

European Data Watch

This section offers descriptions as well as discussions of data sources that are 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 focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to (wagner@leuphana.de). Past “European Data Watch” articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: <http://www.ratswd.de>.

The IAB Establishment Panel – Things Users Should Know*

By Gabriele Fischer, Florian Janik, Dana Müller,
and Alexandra Schmucker

1. Introduction

The IAB (Institute for Employment Research) Establishment Panel is an annual survey of establishments and is unique in Germany, as it represents all industries and establishment sizes nationwide and can also be analysed on a

* We would like to thank Hans Kiesl and Stefan Bender for their constructive criticism. We are grateful to Lutz Bellmann, Harald Bielencki, Udo Brixy, Micha Andreas Crimmann, Peter Ellguth, Nicole Gürtzgen, Tanja Hethey, Martina Huber, Peter Jacobebbinghaus, Susanne Kohaut, Johannes Ludsteck and Dieter Urban for helpful comments on the FDZ Methodenreport (Fischer et al., 2008) which this article is based on. Last but not least we would like to thank Claudia Schiener and her team for translation of the FDZ Methodenreport.

longitudinal basis. The survey began in West Germany in 1993, with the aim of building up a representative information system for continuous analysis of labour demand. It has been carried out in East Germany since 1996, making it a nationwide survey. The IAB Establishment Panel is conceived as a longitudinal survey, i.e. a large majority of the same establishments are interviewed every year. Consequently, it enables both analysis of developments across time through comparison of cross-sectional data on different points in time, and also longitudinal studies of individual establishments.

Now in the IAB Establishment Panel approx. 16,000 establishments are surveyed on a large number of employment policy-related subjects, including employment development, business policy and business development, investment activities, innovations in the establishment, public funding, personnel structure, vocational training and apprenticeships, new and existing personnel, recruitment, wages and salaries, working times in the establishment, further training and general data on the establishment. The survey also includes varying focal topics every year.

The IAB Establishment Panel contains high data quality, achieved by means of the high-quality sample, the high response rate and the sophisticated process of data monitoring and error correction. The survey is carried out by TNS Infratest Sozialforschung GmbH on behalf of the IAB. A general introduction to the IAB Establishment Panel is contained in German in Bellmann (2002) or in English in Koelling (2000). The IAB Establishment Panel is based on a complex study design, which also presents challenges for users of the dataset.

This paper provides an overview of the methodology of the IAB Establishment Panel and complements the existing articles (Bellmann, 2002; Koelling, 2000). It describes in detail the sample and survey design, the weighting process and data access at the Research Data Center (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). This paper is also aimed at users of the IAB Linked-Employer-Employee Dataset (LIAB¹), in which the IAB Establishment Panel is an important component. More detailed information about the IAB Establishment Panel is described in Fischer et al. (2008).

2. Sample Design

The population of the IAB Establishment Panel consists of all establishments with at least one employee liable to social security as of 30 June of the previous year. The basis for sampling is the BA establishment file, which is created on a quarterly basis and contains some two million establishments.

¹ This dataset contains both the IAB Establishment Panel data and the process-produced data on individuals of the (BA). This enables a simultaneous analysis of the supply and demand sides of the labour market (Alda et al., 2005).

The establishment file contains all establishments that notify the social security agencies of their employees as required. The establishments receive an establishment number for these notifications from the respective local Employment Agency responsible for the establishment, as of 2007 from the BA's central establishment number service. These establishment numbers are compiled centrally in the BA establishment file. The establishment number is the relevant unit for the sampling and weighting processes.

2.1 The IAB Establishment Identification Number

The establishment number is initially a necessary indicator for the administration process. Every establishment is allocated an establishment number as part of the notification procedure for social security. An "establishment" in this sense denotes a regionally and economically separate unit, in which employees liable to social security work. Specific rules (Bundesagentur für Arbeit, 1997) govern the allocation of establishment numbers, and are applied in the BA's central establishment number service and former in the individual Employment Agencies. New establishment numbers are generally allocated when:

- an establishment is newly founded,
- a change of establishment owner takes place,
- the economic sector of the establishment changes, or
- the employer submits an application, e.g. to combine several branches.

