FDZ·Methodenreport

Methodological aspects of labour market data

Do we measure employment durations correctly?

The case of German administrative employment data

Ursula Jaenichen

Bundesagentur für Arbeit

Do we measure employment durations correctly?

The case of German administrative employment data

Ursula Jaenichen (Institute for Employment Research)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

Contents

Abstract

German social security data on employment are based on compulsory notifications by employers on employees at least once a year. To compute individual employment durations, the employment episodes reported in successive notifications have to be concatenated. Several issues in this computation raise the question of whether the durations are measured precisely. Using IAB data on individual employment biographies, the relevance of these issues is demonstrated empirically.

Zusammenfassung

Die Daten zur sozialversicherungspflichtigen Beschäftigung beruhen auf den Arbeitgebermeldungen, die mindestens einmal im Jahr erfolgen. Um individuelle Beschäftigungsdauern zu berechnen, müssen die in einzelnen Meldungen angegebenen Episoden aneinander gehängt werden. Mehrere unklare Aspekte in der Berechnung werfen die Frage auf, inwieweit die Dauern korrekt gemessen werden. Mit IAB-Daten zu individuellen Erwerbsbiographien wird die Relevanz dieser Aspekte empirisch untersucht.

Keywords

1   Introduction

The use of administrative data in research is fruitful for various reasons. Such data are often available for the universe of a specific population, their reporting is often compulsory and will be done very carefully if the information is linked to financial claims, like e.g. the payment of pensions or wages. As the data result from administrative processes in real time, and events are recorded whenever they occur, recall error which is often a problem in survey data is avoided.

German social security data on employment and wages are based on compulsory notifications on employees legally required from employers at least once a year. These data represent an important source of information both for Official Statistics and labour market research.

Employment relationships often last for years, while notifications report employment periods of at most one year. Therefore, the employment episodes reported in successive notifications have to be concatenated to arrive at total employment durations. There are several issues in the computation of employment durations raising the question of whether these durations are measured precisely.

One issue is the availability of the variable "reason for notification" which is required for every notification and which typically would allow to judge whether an employment relationship does go on or not. It will be demonstrated that the observed values of this variable do not always correspond to the presence or absence of further notifications with the same establishment[1].

As a second issue, the data may be incomplete in some cases. There are hints on discontinuities in the number of notifications in certain labour market districts during the years 1992 to 2000, in one case related to the notifications from a large multi-establishment enterprise, in another case related to the city of Berlin and still in another case again stemming from probably missing notifications of just one establishment.[2] On the individual level, employment biographies sometimes contain gaps lasting exactly one year and followed by a new employment episode in the same establishment as before the gap (Drews et al. 2007).

The third issue concerns the establishment identifier (EID) which is used to distinguish employers. Different EIDs across successive employment notifications of the same worker are interpreted as job changes from one establishment to another. As this identifier might change for reasons other than the birth or the death of an establishment, its use for the generation of job spells can be qualified. We will check whether the additional use of the information on worker flows between firms (Hethey-Maier and Schmieder 2010, 2013) has a major influence on the resulting job durations.

As an appendix to this report, Stata do-files containing the computation of employment durations are available.

---

[1]   Boockmann and Steffes (2005) observe that the reason for notification does not always indicate an end of employment despite of a different establishment identificator on the worker's next spell.

[2]   Internal report (IAB 2016) on the Employment history data, ITM-Datenreport BeH V10.01.00-160816, p24ff.

2 Data

The main data base of this report is a 2 percent sample of the Integrated Employment Biographies (IEB)[3], a combined individual-level data set from various administrative sources. Compulsory employment notifications form a major part of this data base. Whilst our focus is on employment durations of individual workers, the notifications can also be aggregated by establishment to create linked employer-employee data.

A data record corresponding to the notification for an individual worker contains the identifier of the establishment, an identifier of the worker, the type of notification, begin and end of the employment episode reported, socio-economic characteristics of the worker, information on the type of the job (full- vs. part-time), social security status, occupational status. In 2011, there was a major change in the form for notification accompanied by changes in the reporting of some variables. These changes do not concern the analyses here as only data up to 2010 are used.

As a complementary data source and to obtain information on workers' flows between establishments, the IAB's Establishment History Panel (BHP)[4] is used in the version including the years 1975 to 2010. The BHP aggregates all notifications of employment episodes overlapping June, 30 of a respective year into establishment level data. Only establishments with at least one employee liable to social insurance are included. There is no direct information on whether an establishment is new, already existent or has been closed down since the last year. Similarly, spin-offs or take-overs of establishments are not observed directly. Following the approach of Benedetto et al. (2007) and Hethey-Maier and Schmieder (2010, 2013), it is possible to identify these events by measuring year-to-year worker flows for establishments with new EIDs or with EIDS that are not observed any longer. Concentrating on the largest inflow and the largest outflow of workers, mere ID changes (e.g. because of changes in ownership) can be distinguished from the birth or death of an establishment[5] and also from spin-offs or take-overs.

2.1 Generation of job spells

We define employment spells as periods of continuous employment of a worker in the same establishment. As in the BHP, the employment notifications in the IEB contain the EID to identify establishments. For ongoing employment relationships, a notification is required at the end of a year. The end of an employment relationship always requires a notification by the employer. There are notifications which are not relevant to the length of an employment spell, e.g. a change in the worker's occupational status or a change in the worker's sickness insurance

---

3   The IEB version 11.01.01-150220 containing notifications up to the year 2013 is used. Jacobebbinghaus and Seth (2007) provide information on an anonymized sample of the IEB.
4   The BHP version used is BHP 7510 v1 and covers the years from 1975 to 2010. See Gruhl et al. 2012.
5   Birth or death of an establishment here refers to the presence of workers with social insurance. The birth of an establishments in the BHP means that the EID is observed for the first time in a certain year which implies that there are one or more workers liable to social insurance at June, 30 of that year.

company. In consequence, often multiple notifications for the same worker have to be concatenated to get complete employment durations at the same establishment.

