An Overview on the Linked Employer-Employee Data of the Institute for Employment Research (IAB)

By Jörg Heining, Wolfram Klosterhuber and Stefan Seth

1. Introduction

In 2005, the Institute for Employment Research (IAB) of the German Federal Employment Agency (Bundesagentur für Arbeit, BA) released a unique and innovative Linked Employer-Employee dataset (LIAB, see Alda et al., 2005). The LIAB data combine survey information on establishments with administrative records of employees and thus allow for a simultaneous analysis of the supply and demand sides of the German labor market.

The potential of linked employer-employee data in order to study the interactions between workers and firms is widely acknowledged. For example, Willis (1986) or Abowd/Kramarz (1999) already stressed the importance of linked employer-employee data for empirical research in labor economics. As a consequence, the LIAB data have been high in demand since their first release. Over the years, the LIAB data have been updated and revised several times and new versions and data models became available.

This article provides an overview on the currently existing models of the LIAB data and tries to delimitate them from older versions. The article is structured as follows. Section 2 describes the components of the LIAB data and how they are linked. The different models of the LIAB data are discussed in section 3 whereas section 4 provides some information on data access. A summary is provided in section 5.

2. Components of the LIAB data and Linkage

2.1 Administrative Records on Employees

Administrative records on employees stem from the notification process to the social security institutions in Germany as well as from internal processes of the German Federal Employment Agency. Every employer in Germany is obliged to submit at least once a year a notification on each of his employees to the social security institutions. Information submitted include daily exact infor-

mation on the beginning and the end of employment, gender, educational attainment, (qualitative) information on full-time or part-time work, occupation, place of residence, and the gross wages paid to the employee for the covered period, among others. If an employee is continuously employed all year, the registered begin and end dates of employment are January 1st and December 31st.

Individual information on employment is amended by administrative records generated in the processes of the German labor administration. In particular, the additional information include the periods of unemployment benefit receipt according to the German Social Code Book (Sozialgesetzbuch), II and III, the periods of job search as registered by BA, and the participation in labor market programs and training schemes. As the social security records, this information is available on daily basis, too. The IAB combines social security records as well as other administrative data creating so-called employment biographies for every individual. Thus, as long as an individual is subject to social security contributions, recipient of unemployment benefits, a registered job-seeker, or a participant in labor market programs, it may be tracked over time. The resulting data base is called the *Integrated Employment Biographies* (IEB). The IEB data base comprises information on about 85 million workers from 1975 onwards.

The administrative records on employment in the IEB data base also provide some information on the employer. The most important information in this context is a unique establishment identifier (EID). The EID is assigned by BA at the official start of operation of an establishment. It is assigned on the basis of ownership, industry and municipality. For further information on this assignment process, see Brixy and Fritsch 2002. The EID is also stored in a second data base, the so-called *establishment file of the BA* from which the IAB establishment Panel data are sampled (see below). In addition to the EID, the establishment file also contains some basic information, for example size of the establishment, the branch of economic activity (as indicated by the NACE industry codes), and the location.

In this context, it is important to understand the definition of an establishment and the EID (see Klosterhuber et al. 2013). An establishment is a regionally and economically delimited unit where employees work. However, an establishment is not equitable to a firm since a firm may consist of several establishments. If a firm, for example, consists of several establishments located in different municipalities, every establishment will be assigned a different EID. In contrast, a single EID may belong to multiple work sites owned by the same firm if they are in the same industry and municipality. For the LIAB data, this EID is crucial since it is the key variable to match the information provided by the IAB Establishment Panel data.

2.2 The IAB Establishment Panel

The IAB Establishment Panel (Kölling, 2000; Bellmann, 2002 or Fischer et al., 2008) is an annual representative survey on establishments in Germany on various topics. It provides information on up to approximately 16,000 establishments per year and has been conducted by IAB in West Germany since 1993 and in East Germany since 1996 . Each establishment includes at least one employee subject to social security contributions at the time of the interview (June 30th of each year). Some standard topics like development of employment, business policy and development, investments, et cetera, are queried annually. Additionally, each wave focuses on some special topics, for example types of employment, vocational training and apprenticeship in 2009 or business policy and development in the financial crisis in 2011.

