Schmollers Jahrbuch 128 (2008), 501 – 509
Duncker & Humblot, Berlin

The Establishment History Panel

By Anja Spengler

1. Introduction

Since the beginning of 2007, the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) (cf. Kohlmann, 2005) offers the Establishment History Panel (BHP).

Before 2007, the only establishment-related data sets available at the FDZ were the Establishment Panel (cf. Kölling, 2000) and the Linked-Employer-Employee Data (LIAB) (cf. Alda / Bender / Gartner, 2005). Both data sets are based on disproportionately drawn samples comprising approx. 16,000 establishments. Due to sample size and disproportionate sampling, the analysis possibilities regarding smaller establishments are restricted substantially.

The Establishment History Panel, in contrast to the Establishment Panel, consists of person-related data from the registration for social security contributions which were aggregated on the establishment level. In contrast to the Establishment Panel and the Linked-Employer-Employee Data (LIAB), the BHP is not a random sample but covers the total population.

Each year, the BHP includes between 1.5 and 2.5 million establishments. The data are available for the years 1975 up to 2005. In comparison, Establishment Panel and LIAB data start in 1993 only. The BHP's yearly cross-sections can be merged to form a panel data set because the BHP includes the necessary identifiers.

Comprehensive empirical studies concerning workplace dynamics are possible with this data set. A further considerable advantage of the BHP is that questions based on regional characteristics can be analysed and that the founding or closing date of an establishment can be identified. The possibility to combine the annual data sets into a single panel data set opens up a wide range of research questions.

This article is intended to provide an outline of the data content, data access and possible research questions. The main topic of the following text is a description of the BHP base file. I will also sketch some of the future steps we will take to extend the data set by means of so-called extension files. Finally, I will describe a new FDZ project which uses BHP data.

2. Mandatory Social Security Notification

The mandatory social security notification was introduced in West Germany in 1973 and in East Germany in 1991. The last change to the procedure of registration was in 1999. By law, employers are obliged to notify the authorities of entries or departures of employees liable to social security and (since 1999) of marginal part-time employees. Additionally, they have to yearly report on all employees as of December 31.

These notifications are processed and stored by the IAB in a file entitled "Employee and Benefit Recipient History" (BLH), which consists of the Employee History (BeH) and the Benefit Recipient History (LeH). The BeH contains all pay notifications for social security. The LeH contains information on benefits drawn from the BA in the respective period. The BLH is then used to build the BHP.

All employees who are subject to at least one of the following compulsory insurances are liable to social security: health insurance, long-term care insurance, pension insurance, unemployment and accident insurance. However, not liable to social security and thus not included in the data are civil servants, conscripts, those doing alternative civilian service, self-employed, judges, scholars, students, pensioners, clergy and others.

It should be noted that the Establishment History Panel is biased with respect to establishments in certain industries. This affects especially agriculture and forestry, with a high percentage of self-employed, and the public sector (including social security), with a high percentage of civil servants (cf. Fritsch / Brixy, 2004).

The following information is contained in the notification of social security, among other data, and is incorporated in the BLH:

- social security number (indicating gender and date of birth),
- establishment number,
- code number of the municipality in which the establishment is located,
- industry affiliation,
- profession,
- employment status,
- educational level attained and vocational training concluded,
- nationality and
- gross income.

The Establishment History Panel records all establishments that have, according to the BLH, at least one employee as of June 30 each year. Since 1999, the BLH also includes part-time employees who are not liable to social

security. This gives rise to an increase of the number of establishments in the Establishment History Panel in 1999. It should be noted that the definition of part-time employment has changed several times, and therefore the number of part-time employees has fluctuated during the course of the years.

3. Aggregation

With the aid of the establishment number, the BLH data were aggregated from the employee level to the establishment level. The establishment number is allocated by the local Employment Agency responsible for the district in which the establishment is located. The number is unique and with it is associated the name, address and the economic classification of the establishment.

"Generally, establishment identifies a certain location, i.e. a plant or a workplace. There are two main deviations from this definition. First, if a firm has several plants in one municipality which are affiliated to the same industry, the employees in these plants may be subsumed under the same establishment number. And second, a certain establishment may be split into several establishment numbers if this seems appropriate to the management for any reason." (Fritsch / Brixy, 2004, 185).