The terms "subspell" or "notification" are mostly used as synonyms in this paper. We will denote these as SSP. Employment spells or job durations will be denoted as ESP and are composed of one or more SSP.

To work with the employment data, some corrections to the sample are in order:

- drop SSP without establishment identifier

- drop SSP of pensioned workers

- drop special SSP (not reporting employment periods) of "Altersteilzeitbeschäftigte"[6]

- drop SSP of extra payments (not reporting employment periods)

- drop SSP lasting less than 3 days

- drop accounts with 15 or more such short SSP

- drop accounts with 100 or more SSP

The last three corrections are rather arbitrary and serve to get rid of a small number of persons with a lot of very short SSP.

Further corrections concern apprenticeship spells, spells with zero earnings and spells with identical or overlapping begin and end dates.

Apprenticeship spells and spells of internship can be identified by two variables (employment group and occupational status). Apprenticeship periods typically start off an employment career and last from two to four years. In the data, we find too long periods of apprenticeship as well as periods of apprenticeship for already older workers. While some of these cases might be correct, we suspect most of them being produced by errors in the observed variables. In order not to count apprenticeships as part of employment durations, we simply eliminate all notifications in an individual employment biography up to the last apprenticeship or internship spell.[7]

In the concatenation of successive subspells, we allow for a gap of at most 31 days between subspells in the same establishment.[8] We collapse notifications with overlapping or identical calendar dates in the same establishment. The covariables of a longer ESP are the attributes

---

[6] Altersteilzeitbeschäftigte: special type of subsidized part-time employment of elder workers prior to retirement.

[7] This is a rather crude but also simple procedure. We also developed a more sophisticated algorithm which corrects for misclassification of apprenticeship spells and prevents some cases with too long apprenticeship durations from being dropped. Since the results from that data did not lead to substantial differences in the results of this report, the simpler algorithm was preferred.

[8] We do not experiment with the length of this gap in this report. The assumed gap of 31 days might produce additional inconsistencies if within this time interval, jobs are terminated and restarted, see section 3 for some discussion.

of the first (e.g. begin, occupational status) or the last subspell (e.g. end, reason for notification). Spells with zero wages are dropped if they are not part of a longer employment episode. In the case of overlapping SSP, the attributes of the SSP with the first begin date and of the SSP with the last end date are used as covariates. When analyzing single subspells, in the case of identical begin and end dates, we keep the subspell with the highest earning.

2.2 Number of ESP over time

While the data reach from 1975 to 2010, Graph 1 demonstrates that there are several years or periods characterized by larger changes in the number of ESP contained in the data.

Graph 1: Distribution of job begins and endings across years, 1975-2010
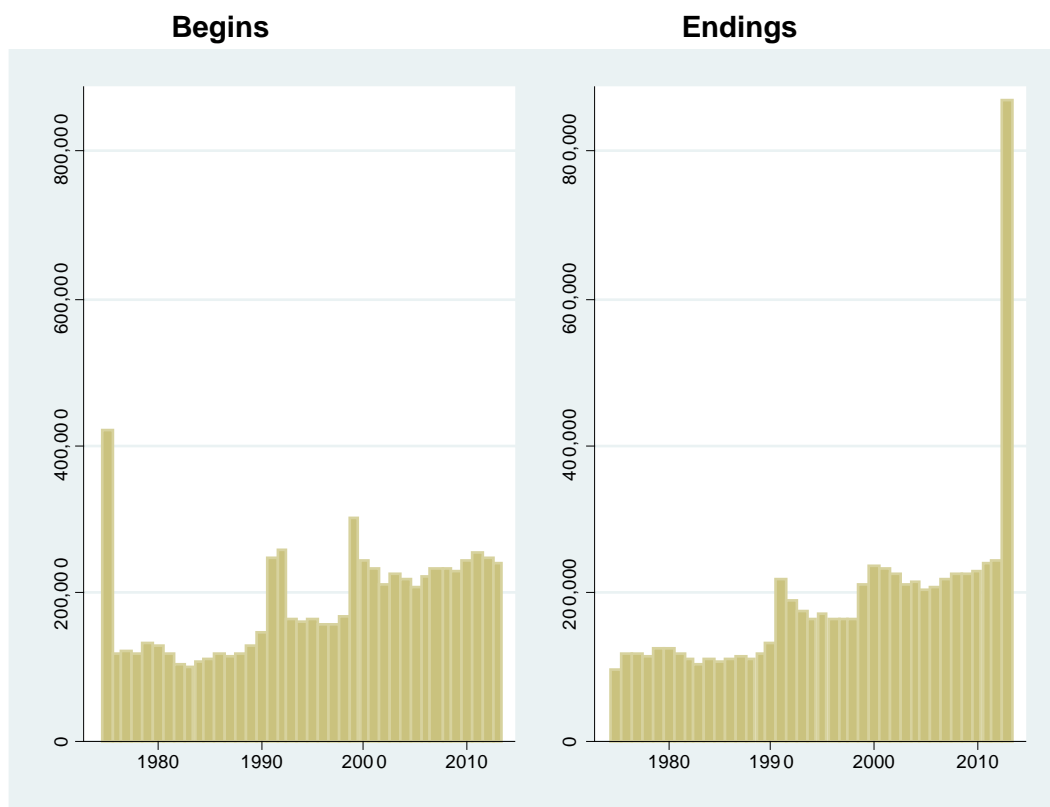
At the left margin, the first register year 1975 contains a huge number of job starts, which covers both already existing and new jobs. At the right margin, the job spells are censored in December 2010, as this is the end of our data window. The effect of the reunification of East and West Germany in the early 1990s is also clearly visible, after a sharp increase the number of new job spells restabilizes after 1993/1994.

