On the Definition of Unemployment and its Implementation in Register Data - The Case of Germany.[1]

May 2007

Thomas Kruppe[2], Eva Müller[3], Laura Wichert[4], Ralf A. Wilke[5]

Abstract

Unemployment information in individual level register data of the German Federal Employment Service depends on institutional settings, administrative procedures and which registers are merged. In this paper we suggest different implementation strategies for common international and German legal unemployment definitions for the Sample of the Integrated Employment Biographies (IEBS). The IEBS belongs to a new generation of German merged register data that is more comprehensive than previous data sets. Our descriptive figures show large differences in the number of spells and the unemployment duration across implementations. This suggests that empirical results of labour market research are likely to depend on the underlying legal definition of unemployment and its implementation in this data.

Keywords: unemployment duration, merged administrative individual data, Germany, Stata

---

[1] This work is a result of the research cooperation between the IAB Nuremberg, the University of Leicester and the ZEW Mannheim within the research project: "Contribution to Quality Monitoring of Integrated Employment Biographies: Development of Alternative Unemployment Classification Proposals". We thank Dirk Oberschachtsiek for helpful comments. Eva Mueller and Ralf Wilke thank the IAB for financial support. Laura Wichert thanks the ZEW for its hospitality.

[2] Institute for Employment Research, Nuremberg, Germany (IAB),
Email: Thomas Kruppe@iab.de

[3] University of Mannheim and Centre for European Economic Research (ZEW),
Email: eva.mueller@bwl.uni-mannheim.de

[4] University of Konstanz, Department of Economics, Box D 124, 78457 Konstanz, Germany,
Email: laura.wichert@uni-konstanz.de

[5] University of Leicester, Department of Economics, University Road, Leicester LE17RH, UK,
Email: raw27@le.ac.uk

1. Introduction

The determinants of unemployment duration are of high interest in social and applied economic research alike. A broad range of empirical and theoretical research in Germany is concerned with this topic. The empirical studies are based on different data such as surveys or administrative registers which have both advantages and disadvantages. See Biewen and Wilke (2005) for a direct comparison of the unemployment information in the German Socio Economic Panel and the IAB Employment sample. While the smaller sample size and errors in the reporting behaviour of the individuals are the main weaknesses of survey data, missing interval information, only a small number of observed variables and inconsistent administrative records are important weaknesses of administrative individual data. Inconsistencies occur due to the merger of different registers. Missing interval information prevents the researcher from fully reconstructing individual employment trajectories with the help of administrative registers. This often makes it impossible to compute the true length of the unemployment periods from this data, since unemployment is inherently a concept or, moreover, a social category, that is highly related to other labour market concepts like employment, inactivity, or being out of the labour market.[6] In applied research the results therefore likely depend on the underlying implementation of unemployment duration in the data. Indeed, there is already empirical evidence for this. See Fitzenberger and Wilke (2004) or Lee and Wilke (2005) for an evaluation of a reform of unemployment compensation on unemployment duration in Germany. In addition to the problem of partial identification of the true length of unemployment periods, we focus in this paper on the question how different legal definitions of unemployment can be implemented in register data.

In the next section, we first direct attention to international and national legal definitions of unemployment and their application. We take Germany as an example and show that the definition of unemployment is not stationary but a social category with different characteristics. We present international standardised unemployment rates, mainly based on the definitions of the International Labour Organisation (ILO), and the main German regulations for the support of unemployed stated in the Third Volume of the Social Code (SGB III).

---

[6] On the genesis of unemployment as a social problem, from the first debates to the foundation of the first compulsory insurance against unemployment, and from the individual status to the formation of a social institution, see Zimmermann 2006.

Arising from this discussion, section 3 formulates six alternative concepts of unemployment. These concepts are then implemented in the Sample of the Integrated Employment Biographies (IEBS). This data is individual merged register data containing periods of employment, claim of unemployment compensation, job seek periods and information about participation in labour market policies. Our empirical exercise is nontrivial given the complicated data structure and the massive amount of information in the data. Moreover, missing interval information and data inconsistencies make unique implementations of unemployment duration impossible. For this reason, we suggest several alternative approaches to the applied researcher. These implementations are available as Stata Do files. Some simple descriptive analysis shows strong empirical evidence for considerable differences in the length of unemployment periods depending on the underlying definition of unemployment and their implementations in the data. In section 4 we discuss further problems with these data that we do not address in this work. Section 5 discusses the main findings of the paper.

2. International and National Concepts to Measure Unemployment in Germany

The international standardised unemployment rates, published by the *Statistical Office of the European Communities* (EUROSTAT) and the *Organisation for Economic Co-operation and Development* (OECD), are mainly based on the definitions of the *International Labour Organisation* (ILO) and calculated using cross sectional random sample survey data sources, namely the *European Labour Force Survey* (LFS). Also, longitudinal data sources can be used to measure unemployment using different concepts, namely the *European Community Household Panel* (ECHP) or the *German Socio Economic Household Panel* (GSOEP). Last but not least, measurement can be based on register data of the German social security system.

The national unemployment rate, officially announced in regular intervals by the *Federal Employment Service* (Bundesagentur für Arbeit (BA)), is based on the number of registered unemployed persons as part of the labour force. The definition used is codified in the *Third Volume of the Social Code* (Sozialgesetzbuch III (SGB III)), which replaced the former *Labour Promotion Act* (Arbeitsförderungsgesetz (AFG)) in January 1998.[7] The *Second Volume of the Social Code* (Sozialgesetzbuch II (SGB II)), introduced in December 2004, broadens the definition of unemployment to all individuals capable of working, as well as the indigent, where the first is interpreted individually and the latter in a household context. This also refers to concepts of labour reserve, hidden unemployment, hidden labour force and discouraged workers.

Comparing the definitions of international standardised unemployment rates with German national unemployment rate gives an impression on the variety of the social category of unemployment.[8]

---

[7] Furthermore, the BA tries to overcome some shortcomings of this definition by widening the scope, additionally using an internal definition of unemployment that includes time in labour market programs ("Dauer faktischer Arbeitslosigkeit").

[8] On the statistical differences of unemployment in European countries see Werner (1984). On the U.S. concept and an adjustment of the Canadian and European unemployment rate on this concept see Sorrentino (2000).

2.1. International Standardised Unemployment Rates

The *International Labour Office* (ILO) defines the unemployed on the *Thirteenth Conference of Labour Statistics* as persons without work, who are currently available for work and seeking work. The concept thus refers to the definition of employment: Persons in paid work could be *"at work, [...] performing some work for wage or salary"*, or could be *"with job but not at work, having already worked in their present job, were temporarily not at work [...] and had a formal attachment to their job"*. Self-employed person are defined as being *"at work, [...] performing some work for profit"* or being *"with an enterprise but not at work, [...] who are not at work [...] for any specific reason"*. To comply with this definition, working for at least one hour is sufficient. The main criterion here is formal job attachment, not the main activity. Consequently, students with a part-time or marginal job are included, as well as persons on leave, laid-off persons, short-time workers, apprentices and members of the armed forces (ILO 1983)[9]. In absence of a formal work attachment - but fulfilling the criteria mentioned above (available and searching for work) - persons are counted as unemployed even independent from their primary activity status (i.e. student).

The unemployment figures published yearly by the ILO include registered unemployment, also published by the BA, as well as unemployment based on the German Microcensus, (ILO 1997). From January 2005 on, the BA, in addition to publishing the number of registered unemployed, estimates and publishes unemployment numbers as defined by the ILO.

