

Alda, Holger; Bender, Stefan; Gartner, Hermann

Working Paper

The linked employer-employee dataset of the IAB (LIAB)

IAB-Discussion Paper, No. 6/2005

Provided in Cooperation with:

Institute for Employment Research (IAB)

Suggested Citation: Alda, Holger; Bender, Stefan; Gartner, Hermann (2005) : The linked employer-employee dataset of the IAB (LIAB), IAB-Discussion Paper, No. 6/2005, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg

This Version is available at:

<https://hdl.handle.net/10419/31906>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The linked employer-employee dataset of the IAB (LIAB)

Holger Alda, Stefan Bender, Hermann Gartner

The linked employer-employee dataset of the IAB (LIAB)

Holger Alda, Stefan Bender, Hermann Gartner

Auch mit seiner neuen Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

Also with its new series "IAB Discussion Paper" the research institute of the German Federal Employment Agency wants to intensify dialogue with external science. By the rapid spreading of research results via Internet still before printing criticism shall be stimulated and quality shall be ensured.

Inhaltsverzeichnis

| | |
|--|----|
| Abstract | 4 |
| 1 Setting up a linked employer-employee dataset at the IAB (LIAB) | 5 |
| 2 Versions of the LIAB | 11 |
| 3 Studies using linked employer-employee panel data: possibilities and examples | 18 |
| 4 Data disclosure and utilisation for external researchers | 21 |
| 5 Summary | 22 |
| Appendix..... | 27 |

The linked employer-employee dataset of the IAB (LIAB)*

Abstract

The linked employer-employee dataset of the IAB (LIAB) is created by matching the data of the IAB establishment panel and the process-produced data of the Federal Employment Services (in particular the IAB). The distinctive feature of these data is the combination of information about individuals and details concerning the firms in which these people work. This facilitates a number of new methods of analysis and the first empirical examination of theoretical concepts. This applies not only to economic questions but for all areas of empirical social research. Studies have so far been conducted on subjects in labour economics, personnel economics and sociology. Some versions of the LIAB are accessible via the research data centre (Forschungsdatenzentrum – FDZ) of the Federal Employment Services (Bundesagentur für Arbeit – BA) at the IAB.

Keywords: Linked employer-employee data

JEL Classification: C81, C23, J23

* The first chapter of this paper are based on Bellmann, Bender and Kölling (2002) and the paper constitutes an update. It was augmented in Chapter 2 by Alda, 2004a. This paper was last updated in the fourth quarter of 2004.

Introduction

Linked employer-employee datasets can be used for all issues concerning interactions between firms and employees (Hamermesh 1999: 25-41). Willis (1986: 589) already pointed out in his paper for the Handbook of Labor Economics that the development of linked employer-employee datasets was crucial for progress in the research fields of wage structure and human capital theories. Abowd/Kramarz (1999) provides an overview of the basic possibilities of linking firm and employee data. They differentiate among other things between links of administrative data alone, those of survey data alone and various combinations, and cite some 100 studies from 15 countries. This is therefore an area of research that is developing rapidly.

In this paper we describe the setting up of the linked employer-employee dataset at the IAB (LIAB), the analytical potential of this dataset and the possibilities of data disclosure and data utilisation for external researchers.

1 Setting up a linked employer-employee dataset at the IAB (LIAB)

One of the few possibilities of producing employer-employee datasets for Germany is to link the process-produced person-specific data (in particular the employee history) of the IAB with the IAB establishment panel.¹ The IAB establishment panel contains annual information about establishment structures and personnel-policy decisions in the period from 1993 up to the current limit. In principle it is possible to link establishment information from the panel with the information about employment, unemployment and measures of the employees covered by social security. In this way flows can be analysed directly in the context of establishment variables – some of them time-dependent – which provide information about the employee structure, recruitment and dismissal practices and the characteristics of the employment relationships.

¹ It is also possible to create a linked employer-employee dataset on the basis of the employee history alone since an aggregation is possible by means of the establishment code. This second possibility is not to be described here. Readers interested in this are requested to look at the publications on this subject (Bauer, Bender and Bonin 2004, Beblo, Bender and Wolf 2004, von Wachter and Bender 2004).

In particular the employee history is of interest for constructing the LIAB (cf. Bender/Haas 2002, Bender/Haas/Klose 2000 and the literature mentioned in the articles). The basis of the employee history is the integrated notification procedure for the health, pension and unemployment insurances, which was introduced in January 1973.² This procedure requires employers to notify the social security agencies about all employees covered by social security. Notifications have to be submitted to the social security agencies within certain periods at the beginning and the end of an employment relationship and on 31 December each year for all employment relationships subject to social security contributions. On 1 January 1999 the regulation to revise the notification procedure (Regulation on Data Collection and Transmission - DEÜV) came into force, bringing in a number of changes (among other things, principles were revised, marginally part-time workers were included, characteristics were changed or added; cf. Neidert 1998 on this subject).

As by definition the employee history only includes employees covered by social security - civil servants and unpaid family workers for example are not included - so approx. 80% of all people employed in western Germany are covered. However, the degree of coverage varies considerably across the occupations and the industries.

The notifications of the employee history include for every employee, among other things,

- age, sex and nationality
- start and end of each employee notification
- the occupation at the three-digit level
- the daily wage, censored at the upper earnings limit for social security contributions³
- schooling/training
- the industry
- regional code (working place)
- the establishment number

² On the structure of the insurance number and on the data office of the pension insurance providers cf. Steeger (2000).

³ Wages above the upper earnings limit for social security contributions can be imputed by estimated wages (see for example Gartner 2004a on this subject).

For a number of questions not only the current employment but also the biographies of a firm's employees are of interest. By sorting the chronological sequence of the employment notifications per insurance number it is possible to generate employment histories per person. Indirectly it is also possible to identify phases of release from work. In such cases the daily pay is zero. At the IAB the employee history file is linked with the benefit receipt data to make the *employee and benefit recipient history* (*Beschäftigten und Leistungsempfänger-Historik* (BLH)). Periods of non-employment are filled in with the notifications of the IAB's *benefit recipient history* (*Leistungsempfänger-Historik*). Gaps in time still remaining after this are not observable with the present data (periods which are not notified). In the benefit recipient history, spells of unemployment can only be observed when the person draws benefits from the Federal Employment Services (Bundesagentur für Arbeit). People registered as unemployed who have no benefit entitlement (e.g. after studying, during receipt of social assistance or following self-employment) can not be observed at present. The information contained in the benefit recipient history includes the following:

- the person's age, sex, nationality, marital status and children (yes/no)
- district code number of the employment office paying the benefit
- type and amount of the benefit
- start and end of each benefit notification
- type of benefit approval
- reason for submitting the notification
- reason for termination the benefit
- the occupation for which the person was trained, at the four-digit level

Benefit notifications are only made in the case of changes that are relevant according to benefit entitlement rules, i.e. unlike employment notifications, a single benefit notification can describe a period longer than one year.