There are further years showing particularly high numbers of beginning or ending job spells. In 1999, the provisions regulating the notification process were completely reformed, probably leading to a higher number of notifications and more reliable content. Also in 1999, the so-called "minijobs" became liable to social insurance and were integrated in the employment statistics. In consequence, there is again a visible increase in the number of new jobs in the years 1999 and 2000. The number of job endings is higher from 1999 onwards, increasing further in 2000 and thereafter levelling out.

Because of the noted discontinuities, employment durations for spells starting before the year 1999 may often be incomplete or censored (Statistik der Bundesagentur für Arbeit 2016, p 16). Left-censoring in the year 1975 is (somewhat) accounted for by evaluating employment spells starting in the year 1976 or later. The "reunification shock" is ignored in the descriptive analyses. We deal with the inclusion of minijobs in 1999 by performing separate analyses for jobs beginning in the years 2000-2009. Within this last period, further relevant innovations in the employment register were the (complete) digitization of the notification process in 2006 and the merge of the formerly separate pension institutions for blue collar and white collar workers in the years 2004/2005. While effects on the number and also the type of notifications are likely, we do not attempt to quantify them in this report.

2.3 Calendar effects in job begins and job endings

Job begins and job endings are not at all equally distributed across calendar months. We aggregate the data on a monthly level to demonstrate the effect of the month December on job endings and the effect of the month January on job begins. Graph 2 shows the result for the years 2000-2010.

Graph 2: Job begins and job endings in different calendar months, 2000-2010

In the month of December, the number of job endings tends to be more than double the number of job endings in March, or the months from June or September, which also have high numbers of job endings. Similarly, in January, the number of newly started jobs is far higher than in other calendar months. The high number of job endings and job begins at the turn of a year probably reflects a typical calendar pattern in job search and recruitment. Firms' budgets or bonus payments at the end of the year may be relevant for this regularity. Nevertheless, some kind of

measurement error might also contribute to the high number of job endings in December.

3   Checks for measurement error in employment durations

3.1   Type of notification

The employer has to classify every notification and indicate a reason (Abgabegrund) for it. Types of notification that do not require earnings information are not contained in the employment data[9]. Still, there are more than 20 different types of notifications. After dropping[10] and aggregating[11], we will look at four different categories of types of notification:

- end of employment

- longer interruption of employment

- end-of-year notification

- intermediate notification

The first category collapses reasons for the end of an employment spell. Among the detailed reasons are the end of an employment relationship (without further specification), the legal end of employment after certain interruptions longer than a month[12], the simultaneous notification of begin and end of employment, a worker's death or the legal end of employment in case of the firm's inability to pay. Longer interruptions of an employment spell in the second category cover periods of sickness, of parental leave or military/alternative civilian service which last longer than a month. The end-of-year notifications are compulsory for ongoing employment relationships. Intermediate notifications do not set an end to the employment spell, but are required for social insurance issues, like e.g. the change in the worker's sickness insurance company or a change in a worker's insurance status.

We expect to find one of the first two reasons on the last SSP of an ESP, and the last two reasons on SSP positioned at the beginning or in the middle of an ESP. In Table 1a and 1b, we crosstabulate the reasons for notification with the position of the SSP contained in an ESP, once for the whole period of 1976 to 2009 and once for ESP beginning in the more recent period 2000 to 2009. Begin and end of the ESP result from concatenating successive subspells by means of the EID. Spells ending after the year 2010 are censored and are excluded. This leaves nearly 16 million SSP and 5.4 million ESP for the analysis.[13]

Tables 1a and 1b show that end-of-year notifications for ongoing employment spells are by far the most frequent type of notification. This is followed by end-of-employment notifications,

---

9   Nevertheless, there are notifications reporting zero earnings.
10   Notifications of one-time-payments and special notifications concerning part-time employment of elder workers are dropped, see section 2.1.
11   Table A-1 in the appendix documents the categorization.
12   E.g. unpaid vacation, absenteism, illegal strikes.
13   The total number of ESP (uncensored plus censored) can be found in Table 5.

while intermediate notifications and longer interruptions account for smaller shares of all notifications. The category "combined" contains multiple notifications in the same establishment having a common end date but different reasons for notification assigned.

Table 1a: Subspells by type of notification and position (all periods: 1976-2009)

| | Last subspell | | | | | |
| Reason for notification | No | | Yes | | Total | |
	abs.	in %	abs.	in %	abs.	in %
End of employment	**387,726**	**3.69**	4,456,884	82.85	4,844,610	30.48
Longer interruption	**124,036**	**1.18**	461,220	8.57	585,256	3.68
End-of-year	9,096,543	86.53	**298,113**	**5.54**	9,394,656	59.12
Intermediate	901,020	8.57	**161,046**	**2.99**	1,062,066	6.68
Combined	3,011	0.03	2,501	0.05	5,512	0.03
Total	10,512,336	100.00	5,379,764	100.00	15,892,100	100.00

Table 1b: Subspells by type of notification and position (period: 2000-2009)

| | Last subspell | | | | | |
| Reason for notification | No | | Yes | | Total | |
	abs.	in %	abs.	in %	abs.	in %
End of employment	**177,461**	**7.78**	1,684,912	88.71	1,862,373	44.56
Longer interruption	**26,938**	**1.18**	104,420	5.50	131,358	3.14
End-of-year	1,717,929	75.35	**53,526**	**2.82**	1,771,455	42.39
Intermediate	356,880	15.65	**56,313**	**2.96**	413,193	9.89
Combined	697	0.03	277	0.01	974	0.02
Total	2,279,905	100.00	1,899,448	100.00	4,179,353	100.00

While unclear, these cases represent a negligible number of spells. In the more recent period 2000-2009, especially end-of-employment notifications and (to a far smaller extent) also intermediate notifications increased their weight in all notifications, while the share of end-of-year notifications clearly decreased. These changes in the structure of SSP are probably due to the changes in the notification system after the year 1999 as mentioned in section 2, we do not try to trace them back to these events, however.[14]

The main message of the Tables 1a and 1b is that, in most cases, the reported reason for notification confirms the observed position of a subspell in an employment spell. Looking at the results for the whole period in Table 1a, in 91.5 percent of the spells (first two entries in the "yes"- column), an end of employment or a longer interruption is reported on the last subspell.