Prior to the aggregation procedure, the activities of multiple job holders have been divided into one main job and one or more secondary jobs. The main job is defined as the one with the highest daily wage. If a multiple job holder earns the same wage in two jobs, the job held for the longest period of time is selected. Marginal part-time employment is considered as main job only if no employment subject to social security is registered for that period of time. Also, secondary activities are permitted in the data in different establishments only. "Excess" jobs in the same establishment have been deleted before aggregation.

4. What's in the Data

Each yearly cross-section of the Establishment History Panel includes all establishments in Germany that had at least one employee liable to social security or – since 1999 – at least one part-time employee on the reference date (June 30 of each year, respectively).

As the BHP is based on process generated data it lacks the disadvantages of survey data. Above all, the information from the Establishment-History Panel can be seen as very reliable, as the employers are obliged to provide this information. This is why typical problems concerning survey data such as memory gaps, refusing to give information or deliberately giving false information do not exist in the BHP. This advantage is particularly important when different years from the BHP are linked to a panel. In a case such as this, refusing to

give information, for example, has a minor effect on the panel mortality of the BHP in comparison with survey data (cf. Bender / Haas / Klose, 2000; Koch / Meinken, 2004).

Also, the data can be expected to be of high quality. The main reason for this is the notification requirement of employers, described above. Moreover, the social security bodies use a common control procedure concerning the reported data. Insurance numbers, addresses, the durations of employment and the wages reported are subject to a special control. (cf. Koch / Meinken, 2004).

As the aggregated variables are consistent over time, it is possible to construct a panel data set by combining some or all BPH waves 1975 – 2005. Between 1.3 and 2.5 million establishments can be found in the Establishment History Panel p.a.:

– 1975: total: 1,291,413 establishments
– 1991: total: 1,743,528 establishments
– 1992: total: 1,901,357 establishments; West: 1,542,882 establishments; East: 288,022 establishments (East / West without Berlin)
– 1998: total: 2,040,358 establishments; West: 1,569,255 establishments; East: 391,710 establishments (East / West without Berlin)
– 1999: total: 2,489,280 establishments West: 1,967,088 establishments; East: 428,078 establishments; establishments with only "marginal" part-time employees – no full-time employees: 419,981 establishments (East / West without Berlin).

5. Groups of Variables

Table 1

Groups of Variables

Categories	Descriptions
Establishment variables	Anonymous establishment number, German state, industry code WS73, industry code WZ93, industry code WZ03
Aggregated variables	total number of employees, number of employees by main occupation, number of employees with daily rate of pay equal zero, number of female employees, number of German citizens employed, number of foreign citizens employed, number of employees by training / qualification groups, number of trainees / apprentices, number of employees by occupational status, number of employees by age groups, mean value and standard deviation of age, quartile values of daily gross wage of full-time employees for different subgroups and of total daily gross wage

The establishment variables allow a differentiated analysis not only by region but also by sector of economy. The regional codes have not changed during the years. The classification of industries, however, has changed. Therefore, industry code WS73 is included in the data for the years up to 2002. From 1999 until 2003, the classification of industries can be identified via WZ93. Since 2003 the industry code WZ03 is valid. Due to these changes carried out in the industry classifications over the years, it is not possible to regard these variables as panel variables over the whole period of time of the BHP. At the moment there is now easy way to combine the different industry codes into one overall code.

An establishment can always be identified by its anonymous establishment number because this number is constant across the yearly cross-sections. As the method of generating the aggregated variables is the same for every cross-section it is possible to analyse, e.g., the number of employees or the development of daily gross wages over time. It must be born in mind, however, that part-time employees have only been included in the data since 1999 so that it is not possible to analyse this group over the years. For the years after 1998, the variable "full-time employees" was adjusted with regard to part-time employed individuals. There is a separate variable indicating the number of part-time employees.

6. Related Data

The FDZ is currently providing three types of establishment data sets:

The **IAB Establishment Panel** is an annual representative survey on various topics such as the determinants of labour demand. It has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany. The IAB Establishment Panel is the central basis for the analysis of labour demand in Germany (cf. Kölling, 2000).