The establishment number from the employment statistics serves as a definition criterion for the establishment, as a selection basis for the IAB establishment panel, and as a classification indicator for the linking of employee and establishment data. The establishment number is assigned to the employer by the employment office responsible and is used by the

health insurance fund as an employer account number. An establishment is understood as the local unit in which the activities of a firm are actually carried out. The Federal Employment Services (Bundesagentur für Arbeit) allocate to each establishment an industrial sector (until 1998: a 3-digit industry class [WZ 73], since 1998 a 5-digit one [NACE Rev. 1]), a regional code number (municipality code of the workplace) and a (5-digit) office number of the public employment service. The allocation of establishment numbers can lead to problems if a firm relocates or if there is a "revival" of the employer. Moreover it is not the same in all cases. In principle a separate establishment number should be allocated for every branch of a firm; however, at the request of a firm several branches of the same industry in the same municipality can be combined under one number as long as the notifications for social insurance for these branches are made by one office. It is also possible for several numbers to be allocated to one and the same establishment at the request of the firm (see Fritsch/Brixy 2004). The lack of clarity in the allocation of establishment numbers should not be overrated however, as studies on this matter reach quite plausible demarcations of the units of firms by means of the establishment numbers from the notification procedure (König 1994).

The IAB establishment panel survey⁴ is based on the employment statistics aggregated via the establishment number as of 30 June of a year. Consequently the panel only includes establishments with at least one employee covered by social security. What is not taken into account are for example one-person firms and firms without any employees covered by social security (e.g. unpaid family workers) and bogus start-ups. The sample is drawn following the principle of optimum stratification according to the stratification cells of the establishment size class (10 categories) and the industry (16 categories⁵). These stratification cells are also used in the weighting and extrapolation of the sample. The survey is conducted by interviewers from TNS Infratest Sozialforschung. For the first wave, 4265 establishments were interviewed in western Germany in the third quarter of 1993. Since then the establishment panel has been conducted

⁴ The approach and structure of the establishment panel are described for example by Bellmann (2002) and Kölling (2000).

⁵ From 2000 onwards the stratification is done according to 20 industries.

annually – since 1996 with over 4700 establishments in eastern Germany in addition. The response rate of units that have been interviewed repeatedly is over 80%. Each year the panel is accompanied by supplementary samples and follow-up samples in order to interview new or reviving establishments and to compensate for non-response due to refusal to participate or establishment numbers that no longer exist. The list of questions includes detailed information about the firms' personnel structure and development and personnel policy. As a result of co-operation with various regional ministries it has been possible since 2001 to obtain a net sample with more than 15,000 establishments, which also facilitates regionalised evaluations at *Bundesland* level⁶.

Table 1 shows the probabilities of selection and other characteristics of the IAB establishment panel according to establishment size, taking as an example the first survey wave conducted in 1993.

Table 1: Probabilities of selection and response rates of the IAB establishment panel in the first wave of the survey conducted in 1993 (western German establishments only)

| Establishment size class | No. of estab's selected | Selection probability | No. of establishments responding | Response rate |
|--------------------------|-------------------------|-----------------------|----------------------------------|---------------|
| 1 to 4 | 1,072 | 0.0011 | 625 | 0.67 |
| 5 to 9 | 431 | 0.0015 | 250 | 0.64 |
| 10 to 19 | 466 | 0.0030 | 299 | 0.71 |
| 20 to 49 | 862 | 0.0089 | 542 | 0.70 |
| 50 to 99 | 535 | 0.0153 | 350 | 0.72 |
| 100 to 199 | 543 | 0.0304 | 376 | 0.77 |
| 200 to 499 | 923 | 0.0862 | 615 | 0.74 |
| 500 to 999 | 479 | 0.1504 | 304 | 0.71 |
| 1,000 to 4,999 | 1,497 | 0.8765 | 924 | 0.72 |
| more than 5,000 | 115 | 0.9127 | 71 | 0.73 |
| total | 6,923 | 0.0043 | 4,356 | 0.71 |

Source: IAB-Betriebspanel 1993, Kölling (2000), p. 294

⁶ In eastern Germany the IAB establishment panel has already been representative at *Bundesland* level since 1996.

There are few changes in the selection probabilities and response rates over the years. However, as there are few large establishments, it is hardly possible to offset a panel mortality occurring there by taking substitute establishments. In the later waves they are replaced by medium-sized establishments. The number of establishments has therefore increased in the course of the panel, but the *number of people* employed in the establishments remains virtually unchanged⁷ (see also Table 2).

The establishment information available in the IAB establishment panel includes the following:

- number of employees (divided according to qualification groups)
- number of temporary employees and agency workers
- working week for full-time workers, and overtime
- the firm's commitment to collective agreements, existence of a works council
- turnover, advance performance and export share
- investment total
- overall wage bill in the June of the survey year
- technological status of the industrial plant
- age of the establishment, legal form and corporate position
- assessment of the overall company-economic situation
- reorganisation measures and company further training activities (at intervals of several years)
- establishment size and industry

This information is collected every year with few exceptions. A detailed overview of the list of questions of the IAB establishment panel and of the availability of variables in individual survey years can be found in the online documentations of the research data centre (FDZ) about the IAB establishment panel⁸.

⁷ The weighting factors of the IAB establishment panel adapt over time to the changes in the number of establishments and the number of employees covered by social security.

⁸ A code book is also available, which contains detailed annual frequency counts for all the variables collected.

2 Versions of the LIAB

A dataset containing all the available information about people who have worked in a panel establishment for at least one day would be several gigabytes in size and would therefore be very difficult to manage for technical reasons. At the IAB different versions of LIAB datasets have therefore been developed which also permit an analysis using standard programmes such as STATA or SPSS.¹ The LIAB files contain only the relevant person-specific data. The variables from the IAB establishment panel which are required in each case have to be added by the researchers themselves by means of the establishment identifier. If the entire establishment panel dataset were already attached to the person-specific data, this too would exceed the computer capacity.

We distinguish between LIAB data models and LIAB versions in the following. LIAB models are virtual data concepts. Concrete extracts from the total stock of all person-specific data are called the LIAB versions, that are available to external researchers for evaluations. *Versions are databases of persons* which are subject to different selection criteria regarding the establishments, the period covered and the list of variables.