---

[14] The integration of minijobs in the employment data in the year 1999 is certainly one of the reasons of having relatively more end-of-employent notifications, see also Table 2.

For the more recent period 2000-2009 (Table 1b), the respective share is about 94 percent. For SSP not on the last position of an ESP, the share of notifications reported as "end of year" or "intermediate" is about 95 percent for the whole period (3rd and 4th entry of "no" - column in Table 1a), and decreases to 91 percent for the more recent years 2000-2009 (Table 1b).

However, there is also a minority of SSP, for which the reported reason for notification does not indicate the end or continuance of an employment spell. Among "last" subspells, a share of 8.5 percent is reported as "end of year" or "intermediate" in Table 1a for the whole period and a share of 5.8 percent in Table 1b for more recent years (3rd and 4th entry of "yes" - column, in bold).[15] Among the SSP which are not on the last position of an ESP, a share of 4.9 percent is reported with the reason "end of employment" or "longer interruption" in Table 1a and an even larger share of 9 percent results for the period 2000-2009 in Table 1b (first two entries of the "no" - column, in bold).

Table 1c: Position of a subspell and type of notification (%)

	Period	
	1976-2009	2000-2009
Plausible	93.89	92.48
Implausible		
First/middle	3.22	4.89
Last	2.89	2.63
Total	100.00	100.00
Total number of SSP	15,892,100	4,179,353

Using the crosstabulation in Table 1a and 1b, we define two categories of implausible constellations of the position of an SSP and the type of notification. The category "Implausible first/middle" in Table 1c consists of SSP which precede the last spell in an ESP and for which the type of notification indicates an end of the employment spell or a longer interruption (entries appearing in bold in the "no" - columns of Table 1a/b). The category "Implausible last" contains SSP which are on the last position of an ESP and which are reported as end-of-year or as intermediate notifications (entries appearing in bold in the "yes" - columns of Table 1a/b). Again, only uncensored spells are included.

One explanation for implausible constellations of the first type, concerning SSP on a first or middle position of an ESP, could be that the concatenation procedure we adopt defines interruptions of 31 days to be harmless to the continuance of an employment spell and thus subspells are concatenated when (in the judgment of the employer) they represent separate em-

---

[15] The Statistics of the Federal Employment Office Agency uses a correction procedure for long-term employment relationships not followed by further notifications and without a notification indicating a regular job ending. The share of these jobs is reported as 0.3 percent of all employment relationships, see Statistik der Bundesagentur für Arbeit (2017).

ployment episodes. In fact, concatenating only subspells in the same establishment which follow each other without any interruption almost halves the number of "Implausible first/middle" SSP, whereas the number of "Implausible last" SSP remains nearly unchanged (without table). This implies that there is a considerable amount of SSP reported as end of employment but actually continued by the same worker in the same establishment.

In Table 2, we look at minijobs, which are contained in the data since 1999. These jobs often have short working hours, but also may be of short-term nature more often than regular jobs. Minijobs account for more than one third of all 4.2 million SSP. In fact, among the subspells on a first/middle position, minijobs represent more than half of the observations in the category "Implausible".[16] In contrast, minijobs account for only 25 percent of the "implausible" SSP observed on the last position of an ESP, but reported as end-of-year or intermediate notifications.

Table 2:   Minijobs – Position of subspell and type of notification (period: 2000-2009)

	Minijobspell					
	No		Yes		Total	
	abs.	in %	abs.	in %	abs.	in %
First/Middle Spells						
Plausible	1,430,611	68.93	644,895	31.07	2,075,506	100.00
Impl. first/middle	96,461	47.19	107,938	52.81	204,399	100.00
Last Spells						
Plausible	1,108,188	61.92	681,421	38.08	1,789,609	100.00
Impl. last	81,972	74.63	27,867	25.37	109,839	100.00
Total	2,717,232		1,462,121		4,179,353	

Another interesting point is whether implausible constellations between the reported notification type and the observed position in an ESP are more likely to occur at the end of the year and thus somehow associated with the large number of jobs ending in December.

Table 3 repeats the crosstabulation in Table 1c but distinguishes between spells ending in the month of December and spells ending in another month of the year. The share of implausible combinations with the reported notification type for "first/middle" SSP ending in the months from January to November (4.8 percent) is comparable to the share of "first/middle" SSP ending in the month of December (5.4 percent). The share of implausible combinations with the reported notification type for spells observed on the last position of an ESP is however much larger for spells ending in December (11.7 percent) than for those spells ending in another month of the year (0.9 percent). This might be due to the fact that end-of-year notifications are by definition ending in December – or conversely, that spells ending from January to November must not be reported as end-of-year notifications.[17] Nevertheless, SSP ending in December

---

[16]   There is no difference in the rules for assigning the reason for notification between minijobs and regular jobs.

[17]   In fact, the respective cell is empty in the data. The entry of 0.9 percent of implausible "last" SSP ending from January to November is caused by SSP with some kind of intermediate notification.

could also be more often observed as "last" spells if they are part of ESP that are only incompletely observed.

Table 3: December endings - Position of subspell and type of notification (period: 2000-2009)

| | SSP ending in December | | | | | |
| | No | | Yes | | Total | |
	abs.	in %	abs.	in %	abs.	in %
Plausible	3,304,667	94.34	560,448	82.86	3,865,115	92.48
Impl. first/middle	167,773	4.79	36,626	5.41	204,399	4.89
Impl. last	30,505	0.87	79,334	11.73	109,839	2.63
Total	3,502,945	100.00	676,408	100.00	4,179,353	100.00

3.2 One-year gaps between subspells in the same establishment

There is already some evidence that the employment data (while ideally representing the universe) may suffer from incompleteness in some regions or years. A phenomenon relevant for employment durations are gaps of one year within series of yearly notifications (see Drews et al. 2007).