No new establishment number is allocated if an establishment moves to a different local authority district or a different employment agency district, unless the employer or health insurer files an application. Establishments in possession of an establishment number but not currently employing any employees liable to social security are referred to as "dormant". The old establishment number can be used again.

In most cases, the establishment number is a suitable basis for identifying an establishment as a survey unit. However, this is not always the case. The above criteria for allocating establishment numbers imply, for example, that a single company may hold several establishment numbers if its units are in different employment agency districts or constitute differing, separate economic units. Section 3.3 contains further details on the establishment number and the identification of the establishments to be surveyed.

2.2 The Sampling Process

The sample for the IAB Establishment Panel is drawn from the establishment file as of 30 June of the previous year, because the information on social

security notifications is only available approximately six months after the respective reference date (Brixy / Fritsch, 2004). However, the precise number of employees liable to social security is not available for all establishments as of the reference date. For approximately 5 % of employees liable to social security, the information is unavailable as of the reference date, due to differing notification dates (Bundesagentur für Arbeit, 2005). In these cases, information is adopted on the basis of the most recent available notification.

For the purpose of drawing the sample, the target values for the net sample are established in dependence on the financed interviews on the federal state (Bundesländer) level and for the manufacturing industry in East Germany. For these 17 partial samples (since 2003), a target degree of completeness is set for every cell of the stratification matrix (17 industries and 10 establishment size² classes) according to the “principle of optimum stratification” (Buttler / Fickel, 2002 p. 147 f.). As a result, large establishments, small federal states and small industries and the manufacturing industry in East Germany³ are overrepresented. This disproportionate stratification has consequences for analyses, which are described in Chapters 4 and 5. The size of the first interviewee sample is then calculated on the basis of the various expected response rates of the individual samples and survey methods.

The stratification matrix has altered over time for the industries, as a result of changes in the system of economic sector classification (from WS73 to WZ93) and modified aims with relation to content.⁴

Up to 2006, separate samples were drawn for each federal state (Bundesland) and for West and East Germany. Up to and including 2006, East Germany included East Berlin and West Germany included West Berlin. It is, however, no longer possible to allocate establishment numbers geographically precisely to West or East Berlin in the establishment file, and Berlin has thus been treated as a single entity since 2007. This has effects on the definition of West and East Germany. As of 2007, West Germany consists of all western German federal states apart from Berlin, with the entire city of Berlin defined as part of East Germany. This applies for the sampling process and thus also for the projection.

² For the sample and weighting of the IAB Establishment Panel, the term “establishment size” denotes the number of employees liable to social security in the establishment.

³ This extension sample is financed by the Institute for Economic Research Halle.

⁴ At the beginning of the panel in 1993, the stratification matrix consisted of 16 economic sectors, increasing to 20 in 2000 and falling to 17 in 2004. Particularly the shift from WS73 to WZ93 in 2000 resulted in significant changes in the classification of establishments, meaning that comparisons across the entire survey period are only possible to a limited extent. Nor can a more aggregated classification achieve comparability. The change in the system of classification for economic sectors from WZ93 to WZ2003 does not restrict comparability of industries, as the changes took place below the level of classification used in the IAB Establishment Panel.

2.3 The IAB Establishment Panel Partial Samples

The longitudinal character of the IAB Establishment Panel is also reflected in the sample. Firstly, the IAB attempts to survey as many establishments as possible over an extended period. Secondly, the IAB Establishment Panel sample must also depict the dynamics of establishment closures and “new” establishments. The annual gross sample thus consists of four respective partial samples:

1. responding establishments from the previous year (“continuers’ sample”),
2. non-respondents from the previous year willing to being surveyed again,
3. “new” establishment numbers and
4. extension sample.⁵

These partial samples are necessary to depict continuity and change in the establishment population. Sample 1, the responding establishments from the previous year, consists of the establishments that are part of the existing stock of establishments from one year to the next. This sample ensures the longitudinal character of the IAB Establishment Panel.