The *Organisation For Economic Co-Operation And Development* (OECD) abolished computing its own unemployment rates for the member states of the European Union in 1996 and took over the rates published by EUROSTAT. Nevertheless, the OECD is still calculating the numbers of unemployed persons for non-EU states. Both EUROSTAT and the OECD are using the ILO definition of unemployment, with the difference that EUROSTAT only includes private households in their European Labour Force Survey. This was also adopted by the OECD in 1999 for countries not covered by EUROSTAT for the purpose of comparison. This exclusion of people living in institutions leads, for example, to the fact that members of the armed forces living in barracks are included in the ILO definition but excluded in estimation by EUROSTAT and the OECD as part of the total labour force, the reference category of the

[9] The proper statistical handling of persons on extended leave and seasonal workers were under discussion (ILO 1996, 1998)

standardised unemployment rate. Due to this change in OECD methodology and data, there were small differences in some countries between older and newer estimations (OECD 1999).

EUROSTAT publishes the number of unemployed person based on harmonized national data from the European Labour Force Date. In the case of Germany, this was integrated in a yearly census, called "Microcensus". Due to the fact that the results have to be reported monthly from 2005 on, the *Federal Statistic Office of Germany* (Destatis) modified the Microcensus from a fixed week every year to a continuous census with twelve randomized samples. This is accomplished through an additional survey, which is conducted until 2007, with the aim of testing and guarantying the quality of the new design of the census.

2.2. Registered Unemployment in Germany

In Germany, support of unemployed persons is regulated by the *Third Volume of the Social Code (SGB III)*. It contains, among other things, legislation on the eligibility of unemployment benefits as part of the social security system. It distinguishes between persons in or not in employment, jobseekers, unemployed persons, disabled persons and re-entrants, while these definitions are only valid in the context of the SGB III:

Persons are counted as **not in employment**, if they are not employed or are employed for less than 15 hours per week, and are searching for employment of at least 15 hours with compulsory social insurance contributions. This means that the person has to and is willing to endeavour all possibilities to end the period of having no job, including being at the disposal for a placement through the Federal Employment Service. This availability to work is more precisely defined as being capable of work and willing to take up a reasonable employment under usual or standard conditions.

Jobseekers are defined as persons looking for dependent employment[10] with duration of more than 7 calendar days at home and abroad. They have to register at the Federal Employment Service for placement, must be capable and allowed to pursue the achieved occupation and to reside in Germany.

[10] "Dependent employment" as used here is equivalent to "wage and salaried employment" as used by the International Labour Organisation.

Unemployed persons are *jobseekers*, who are temporarily *not in employment* but searching for employment with compulsory social insurance contributions and have registered at the Federal Employment Service personally. This definition also includes persons who are not receiving benefits. On the other hand, the criteria of receiving unemployment benefits is not sufficient to be counted as unemployed, which is the case for persons who cannot take up work because of illness up to six weeks.

Implicit from this definition we can conclude that persons who are not counted as unemployed are:

- more than marginally employed (more than 15 hours a week),
- not capable of or allowed to work (like individuals younger than 16 years),
- not available without convincing reason,
- not registered personally at the Federal Employment Service, do not show up again for longer than three months, or did not keep an appointment for several times without cogent reason ,
- taking part in active labour market policy measures (exception: short measures called "Trainigsmaßnahmen" before 2004),
- resident only in a foreign country,
- unable to work because of illness which lasts longer than six weeks,
- doing military or compulsory service or are arrested,
- pupils, students or school leavers, who are looking only for professional training / formation,
- foreigners without compulsory work permit (and persons seeking / granting asylum) including the members of their family (with or without receipt of benefits), if the labour market is closed for them,
- on leave (i.e. parental leave)
- short-time working (with up to zero hours of working time)
- 58 years or more and not fulfilling the definition of unemployment in that sense, as they are allowed being not willing to endeavour all possibilities of ending the period of having no job or cannot take up work immediately.

Beside an enlargement of active labour market policies, the move from *Labour Promotion Act* to the *Third Volume of the Social Code* implied important institutional changes referring to registered unemployment (BA 1999a, p.54):

Since the 1st of January 1998, any unemployed person receiving benefits has to personally renew the registration every three months. Not doing so automatically leads to a removal from the register. Exemptions are only given to older (over 55 years) and hard-to-place workers. One side effect of this regulation is the lowering of long-term unemployment due (mainly) to short interruption, which occurs in the case of a belated renewal.

Furthermore, the definition of reasonable employment has been tightened up. While formerly a reasonable employment was defined by the occupation that the unemployed was qualified for, and the net wage a person earned before getting unemployed, now any work is considered reasonable. This leads to a downwards displacement in filling of vacancies from more to less qualified occupations and, therefore, in the composition of the qualifications of the registered unemployed, where the lower the qualification the higher the risk of becoming and staying unemployed.

Any registered unemployed individual is obliged to search for a job on his or her own. The Federal Employment Service is allowed to ask for evidence of an active job search, such as letters of application, reading of newspapers or participation in interview with firms. If this evidence can not be given, an exclusion from the status of a registered unemployed and therefore the exclusion of the possibility of receiving unemployment benefits could come into force. In this context, it is worth noting that training measures can explicitly be used to check for the capability, readiness and willingness of taking up work.

Previous important changes took place in 1985, when two groups were excluded from being counted as unemployed: those who were unable to work because of illness, and persons with an age of over 58 years, who could receive an early retirement pension after the unemployment period.

The **unemployment rate** has been traditionally calculated as the share of the unemployed related to the civilian dependent labour force in Germany. Nowadays, the international more comparable form using the total civilian labour force as reference category is also reported for overall rates, but rates on selected groups of the labour market still generally exist with respect to civilian dependent labour force[11]. The reason for this is a technical one: While there is

[11] **Unemployment rate (BA, no.1):** Unemployment as percentage of civilian *dependent* labour force (Employed persons with compulsory social insurance contributions (without vocational training), marginal employed persons, civil cervants and unemployed persons) = Unemployed / (civilian *dependent* labour force) * 100

monthly data on dependent employment and unemployment available through the social security insurance register, the estimation of the number of overall employment (including self-employment and family workers) is renewed only once a year through the Statistical Office of Germany *(Statistisches Bundesamt),* and is thus much less accurate. Bringing together different national data sources, the Statistical Office corrected their estimates of the overall employment in 1999 as well as in 2006. For 1998, the 1999 revision calculated an increase of about 2 million employed person for that year and also revised the numbers for previous years (BA 1999).

In the context of the *Second Volume of the Social Code* (Sozialgesetzbuch II (SGB II)), introduced in December 2004, the definition of unemployment is broadened over the SGB III to anybody who is needy and capable of work, where both are interpreted in a household context. This refers also to concepts of labour reserve, hidden unemployment, hidden labour force and discouraged workers. In contrast to the former legislation, anybody who is in need of and wants to receive social benefits is presumed to be unemployed if he or she is able to work for at least three hours a day. This is also tested for any person living in the household who could be obliged to pay alimony or palimony. Therefore, discouraged workers are now a category depending on a reasonable maintaining in the household context. This change in legislation has had a clear statistical impact on the official unemployment numbers which are significant higher from 2005 ongoing.

Table 1 highlights the main differences in the survey of the international standardised unemployment definition based on the ILO criteria and the national unemployment definition arising from SGB III respective their operationalisation.

Unemployment rate (BA, no.2): Unemployment as percentage of civilian labour force (civilian dependent labour force *plus* self-employment and family workers) = Unemployed / (civilian labour force) * 100

Table 1: Survey method of ILO- and FES statistic in Germany[12]

	ILO-Statistic (Federal Statistical Office)[13]	Registered Unemployment (FES)
Survey	- Computer Assisted Telephone Interview (CATI) - Random sample 30.000 Interviews - 6 month panel with monthly interview - Monthly expansion-based estimate for total population	- Individual registration at FEA checked by case worker - Complete count - Information can be out of date - Monthly reference day
Definition of "active job search"	- Search on employment of at least 1 hour/week and - Specific search during the last four weeks	- Search on employment of at least 15 hours/week and - The case worker concludes, that the person is using all possibilities of job search
Availability	- Taking up work is possible within two weeks	- Person is willing and able to take up work immediately
Without employment	- No employment or employment less than 1 hour/week	- - No employment or employment less than 15 hour/week

The consequence of these different definitions and survey methods can be illustrated on the ongoing reporting on unemployment: While the FES publishes a reduction of 869,000 to a

[12] Mainly based on Hartmann / Rieck (2005).