The IAB Establishment Panel is sampled from BA's establishment file, which contains the universe of establishments with at least one employee liable to social security as of 30 June of the previous year. The sample is stratified according to establishment size, industry and federal state. The sample is multiply disproportional. In order to account for panel mortality and to reflect the foundation of new establishments, supplementary samples amend the basic sample for the IAB Establishment data.

2.3 Establishment History Panel

Additional information on establishments from the Establishment History Panel data (see Spengler, 2008) may be added either to the individual or establishment data using the EID. The Establishment History Panel data contains aggregations of individual social security records by EID. It is composed of cross-sectional datasets for West-Germany since 1975 and for East Germany since 1991. Every cross section contains all establishments in Germany with at least one employee subject to social security on June 30th. Since 1999, establishments consisting solely of one marginal part-time employee are also included.

The BHP data contains information about the branch of industry and the location of the establishment. Furthermore, there is the number of employees liable to social security per establishment, as well as marginal part-time employees (since 1999), both in total and broken down by gender, age, occupational status, and qualification, among others.
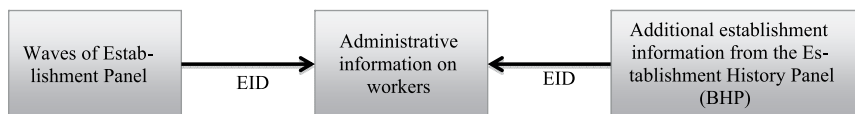

Figure 1: The Components of the LIAB data.
The single components of the LIAB data may be linked by using the EID.

3. Models of the LIAB data

The longitudinal structure of the individual employment biographies in com-
bination with the IAB Establishment data allows the design of two different
data models. While individual information in the LIAB Cross-sectional Model
is only available on June 30th of the respective year, the LIAB Longitudinal
Model exploits the longitudinal structure of workers' employment histories. A
third model of the LIAB data, the LIAB Mover Model, is also based on the
longitudinal structure of the individual information, but focuses on worker tran-
sitions between establishments.

3.1 The LIAB Cross-sectional Model

The LIAB Cross-sectional Model comprises all establishments with a valid
interview from the IAB Establishment Panel data in a given survey year.
Furthermore, all individuals, identified in the IEB database, which have been
employed on the June 30th in the respective year and establishment, are in-
cluded. For these individuals, only spells with begin and end date including
June 30th of a given survey year are extracted from the IEB database. For ex-
ample, an establishment provided a valid interview for the IAB Establishment
Panel data in 2006. Then, all individuals employed at this establishment on
June 30th 2006 are included in the LIAB Cross-sectional Model. In addition to
the individual employment information, the LIAB Cross-sectional Model con-
tains also spells indicating unemployment benefit receipt or job-search as long
as June 30th of the respective year is comprised.

Given this design, the current version of the LIAB Cross-sectional data
(LIAB QM2 9310, see Heining et al., 2013) covers the period between 1993
and 2010. In total, the LIAB QM2 9310 data include information on 49,844
establishments from the IAB Establishment Panel data, which may be com-
bined with administrative records of 10,314,524 workers. Annual numbers of
linked establishments vary between 4,188 and 14,981 and between 1,629,542
and 2,584,520 for linked workers.

Differences between the current LIAB QM2 9310 and older versions consist
mainly in the covered time period and in the availability of additional worker
information on registered job-search.

3.2 The LIAB Longitudinal Model

Worker information in the longitudinal LIAB model is not referenced to a
specific date. Instead, worker's coherent employment biographies are amended
by information from the IAB Establishment Panel data. As a consequence, the
sampling process for both establishments and individuals differs from the
LIAB Cross-sectional data.

For the LIAB Longitudinal Model, usually a selection of establishments is drawn from the IAB Establishment Panel data for a coherent time period. Criteria for this sampling varied over the several versions of the LIAB Longitudinal Model. They included, for example, the existence of a valid interview, a coherent participation in the survey or a minimum amount of employees.