Looking at ESP starting in the years 1976 to 2009, we find 62,820 gaps of single years without notification when at the same time there are notifications for the same establishment before and after the gap and when there are no IEB-episodes indicating some kind of non-employment during the gap.[18]

To see whether these one-year-gaps can help to explain the inconsistencies observed in the last section, we look again at the reason for notification reported on the subspells located directly before a "one-year-gap". These SSP are "last" spells by definition.

Table 4: Type of notification for "last" subspells preceding one-year gaps in employment spells

| | Period | | | |
| | 1976-2009 | | 2000-2009 | |
	abs.	in %	abs.	in %
Plausible	6,257	9.96	1,984	19.51
Impl. last	56,563	90.04	8,186	80.49
Total	62,820	100.00	10,170	100.00

From Table 4 we can see that SSP preceding one-year-gaps are reported with a notification type other than end-of-employment or interruption in more than 80 percent (90 percent for the

---

[18] There were 3.425 one-year gaps between employment spells at the same establishment which were in some way intersected by other IEB-spells, reflecting periods of unemployment, of benefit receipt or of programme participation.

whole period) of the cases. This means that the type of notification confirms the notion of in-complete data associated with one-year-gaps. However, filling all gaps might possibly create new measurement error if the one-year-gaps are caused by "real" interruptions of employment spells. Further, the relevance of gaps lasting exactly one year is rather small. Compared to the total numbers of implausible types of notifications in Tables 1a and 1b, the "Implausible last" spells in Table 4 account for 7 percent of all "Implausible last" in the years 2000-2009 and for 12 percent of the "Implausible last" observed over all years from 1976-2009.

3.3   Changes in the establishment identifier

The establishment identifier is critical for the computation of job durations, because single no-tifications report employment periods of at most one year, and thus have to be concatenated to arrive at total job durations.

The EIDs are assigned by a central authority belonging to the Federal Employment Agency. They used to be assigned by local authorities in the labour offices until 2008. In the definition of the employment register, an establishment is a unit with employees, which is distinguished by region and industry. While some firms have several establishments, and there are also establishments representing more than one branch or office, the majority of firms consists of one establishment only. In most cases, therefore, an establishment identifier appearing in the register for the first time will denote a new establishment as part of an existing firm or a newly founded firm. Similarly, extinguished identifiers are usually indicative of a firm's closure or take-over by another firm. As only establishments with at least one employee are included in the data, there may also be new or disappearing EIDs because of establishments changing from zero to a positive number of employees or because of establishments remaining without any employee.

Relevant to the length of employment durations, there are occasions in which the EID changes and it is not clear whether this sets an end to existing employment relationship or not. Some-times, establishments will get a new identifier, for example when changes in ownership, legal status, production focus, or region occur. Another case for new EIDs are spin-offs from existing firms or outsourced parts of an establishment. These events might also go along with the foun-dation of new firms, but they differ from other foundations as larger parts of the workers do not change their jobs, but keep working together in another establishment. Unfortunately, such changes in the establishment identifier are not commented in the data. Without the use of additional assumptions it is impossible to tell whether the observed flows happen inside a firm composed of different establishments or whether these flows e.g. imply the foundation of new companies or the integration into another already existing firm.[19]

Following the approach of Benedetto et al. (2007) and Hethey-Maier and Schmieder (2010, 2013), we use information on workers' flows which is available as an extension file to the BHP.

---

[19]   Goldschmidt and Schmieder (2017) use the occupational and industry information in addition to information on worker flows to identify the outsourcing of groups of workers into certain service industries or temporary agen-cies.

The computation of flows is based on yearly cross-sections of all social security records notifications aggregated on the establishment level. Then, for any new EID (1) or any EID not observed in the following year (2), the maximum number of inflow workers from other firms (1) or the maximum number of outflow workers to other firms (2) is calculated.

Assume there is a larger group of workers working under one EID in the first year and the same group working under a different EID in the second year. This observation might be interpreted as a simple change in the EID if the first EID disappears from the data while the EID of the second year is a new one. The same flow will be classified as a spin-off or as outsourcing if the EID of the first year continues to exist. It will be considered as a takeover or a result of restructuring, if the EID of the first year disappears, while the EID of the second year had already been existent in the data (Hethey-Maier and Schmieder 2013).

Pure EID changes do not seem to happen very often in the employment register. Their share in all new or all disappearing establishments Ids is reported to be 0.82 and 0.89 percent by Hethey-Maier and Schmieder (2013, 488). In our version of the administrative data (BHP1975-2010), we find the respective shares to be often larger than 1 percent, however no larger than 2 percent in any of the years. The occurrence of spin-offs or takeovers of groups of coworkers is somewhat more frequent. Spinoffs and takeovers together account for about 3 percent both for new and for disappearing EIDs.

To get a conservative estimate of how job durations are influenced by these EID changes which are not clearly indicative of death or birth of an establishment, we collapse pure EID changes, outsourcing, spin-offs, restructuring and takeovers into one group. We then let employment durations continue despite a change in the EID when a larger worker flow between old and new EID has been observed somewhen in the data.[20]

Table 5:   Number of ESP across data versions

Data version		Period	
		1976-2009	2000-2009
0	no use of flow information	5,856,948	2,269,304
0.8	cutoff 80 percent of workers	5,698,050	2,198,885
0.7	cutoff 70 percent of workers	5,661,370	2,183,246

To check in how much job durations depend on the identifier definition, we compare three versions of the individual job tenure data. As a "null" version, we concatenate successive SSP

---

[20] We also allow for indirect links between EIDs. Imagine a worker starting a job in establishment A and continuing in establishment C. We might identify the relationship between A and C by first observing a larger flow of workers to establishment B and then another flow from B to C. In practice, EIDs were grouped into pools where one ID could link to one or several other ID, depending on whether larger worker flows had occured somewhen in the BHP. The largest pool results in 368 different establishment Ids, when using the cutoff point 0.80.

with the same establishment Id only. In the "0.8" version, a "larger" worker flow between two establishments is defined as a flow of workers being observed in two successive years and accounting for at least 80 percent of one of the two establishments, or of both. As a robustness check, there is also a "0.7" version of the data, with the cutoff-point analogously defined.