[13] 2005-2007

total number of 4,108,000 unemployed persons from March 2006 to March 2007[14], the Federal Statistical Office reported a reduction of about 690,000 to a total of 3.03 million unemployed individuals. But the Federal Statistical Office also states:

"As the telephone survey is based on a random sample of about 30,000 persons per month, the random sampling error has to be taken into account when interpreting the results. For the number of unemployed observed for March 2007, that error may amount to a maximum of +/–190,000. This means that with an observed result of 3.03 million, the actual number of unemployed persons was very probably within the range of 2.84 and 3.22 million in that month."[15]

2.3 The Concept of the Potential Labour Force

The composition of the unemployed is directly connected to the concepts of labour force. In the past 20 years, the concept of the *"Stille Reserve"*[16] to estimate the number of the potential labour force was always a matter of discussion, whereas today, mainly the statistical handling of persons in active labour market policy measures is questioned critically. The high number of participants in job creation, further training and retraining measures leads to a distortion of the unemployment rate through hidden unemployment. Aside from this, early retirement and underemployment, partially due to labour market policy measures like short-time work and partially due to unwanted part-time work wishing full-time work, are a major object of the discussion of extended unemployment rates.

The IAB brings these aspects together in defining the *"Stille Reserve"* as a whole as persons not in employment, who are looking for a job without being registered as unemployed. One part is comprised of jobless people in labour market policy programs, especially in full-time measures of further vocational training (including vocational rehabilitation measures and German language courses) and in early retirement. Taking out this part of the group reveals

---

[14] Source: Pressrelease 023 29.03.2007
http://www.arbeitsagentur.de/zentraler-Content/Pressemeldungen/2007/Presse-07-023 (30.5.2007)

[15] Source: http://www.destatis.de/presse/englisch/pm2007/p1810031.htm (30.5.2007)

[16] Literal translation: hidden reserve or hidden labour force.

the *"Stille Reserve"* in its traditional meaning, which cannot be quantified exactly (Fuchs 2003; Brinkmann/Klauder/Reyher/Thon 1987).

The following section discusses alternative implementations of these concepts using German register data, based on individual status information: Starting from the duration of unemployment as calculated by the Federal Employment Service (Bundesagentur für Arbeit (BA)) six alternative concepts are developed. Contrasted by two additional benchmark concepts, empirical evidence shows that there is a high impact on the results depending on the underlying definition of unemployment.

3 The Duration of Unemployment

The individual status of unemployment as discussed in section 2 is time-dependent. Different underlying definitions of unemployment can lead to different interpretations of individual employment histories in the data. We now focus on the duration of unemployment. Referring to this one sub-category of the unemployed is the number of long-term unemployed persons, which is generally seen as an indicator of the persistence of unemployment. Up to 1985 all persons who worked not more than 13 weeks in a row after their unemployment registration were counted as long-term unemployed after 12 months of unemployment. Since 1985, any interruption (employment, illness of more than 6 weeks, further training etc.) leads to a restart of the time counting for the 12 month period. In practice, this change had a clear statistical impact. A recalculation of the long-term unemployment using the criteria from 1985 for the period before this change (1977 to 1984) shows a statistical reduction of between 12 and 23 percentage points through the new definition (Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen 1998). This must be seen as an additional, cumulative factor beside a general underestimation of long-term unemployment (Auer 1984, Karr 1997). While in 1994 the national rate of long-term unemployment for West Germany was reported by the IAB as 32.5 percent, a special analysis of the Microcensus data showed a rate of 45.8 percent (Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen 1998). While any interruption is still counted as an outflow from and an inflow into unemployment, the introduction of the SGB III changed again the definition of the long-term unemployed. The duration is now a retrospectively calculated sum of periods rather than a fixed status at a point of time. This definition is used to give access to benefits and active labour market policy measures.[17] Not deducted from the duration of unemployment is any period of

- participating in active labour market policy measures,
- illness or maternity protection,
- childcare or long-term care of a family member in need,
- employment or self-employment up to six months,

[17] Recently, the BA developed a concept of unemployment including participation periods in labour market programms, which also indicates a retrospective sum of periods rather than a status at a point of time and is calculated in two versions as the time of being unemployed including periods of participating in active labour market policy measures. While one version also includes periods of subsidized work, this is not the case in the second version.

- periods without the legal basis to take up employment
- short interruptions without certification of a reason.

The following section deals with several concepts, which periods should be included into the calculation and which periods should be counted as interruption.

3.1. Concepts for the Implementation of Unemployment Duration

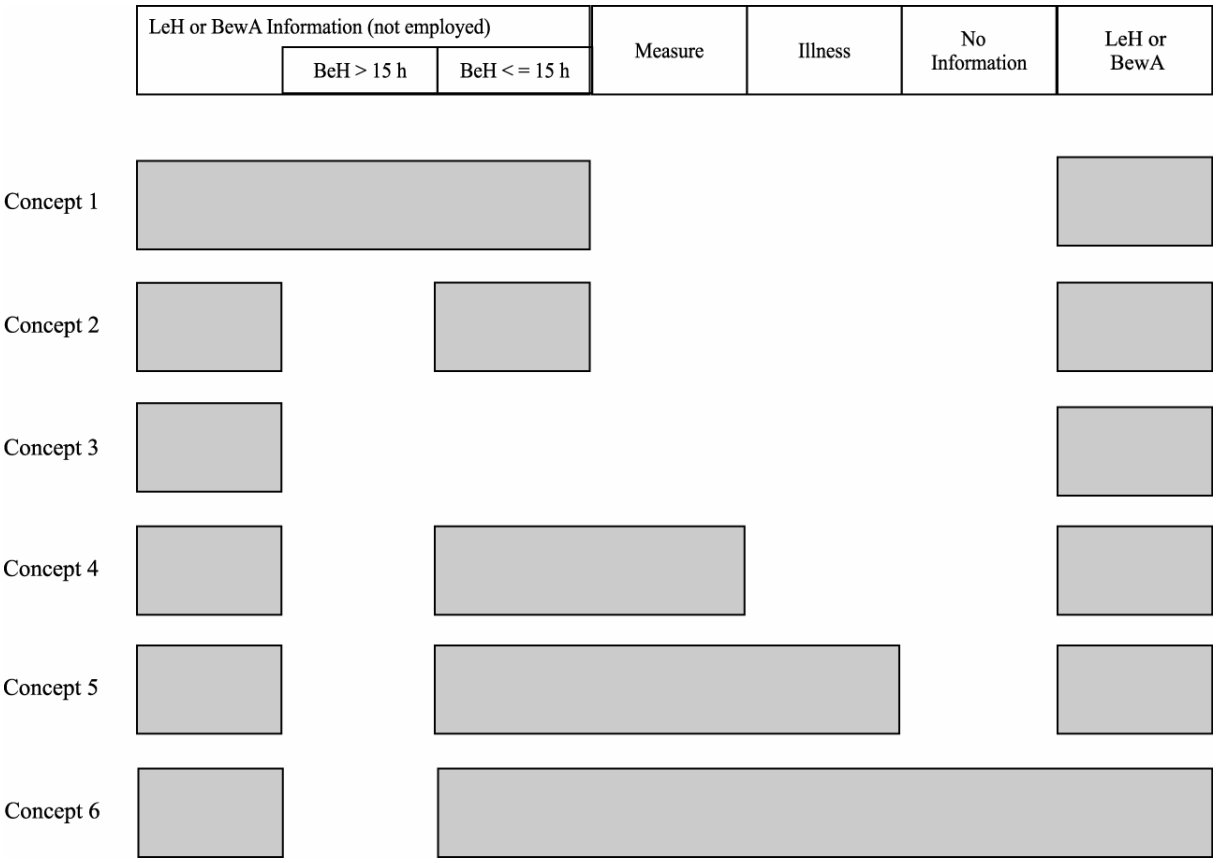
We now establish general links between the different definitions of unemployment and the actual calculation of unemployment duration. Based on the general information available in German register data, unemployment duration can be measured according to one the following concepts:

