronologically. Numerically non-ordered sequences can occur due to separation and re-sorting, but also due to non-chronological information provided by the respondent.

8   Flag Variable monthly gross-to-income: cocc312b_g1

8.1   Argumentation

The questionnaire asks for the gross monthly income. There are always incomes that seem implausible either in themselves or at the latest when compared with the occupation and the number of hours. To facilitate and standardise the evaluation, an additional variable was added that reflects the project team's assessment of plausibility.

8.2   Procedure

The assessments were based on other income information (receipt of allowances), industry, occupation, occupational status, hours, company size, public sector employment, country of work and satisfaction with income. However, most assessments are based only on hours and occupation. Based on this, the information on gross monthly income was divided into the following four categories.

Table 4: Details of the gross monthly income

Details of the gross monthly income	
0	Plausible income Income appears plausible under consideration of employment characteristics and can be interpreted as valid information. However, other measurement errors can of course not be excluded, there is merely no indication of corresponding errors. Incomes that are not considered in more detail also receive this flag value of 0.
1	Presumed annual income Income appears exceptionally high, taking into account the employment characteristics, and is in the range of an expected annual income. As the indication of an annual rather than a monthly income is a known error, we usually recommend taking it and dividing it by 12.
2	Unusual income Income appears to be unusually low or high, taking into account the employment characteristics, without being able to be identified as annual income. However, the value is theoretically still conceivable. We recommend excluding these cases from evaluations.
3	Implausible incomes Incomes appear absurdly low or high when taking into account the employment characteristics and as such implausible. We strongly recommend excluding these cases from evaluations.

8.3   Hints for Use

It is important to emphasise that not all incomes were examined using the categories, but only conspicuously high and low incomes (also for allowances). For a further examination, we still recommend an additional consideration of exceptional gross hourly wages.

9   Source Type Calendar Episode: Source

9.1   Argumentation

In the course of processing the episodic information from the calendar and comparing it with the employment table, inconsistencies or gaps became apparent in some places. This is conceivable due to errors in entering the month-specific calendar information (especially on mobile devices) or different reminders by the respondents for the calendar at the beginning of the questionnaire, the employment tableau and the further academic qualifications in the further course of the questionnaire. In some cases, it was therefore decided to supplement (add) individual months or entire calendar episodes accordingly the variable.

9.2   Procedure

The additions in the calendar were made in the following cases:

- there is a gainful activity in the gainful activity table that is not covered by an episode in the calendar (also not taking into account episodes that do not fit in terms of content),
- there is an academic qualification in the qualification table that is not covered by an episode in the calendar (also not taking into account episodes that do not fit in terms of content),
- a continuous gainful activity in the employment table has a corresponding episode in the calendar, which, however, has a gap (empty month) that has no comprehensible reason (parental leave within the episode, for example, would be a comprehensible reason for an employment gap),
- an employment in the employment table is indicated as continuous, but was not indicated for the survey month, but only up to the month before.

The addition was made automatically in a first step for gainful employment/academic qualification for which no balancing with other parallel episodes in the calendar was necessary (empty calendar). In all other cases, a case-based assessment was made as to whether a supplement was necessary.

9.3   Hints for Use

The additions for the calendar are only a suggestion of the project team and can be revised by the users (analogous to activities added in the employment table). Original and subsequently added episodes have not been harmonised, so that under certain circumstances there may be three episodes (the original before and after the gap, as well as the subsequently added gap), although it may be a continuous episode (e.g. in comparison with the employment table). This applies, for example, to respondents who did not indicate their studies in the calendar during the months of the semester break. Instead of four episodes per year (two study phases, two semester breaks), a continuous study status would be more intuitive.