Selecting ESP starting in the years from 1976 to 2009, the data version using 0.8 as the cutoff contains 5,698,050 compared to 5,856,948 ESP in the data version 0 when only the establishment identifier is used. The difference between the data versions 0.8 and 0.7 is even smaller.

There are 122,371 ESP that are composed of more than one EID (without table), accounting for only 2 percent of all ESP in the 0.8 data version and including ESP starting in the years 1976-2009. However, on the person level, about 9.2 percent of all persons have at least one ESP consisting of subspells with different EIDs (Table 6).

Table 6:   Number of different EIDs per ESP and per Person (0.8 data version)

Number of EIDs	Per Tenure Spell (ESP)		Per Person	
	abs.	in %	abs.	in %
1	5,575,680	97.85	1,130,488	90.77
2 or more	122,370	2.15	114,931	9.23
Total	5,698,050	100.00	1,245,419	100.00

3.4   Type of notification with broader EID concept

Looking at Table 7a, we can see that the share of "implausible" combinations of notification type and the position of an SSP in an ESP found in section 3.1 are also not strongly influenced by the use of a broader EID concept. The share of "first/middle" SSP reported as end of employment or longer interruption increases from 3.2 percent in the version using the original EID (table 1a) to about 3.9 percent in the 0.7 data version. The share of "last" SSP with a reported end-of-year or intermediate notification decreases slightly from 2.9 percent in the 0 data version to 2.7 percent in the 0.8 data version and to 2.6 percent in the 0.7 data version.

Table 7a: Reason for notification and position of subspell with broader EID concept, all periods

	Data Version					
	0		0.8		0.7	
	abs.	in %	abs.	in %	abs.	in %
Plausible	14,921,179	93.89	14,515,264	93.58	14,440,608	93.49
Impl. First/Middle	511,762	3.22	580,819	3.74	598,801	3.88
Impl. Last	459,159	2.89	415,477	2.68	406,327	2.63
Total	15,892,100	100.00	15,511,560	100.00	15,445,736	100.00

If we look at Table 7b, we can see that the share of "first/middle" SSP that are reported as end-of-employment or longer interruption is somewhat higher for more recent years than for the

whole period, accounting for 4.9 to 5.5 percent of all subspells.[21] Across data versions, the share is highest for the 0.7 version, possibly indicating that "more" concatenation leads to a (very modest) increase in measurement error. The number of "last" spells reported as end-of-year or intermediate notifications decreases across data versions but the share of these spells is rather stable, varying between 2.6 and 2.4 percent.

Table 7b: Reason for notification and position of subspell with broader EID concept (period: 2000-2009)

	Data Version					
	0		0.8		0.7	
	abs.	in %	abs.	in %	abs.	in %
Plausible	3,865,115	92.48	3,718,743	92.15	3,688,312	92.07
Impl. First/Middle	204,399	4.89	217,690	5.39	221,443	5.53
Impl. Last	109,839	2.63	99,090	2.46	96,425	2.41
Total	4,179,353	100.00	4,035,523	100.00	4,006,180	100.00

3.5 December endings and broader EID concept

Does the use of a broader EID concept which is based on labour market flows influence the large number of endings of ESP observed in the month of December? As Table 8 demonstrates, the reduction in the share of ESP ending in December between data versions 0 (using the original EID) and 0.7 (using the broader concept and the cutoff 70 percent) is less than one percentage point, from 19.9 to 19.0 percent. Looking at the last period, the reduction is only about 0.6 percentage points. Thus, the number of ESP ending in December is not substantially reduced by using the broader EID concept.

Table 8: Employment spells (uncensored) ending in the month of December across data versions

	Period			
Data Version	1976-2009		2000-2009	
	abs.	in %	abs.	in %
0	1,071,489	19.92	326,460	17.19
0.8	999,653	19.16	307,165	16.70
0.7	983,110	18.98	302,122	16.55

From Table 9 we can see that also for SSP ending in December implausible notification types for "first/middle" spells increase slightly across versions (see Table 7b for the number of spells ending in any month). As already noted, there is a large number of "last" SSP ending in December and reported as end-of-year or intermediate notifications. While there is a slight decline across data versions, the percentage share of these spells remains fairly stable at 11 to 12

---

[21] A similar difference between periods was observed in Table 1a/b.

percent.

Table 9: **Position of a subspell and type of notification for employment spells ending in December across data versions (Period: 2000-2009)**

| | Data Version | | | | | |
| | 0 | | 0.8 | | 0.7 | |
	abs.	in %	abs.	in %	abs.	in %
Plausible	560,448	82.86	518,238	82.52	507,996	82.46
Impl. First/Middle	36,626	5.41	38,147	6.07	38,455	6.24
Impl. Last	79,334	11.73	71,618	11.40	69,578	11.29
Total	676,408	100.00	628,003	100.00	616,029	100.00

3.6 Job durations

As a last step of analysis, we will present the results of a duration model which illustrates the quantitative dimension of using a broader EID concept for the computation of job durations. We select ESP starting in the years from 2000 to 2009. We select job spells beginning in West Germany. We do not restrict by age, the observed age ranges from 12 to 76 years. Different from the preceding analyses, censored spells are included.