Two data models are described in the following:

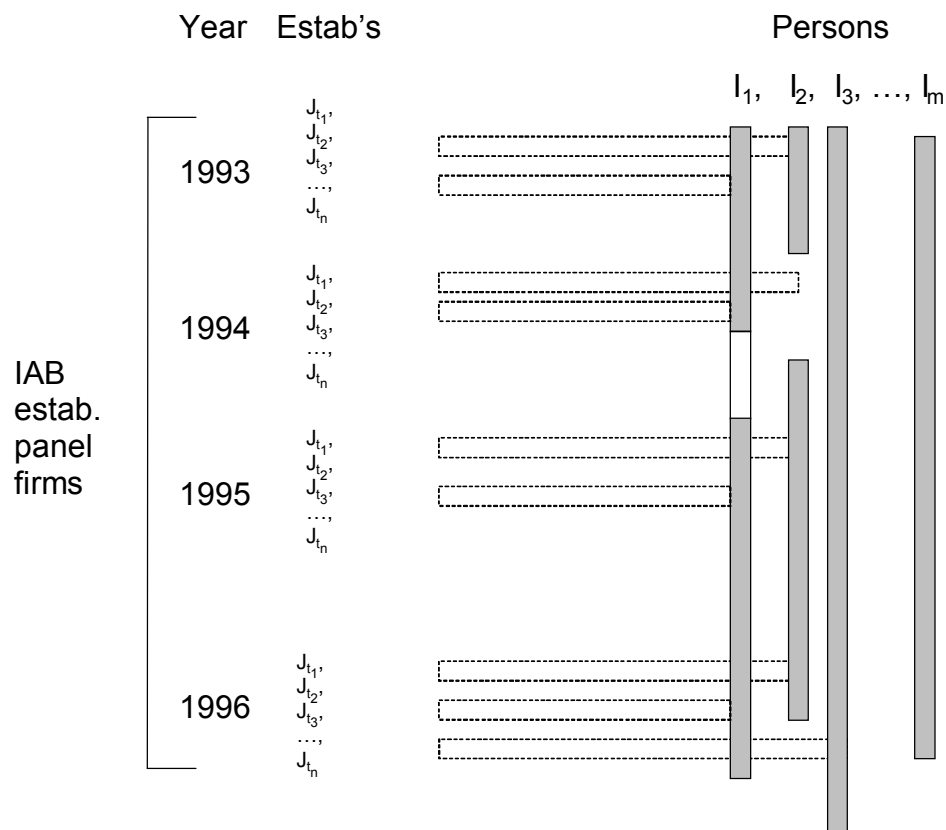
- For the first model, the *LIAB cross-sectional model*, the employment spells valid on 30 June of any year of the people who were working in one of the firms in the IAB establishment panel are extracted from the employee history. 30 June is the reference date of the questions in the establishment panel. With these cross-sections of years it is possible to keep the number of individual notifications relatively small.
- For the second model, the *LIAB longitudinal model*, the course of all the employment and benefit notifications of people who have worked in a panel establishment are filed for a certain period of time. These notifications are available from 1991 until the current limit (this is currently 2001). There can be a considerable number of individual notifications (spells) for each employee. In order to keep the dataset easy to manage, the panel establishments have to be limited to a selection.

The expressions cross-section and longitudinal section refer to the individuals. As the person-specific data are based on the establishment panel,

the cross-sections can also be linked to construct a panel via the establishment or person identifier.

Figure 1 clarifies the two models mentioned. The horizontal lines represent firms interviewed in the IAB establishment panel in the respective year, the vertical ones represent individuals. Column sections shaded in grey indicate periods of employment and pale sections indicate registered unemployment. An interrupted column means that the person was neither in employment subject to social security contributions nor claiming benefits from the Federal Employment Services (BA).

Figure 1: Establishments and employees in LIAB data



A cross-section of a year consists of the points where the horizontal bars (the establishments) of a year cross the vertical bars (their respective employee notifications as of 30 June). For 1996 the longer lower horizontal bar represents an eastern German establishment that was interviewed for the first time in that year. Individual I_1 can therefore be observed in each cross-section from 1993 to 1996, possibly in different establishments. Receipt of benefit is not identifiable at this data level. Person I_2

has an interrupted history of employment subject to social security contributions and can not be observed in 1994. Individual I_3 is employed in an eastern German establishment and can therefore be observed for the first time in 1996. The bar for I_m is intended to make clear graphically that there are also employees who do not appear at all in any of the firms in the IAB establishment panel.

With the longitudinal data with all of an individual's employee and benefit notifications it is possible to determine also points which do not cross and the corresponding employment situations for such individuals, for example registered unemployment with receipt of benefit. Thus for person I_1 the spell of unemployment can be identified and for person I_2 it is known that he/she had a break from employment in 1994. This can not be observed with cross-sections. The existence of a point of intersection is the selection criterion for including the person in the dataset. It does not necessarily have to concern 30 June of a year; the condition for the inclusion of a person and his/her notifications is that he/she was employed in one of these establishments for at least one day in a period to be specified.

The first versions of the data models will be available in the visiting researcher workplaces in the FDZ⁹. The social security number and the establishment number are replaced in each dataset by consecutive and *unique* person and establishment identifiers (unsystematic identifiers¹⁰). There are no further restrictions regarding the variables and their contents. Data reports (Alda 2004a,b,c,d) with exact descriptions of the first two versions are available on the Internet under <http://fdz.iab.de>. Following a general introduction they inform about the degree of coverage across establishments and employees, then they explain the person-specific data and made finally some version-specific basic computations. In the data reports, the variables of individuals and their range of values are also described in detail. A LIAB data manual will be made available to visiting researchers during their stay at the FDZ, which makes it easier to deal with the conditions there.

⁹ Cf. Kohlmann (2004) and Allmendinger & Kohlmann (2005) on the general way that the FDZ functions.

¹⁰ Unsystematic identifiers are anonymised numbers intended to prevent identification of the establishments or the individuals.

At the beginning of 2005 the following two versions are available in the FDZ of the Federal Employment Services at the IAB:

LIAB cross-sectional model, version 1: This dataset contains the respective notifications of the people who were working in one of the interviewed panel establishments on 30.6 of the particular year. These are therefore annual cross-sections of the employee data. The data contain the variables from the employee history file and the date when the employee first joined the establishment and first entered the labour force. The files of individual cross-sections are between 140 and 205 MB in size¹¹. Table 2 shows the unweighted number of establishments and of employees covered by social security in this version of the LIAB cross-section. The content and structure of the data are described in the IAB-Datenreport Nr. 3 (Alda 2004c).