1. **Concept 1**: Each uninterrupted unemployment period shown by the administrative record.
2. **Concept 2**: Concept 1, corrected for periods of dependent employment over 15 hours a week.
3. **Concept 3**: Concept 2, corrected also for periods of dependent employment less than 16 hours a week.
4. **Concept 4**: Concept 2, with added periods of participating in (any) active labour market policy measures.
5. **Concept 5**: Concept 4, with added periods of illness, identified by a variable on the reason for leaving and entering the registered status of unemployment.
6. **Concept 6**: Concept 5, with added period(s) without information on the employment status of individuals, presuming that most people have to search for employment and are willing to work under good conditions, even if they are not registered. (Problem: This concept includes also self-employed, civil servants, etc. - imputation could be a solution.)

In the following we illustrate these rather theoretical concepts by using a fictitious individual employment history that may occur in real data. In particular, we illustrate how the foregoing concepts are implemented to determine the status of unemployment and to calculate its duration. The following figure presents some - partially parallel - spells of different labour market states over seven time periods. In real data these spells were in fact merged from different administrative registers and would have different lengths. The figure shows the resulting unemployment spells derived from the above six concepts. It is evident that the resulting number of unemployment spells and their duration differ across the concepts as we obtain two to three

unemployment spells and the cumulative unemployment duration ranges from two (concept 3) to six (concept 6) periods. The length of the last unemployment duration varies between one (concept 1 to concept 5) and five (concept 6) periods.

Figure 1: Illustration of six unemployment concepts with the help of a fictitious employment history.



3.2. Empirical Evidence

In our empirical exercise we are implementing the above six concepts in merged administrative data from Germany. In addition we compare our results to two benchmark proxies for unemployment which were used in the former literature on unemployment using this data. In our empirical analysis we use the IEBS which are made available for scientific use by the research data center of the FES (FDZ).[18] They include information on dependent employment (Source: BeH), registered unemployment and on job search (BewA), unemployment benefits (LeH) and participation in active labour market policies (MTG) as a representative 2.2 percent

[18] For Information on FDZ and their data see http://fdz.iab.de.

sample covering about 80 percent of the labour force (Jacobebbinghaus/Seth 2007).[19] Whereas information on employment are included from 1990 until 2003 and unemployment benefits from 1990 until 2004, information on participation in active labour market policy measures are only reliable in the period 2000-2004. We use data between 2000 and 2003 exclusively in order to focus on a period for which all registers are available. The key properties of our sample are described in Table 2.

Table 2: Number of episodes and parallel episodes in the IEBS period 2000-2003.

	BeH	LeH	MTG	BewA
Number of episodes	4,937,224	1,059,985	313,786	1,516,699
Parallel episodes				
BeH	275,630	199,901	107,918	443,013
LeH	199,901	1,877	136,531	776,205
MTG	107,918	136,531	19,068	170,728
BewA	443,013	776,205	170,728	4,992

While implementing the six concepts using this data set, we face certain problems such as data inconsistencies and missing interval information. Data inconsistencies due to unfeasible overlapping of register information have already been extensively analysed by Jaenichen et al. (2005) and Bernhard et al. (2006). For this reason we do not directly address them in our implementations but we highly recommend performing the data corrections suggested by Bernhard et al. (2006) before applying our implementations to the data. The main focus of this work is to address the problem of missing interval information. We will, however, take into account contradictory information about the length of training measures which are funded by the FES.

[19] For a detailed description see Hummel et al. (2005) or Jabobebbinghaus and Seth (2007).

Since there are several ways to deal with the present data problems, we present different approaches to implement the above six concepts in the German register data. In a first step the researcher has to decide which administrative information is used as the core information for an unemployment period, since BewA and LeH provide sets of unemployment information which are related to the receipt of unemployment compensation or to job seeking activities. For this reason it has to be defined from which source the unemployment information is taken: BewA, LeH or from both (BewA+LeH). The choice should be made according to the specific research question at hand. When we are interested in some economic effects of unemployment benefits, the LeH may be sufficient. Using only BewA may be useful for analysing registered unemployment. To get most comparable results of unemployment, using information from both BewA and LeH may be the right decision. With this paper we provide program code for both approaches, but we decided to focus the following analysis to unemployment information taken from the BewA only. This is only done for reasons of simplifaction and does not mean that using the BewA information is more suitable for empirical analysis.

In order to implement the above six concepts, the researcher has to make further decisions:

1. **How to deal with parallel employment information?**
   In case of parallel full-time or part-time employment and unemployment information, the researcher has to decide if the information is assessed as employment or unemployment. While there are some regulations that allow registration as unemployed parallel to dependent employment, this, for example, could be interpreted as underemployment (see also chapter 2.3).

2. **How to deal with training measures, illness or out of the labour force information?**
   In a second step the researcher has to make a decision about the unemployment-status of periods of training measures, illness or out of the labour force. Here we face similar questions as when analysing employment. In general (short) illness is not shown by register data on employment spells and therefore it is not counted as an interruption. While some “training measures” in Germany are used to check for the readiness to engage in work, others are used to train unemployed persons to write a letter of application or give them practical advice in direct connection to a subsequent job. Further vo-

cational training can range from short modules of several weeks to long term measures lasting two years or more, providing a recognized vocational qualification.[20]

In the first step – how to deal with parallel information - the researcher has to choose among three possibilities:

1. **Implementation A** (based upon concept 1):
   Each uninterrupted unemployment period shown by the administrative record, i.e., each uninterrupted BewA spell, including those with parallel BeH-spells, are considered as unemployment spells.

2. **Implementation B** (based upon concept 2):
   It includes each uninterrupted unemployment period shown by the administrative record, corrected for periods of dependent full time employment, i.e., all uninterrupted BewA spells except for those with parallel BeH-spells coming from a full time employment. If a parallel BeH-spell comes from a full time employment, the spell is considered as an employment spell, if it comes from a part time employment, it is considered as an unemployment spell (the variable used to identify full time or part time employment is "*Erwerbsstatus*").

3. **Implementation C** (based upon concept 3):
   Each uninterrupted unemployment period shown by the administrative record, corrected for all periods of employment. Thus, only BewA- spells without any kind of parallel BeH-spell are considered as unemployment spells.[21]