We estimate a piecewise constant exponential model. The model is of the proportional hazard type. Duration dependence is modelled by splitting the observed durations into time intervals. The estimated coefficients measure the influence of these time intervals and of the other co-variables on the hazard rate. The hazard rate is constant within but is allowed to vary across intervals.

To allow for a trend in durations even in this fairly short period, we include a set of dummy variables for the begin year of a spell. We pool the different versions of the data and include two dummy variables indicating the EID concept: the reference group are durations concate-nated using the EID only, the 0.8 (0.7) data version additionally collapses establishment IDs when larger worker flows between them were observed at any time point in the data, using 80 (70) percent of the establishment workforce as cutoff points.

We further control for industry, firm size, ESP composed of mainly minijobs, gender, age cat-egory, foreign nationality and educational/vocational degree.

As a central result (Table A-2 in the appendix), the coefficients[22] for the data versions using different cutoff points are significant, but very small. The hazard ratios in Table 10 show that the multiplicative effect on the hazard rate is very close to one in the data version using the broader EID concept. Also in Table 10, the predicted median duration[23] calculated from the

---

[22] In the model, the hazard rate is defined as $\exp(X'\beta)$, multiplied by a time-interval dependent baseline hazard. Therefore, the coefficient β^k of a covariable X^k has a non-linear but monotonous association with the hazard rate. A larger coefficient implies a larger risk of ending the job and a shorter expected job duration.

[23] Median durations indicate a value of the survivor function of 0.5. The results in table 10 represent averages of predicted median durations for every employment spell.

estimated coefficients show that on average, the effect of using a broader EID concept is only about 1.5 days when changing from data version 0 to data version 0.8 and about another 0.8 days when changing from data version 0.8 to data version 0.7.[24]

Table 10: Data versions and job durations

	Coefficient	Hazard ratio	Predicted median duration (days)
Data version 0 (original EID)			375.23
Data version 0.8 (cutoff 80 percent)	-.0026	0.9974	376.75
Data version 0.7 (cutoff 70 percent)	-.0039	0.9961	377.51

4   Conclusion

Administrative data like the German employment biographies contain valuable information which is difficult to collect in surveys. They are also highly reliable data sources in so far as the information is important for financial transactions like e.g. the payment of pensions or unemployment benefits.

This report has a closer look at various issues arising when generating employment durations from employers's notifications to social security in Germany. As these notifications provide the begin date and the end date of an employment period, it is straightforward to concatenate these notifications using the id of the worker and the id of the establishment. Employers have to assign a reason to every notification that indicates whether the employment episode is ended or is continued. When cross-checking, the reason for notification does not always confirm the end or the continuance of a subspell as resulting from concatenation.

As a further observation, the concatenated employment spells are far more likely to end in the month of December than in other months of the year. Subspells on the last position of a concatenated employment spell that end in December are far more likely to be reported as continued employment than are "last" spells ending in other calendar months. Since existing documentations of the data already give hints on missing notifications, it is checked whether incompleteness in the form of one-year-gaps in individual employment spells with the same employer can explain some of the inconsistencies found. It is found that these gaps do not occur very often and therefore their contribution to the total number of "last" subspells carrying an implausible reason of notification is rather modest.

---

[24]   In the estimated models 1 and 2 the coefficients of the data versions with different cutoff points are negative, supporting the intuitive idea of getting longer durations when some of the original spells are concatenated. This is not a necessary result however, as the effect depends on how the duration composition of spells changes across data versions.

Another possible source of error are changes in the establishment identifier which do not go along with the death or birth of an establishment. Such changes are not documented in the data but can be identified by observing major year-to-year flows of workers between establishments. Job durations can be regarded as going on if the worker finds himself in the same group of coworkers. However, even using a broad concept of EID changes (including also takeovers and outsourcing), integrating such changes in the computation of employment durations does not have but a small influence on the data.

Unfortunately, the inspection of different sources of measurement error does not give clear advice of how to deal with them. While correction rules might eliminate some part of the observed implausibilities, they might also add new error to the data. As a bottom line, administrative data are highly reliable but a bit less than perfect.

5   Programs

As an appendix to this report, Stata do-files containing the computation of employment durations are available. The do-files use the Sample of Integrated Labour Market Biographies (SIAB) 1975-2014, which is slightly different from the data used in this report. The SIAB is also based on the Integrated Employment Biographies and is available at IAB's Research Data Center (see Antoni et al. 2016).[25]

The accompanying Stata do-files are structured as follows:

- **tenure.do**: computation of employment durations based on the original establishment ID

- **tenure_pool.do** and **pool.do**: computation of employment durations using aggregated IDs which are eventually composed of two or more original establishment IDs. Original IDs are pooled if between two of them

  - ID-changes,

  - spin-offs or

  - takeovers

  are observed in the Establishment History Panel in any year since 1976. [26]

---

[25]   Another programming example for computing employment durations in the presence of parallel and overlapping subspells is provided by Eberle and Schmucker (2017).

[26] The do-files use the full sample of the BHP 7514 extension files „entry and exit" to compute a pooled ID. External researchers can use the BHP entry and exit extension files prepared for the SIAB data. These have to be specifically requested within the data application form. The BHP extension files for SIAB only provide information on establishments with workers belonging to the SIAB sample. Mergers and ID changes taking place during a period where the respective establishments do not employ any workers included in the SIAB sample are therefore not covered in the extension file for the SIAB data. As a result, the number of ESP is slightly higher and mean durations are slightly shorter when using the SIAB-specific BHP extension files (compared to using the full BHP sample).

Download of the Stata do-files:

http://doku.iab.de/fdz/reporte/2018/MR_10-18_EN_programs.ZIP

References

Antoni, M., Ganzer, A., and vom Berge, P. (2016): Sample of Integrated Labour Market Biographies (SIAB) 1975-2014. FDZ-Datenreport 04/2016, Nürnberg.