Table 2: Number of establishments and of employees covered by social security in version 1 of the LIAB cross-sectional model

| Year | 1993 | 1994 | 1995 | 1996 | 1997 |
|---------------------|-----------|-----------|-----------|-----------|--------------------------------------|
| Germany west | | | | | |
| Establishments | 4,189 | 4,038 | 3,933 | 3,846 | 3,825 |
| Employees | 2,584,305 | 2,225,384 | 1,885,781 | 1,637,429 | 1,468,918 |
| Germany east | | | | | |
| Establishments | | | | 4,210 | 4,549 |
| Employees | | | | 774,375 | 653,544 |
| Year | 1998 | 1999 | 2000 | 2001 | 2002 |
| Germany west | | | | | |
| Establishments | 4,057 | 4,235 | 8,089 | 9,478 | The data are not yet available |
| Employees | 1,422,637 | 1,373,706 | 1,693,393 | 1,939,946 | |
| Germany east | | | | | |
| Establishments | 4,695 | 5,118 | 5,272 | 5,400 | |
| Employees | 662,694 | 646,728 | 548,785 | 587,279 | |

Source: LIAB cross-sectional model, version 1, unweighted values

¹¹ The information regarding size as well as all subsequent size-related information refers to STATA 8.2. SE.

LIAB longitudinal model, version 1: For approximately 2100 eastern German and about the same number of western German establishments, all the employee and benefit recipient notifications are available from 1991 (western Germany) or 1992 (eastern Germany) until 2001. A person is included, if he or she is employment lasting at least one day in one of the longitudinal establishments in the period 1996 to 2001. As longitudinal establishments are defined firms participating in the IAB establishment panel from 1999 until 2001. The size of the file is 1520 MB in the case of the western German person-specific dataset, and 1150 MB in the case of the eastern German one.

Table 3 shows some figures concerning this version of the LIAB longitudinal model. We differentiate between the levels of establishments, persons and notifications. There are several notifications for individual persons and these depict a person's employment history. 2131 establishments in western Germany and 2116 establishments in eastern Germany form the basis of this LIAB version. These good 2100 establishments in each part of the country have notifications of somewhat more than a million employees covered by social security in western Germany and 700,000 in eastern Germany. The number of notifications is over 14 million for western Germany and just under 11 million for eastern Germany. Over 1.3 million of the notifications in western Germany are benefit recipient notifications, in eastern Germany the figure is more than twice as high (just under 3 million). 42,105 notifications are included in both the dataset for western Germany and that for eastern Germany. This is, because some individuals moved from western German to eastern German establishments in the IAB establishment panel or vice versa. For analysing Germany as a whole these notifications have to be deleted from one of the two datasets. As there is also migration outside of firms in the establishment panel, both datasets contain notifications from the area of application of the other (western Germany: approx 350,000; eastern Germany: approx. 1 million).

Individuals may have more than one employment relationship. Table 3 also shows the number of people who have more than one employment relationship at the same time for at least one day (western Germany: approx. 240,000; eastern Germany: approx. 150,000).

Since people change their employment relationships, the dataset does not only include the slightly more than 2000 firms of the establishment panel which led to the selection, but all in all over half a million (western Germany) and 370,000 (eastern Germany) establishments. As people can also change to panel establishments which were not used for the selection of the sample, these are also shown in Table 3 (western Germany: 29,000; eastern Germany: 20,000).

Table 3: Characteristics of the first version of the LIAB longitudinal model

| Characteristic | Dataset | |
|---|-------------------|-------------------|
| | West | East |
| Number of establishment codes (basis) | 2,131 | 2,116 |
| Number of people in the dataset | 1,078,735 | 729,440 |
| Total number of notifications | 14,473,545 | 10,955,976 |
| Notifications of benefit receipt | 1,375,873 | 2,964,702 |
| Notifications of people who are included in both datasets | 42,105 | 42,105 |
| Notifications from the respective other part of the country | 346,207 | 950,676 |
| Number of people who have more than one employment relationship simultaneously for at least one day | 236,737 | 145,604 |
| Total number of establishment numbers | 562,548 | 370,990 |
| Number of additional IAB establishment panel establishment numbers | 29,025 | 19,925 |
| Size of dataset in MB* | 1,520 | 1,150 |
| Number of variables | 31 | 31 |

*Size of the corresponding STATA 8.2. SE datasets.

If we weight the data set¹² it covers 55.3 percent of western German and 35.0 percent of eastern German establishments; without weights it covers 58.4 percent of western German and 47.3 percent of eastern German establishments. With weights, 56.9 percent of western German and 49.99 percent of eastern German employees work in these establishments (without weights, 47.5 percent (western Germany) and 57.0 percent (eastern Germany)). Thus the first version of the LIAB longitudinal model

¹² Panel weights were used for the calculation instead of cross-sectional weights. These are explained in the IAB-Datenreport Nr. 4 (Alda 2004d).

reaches a degree of coverage across establishments and employees, that is normal for longitudinal establishments of the IAB establishment panel. The content and structure of the data are described in the IAB-Datenreport Nr. 4 (Alda 2004d).

During the course of the first project phase of the research data centre (FDZ) the following versions of the individual data models will *probably* be made available in addition:

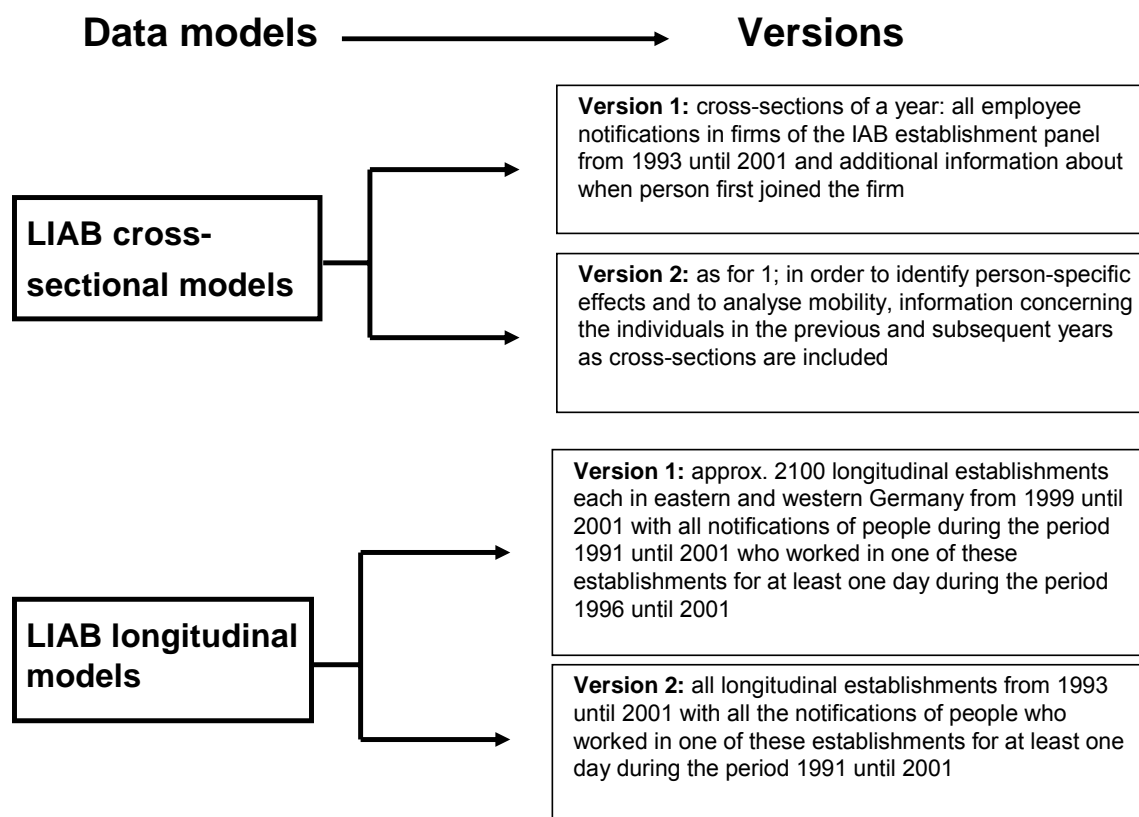
LIAB cross-sectional model, version 2: This is structured like version 1. A single cross-section of a year contains all the employee notifications of the particular survey year and – if available¹³ – the notifications as of 30 June of the previous year and of the subsequent year for the same individuals. This information is helpful for identifying establishment-specific and individual-specific effects or for analysing mobility.