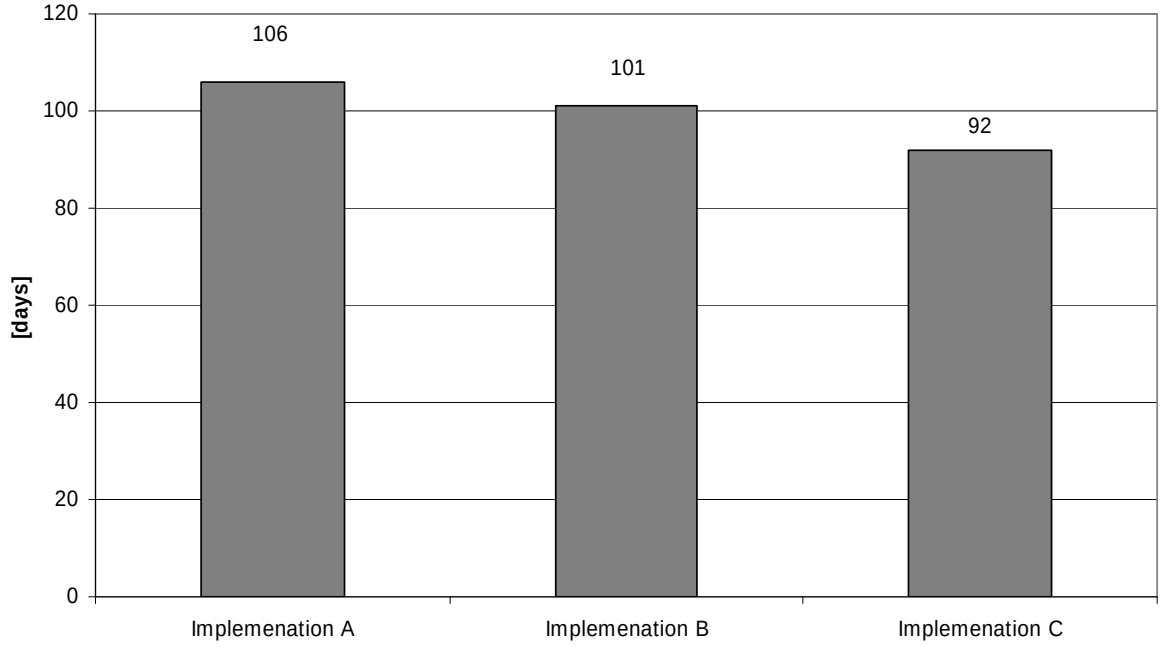
---

[20] Furthermore, if one includes periods of labour market policy measures, one has to opt for their correct lengths. Based on Waller (2007), we implement two alternative approaches to define the end date of a labour market policy measure which we call a) the "naïve concept" and b) the "standard concept". See Appendix A for more details on these impelementations. In this paper we focus on alternative b). The results of implementation a) can be taken from an internet appendix.

[21] This implementation may be considered as somewhat crude because in many cases such overlapping periods correspond to employment subsidies with focus on the second labour market (e.g. ABM, SAM). It would also be plausible to define these cases as unemployment, but we decided to not distinguish at all between different types of labour market measures in order to keep the number of implementations manageable. As with any other analysis, a decision about this is in the responsibility of the researcher.

The following figure presents the differences in the median unemployment lengths depending on which implementation is chosen:

Figure 2: Median unemployment length for males for different treatments of parallel employment spells (Source: BewA)



Additional information about the participation in active labour market policy measures can be found in the "*Maßnahme-Teilnehmer-Grunddatei*" (MTG). This information offers the possibility of defining further refinements of unemployment (please note that the MTG only contains information about the main policy measures funded by the Federal Employment Agency). In the second step there are again three alternatives:

1. **Implementation I** (based upon concept 4 of the previous chapter):
   Spells with information about the participation in active labour market policy measures are also considered as unemployment spells.

2. **Implementation II** (based upon concept 5):
   Spells with information about illness are also considered as unemployment spells.[22]

[22] Sick leave during an employment period does not interrupt the employment spell.

These periods can be directly identified from the BewA or they are identified using the variable on the reason for leaving and entering the registered status of unemployment.

3. **Implementation III** (based upon concept 6):
   Periods without any information about the employment status of an individual are considered as unemployment periods. We have to keep in mind, though, that people who cannot even be found in the data set could be unemployed according to this concept.

The different implementation strategies are summarized in table 3.

Table 3: Summary of Implementation strategies for unemployment in the IEBS.

Decision	Alternative	Brief explanation
1	A	Unemployment = each uninterrupted BewA or LeH period (also including any parallel employment information)
	B	Unemployment = each uninterrupted BewA or LeH period except for parallel periods of dependent full time employment.
	C	Unemployment = each uninterrupted unemployment period without any kind of parallel employment information.
2	I	Unemployment = each uninterrupted BewA, LeH or MTG period except for parallel periods of dependent full time employment.
	II	Unemployment = each uninterrupted BewA, LeH, MTG or "illness" period except for parallel periods of dependent full time employment.
	III	Unemployment = each uninterrupted BewA, LeH, MTG or "illness" period and all periods without any information about the employment status (except for parallel periods of dependent full time employment).

Depending on the question, the one has to choose which concept best corresponds with his analyses. In order to illustrate the importance of each of the decisions, we present exemplarily the different outcomes of the median length of unemployment for males.[23] For this purpose we hold constant all the remaining decisions, except the one in question:

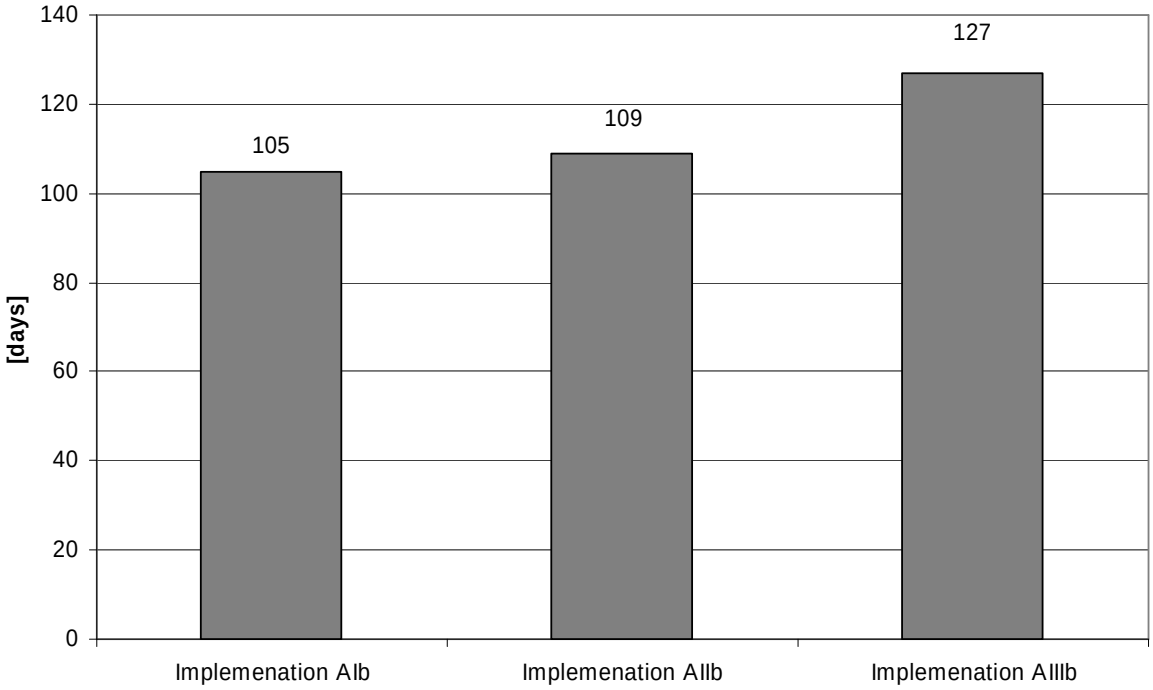
- Information is taken from BewA
- We consider unemployment spells with any kind of parallel employment information as unemployment (Implementation A);
- Training measures are also treated as unemployment periods (Implementation I);
- The duration of the length of a labour market policy measure is determined using the "standard concept" (Implementation b). See Appendix A for more details.

This makes the implementation AI (BewA) the default implementation in what follows when we report descriptive figures. For this purpose we also hold the implementation of the length of training measures constant and choose the "standard" implementation b as default. By doing this we can easily illustrate the changes in unemployment duration when we deviate from implementation AIb.

The following figure shows the median length of unemployment for the possibilities AIb, AIIb and AIIIb:

[23] Moreover, we use seven days as the maximum tolerated interruption length, because there is some evidence that short interruptions could be caused by minor failures in the data generating process. Also, official calculation of unemployment duration tolerates interruptions caused by short employment. Since results are likely to depend on the chosen value, it can be set as a parameter in our STATA code.