For the selected establishments, all individuals, employed at least one day at one of the establishments in a certain period, are identified in the IEB database. This period usually overlaps the timeframe used to select the establishments from the IAB Establishment Panel data. In a next step, the employment biographies of the selected individuals are extracted from the IEB database. Because of storage issues, employment biographies may be censored in the LIAB Longitudinal Model, i.e., the employment biographies may not cover the full employment history of an individual as registered in the IEB database. In order to compensate the censoring problem, additional variables providing information on the (un-)employment status before or after the censoring are added to the LIAB Longitudinal Model.

Current version of the LIAB Longitudinal Model (LIAB LM 9310)

For the current version of the LIAB Longitudinal Data (see Klosterhuber et al., 2013) all panel cases for the time period 2000 to 2008 are selected. They consist of establishments contained in the starting year, "new" establishments in the subsequent waves and establishments going out of operation during the panel period (see Fischer et al., 2008). The employment period for individuals in these establishments ranges from 1999 to 2009 and the employment biographies cover the years 1993 to 2010. Figure 2 illustrates the sampling procedure.
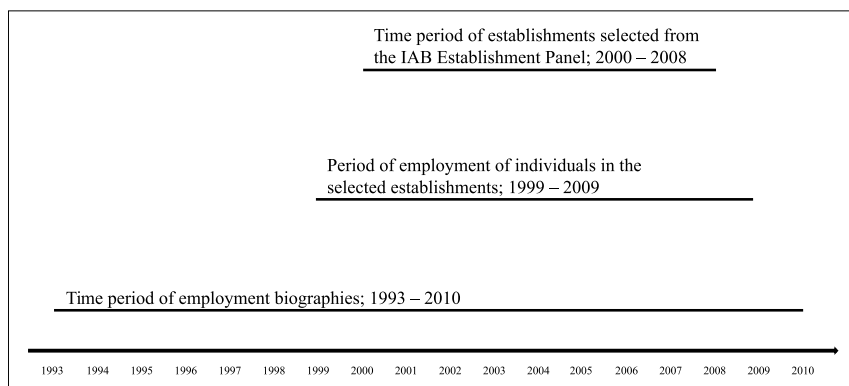

Figure 2: Sampling of the LIAB Longitudinal Model 9310 (LIAB LM 9310).

An important difference between the current versions of the LIAB Longitudinal Model and the LIAB Cross-sectional Model is the fact that in the longitudinal version, only a restricted set of information on employees is available. In particular, information on workers in the LIAB Longitudinal Model cover only employment spells and spells of benefit receipt according to Book III of the German Social Code. Information on job-search is not included in the LIAB LM 9310.

Given this sampling procedure, 1,883,198 workers in total are included in the LIAB LM 9310, which may be linked to 146,781 establishments from the IAB Establishment Panel data.

Older versions: LM1 to LM 3

As well as other data products available at the Research Data Centre (FDZ), the various models of the LIAB data have underwent several re-designs. These apply mainly to the longitudinal versions since the basic sampling framework for the Cross-sectional model stayed unchanged over time. In total, there are three predecessors to the current LIAB LM 9310 data. They differ in the sampling period for the establishments from the IAB Establishment Panel data, the times of employment in these establishments and the censoring dates for the employment biographies. Moreover, the sampling criterion for the establishments was different, too. Table 1 provides an overview of the old LIAB Longitudinal models.

Table 1

Older versions of the LIAB Longitudinal Model

	(Sampling) Period		
	Longitudinal Model 1 (LIAB LM1)	Longitudinal Model 2 (LIAB LM2)	Longitudinal Model 3 (LIAB LM3)
Establishments	1999–2001	2000–2002 plus all establishments included in the LM1 data	1996–2005
Workers	1999–2001	1997–2003	1993–2006
Employment Biographies	1990–2006	1993–2006	1993–2006

3.3 LIAB Mover Model

Although in principle a longitudinal model, the LIAB Mover-Model (see Heining et al. 2012) is based on a different sampling procedure and was specifically designed for the study worker transitions between establishments.