LIAB longitudinal model, version 2: This database contains all person-specific data for the longitudinal-section establishments of the IAB establishment panel from 1993 until 2001. These are the western German establishments that participated constantly in the survey from 1993 to 2001. The dataset comprises the entire employment histories from 1990 to 2001 for the people who were employed in a longitudinal establishment for at least one day. Due to its size, this version is quite unwieldy and can only be analysed with a few programmes, such as SAS or TDA.

Figure 2 provides a summary of the LIAB versions. Further detailed information – also concerning methods of access, test data and much more – can be found on the Internet pages of the FDZ.

¹³ The condition is the carrying out of employment subject to social security contributions in the respective previous and/or subsequent year, or a valid notification of benefit receipt.

Figure 2: Product genesis of the linked employer-employee data of the IAB



3 Studies using linked employer-employee panel data: possibilities and examples

The linked employer-employee data make it possible to deal with a number of new research questions. This can be explained using examples of work conducted so far. Among other things the LIAB data have already been used in studies on gender-specific wage inequality, on individual-specific and company-specific determinants of failure to complete company vocational training, and on the effects of technological and organisational change on mobility. In the appendix we have documented the studies available so far which have used the LIAB data. Information is also included in the appendix as to which of the LIAB data models was used or could have been used to answer the research questions.

Gartner and Stephan (2004) examine why the wage gap between men and women is smaller in establishments which have a works council or are bound by collective wage contracts than it is in other establishments. Possible reasons for this are: firstly, women in establishments which have a

works council or are bound by collective contracts are endowed relatively better with human capital than women in other establishments. Secondly, their human capital is remunerated relatively better. Thirdly, as women often work in the lower wage groups and it is these wages in particular that are raised more considerably by works councils and collective contracts, they benefit more strongly from the higher wages. The wage gap would then therefore be lower because the variance of the wages is smaller. Fourthly, even if the variance of the wages were unchanged, the position of women within the wage distribution can be better in establishments which have a works council or are bound by collective contracts. These differences in the variance of the wages and in the position in the wage distribution can, however, be attributed on the one hand to differences within firms and on the other hand to differences between firms. Using the information about the individuals and the establishments from the LIAB cross-section for 2001, Gartner and Stephan break down the wage gap into various components. For this they use a Juhn-Murphy-Pierce procedure. With this procedure the residuals of the individuals are usually decomposed into individual components. As fixed firm-specific effects can also be estimated using an employer-employee dataset, they also break down the firm-specific effects. The key results are: effects of position and variance within the firm explain the majority of the differences in wage inequality. Sorting effects between the firms are of little significance. Less considerable differences in gender-specific endowment explain a small part of the lower wage gap in establishments bound by collective contracts. In establishments with a works council there is an opposite effect – the difference in men's and women's endowment is greater there than in establishments without a works council.

Another analysis of the gender-specific wage gap that uses the LIAB can be found in Hinz and Gartner (2004). They examine the development of the wage gap within industries, occupational groups and establishments using the individual cross-sections from 1993 to 2001. Achatz, Gartner and Glück (2004) use the LIAB cross-section of 2000 to estimate wage equations which include both individual-specific and firm-specific characteristics. The above-mentioned analyses of wage inequality between men and women can be conducted using the versions of the LIAB cross-sectional model.

Another set of research questions explicitly includes the flow information of individuals in the study, such as for example the studies by Alda (2004e), von Wachter and Bender (2004), Schwerdt and Bender (2003) or Grotheer, Struck, Bellmann & Gewiese (2004). Alda (2004e) examines the determinants of failure to complete company vocational training. In a first step all *first* and *newly begun* employment relationships of trainees are filtered out of the flow information. An employment spell lasting less than two years is defined as a discontinuation of vocational training. The data of the trainees are linked with the IAB establishment panel and the determinants of failure to complete company training are ascertained using a probit estimate in the cross-section for 1998. It emerges that in western Germany a considerable proportion of failure to complete company vocational training is determined by personal factors, whilst in eastern Germany firm-specific factors are of greater significance.

The available flow information at the level of individuals has a further advantage: it has also been possible to use it to examine whether people who fail to complete their vocational training also abandon other training efforts later on. This can only be ascertained for a minority of the observed cases of failure to complete training. The individuals concerned then often go into activities for unskilled and semi-skilled workers. Even after successful completion of company vocational training (in this case following an interruption) a sub-population has quite serious problems in achieving satisfactory integration in the labour market. They are unemployed for a more or less long period immediately after their training. This applies above all to eastern Germans who have completed training, whereas western Germans are more likely to practise an occupation that does not require formal qualifications in spite of their having completed vocational training. Using the LIAB data it is possible to examine specifically which establishments generate such employment history patterns and which establishments the people concerned go into.