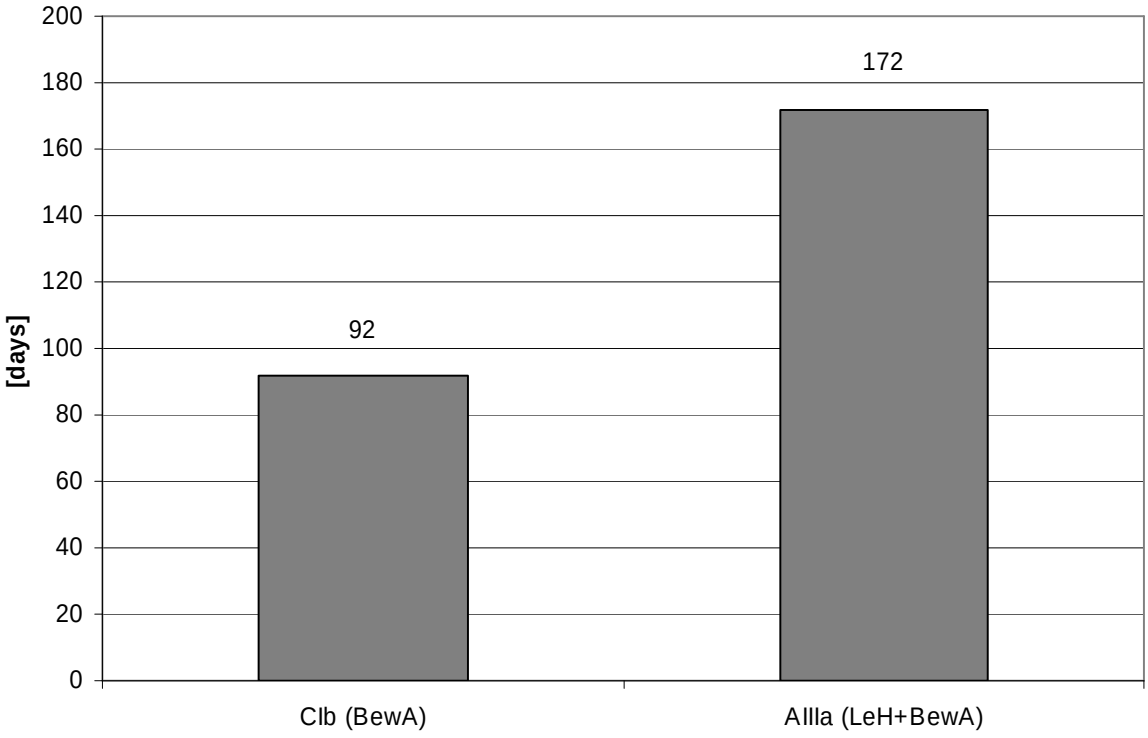
Figure 3: Median unemployment length for males for different treatments of training and illness periods (Source: BewA)



The distinction (A/B/C) offers three different implementations of unemployment. Based upon these three definitions and using information from the MTG, we can combine the three refinements (A/B/C, I/II/III, and a/b) among each other in order to get, e.g., A-I-a or C-II-b. This leaves altogether 21 different definitions of unemployment based upon the concepts 1 to 6 from the previous chapter. These 21 definitions can be computed using LEH, BewA or both as the core information of the unemployment period. Our implementations therefore yield 63 definitions of unemployment in the IEBS.

Even though the differences in the median length in the examples presented above are already remarkable, they still are comparably small to those if more than one of the decisions is changed. Figure 4 presents the maximum difference in median duration for males over all implementations discussed in this section (see also figure 6 in Appendix C where we alternate over a variety of stratifications such as sex, age, east/west, etc.).

Figure 4: Minimum and maximum median length for males of all the discussed implementations



Before we discuss the descriptive statistics for the different implementations, two additional proxies for unemployment in the IAB data are introduced. These proxies are based on the information content in the IAB employment sample (IABS) and they have already been used in the econometric literature (see e.g. Fitzenberger and Wilke, 2004 or Lee and Wilke, 2005). For this reason we refer to these concepts as benchmarks in what follows. Moreover, necessary assumptions for the programming and differences in the definition of censoring for each of the concepts mentioned above are presented.

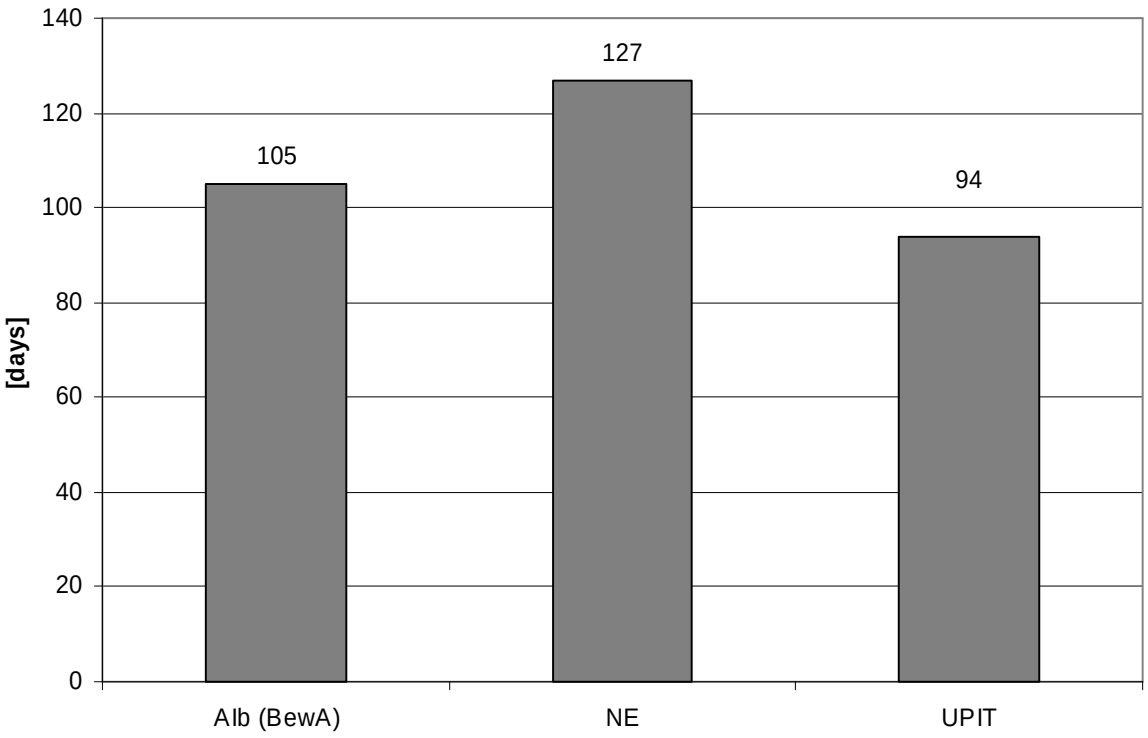
1. **Nonemployment (Fitzenberger / Wilke) NE**: All periods of nonemployment after an employment period which contain at least one period with income transfers by the German Federal Employment Service. A period of nonemployment is right censored if the last nonemployment spell is not followed by an employment spell. (If an income transfer spell (LeH) is parallel to an employment spell (BeH), the period is treated as an employment spell.)

2. **Unemployment with permanent income transfer (Lee / Wilke) UPIT**: All periods of nonemployment after an employment period with a continuous flow of unemploy-

ment compensation from the German Federal Employment Service. The maximum interruption in compensation transfers is one month (30 days). The limit of one month is chosen because unemployed who do not receive unemployment compensation for more than one month loose their social insurance protection. This implies that there is a strong financial incentive for not having long gaps and if they are observed, they may be related to out of the labour market periods. An observation is marked as right censored if the interruption in transfer payments is longer than 30 days or if there is no other observation after a compensation payment spell. (If an income transfer spell (LeH) is parallel to an employment spell (BeH), the period is treated as an employment spell.)