In the **Linked-Employer-Employee Data (LIAB),** data from the IAB Establishment Panel are combined with longitudinal information on employees who are subject to social security. Several versions of LIAB datasets are available, which differ in the selection of establishments and the time span covered (cf. Alda / Bender / Gartner, 2005).

As previously stated, process-generated data on establishments are primarily to be found in the BHP. However, variables on establishments are also included in the basic file of the IAB Employment Samples and in the BA Employment Panel. Yet most of these variables indicate proportions of specific employee groups within the establishment only, as both datasets focus on individual data and the information about establishments is meant to be supplementary only.

Generally, the cross-sectional BHP files could be merged with other data sets available at the FDZ. This would allow an even broader range of research

questions to be addressed. (For further information about available data sets, please visit the FDZ website at http://fdz.iab.de.)

7. Data Access

At present, several separate random samples are being taken for researchers wishing to work with the Establishment History Panel data. In future, a uniform random sample for all researchers will be available.

The design of the sample drawn depends on the requirements of the specific research project. Generally, an application must be made before data access is possible. This application will have to be approved by the Federal Ministry of Labour and Social Affairs and a contract must be signed. The data access to the BHP is regulated by law in ì 75 book 10 of the German Social Welfare Act (SGB). All person-specific information collected by the FEA in order to provide unemployment insurance, measures of labour market policy or job placement are so-called social data (*Sozialdaten*) and have to be especially protected. Access to social data requires the following conditions to be met:

– Scientific research regarding social security
– Prevailing public interest
– Permission of the Federal Ministry of Labour and Social Affairs

External researchers can make use of the BHP during visits to the Research Data Centre (FDZ) or by means of remote data access. However, as the BHP is difficult to handle it is considered indispensable by the FDZ that researchers visit the FDZ once prior to using the data via remote data access.

Remote data access means that the researcher prepares program code (Stata Do-files, for instance) with the help of the data documentation and test data. The researcher then sends the programs to the FDZ by e-mail. FDZ staff run the code, check the files that contain the results and erase information suited to identify individuals. The remaining results are sent back to the researcher (Jacobebbinghaus, Seth, 2007).

8. KombiFiD – Combined Firm Data for Germany

The FDZ is currently cooperating in a project named “KombiFiD” with the Research Data Centre of the Federal Statistical Office of Germany, the Research Data Centre of the Statistical Offices in the German States and the Institute of Economics of the University of Lüneburg (cf. Bender, Wagner, Zwick, 2007). The BHP is one of the data sets relevant for this project. Within the KombiFiD project, not only data of different producers will be merged but also the legal foundations to establish a permanent connection between these

data should be created. During the project, the following data sets will be combined, among others:

– Business register [Unternehmensregister (URS)]
– Cost structure survey [Kostenstrukturerhebungen] (cf. Fritsch / Görzig / Hennchen / Stephan, 2004)
– Salary and wage structure survey [Gehalts- und Lohnstrukturerhebung] (cf. Stephan, 2001)
– Establishment level data based on the employment statistics [e.g. BHP]

The data sets will be merged based on the URS because the URS has the necessary identifiers. The editing and combining of the data will be the tasks during the first project phase. In the following second phase, the data will be offered to the scientific community.

The goals of the project are, on the one hand, the combination of selected data sets beyond the limits of the individual labour market data producers and, on the other hand, offering them to scientific research projects. For this purpose, selected data for the years 1995 to 2006 of the statistical offices and the IAB will be combined. The new data will open up completely new perspectives to research and applied economics.

KombiFiD's successful conclusion and the establishing of a permanent regulatory framework concerning data merging will also reduce stress for establishments: today many basic statistics have to be repeatedly reported by establishments. The unique opportunity of merging the establishment' information thus offers a high potential of rationalisation.

9. Outlook

The BHP includes detailed information about various sectors. Therefore, the data set is qualified for analysis concerning different fields of research, e.g. foundings and closings of businesses, regional differentiation of economic circumstances and its effect on workplace dynamics. Furthermore, studies about different wage levels as well as a differentiation according to industrial sectors are possible.