Bauer and Bender (2004) use the LIAB to examine the effects of organisational and technological change on mobility (gross job and worker flows). For this they first selected all western German establishment panel firms between 1993 and 1996. In a second step the employment biographies from the employee history file of the IAB were then added. All the employees covered by social security who had worked in one of the estab-

lishment panel firms for at least one day between 1992 and 1996 were put into the selection. The person-specific information was then only used for 30 June of the particular year, however. From the person-specific details in the employee file it was possible to differentiate three qualification levels (unskilled, skilled and highly skilled) and to calculate diverse job and worker flow measures for them.

The empirical findings show that organisational change favours highly qualified employees due to the net employment growth rates of unskilled and skilled employees falling as they have higher job destruction and separation rates. In contrast, these measured values for the highly qualified are not affected. New information technologies increase the churning rate for skilled and highly skilled employees.

The research potential of the employer-employee dataset at the IAB, which has been illustrated with the examples, makes the data interesting for external researchers, too.

4 Data disclosure and utilisation for external researchers

The LIAB versions described above are made available by the research data centre of the Federal Employment Services at the Institute for Employment Research (IAB). The user is expected to be familiar with the use of statistical software packages (SAS, SPSS, STATA and TDA are available). As working material can be used the documentation of the IAB employment sample, the IAB establishment panel and the LIAB data reports (Alda 2004a, b, c, d). In addition, for each version of the LIAB, anonymised test data are available with can not be de-anonymised and are partly invented (absolutely anonymised data).

The individual LIAB original data may be accessed following authorisation of a research application from the social services field according to § 75 of the Social Code volume X (Disclosure of social data for research and planning). The application may be submitted directly to the FDZ. The authorisation procedures are standardised and linked to the guidelines of the Federal Data Protection Act. These guidelines demand for instance that

concrete data sets and required variables be named.¹⁴ After the application has been authorised the test data are transferred to external researchers.

By means of the clearly defined LIAB data sets it is possible to make agreed procedure reports to the data protection authorities. These reports help for example to clarify questions of data security for the individual applications from external researchers in the visiting researcher model of the FDZ. In this way the authorisation of access to data in accordance with § 75 of the Social Code volume X can concentrate to the greatest possible extent on checking the (external) research interests and in this way is simplified, standardised and made more transparent.

External researchers can gain access to slightly anonymised (social) data by visiting the FDZ in Nuremberg. In each LIAB version the social security number and establishment code are replaced by consecutive and clear person identification and establishment identification codes. There are no further restrictions with regard to the variables and their contents; regional characteristics are also available in their original degree of differentiation.

5 Summary

The linked employer-employee data set at the IAB (LIAB) is created by merging the data of the IAB establishment panel and the employee and benefit recipient history of the IAB. The distinctive feature of these data is the combination of information about the establishments and information about the people working in these establishments. This facilitates a number of new methods of analysis and the empirical examination of new questions. So far studies have been conducted on subjects in labour economics, personnel economics and sociology.

The LIAB data can be accessed by visiting the research data centre (Forschungsdatenzentrum - FDZ) of the Federal Employment Services at the

¹⁴ Other guidelines are for example security in the holding of data and a description of the reasons for the research (in order to ascertain the existence of a research intention from the field of social services and the other requirements associated with this). Researchers interested in this can obtain information concerning access requirements on the Internet. Various forms of support are provided for making the application.

IAB in Nuremberg. The IAB have developed data models which allow the user to process the data using standard software such as STATA or SPSS.

During the past years a number of studies have already been conducted using this dataset. The dataset may now also be used by researchers outside the IAB. This lead us to expect a multitude of further interesting analyses using the IAB's linked employer-employee data.

References

- Abowd, J. & Kramarz, F. (1999): The Analysis of Labor Markets Using Matched Employer-Employee Data. In: Ashenfelter, O. & Card, D. (eds.): Handbook of Labor Economics, Amsterdam, 2629-2733.
- Achatz, J.; Allmendinger, J. & Hinz, T. (2001): Sex Segregation in Organizations. A Comparison of Germany and the US. Working paper presented at ASA-conference in Washington D.C., USA, august 2000.
- Achatz, J; Gartner, H. & Glück T. (2004): Bonus oder Bias? - Mechanismen geschlechtsspezifischer Entlohnung. IAB Discussion Paper No. 5.
- Alda, H. (2004a): IAB-Datenreport Nr. 1: Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten, LIAB des Instituts für Arbeitsmarkt- und Berufsforschung.
- Alda, H. (2004b): IAB-Datenreport Nr. 2: Die Verknüpfungsqualität der LIAB-Daten.
- Alda, H. (2004c): IAB-Datenreport Nr. 3: Datenbeschreibung der Version 1 des LIAB-Querschnittmodells.
- Alda, H. (2004d): IAB-Datenreport Nr. 4: Datenbeschreibung der Version 1 des LIAB-Längsschnittmodells.
- Alda, H. (2004e) : Individuelle und betriebliche Determinanten von vorzeitigen Beendigungen der betrieblichen Berufsausbildung, Bericht an das Bundesinstitut für Berufsbildung, Nürnberg.
- Alda, H., Bellmann L. & Gartner H. (2004) Wage Structure and Labor Mobility in the West German Private Sector 1993 – 2000, working paper, presented at Empirical Personnel Economics Meeting NBER, april 17, 2004.
- Allmendinger, J. & Kohlmann, A. (2005): Datenzugang und Datenverfügbarkeit im Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt- und Berufsforschung, in: Allgemeines Statistisches Archiv (im erscheinen).

- Bauer, T. & Bender, S. (2001): Flexible Work Systems and the Structure of Wages: Evidence from Matched Employer-Employee Data. IZA-Discussionpaper No. 353.
- Bauer, T. & Bender, S. (2004): Technological Change, Organizational Change, and Job Turnover, *Labour Economics*; No. 2, 11, S. 265-291.
- Bauer, T., Bender, S. & Bonin, H. (2004): Dismissal Protection and Worker Flows in Small Establishments, IAB-Discussionpaper No. 1.
- Beblo, M., Bender, S. & Wolf, E. (2004): The Wage Effects of Entering Motherhood: A within-firm matching approach, mimeo Mannheim, Nürnberg.
- Bellmann, L. (2002): Das IAB-Betriebspanel - Konzeption und Anwendungsbereiche, *Allgemeines Statistisches Archiv* 86, 177-188.
- Bellmann, L., Bender, S. & Kölling, A (2002): Der Linked Employer-Employee-Datensatz aus IAB-Betriebspanel und Beschäftigtenstatistik der Bundesanstalt für Arbeit (LIAB), in: G. Kleinhenz (Hrsg.), IAB-Kompendium Arbeitsmarkt- und Berufsforschung, Beiträge aus der Arbeitsmarktforschung 250, Nürnberg, 21-29.
- Bellmann, L.; Bender, S. & Schank, T. (1999): Flexibilität der Qualifikationsstruktur aus betrieblicher Sicht: Substitutionalität oder Komplementarität. *Jahrbücher für Nationalökonomie und Statistik* 219:1 & 2, 109-126.
- Bellmann, L.; Caliendo, M.; Hujer, R.; Radic, D. (2002): Beschäftigungswirkungen technisch-organisatorischen Wandels - eine mikroökonomische Analyse mit dem Linked IAB-Panel. In: *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, Jg. 35, H. 4., S. 506-522.
- Bellmann, L. & Schank, T. (2000): Innovations, Wages and the Demand for Heterogenous Labour: New Evidence from a Matched Employer-Employee Data-Set. IZA-Discussionpaper No. 112.
- Bender, S. & Haas, A (2002): Die IAB-Beschäftigtenstichprobe. In: G. Kleinhenz (Hrsg.), IAB-Kompendium Arbeitsmarkt- und Berufsforschung, Nürnberg, 3-12.
- Bender, S.; Haas, A. & Klose, C. (2000): The IAB Employment Subsample 1975-1995, *Schmollers Jahrbuch* 120, S. 649-662.
- Bender, S.; Konietzka, D. & Sopp, P. (2000): Diskontinuität im Erwerbsverlauf und betrieblicher Kontext. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Jg. 52, H. 3, S. 475-499.