The following figure presents the median length of unemployment for males according to the example discussed above (males, AIb (BewA)) and the two benchmark proxies NE and UPIT:

Figure 5: Median unemployment length for males compared to the two proxies NE and UPIT



Both proxies for unemployment are conditional on a preceding employment period. This is an important difference compared to the concepts described earlier in this section. Conditioning on a foregoing employment period has two main implications: First, it restricts the set of unemployment periods to those who were just transiting into unemployment, while it excludes individuals with a loose labour market connection and any employment spell. Second, by conditioning on an observed employment period in the data, the sample of unemployment spells becomes more representative with respect to transitions from socially insured employment into unemployment. Thus, problems of sample selection do not apply.

A long list of descriptive statistics for all different definitions based on the IEBS from 1.1.2000 to 31.12.2003 can be found in the appendix. Tables 5 and 6, however, only stratify with respect to gender in order to keep the length of this paper tractable. Additional stratifications with respect to age groups and East/West Germany are available as an internet appendix http:/doku.iab.de/fdz/reporte/2007/.....zip). This appendix also presents detailed results for implementations a and b and which underlying register (BewA, LeH or both) was chosen.

In the following we briefly discuss how the different definitions of unemployment are related in terms of unemployment duration, since the different implementations imply a certain ordering. Table 4 shows the resulting ordering for a given unemployment period.

Table 4: Ordering of the different implementations with respect to the length of an unemployment spell

Length of unemployment				
A	≥	B	≥	C
III	≥	II	≥	I
a	≥	b		
NE	≥	UPIT		

1. Implementation A results in a longer duration compared to implementation B and C, the latter yields the shortest duration. This is due to the fact that the implementation A includes all unemployment periods shown in the administrative data, regardless of any kind of parallel employment information. Following implementation B or implementation C, a parallel full time employment or any kind of parallel employment, respectively, leads to an interruption and thus to a shorter unemployment spell.

2. Implementation I is shorter than II because it excludes periods of illness. Implementation III yields the longest unemployment duration because it also includes out of the labour force periods. In many cases implementation III merges two or more durations according to implementations I and II.

3. Implementation a treates the end of a policy measure as the end of unemployment while according to implementation b, unemployment ends when an employment starts. Therefore, the spell length in the latter case is shorter.

4. Since the NE proxy relies on weaker requirements than UPIT in terms of tolerated interruption length, the observed unemployment spell is longer in the first than in the latter case.

Marginal distributions of unemployment duration for the implementations may not possess this ordering because of different number of observations. It is less clear to rank implementations according to their number of observations. It is, however, evident from the tables in the appendix that there are in any case several hundred thousand observations which are more

than enough for an empirical application. Since NE and UPIT require foregoing employment periods, they have the least observations.

Finally we remark on the notion of censoring in the computed unemployment periods. Contrary to the two benchmark proxies NE and UPIT, censoring is only at the start and the end of the observation period (1.1.2000 and 31.12.2003). Note that this notion of censoring differs to those of NE and UPIT: There, left censoring is impossible because it requires the observed transition from employment to unemployment. Right censoring is not only due to the end of the data, it is also at the end of income transfers if there is no observed transition to employment. This reflects that unobserved periods impose some uncertainty in the length of the true unemployment period which is not observed. Since in a variety of applications there are still not estimators ready for use which are able to deal with left censoring in an appropriate manner, left censoring may cause difficulties in applied empirical work.

4. Further Topics

In addition to missing or conflicting spell information, the applied researcher faces the problem that registers of the social insurance system are typically not representative with respect to the whole population. In order to appear in an administrative record, one needs to be in contact with one of the relevant data generating administrations. In Germany, this is unlikely the case for self employed civil servants and other individuals who never contributed to the national insurance system. It implies that during a period of unemployment these individuals are likely to be ineligible for income transfers from the FES. While these unemployed consider themselves unemployed, they may not register as job seekers as they do not consider this beneficial for them. For this reason, the BewA is not representative because appearance in this register depends on a subjective choice without evident financial implications. Hence, such sample selection issues should also to be taken into account when working with this data. A non-representative sample of unemployment spells cannot be considered as a random sample and results of statistical analysis may be biased. This is in particular the case if selection is not independent of the variable we try to explain. This may prevent the researcher from consistently estimating the causal effect of policy interventions using this data. Since the degree and the kind of selectivity are unknown, we cannot suggest a general solution to this problem. We think, however, that for some specific empirical questions one may create samples which are (almost) representative. This is, for example, the case if one conditions the sample of unemployment spells on those which have foregoing employment durations with a minimum length

before unemployment[24] and to relate unemployment periods to the receipt of income transfers. The two benchmark definitions NE and UPIT proceed in this way. Questions related to the receipt of unemployment benefits can be analysed quite well with this data. Their application fields are, however, limited to quite specific empirical problems and in many problems it may not make sense to use them in applied analysis.

5. Discussion

In this paper we discuss several theoretical and legal concepts of unemployment. As a consequence, the labour market state and the duration of an unemployment period depend on the nature of the concepts. This is likely to cause difficulties for applied research on unemployment, since results may depend on the definition of unemployment.

Given the theoretical notions of unemployment we focus on the question how the labour market state unemployment and the duration of unemployment can be defined in real world data. In our empirical work we use the Sample of the Integrated Employment Biographies (IEBS), which is the German merged administrative individual data. In addition to two well known benchmarks, we develop more than 60 different implementations of unemployment in this data. A short descriptive analysis shows considerable differences in the number of unemployment spells and in the length of unemployment periods, which provides evidence for the importance of our work. Our implementations are available for users of this data from the Research data centre of the FES (fdz.iab.de). They are provided as ready for use Stata do files.

Given the theoretical concepts and the data at hand, we face several difficulties: In some cases the legal definitions of unemployment cannot be exactly implemented because not all necessary information is available in the data, e.g., the true unemployment duration according to a theoretical concept cannot be exactly identified from the data. For this reason we use several variants of the implementation that take this difficulty into account. Our implementations therefore provide alternative unemployment classifications in the data.

---

[24] To be eligible for the receipt of unemployment compensation from the FES, one needs to be at least once employed for a minimum duration during a claim period which can be more than 5 years depending on the age of the unemployed and the specific legal conditions in force. This condition may not be satisfied for young unemployed just starting their career. For this reason the sample of NE and UPIT is also not representative with respect to all transitions from employment to unemployment.

Unobserved periods in the employment trajectories are an important weakness of this data, which may cause a variety of problems in applied analysis. It requires careful judgement of whether an empirical question can be solved despite the difficulty and which definition of unemployment seems to be most appropriate. Indeed, our descriptive figures reveal remarkable differences in the number of observations and in the length of unemployment across the implementations. This provides additional motivation for the content of this work. Future work may address the question whether missing information or differences in the length of unemployment periods across the implementations are random or whether they are correlated with observable variables.

Our work delivers more than 60 different implementations of unemployment in the data. This is already quite comprehensive and we think that we have addressed several important topics. There is, however, further scope for interesting extensions. For example, one could define finer classifications which distinguish between the type of the labour market measures. This is important because labour market programs are heterogeneous in their purposes.

6. References

Auer, P. (1984). Maßnahmen zur Reintegration von Langzeitarbeitslosen. Discussion Paper IIM/LMP 84-20, Social Science Research Center, Berlin.

BA – Bundesanstalt für Arbeit (1999). Arbeitsmarkt 1998. Amtliche Nachrichten der Bundesanstalt für Arbeit, Jg. 47, Sondernummer vom 12.06.1999, Nürnberg.