The non-respondents from the previous year (sample 2) raise the number of cases in cross-sectional terms. In a survey of establishments over an extended period of time such as the IAB Establishment Panel, a concentration solely on the establishments continuing to exist from one year to the next would lead to selection effects. Establishments that have existed over an extended period differ in many operative characteristics from newly founded establishments. In order to depict this dynamic, “new” establishment numbers (sample 3) are added to the IAB Establishment Panel sample every year. These establishment numbers had at least one employee liable to social security as of the reference date, but not in the previous year. Such an establishment number does not necessarily denote a newly founded establishment. It can be an establishment that has existed for some time, but has only recently taken on an employee liable to social security. This definition deviates from that of newly founded establishments.⁶ Despite these difficulties arising from the establishment file system, this procedure is the only feasible option for depicting the activity of founding new establishments within the parameters of the sampling process.

In addition to the samples described above, it is necessary to add further existing establishments as of the reference date, in order to make up for losses and achieve the required number of cases in the individual federal states (sample 4).

⁵ With the exception of Hamburg, all German federal states currently fund regional extension samples to the IAB Establishment Panel.

⁶ The attribute “new” is always written in quotation marks below for these establishment numbers, so as to prevent false conclusions. Genuine newly founded establishments can be identified via the foundation year surveyed in the questionnaire.

3. Survey Design, Field Work and Editing

3.1 Questionnaire Design

As well as the sample, the questionnaire also has to take the longitudinal character of the IAB Establishment Panel into account. The main aim is to gather certain information on a regular basis in order to measure developments. Simultaneously, up-to-date questions relevant for labour market policy also have to be included in the survey. The IAB Establishment Panel questionnaire contains numerous questions that are asked in every wave, so as to depict changes consistently over time. This basic programme of questions is generally identical over the years.⁷

Up to the 2007 wave, this basic programme consisted of the subject blocks Employment Development, Business Policy and Development, Vocational Training, Personnel Structure and Personnel Movements, Investments, Wages and Salaries and Adherence to Collective Agreements. In addition, specific subject blocks are also regularly included in the questionnaire at certain intervals, e.g. subjects such as further training, working time, public funding and innovations.

The design of the questionnaire was changed slightly in 2007. The basic programme was supplemented by basic indicators from the previously multi-year subject blocks further Training, innovation and working Times. These areas will be surveyed annually from the 2008 wave. For all subject blocks regularly surveyed in the IAB Establishment Panel, the basic information is now available every year, thereby raising the analysis potential, as content-based links are also possible on a cross-sectional basis. At two-year intervals, these basic indicators are supplemented or extended by additional questions.

Alongside the fixed elements of the questionnaire, varying current focus subjects are included every year. These include, for example, questions on demand for qualified employees, on employment of older workers or securing employment and location.

The design of the IAB Establishment Panel was developed and extensively tested in the early 1990s. This process also involved parallel development activities taking place on the Hanover Firm Panel, which were carried out on behalf of the Forschungsstelle Firmenpanel at the University of Hanover (Gerlach/Hübler/Meyer, 2003) and the Institute for Applied Economic Research

⁷ Several questions, however, have had to be changed at some point. A synopsis of all panel variables are available on the FDZ website <http://fdz.iab.de>. Furthermore the FDZ Datenreport 2/2006 (Alda et al., 2006) contains a description of adaptations of sample variables to remain consistent over time. Additionally, PanelWhiz, a Stata tool for creating panel datasets quickly and easily, is due for launch shortly (for further information on PanelWhiz: <http://www.panelwhiz.eu/>).

in Tübingen (IAW). Before the survey was extended to cover the East German federal states, test surveys were carried out by TNS Infratest and the Institute for Socio-economic Structural Analysis (SÖSTRA) on behalf of the IAB. Furthermore, newly developed or reformulated questions are tried out in pre-tests. The specially trained IAB-interviewers not only test the real interview situation, but also deal with any problems in understanding or interpretation with the interviewee. The findings from these pre-tests are integrated into the development of the final questionnaire.