As for the Longitudinal Model, the IAB Establishment Panel data is the starting point of the sampling for the LIAB Mover Model. First, all establishments of the IAB Establishment Panel data are dropped for which the number of employees in the survey deviates by more than 50 per cent of the number of employees according to the individual data for at least one year. From the remaining establishments those are selected who employ at least one mover. A mover is defined as an employee who is employed in their main job at least at two different establishments of the IAB Establishment Panel with a valid survey on two different reference dates of the IAB Establishment Panel (June 30th of the respective year).

In addition, the movers up to 500 employees per establishment were chosen randomly. This means that every employee was chosen for establishments below 500 employees and a maximum of 500 employees were choosen for establishments with more than 500 employees. The remaining employees either stay in the establishment or move to an establishment, which did not take part in the survey.

Given this sampling, the LIAB Mover Model contains information on 2,361 to 8,879 establishments per year. The number of individuals lies between 3,175,801 and 3,815,061 per year. In total, 24,709 establishments and 4,666,926 workers are covered by the LIAB Mover Model.

4. Access to the LIAB data

All models of the LIAB data may be accessed on-site at the sites of the Research Data Centre of the Federal Employment Agency with subsequent remote execution. Details on the application process for the dataset and possibilities for data processing are provided on the website of FDZ. Access to the outdated models of the LIAB data such as LM1 to LM3 is only provided for purposes of replication studies.

5. Summary

The Linked Employer-Employee Data of IAB consist of a unique combination of rich survey data on establishments with high-quality administrative records on workers. Over the years, the LIAB data have been accessed by many researchers and have been applied to a wide area of topics.

The FDZ will continue to update existing models but will also eagerly pursue the development of new data models for the LIAB data. Key efforts in this context will be for example the representation of the firm level in the LIAB data and the adding of information on financial statements. This continuous improvement will ensure that the LIAB data will remain a powerful and rich resource for empirical research in economics, sociology and statistics.

References

Abowd, J. / *Kramarz*, F. (1999): The Analysis of Labor Markets Using Matched Employer-Employee Data, in: Ashenfelter, Orley; Card, David (eds.), Handbook of Labor Economics, Amsterdam, 2629 – 2733.

Alda, H. / *Bender*, St. / *Gartner*, H. (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB), Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften 125 (2), 327 – 336.

Bellmann, L. (2002): Das IAB-Betriebspanel. Konzeption und Anwendungsbereiche, Allgemeines Statistisches Archiv 86 (2), 177 – 188.

Brixy, U. / *Fritsch*, M. (2002): Die Betriebsdatei der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: M. Fritsch & R. Grotz (Hrsg.), Das Gründungsgeschehen in Deutschland. Darstellung und Vergleich der Datenquellen, Heidelberg, 55 – 78.

Fischer, G. / *Janik*, F. / *Müller*, D. / *Schmucker*, A. (2008): The IAB establishment panel * from sample to survey to projection. (FDZ-Methodenreport, 01 / 2008 (en)), Nürnberg.

Heining, J. / *Jacobebbinghaus*, P. / *Scholz*, Th. / *Seth*, St. (2012): Linked-Employer-Employee Data from the IAB: LIAB-Mover-Model 1993 – 2008 (LIAB MM 9308). (FDZ-Datenreport, 01 / 2012 (en)), Nürnberg.

Heining, J. / *Scholz*, Th. / *Seth*, St. (2013): Linked-Employer-Employee data from the IAB: LIAB Cross-sectional model 2 1993 – 2010 (LIAB QM2 9310). (FDZ-Datenreport, 02 / 2013 (en)), Nürnberg.

Klosterhuber, W. / *Heining*, J. / *Seth*, St. (2013): Linked-Employer-Employee data from the IAB: LIAB longitudinal model 1993 – 2010 (LIAB LM 9310). (FDZ-Datenreport, 08 / 2013 (en)), Nürnberg.

Kölling, A. (2000): The IAB-Establishment Panel, Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften 120, 291 – 300.

Spengler, A. (2008): The Establishment History Panel, Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften 128 (3), 501 – 509.

Willis, R (1986): Wage Determinants, in: O. Ashenfelter / R. Layard (eds.) Handbook of Labor Economics, Amsterdam, 525 – 602.