- Bender, S.; Preißler, J. & Wübbecke, C. (2000): Betriebliche Determinanten des Generationenaustausches in westdeutschen Betrieben - eine Untersuchung auf der Basis des IAB-Employer-Employee-Datensatzes für die Jahre 1994 und 1995. In: R. George & O. Struck (Hrsg.): Generationsaustausch im Unternehmen, München u. a.: Hampp, S. 209-225.
- Brand, R.; Bender, S. & Kohaut, S. (1999): Möglichkeiten der Erstellung eines Scientific-Use Files aus dem IAB-Betriebspanel. In: Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik, Stuttgart: Metzler-Poeschel, S. 148-167. Schriftenreihe Spektrum Bundesstatistik Nr. 14.
- Brixy, U.; Kohaut, S. & Schnabel, C. (2004): Do newly founded firms pay lower wages? First evidence from Germany. Erlangen u. a.: 17 S.; IAB Discussion Paper No. 4/04.
- Fritsch, M./Brixy U. (2004), The Establishment File of the German Social Insurance Statistics, in Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 124, H. 1, 183-190.
- Gartner, H. (2004a): Die Imputation von Löhnen oberhalb der Beitragsbemessungsgrenze in der IAB-Beschäftigtenstatistik. Forthcoming as IAB-Methodenreport, Nürnberg.
- Gartner, H. (2004b): Gender wage inequality and rent sharing - evidence from a German employer-employee-dataset, (forthcoming), Nürnberg.
- Gartner H. & Stephan G. (2004): How Collective Contracts and Works Councils Reduce the Gender Wage Gap, IAB Discussion Paper No. 7/04.
- Grotheer, M.; Struck, O.; Bellmann, L. & Gewiese, T. (2004): Determinanten der Beschäftigungsstabilität - Chancen und Risiken von 'Entrants' im ost-westdeutschen Vergleich, in: Struck, O. & Köhler, C. (Hrsg.), Beschäftigungsstabilität im Wandel? : empirische Befunde und theoretische Erklärungen für West- und Ostdeutschland, München u. a.: Hampp, S. 125-156.
- Hamermesh, D. (1999): Leaping into the Future of Labor Economics: The Research Potential of Linking Employer and Employee Data. Labor Economics 6, 25-41.
- Hinz T. & Gartner, H. (2004): Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben. Forthcoming in: Zeitschrift für Soziologie.
- Hinz, T. & Schübel, T. (2001): Geschlechtersegregation in deutschen Betrieben. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 34:3, 286-301.
- Kohlmann, A. (2004): Datenzugang und Datenverfügbarkeit im Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt- und Berufsforschung, in: VDR-Schriften Bd. 55.

- Kölling, A. (2000): The IAB-Establishment Panel. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 120, H. 2, S. 291-300.
- Kölling, A. (2002): Fachkräftebedarf als betriebliches Matching-Problem. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, Jg. 35, H. 4, S. 566-579.
- Kölling, A. & Schank, T. (2001): Skill Biased Technological Change, Labour Hoarding and the Wage Structure. Working paper presented at EALE 2001 in Jyväskylä, Finland, september 2001.
- Kölling, A.; Schnabel, C. & Wagner, J. (2002): Establishment age and wages - Evidence from German linked employer-employee data. Universität Erlangen, Nürnberg, Lehrstuhl für Arbeitsmarkt- und Regionalpolitik. Diskussionspapiere Nr. 13.
- König, A. (1994): Betriebliche Beschäftigungsdynamik und personeller Strukturwandel - Eine Longitudinalanalyse; Frankfurt & New York.
- Neidert, A. (1998): Neues Meldeverfahren zur Sozialversicherung ab 1999, in: Deutsche Rentenversicherung 5/1998, 315-330.
- Schank, T. (2003): Die Beschäftigung von Un- und Angelernten - Eine Analyse mit dem Linked Employer-Employee Datensatz des IAB, in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 36/3: 257-270.
- Schank, T., Schnabel, C. & Wagner, J (2004): Exporting firms do not pay higher wages, ceteris paribus. First evidence from linked employer-employee data IZA Diskussionspapier 1185.
- Schwerdt, W & Bender, S. (2003): Was tun Lehrlinge nach ihrer Ausbildung? - eine Analyse mit dem Linked Employer-Employee-Datensatz des IAB, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, Jg. 36, H. 1, S. 46-59.
- Steger, W. (2000): 25 Jahre Datenstelle der Rentenversicherungsträger (DSRV), in Deutsche Rentenversicherung 10-11/2000, 648-684.
- von Wachter, T. & Bender, S. (2004): In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers, IZA-Discussionpaper, No. 1348.
- Willis, R. (1986): Wage Determinants. In: Ashenfelter, O. & Layard, R. (eds.): Handbook of Labor Economics, Amsterdam, 525-602.