3.2 Survey Method

The survey is generally carried out in the form of face-to-face interviews in the establishments, by interviewers. Interviewer continuity is a decisive factor for the success of the survey. TNS Infratest therefore aims to send the same person to carry out interviews in each respective establishment. As the questionnaire contains a relatively large number of questions on figures, which the interviewees cannot answer spontaneously, the questionnaire may be left behind in the establishment. This means that data not available during the interview is researched afterwards and added to the questionnaire. This option is mainly used by larger establishments. The majority of establishments, however, are surveyed on an entirely face-to-face basis (2006: 73 %). Written surveys have taken place in some cases since 2000. This became necessary because extension samples had to be realised in certain federal states in order to make analysis possible on the federal state level. The financial means were not sufficient for the large number of face-to-face interviews, however, which is why a large part of the sample is surveyed by post in these federal states.

3.3 Field Phase

As the interviews take place in the third quarter, the questions on stock figures generally refer to the reference date of 30 June, and the questions on flow figures (e.g. newly recruited or exiting personnel) to the first half of the survey year. This is designed to minimise memory errors. Both the face-to-face interviews and the written postal surveys are generally completed by mid-October. In parallel, the surveyed data undergoes checks and errors are eliminated (see section 3.5).

Before the interview starts, the interviewers have to ensure that the interview is carried out for the correct establishment unit. It is essential for longitudinal analyses that the interviews always refer to the same establishment unit. This is the only way to ensure that changes over time (such as employment development) depict actual changes and are not based on deviations in the surveyed establishment unit. The correct establishment unit is identified at

the beginning of the interview, on the basis of the establishment name, the number of employees liable to social security on the reference date (30 June of the previous year) and – in the event of deviations – the establishment number.

In the case of first-time respondent establishments, the employee figures are compared with the number of employees liable to social security as of the reference date (30 June of the previous year) according to the establishment file. If it proves impossible to carry out the interview with precisely the unit drawn in the sample because this does not constitute an economically viable unit or there is no specific information available on this unit in the establishment, the survey may deviate from the originally selected unit, in accordance with fixed rules. The interviewed establishment unit must, however, bear some relation to the establishment number drawn. This may be the case if a larger establishment unit is surveyed which contains the establishment number, or if a smaller establishment unit is surveyed for which the establishment number partially applies. These deviations are documented in the address protocol, so that the same (deviating) unit can be surveyed again in the following year. If it proves impossible to find a link to the establishment number drawn for first-time respondent establishments, no interview is carried out.

In the case of continuer respondent establishments, the correct establishment unit is determined on the basis of the employee figures from the previous year's questionnaire. If deviations occur, the establishment number is again used to determine whether at least part of the establishment or a larger unit bears the establishment number in question. If this is the case, the survey is carried out and the relationship between the unit surveyed and the establishment number is documented in the address protocol. These establishments can then be evaluated for the cross-sectional survey, but are no longer available for panel analyses. The establishments surveyed by post receive the relevant information for determining the correct survey unit along with the questionnaire, and are requested to refer all data to this unit.

3.4 Non-Response to Interviews and Questions

The response rate to the surveys has varied between 63 % and 73 %.⁸ The variations in the response rates are mainly due to differing sizes of the extension samples. As the response rates among establishments surveyed for the first time are significantly lower than those of continuer establishments⁹, the total response rate is much lower in the years with large extension samples. Alongside differences between first-time and continuer respondents, the response rate also differs according to the survey method. Establishments are far

⁸ This calculation does not include interviews unsuitable for evaluation.

⁹ E.g. the response rate for the new entries sample in 2006 was 30 %.

more willing to participate in the face-to-face interview than in the written survey. For example, in 2006 the response rate among first-time establishments contacted by post was 12.9% – significantly lower than the establishments first interviewed orally at 36.3%. Similarly, in the same year the continuer establishments were less willing to participate in the written survey at 61.9% than in the on-site interviews. The response rates for the face-to-face interviewed continuer establishments, however, are stable at between 81% and 84%.