Benedetto, G., Haltiwanger, J., Lane, J. and K. McKinney (2007), Using Worker Flows to Measure Firm Dynamics, Journal of Business & Economic Statistics, 25 (3), 299-313.

Boockmann, B. and Steffes, S. (2005), Individual and Plant-Level Determinants of Job Durations in Germany. ZEW Dicussion Paper 05-89.

Drews, N., Groll, D. and P. Jacobebbinghaus (2007), Programmierbeispiele zur Aufbereitung von FDZ Personendaten in STATA. FDZ-Methodenreport 6/2007, Nürnberg.

Eberle, J. and Schmucker, A. (2017), Creating Cross-sectional Data and Biographical Variables with the Sample of Integrated Labour Market Biographies 1975-2014, FDZ-Methodenreport 06/2017, Nürnberg.

Goldschmidt, D. and Schmieder, J. (2017), The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure, The Quarterly Journal of Economics, 132 (3), 1165–1217.

NBER Working Paper No. 21366 (New version: January 2017).

Gruhl, A., Schmucker, A. and S. Seth (2012), The Establishment History Panel 1975-2010, Handbook Version 2.2.1, FDZ-Datenreport 4/2012, Nürnberg.

Hethey-Maier, T. and Schmieder, J. (2010), Using Worker Flows in the Analysis of Establishment Turnover. Evidence from German Administrative Data. FDZ-Methodenreport 06/2010, Nürnberg.

Hethey-Maier, T. and Schmieder, J. (2013), Does the Use of Worker Flows Improve the Analysis of Establishment Turnover? Evidence from German Administrative Data. Schmollers Jahrbuch 133, 477-510.

IAB (2016), ITM-Datenreport BeH V10.01.00-160816, Internal Report, Nürnberg.

Jacobebbinghaus, P. and Seth, S. (2007), The German Integrated Employment Biographies Sample IEBS. Schmollers Jahrbuch, 127 (2), 335-342.

Statistik der Bundesagentur für Arbeit (2017), Grundlagen: Methodenbericht – Revision der Beschäftigungsstatistik 2017, Nürnberg.

Statistik der Bundesagentur für Arbeit (2016), Grundlagen: Methodenbericht – Die Dauer von Beschäftigungsverhältnissen, Nürnberg.

Appendix

Table A-1: Categorization of detailed reasons for notification

Type of notification	Code	Detailed reason for notification
End of employment:	30	End of employment
	34	Interruption of employment for more than one month (§ 7 Abs. 3 Satz 1 SGB IV)
	40	Simultaneous registration and deregistration
	49	Death
	72	Legal end of employment (insolvency)
Longer interruption of employment:	51	Receipt of compensating benefits (e.g. health insurance allowance)
	52	Parental leave
	53	Military/civilian service
End-of-year notification:	50	Annual notification
	70	Annual notification for employees exempted from work (insolvency)
Intermediate notification:	31	Change of health insurance company
	32	Change in contribution group
	33	Other reasons
	35	Interruption of employment for more than one month because of industrial conflict/dispute
	36	Change of payroll accounting system/currency changeover
	57	Special notification according to § 194 SGB VI (effected approximately 3 months before entry into pension to facilitate pension calculation)
	59	Notification of wage/salary for casual employees
	71	Notification of the last day before an insolvency/ exemption

Table A-2: Results from job duration model

	Coefficient	z
Time pieces		
0-30 days	-5.165	-1124.21
31-90 days	-5.314	-1170.05
91-182 days	-5.408	-1190.65
183-364 days	-5.664	-1249.16
365-547 days	-6.116	-1308.91
548-729 days	-6.227	-1302.15
730-912 days	-6.399	-1291.50
913-1094 days	-6.566	-1268.03
1095 days or more	-6.856	-1478.89
Data version (reference "0")		
0.7	-0.004	-3.44
0.8	-0.003	-2.29
Year of spell begin (reference "2000")		
2001	0.002	0.98
2002	0.017	8.25
2003	-0.063	-31.95
2004	-0.014	-6.85
2005	-0.008	-4.05
2006	0.011	5.34
2007	0.010	4.84
2008	-0.011	-5.47
2009	-0.030	-13.79
Industry (reference "Agriculture")		
Mining, energy, traffic and information	-0.335	-88.28
Manufacturing	-0.555	-151.29
Construction	-0.184	-47.61
Trade and retail	-0.424	-117.73
Business services	-0.184	-51.93
Personal and domestic services	-0.231	-62.65
Social and public services	-0.449	-123.72

Table A-2 continued

	Coefficient	z
Establishment size (reference "Missing")		
1-49	-0.124	-57.63
50-99	-0.078	-31.31
100-249	-0.129	-52.34
250-499	-0.203	-73.98
500-999	-0.278	-92.14
1000-2999	-0.304	-97.06
3000 and more	-0.431	-117.35
Minijob	0.364	316.65
Part-time	0.091	51.67
Female	-0.135	-131.95
Age group (reference " < 25")		
25-34	-0.141	-105.56
35-44	-0.352	-249.32
45-54	-0.390	-243.36
55+	-0.365	-193.17
Foreign	0.176	125.94
Skill level (reference "Missing")		
No vocational training with at most intermediate degree	0.185	127.99
Vocational training with at most intermediate degree	-0.100	-76.79
Abi/equivalent without vocational training	0.263	99.70
Abi/equivalent with vocational training	-0.229	-86.15
Technical college degree	-0.358	-95.67
University degree	-0.227	-83.97
Likelihood-Ratio-Test (chi^2 39)	573230.17	
Number of Tenure Spells (ESP)	5,528,914	
Uncensored	4,605,451	

Note: Covariates are measured on the first subspell. Minijobs are ESP for which at least 80 percent of the subspells are minijobs.

Imprint

Corresponding author:
Ursula Jaenichen
Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nürnberg
Phone: +49-911-179-5415
Email: Ursula.Jaenichen@iab.de