Bernhard, S., Dressel, C., Fitzenberger, B., Schnitzlein, D. and Stephan, G.(2006). Überschneidungen in der IEBS: Deskriptive Auswertung und Interpretation. FDZ Methodenreport, 04/2006, Nürnberg.

Biewen, M. and Wilke, R.A. (2005). Unemployment Duration and the Length of Entitlement Periods for Unemployment Benefits: Do the IAB Employment Subsample and the German Socio-Economic Panel Yield the Same Results? Allgemeines Statistisches Archiv, 89(2), 209-236.

Brinkmann, Ch., Klauder, W., Reyher, L. and Thon, M. (1987). Methodische und inhaltliche Aspekte der Stillen Reserve, in: Mitteilungen aus der Arbeitsmarkt und Berufsforschung (MittAB), Jg. 20. Nürnberg.

Fitzenberger, B. and Wilke R.A. (2004). Unemployment Durations in West-Germany Before and After the Reform of the Unemployment Compensation System During the 1980s. ZEW Discussion Paper No. 04-24. ZEW, Mannheim.

Hartmann, M. and Rieck, T. (2005). Erwerbslosigkeit nach dem Labour-Force-Konzept – Arbeitslosigkeit nach dem Sozialgesetzbuch: Gemeinsamkeiten und Unterschiede, in: Wirtschaft und Statistik, 4/2005, Wiesbaden.

Hummel, E., Jacobebbinghaus, P., Kohlmann, A., Oertel, M., Wübbeke, C. and Ziegerer, M. (2005). Stichprobe der Integrierten Erwerbsbiografien, IEBS 1.0., Handbuch-Version 1.0.0 FDZ Datenreport No. 6/2005. Nürnberg.

ILO – International Labour Office, Bureau of Statistics (1983). Resolution concerning economically active population, employment, unemployment and underemployment. In: Bulletin of Labour Statistics, 1983-3, Geneva.

ILO – International Labour Office, Bureau of Statistics (1996). Resolution of the Prag Meeting. In: Bulletin of Labour Statistics, 1996-1, Geneva.

ILO – International Labour Office, (1997). Yearbook of labour statistics, Geneva.

ILO – International Labour Office, Bureau of Statistics (1998). General Report on the Sixteenth International Conference of Labour Statistics, http://www.ilo.org/english/120stat/techmeet/16thicls/report4.htm (25.08.98), Geneva.

Jacobebbinghaus, P. and Seth, S. (2007). The German Integrated Employment Biographies Sample IEBS. In: Schmollers Jahrbuch, forthcoming.

Jaenichen, U., Kruppe, T., Stephan, G., Ullrich, B. and Wießner, F. (2005). You can split it if you really want * Korrekturvorschläge für ausgewählte Inkonsistenzen in IEB und MTG. FDZ Datenreport Nr. 04/2005, Nürnberg.

Karr, W. (1997). Die Erfassung der Langzeitarbeitslosigkeit – Ein kaum beachtetes Meßproblem, in: IAB Kurzbericht, Nr. 5, Nürnberg.

Lee, S. and Wilke, R.A. (2005). Reform of Unemployment in Germany: A nonparametric Bounds Analysis using Register Data. ZEW Discussion Paper 05-29.

Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (Edt.)(1998). Arbeitslose, Langzeitarbeitslose und ihre Familien, Landessozialbericht, Vol. 8, Düsseldorf.

OECD - Organisation For Economic Co-Operation And Development (1999). Quarterly Labour Force Statistics, No. 1/1999, Paris.

Sorenntino, C. (2000). International unemployment rates: how comparable are they? Monthly Labor Review, June 2000, Vol. 123, No. 6, Washington.

Waller, M. (2007). Do Reported End Dates of Treatments Matter for Evaluation Results? – An Investigation Based on the German Integrated Employment Biographies Sample, FDZ Methodenreport, 1/2007, Nürnberg.

Werner , H. (1984). Unterschiede in der Erfassung der Arbeitslosigkeit. Mitteilungen aus der Arbeitsmarkt und Berufsforschung Nr. 3, Nürnberg.

Zimmermann, B. (2006). Arbeitslosigkeit in Deutschland – Zur Entstehung einer sozialen Kategorie. Campus, Frankfurt/New York.

Appendix A: Definition of the length of a labour market policy measure

If periods of active labour market measures count as unemployment, the researcher has to decide in a fourth step how the length of such measures is defined in the data. Since there are different sources of information in the MTG, Waller (2007) suggests two concepts to deal with possible contradictory information[25]:

1. **Implementation a** ("naïve concept"):
   It relies on the information about the participation in labor market policy measures, i.e., we use the end date of the measure given in the data as the real end measure, even if there are parallel employment spells.

2. **Implementation b** ("standard concept"):
   It relies on the information from employment spells, i.e., if an employment spell starts before the end date of the labour market measure given in the data, we consider the observation as an employment spell.

The results presented in this paper correspond to Implementation b, the "standard concept". We also worked out the results for Implementation a, which can be taken from an internet appendix.

[25] The following two implementation approaches are therefore motivated by specific data inconsistencies and not by missing spell information.

Appendix B

Table 5: Descriptive Statistics for marginal unemployment duration distribution using the BewA stratified by gender

Implementation	num_obs	mean	median	0.25-Quantile	0.75 Quantile
A	587278	215.8124	115	51	268
males	341310	201.2612	106	47	244
females	245968	236.0039	125	56	298
B	589072	211.0129	109	46	261
males	342735	196.1846	101	43	238
females	246337	231.644	122	52	291
C	593740	185.156	94	37	223
males	348339	177.1649	92	36	212
females	245401	196.4991	100	40	244
AIb	635683	216.0664	113	46	273
males	366743	203.122	105	44	251
females	268902	233.7174	122	51	301
missings	38	239.3684	184	92	365
BIb	636680	212.0321	108	43	268
males	367559	198.8578	100	40	245
females	269083	230.0239	121	47	295
missings	38	239.3684	184	92	365
CIb	634145	190.9311	94	34	237
males	370163	182.976	92	33	222
females	263944	202.0805	99	36	254
missings	38	239.3684	184	92	365
AIIb	625774	228.1369	117	48	284
males	357501	216.1743	109	45	269
females	268235	244.0791	123	51	311
missings	38	239.3684	184	92	365
BIIb	628301	223.8717	112	44	277
males	359302	211.5826	105	42	261
females	268961	240.2863	122	48	306
missings	38	239.3684	184	92	365
CIIa	609526	231.9031	108	41	290
males	350611	223.2859	103	39	275
females	258877	243.5728	117	43	311
missings	38	239.3684	184	92	365
CIIb	632355	204.0694	97	35	250
males	364231	196.3675	94	34	240
females	268086	214.5285	103	37	273
missings	38	239.3684	184	92	365

AIIIb	489461	281.7411	136	53	365
males	278833	270.5537	127	50	350
females	210590	296.5614	151	57	385
missings	38	239.3684	184	92	365
BIIIb	204074	217.6278	109	31	275
males	106654	218.6805	108	29	278
females	97382	216.4663	111	33	275
missings	38	239.3684	184	92	365
CIIIb	204074	217.6278	109	31	275
males	106654	218.6805	108	29	278
females	97382	216.4663	111	33	275
missings	38	239.3684	184	92	365

Table 6: Descriptive Statistics for marginal unemployment duration distribution for the benchmark concepts using the LEH stratified by gender

Implementation	num_obs	mean	median	0.25-Quantile	0.75 Quantile
NE	318596	246.2681	136	52	352
males	195928	240.1096	127	51	337
females	122668	256.1047	151	55	365
UPIT	284369	187.9828	100	40	243
males	174360	180.419	94	39	226
females	110009	199.9711	110	42	273

Appendix C

Figure 6: Minimum and maximum median length of all the implementations over all stratifications (aged >49 years)