In contrast to other establishment data sets, the Establishment History Panel is based on process generated data and on the total population. Therefore advantages that accrued are the absence of typical survey data problems such as memory gaps, refusing to give information or deliberately giving false information. The data of the BHP is to be seen as of very high quality. The main reason for this is the notification requirement of employers. The aggregated data of the BHP is available for the years 1975 up to 2005. The BHP cross-sections can be linked to a panel because the BHP data includes the necessary identifiers.

The preceding text relates to the basic version of the Establishment History Panel. Extended versions will also be available. So-called extension files will include diverse supplementary information, e.g. about foreign employees or differentiated job-related characteristics. As is the case with the basic version, the extensions will also be drawn on the reference date (June 30 each year). Apart from cross-section characteristics, dynamic characteristics such as in- and outflows can be analysed with the aid of the extensions. It is further planned that we will regularly offer extensions, depending on requests from users. Of course, it will be possible to link the basic file with the extension files.

To date, it is only inadequately possible to combine the Establishment History Panel and individual-level data. For a comprehensive empirical analysis of numerous scientific objectives, data sets are necessary which include information about the employees as well as information about the establishment in which the employees are employed (cf. Kaiser / Wagner, 2007). At the same time, key establishment data, including information collected extensively for statistical purposes, have a limited availability. That's why it is planned, in conjunction with the BHP, that individual-level data and establishment data be combined. The result will be a great expansion of analysis possibilities. Two individual-level data sets of the FDZ can be merged with the BHP. These data sets are the IAB Employment Sample (IABS) and the Integrated Employment Biographies Sample (IEBS). Like the BHP, the IABS and the IEBS are based on the employee history and the benefit recipient history compiled in the IAB. The IEBS integrates other available data sources, too, such as the participants-in-measures and applicants-pool databases. The IABS has daily data on employment subject to social security as well as on receipt of unemployment benefits, unemployment assistance and maintenance allowance. Additionally, the data set contains a number of establishment characteristics. The IEBS comprises event history data on employees liable to social security, benefit recipients, persons who are searching for employment, unemployed persons and participants in measures of active labour market policy. The link between the BHP and the IABS or the IEBS would enable to perform analyses on the employee and benefit recipient history also in conjunction with establishment characteristics. Furthermore it allows more detailed overviews of employment histories as well as comprehensive analyses on active labour market policies.

Within the framework of the KombiFiD project, the establishments of the BHP will be aggregated to the company level. Moreover, it is planned to check the possibility of including data of other data producers into the KombiFiD project.

References

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), Schmollers Jahrbuch 125 (2), 327 – 336.

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

Bender, S. / *Wagner,* J. / *Zwick,* M. (2007): KombiFiD – Kombinierte Firmendaten für Deutschland, Konzeption einer Machbarkeitsstudie für eine Zusammenführung von Unternehmensdaten der Statistischen Ämter, des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit und weiterer Datenproduzenten, FDZ-Methodenreport, No. 5 / 2007.

Bundesanstalt für Arbeit (Hg.) (1997): Handbuch für die Betriebsnummernstellen der Arbeitsämter zur Aktualisierung der Betriebsdatei mit dem IT-Verfahren coBer-Betriebe / coBer-coStat.

Dundler, A. / *Stamm,* M. / *Adler,* S. (2006): The Establishment History Panel. FDZ Datenreport, No. 3 / 2006.

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

Fritsch, M. / *Görzig,* B. / *Hennchen,* O. / *Stephan,* A. (2004): Cost Structure Surveys for Germany, Schmollers Jahrbuch 124 (4), 557 – 566.

Jacobebbinghaus, P. / *Seth,* S. (2007): The German Integrated Employment Biographies Sample IEBS, Schmollers Jahrbuch 127 (2), 335 – 342.

Kaiser, U. / *Wagner,* J. (2007): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, forthcoming in: Perspektiven der Wirtschaftspolitik

Koch, I. / *Meinken,* H. (2004): The Employment Panel of the German Federal Employment Agency, Schmollers Jahrbuch 124 (2), 315 – 325.

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

Kohlmann, A. (2005): The Research Data Centre of the Federal Employment Service in the Institute for Employment Research, Schmollers Jahrbuch 125 (3), 437 – 447.

Stephan, G. (2001): The Lower Saxonian Salary and Wage Structure Survey – Linked Employer-Employee Data from Official Statistics, Schmollers Jahrbuch 121 (2), 267 – 274.