Schmollers Jahrbuch 125 (2005), 327 – 336
Duncker & Humblot, Berlin

European Data Watch

This section will offer descriptions as well as discussions of data sources that may be of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We will start with data from German speaking countries that allow international comparative research. While most of the data will be at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) will be included, too. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@uni-lueneburg.de⟩.

The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB)[1]

By Holger Alda, Stefan Bender and Hermann Gartner

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) provide an overview of the basic possibilities of linking firm and employee data. They differentiate among other things

[1] The first chapter of this paper are based on Bellmann / Bender / Kölling (2002) and this paper constitutes an update.

between links of administrative data alone, those of survey data alone and various combinations, and cite some 100 studies from 15 countries. The area of research on linked employer-employee datasets is developing rapidly.

In this paper we describe the setting up of the linked employer-employee dataset at the IAB (LIAB), the research 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.[2] The IAB establishment panel contains annual information about establishment structures and personnel-policy decisions in the period from 1993 onwards. 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.

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.[3] 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).

[2] 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 / Bonin 2004, Bebl o / Bender / Wolf 2005, von Wachter / Bender 2004).

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

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 contributions,[4]
- schooling / training,
- the industry,
- regional code (working place),
- the establishment number.

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 spells out of labor force. 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 of the employment office paying the benefit,
- type and amount of the benefit,

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

– 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* survey[5] 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. 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 estab-

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

lishment size class (10 categories) and the industry (16 categories[6]). 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 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* level[7].

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

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

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

documentations of the research data centre (FDZ) about the IAB establishment panel (http: // fdz.iab.de).[8]

A *LIAB 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 permit an analysis using standard programmes such as STATA or SPSS. The extended version of this paper (Alda / Bender / Gartner, 2005) provides an overview of the different LIAB data models and LIAB versions (see www.iab.de). Data reports (Alda 2004a,b,c,d) contain an exact descriptions of the first two LIAB versions. They are available under http: // fdz.iab.de.

The first versions of the data models will be available in the visiting researcher workplaces in the FDZ[9]. The social security number and the establishment number are replaced in each dataset by consecutive and *unique* person and establishment identifiers (unsystematic identifiers[10]). There are no further restrictions regarding the variables and their contents. 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.

2. 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.

Gartner / 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. They use a Juhn-Murphy-Pierce procedure for the decomposition of the gender wage gap. Another analysis of the gender-specific wage gap that uses the LIAB can be

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

9 Cf. Kohlmann (2004) and Allmendinger / Kohlmann (2005) for the general functions of the FDZ.

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

found in Hinz / Gartner (2005). 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 / Glück (2004) use the LIAB cross-section of 2000 to estimate wage equations which include both individual-specific and firm-specific characteristics.

Another set of research questions explicitly includes the flow information of individuals in the study, such as for example the studies by Schwerdt / Bender (2003) or Grotheer / Struck / Bellmann / Gewiese (2004). Alda (2004e) examines the determinants of failure to complete company vocational training. He linked data of trainees with the IAB establishment panel and ascertain the determinants of failure to complete company training using a probit estimation.

Bauer / Bender (2004) use the LIAB to examine the effects of organisational and technological change on mobility (gross job and worker flows). 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.

In Alda / Bender / Gartner (2005) we discuss the examples in more detail and document the further studies available so far which have used the LIAB data. 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.

3. Data disclosure and utilisation for external researchers

The LIAB versions 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 which 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.[11] 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.

4. 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 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.

[11] 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.

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. / *Gartner,* H. / *Glück,* T. (2004): Bonus oder Bias? – Mechanismen geschlechtsspezifischer Entlohnung. IAB-Discussion-Paper No. 5.

Alda, H. (2004a): FDZ-Datenreport Nr. 1: Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten, LIAB des Instituts für Arbeitsmarkt- und Berufsforschung.

– (2004b): FDZ-Datenreport Nr. 2: Datenbeschreibung der Version 1 des LIAB-Querschnittmodells.