Appendix

Studies using the LIAB

| Author | Title | LIAB Version |
|--|--|---|
| Achatz J., Gartner, H. & Glück T. (2004) | Bonus oder Bias? – Mechanismen geschlechtsspezifischer Entlohnung | LIAB cross-section 2000 |
| Alda, H. (2004) | Individuelle und betriebliche Determinanten von Ausbildungsabbrüchen | LIAB longitudinal data, 1996-2000 |
| Alda, H. (2005) | Betriebs- und Personeneffekte auf die Entlohnung | LIAB longitudinal data, 1996-2001 |
| Alda, H., Bellmann, L. und Gartner, H. (2004) | Wage Structure and Labor Mobility in the West German Private Sector 1993 – 2000 | LIAB cross-section, 1993-2000 |
| Andrews, M., Schank, Th. & Upward, R. (2004) | Practical estimation methods for linked employer-employee data | Possible with LIAB cross-section, 1993-1997 |
| Bauer, T. & Bender, S. (2004) | Technological Change, Organizational Change, and Job Turnover | Possible with LIAB cross-section |
| Bauer, T. & Bender, S. (2001) | Flexible Work Systems and the Structure of Wages: Evidence from matched Employer-Employee Data | Possible with LIAB cross-section |
| Bellmann, L. & Schank, T. (2000) | Innovations, Wages and the Demand for Heterogeneous Labour: New Evidence from a Matched Employer-Employee Data-Set | LIAB cross-section, 1995 |
| Bellmann, L., Caliendo, M., Hujer, R. & Radić, D. (2002) | Beschäftigungswirkungen technisch-organisatorischen Wandels: Eine mikroökonomische Analyse mit dem Linked IAB Panel | LIAB cross-section, 1998 |
| Bellmann, L.; Bender, S. & Schank, T. (1999) | Flexibilität der Qualifikationsstruktur aus betrieblicher Sicht: Substitutionalität oder Komplementarität | LIAB cross-section, 1994-1997 |
| Bender, S.; Konietzka, K. & Sopp, P.M. (2000) | Erwerbsverlauf und betrieblicher Kontext | Possible with LIAB longitudinal data |
| Bender, Stefan; Preißler, Josef & Wübbecke, Christina (2000) | Betriebliche Determinanten des Generationenaustausches in westdeutschen Betrieben - eine Untersuchung auf der Basis des IAB-Employer-Employee-Datensatzes für die Jahre 1994 und 1995. | LIAB cross-section, 1994-1995 |
| Brixy, U.; Kohaut, S. & Schnabel, C. (2004) | Do newly founded firms pay lower wages? first evidence from Germany. | LIAB cross-section, 1997-2001 |
| Gartner, H. (2004b) | Gender wage inequality and rent sharing - evidence from a German employer-employee-dataset | LIAB cross-section 2001 |
| Gartner, H. & Stephan, G. (2004) | How Collective Contracts and Works Councils Reduce the Gender Wage Gap | LIAB cross-section 2001 |
| Grotheer, Michael; Struck, Olaf; Bellmann, Lutz & Gewiese, Tilo (2004) | Determinanten der Beschäftigungsstabilität - Chancen und Risiken von 'Entrants' im ost-westdeutschen Vergleich | LIAB longitudinal data, version 2 |

| | | |
|---|---|---|
| Hinz, T. & Gartner, H. (2004) | Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben | LIAB cross-section 1993-2001 |
| Kölling, A., Schnabel, Claus & Wagner, Joachim (2002) | Establishment age and wages - Evidence from German linked employer-employee data | LIAB cross-section 1996 |
| Kölling, A. & Schank, Th. (2003) | Skill-Biased Technological Change, International Trade and the Wage Structure | Possible with LIAB cross-section, 1993-1997 |
| Kölling, A. (2002) | Fachkräftebedarf als betriebliches Matching-Problem | Possible with LIAB cross-section 1999 |
| Schank, Th., Schnabel, C. & Wagner, J. (2004): | Exporting firms do not pay higher wages, ceteris paribus. First evidence from linked employer-employee data | LIAB cross-section 1995-1997 |
| Schank, Th. (2003) | Die Beschäftigung von Un- und Angelernten | Possible with LIAB cross-section, 1993-1997 |
| Schwerdt, W. & Bender, S. (2003) | Was tun Lehrlinge nach ihrer Ausbildung? - eine Analyse mit dem Linked Employer-Employee-Datensatz des IAB | Possible with LIAB longitudinal data |

In dieser Reihe sind zuletzt erschienen

Recently published

| No. | Author(s) | Title | Date |
|--------|---|---|---------|
| 1/2004 | Bauer, Th. K., Bender, St., Bonin, H. | Dismissal Protection and Worker Flows in Small Establishments | 7/2004 |
| 2/2004 | Achatz, J., Gartner, H., Glück, T. | Bonus oder Bias? Mechanismen geschlechtsspezifischer Entlohnung | 7/2004 |
| 3/2004 | Andrews, M., Schank, Th., Upward, R. | Practical estimation methods for linked employer-employee data | 8/2004 |
| 4/2004 | Brixy, U., Kohaut, S., Schnabel; C. | Do newly founded firms pay lower wages? First evidence from Germany | 9/2004 |
| 5/2004 | Kölling, A., Rässler, S. | Editing and multiply imputing German establishment panel data to estimate stochastic production frontier models | 10/2004 |
| 6/2004 | Stephan, G., Gerlach, K. | Collective Contracts, Wages and Wage Dispersion in a Multi-Level Model | 10/2004 |
| 7/2004 | Gartner, H. Stephan, G. | How Collective Contracts and Works Councils Reduce the Gender Wage Gap | 12/2004 |
| 1/2005 | Blien, U., Suedekum, J. | Local Economic Structure and Industry Development in Germany, 1993-2001 | 1/2005 |
| 2/2005 | Brixy, U., Kohaut, S., Schnabel, C. | How fast do newly founded firms mature? Empirical analyses on job quality in start-ups | 1/2005 |
| 3/2005 | Lechner, M., Miquel, R., Wunsch, C. | Long-Run Effects of Public Sector Sponsored Training in West Germany | 1/2005 |
| 4/2005 | Hinz, Th., Gartner, H. | Lohnunterschiede zwischen Frauen und Männern in Branchen, Berufen und Betrieben | 2/2005 |
| 5/2005 | Gartner, H., Rässler, S. | Analyzing the Changing Gender Wage Gap based on Multiply Imputed Right Censored Wages | 3/2005 |

Impressum

IAB Discussion Paper
No. 6 / 2005

Herausgeber

Institut für Arbeitsmarkt- und Berufsforschung
der Bundesagentur für Arbeit
Weddigenstr. 20-22
D-90478 Nürnberg

Redaktion

Regina Stoll, Jutta Palm-Nowak

Technische Herstellung

Jutta Sebald

Rechte

Nachdruck – auch auszugsweise – nur mit
Genehmigung des IAB gestattet

Bezugsmöglichkeit

Volltext-Download dieses Discussion Paper
unter:

<http://doku.iab.de/discussionpapers/2005/dp0605.pdf>

IAB im Internet

<http://www.iab.de>

Rückfragen zum Inhalt an

Hermann Gartner, Tel. 0911/179-3386,
oder e-Mail: hermann.gartner@iab.de