– (2004c): FDZ-Datenreport Nr. 3: Datenbeschreibung der Version 1 des LIAB-Längsschnittmodells.

– (2004d): FDZ-Methodenreport Nr. 1: Die Verknüpfungsqualität der LIAB-Daten.

– (2004e): Individuelle und betriebliche Determinanten von vorzeitigen Beendigungen der betrieblichen Berufsausbildung, Bericht an das Bundesinstitut für Berufsbildung, Nürnberg.

Alda, H. / *Bender,* S. / *Gartner,* H. (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB), forthcoming as IAB-Discussion-Paper.

Allmendinger, J. / *Kohlmann,* A. (2005): Datenzugang und Datenverfügbarkeit im Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt- und Berufsforschung, in: Allgemeines Statistisches Archiv (forthcoming).

Bauer, T. / *Bender,* S. (2004): Technological Change, Organizational Change, and Job Turnover, in: Labour Economics 2 (11), 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. (2005): The Wage Effects of Entering Motherhood: A within-firm matching approach, mimeo Mannheim, Nürnberg.

Bellmann, L. (2002): Das IAB-Betriebspanel – Konzeption und Anwendungsbereiche, in: 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 (ed.), IAB-Kompendium Arbeitsmarkt- und Berufsforschung, Beiträge aus der Arbeitsmarktforschung 250, Nürnberg, 21 – 29.

Bender, S. / *Haas,* A. / *Klose,* C. (2000): The IAB-employment-subsample 1975 – 1995, in: Schmollers Jahrbuch 120 (4) 649 – 662.

Bender, S. / *Haas,* A. (2002): Die IAB-Beschäftigtenstichprobe, in: Kleinhenz, G. (ed.), IAB-Kompendium Arbeitsmarkt- und Berufsforschung, Nürnberg, 3 – 12.

Fritsch, M. / *Brixy,* U. (2004): The Establishment File of the German Social Insurance Statistics, in Schmollers Jahrbuch 124 (1), 183 – 190.

Gartner, H. (2004): Die Imputation von Löhnen oberhalb der Beitragsbemessungsgrenze in der IAB-Beschäftigtenstichprobe. forthcoming as FDZ-Methodenreport, Nürnberg.

Gartner, H. / *Stephan,* G. (2004): How Collective Contracts and Works Councils Reduce the Gender Wage Gap, IAB-Discussion-Paper No. 7

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. (eds.), Beschäftigungsstabilität im Wandel?: empirische Befunde und theoretische Erklärungen für West- und Ostdeutschland, München, 125 – 156.

Hamermesh, D. (1999): LEEping into the Future of Labor Economics: The Research Potential of Linking Employer and Employee Data, in: Labor Economics 6, 25 – 41.

Hinz, T. / *Gartner,* H. (2005): Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben, in: Zeitschrift für Soziologie (forthcoming).

Kohlmann, A. (2004): Datenzugang und Datenverfügbarkeit im Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, in: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau: Bericht vom ersten Workshop des FDZ-RV am 28. und 29. Juni 2004 in Würzburg, DRV-Schriften Nr. 55, Frankfurt am Main, 77 – 88.

Kölling, A. (2000): The IAB-Establishment Panel, in: Schmollers Jahrbuch 120 (2), 291 – 300.

König, A. (1994): Betriebliche Beschäftigungsdynamik und personeller Strukturwandel – Eine Longitudinalanalyse, Frankfurt am Main / New York.

Neidert, A. (1998): Neues Meldeverfahren zur Sozialversicherung ab 1999, in: Deutsche Rentenversicherung 5 / 1998, 315 – 330.

Schwerdt, W. / *Bender,* S. (2003): Was tun Lehrlinge nach ihrer Ausbildung? – eine Analyse mit dem Linked Employer-Employee-Datensatz des IAB, in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 36 (1), 46 – 59.

Steeger, 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.