To judge the survey quality, one must look at the item non-responses to specific questions as well as the unit non-responses to the entire survey. Questions with high item non-response rates are either hard to understand, hard to answer, or participants frequently refuse to respond to them.¹⁰ Across the waves of the IAB Establishment Panel, the sensitive variables have always the highest item non-response rates, but these are relatively stable e.g. business volume (19%)¹¹, total wages and salaries (14%), share of advance performance and cost of debt in total sales (17%) and total investment grants (14%). In the written survey, the “no response” rates are considerably higher than in the oral interviews. The lower rate of non-response items and the higher response rates in the face-to-face interviews underline the data quality arising from the survey method applied.

3.5 Editing

In parallel to the field phase, the data are comprehensively checked. As well as monitoring the interviewers, this process mainly optimises the data quality. To do so, the responses to individual questions are checked for consistency and plausibility. This process makes additional use of both questions linked to the question being checked and questions from the previous year, provided these are available. If errors / implausibilities cannot be cleared up on the basis of the questionnaire data, the establishment is contacted by telephone and the problem is solved in conjunction with the interviewee. The editing process also includes another check of the relationship to the correct establishment unit. Should the editing process establish that first-time respondents do not have any relation to the establishment number and the establishment was thus wrongly surveyed, these interviews are excluded. The same applies to interviews with high rates of missing or erroneous responses and questionnaires that arouse suspicion of falsified interviews.

¹⁰ These values are registered as “no response” and coded “-9”.

¹¹ The number means the proportion of item non-response for 2007.

4. Cross-Sectional Evaluations

The survey design of the IAB Establishment Panel enables evaluations on a cross-sectional and longitudinal basis. The two types of evaluation are based on different logics: a cross-sectional study looks at the establishments in the survey in the relevant year. In a longitudinal study, however, researchers can look at developments of individual establishments involved in the survey over a longer period. Additionally, the dynamics of foundations and closures¹² can also be integrated into the analysis. Both the logic of the definition of cross-sectional and longitudinal cases and the respective weighting factors differ from one another. As most analyses are generally either cross-sectional or longitudinal, the two types of evaluation are presented separately below.

4.1 Definition of the Cross-Sectional Cases

The cross-sectional cases represent the population of the respective wave. The criteria for a cross-sectional case are that a valid questionnaire has been completed and that the establishment had at least one employee liable to social security as of the reference date of 30 June of the previous year. In the datasets of the individual waves, the cross-sectional cases can be identified respectively via the string variable `querxxxx` (xxxx describing the year of the respective wave, e.g. `quer2006` for the 2006 wave). These contain the value “Q”. These cases form the basis for the cross-sectional weighting in each wave.

4.2 Cross-Sectional Weighting

The multiply disproportionate structure of the sample is corrected with the aid of a weighting procedure and adapted to the structure of the population. The weighting takes place in the form of a projection onto the number of establishments in the population. This applies for the whole of Germany, for East and West Germany, and also for the individual federal states and the establishments in the manufacturing industry in East Germany.

The required structures for weighting the cross section are produced from the BA establishment file as of the reference date (30 June of the previous year). This consists of the population per federal state across the industry and size classifications, in accordance with the stratification matrix for the sampling process (from 2003: 170-cell matrix). As a valid projection requires sufficient net cases per weighting cell, not all federal states can be weighted using this 170-cell matrix. In the federal states with a lower number of cases, the

¹² For a definition of “new” establishments in the IAB Establishment Panel, please see section 2.2.

projection takes place onto the respective marginal distributions, i.e. the margin of the establishment size classifications and the margin of the industries.¹³

The projection includes all cross-sectional cases with the surveyed actual values on employment, industry allocation and federal state. The establishments are thus allocated to the separate cells of the stratification matrix according to their responses to the questionnaire which insures that they get the weight according to the response of the interview unit. The sampling cell and the weighting cell of a single establishment may therefore differ from one another.

In the first step of the cross-sectional weighting process, the “new” establishment numbers from the new entries sample are projected onto the appropriate target values from the establishment file.

In a second stage, all surveyed establishments are projected per federal state onto the target structure according to the BA. As mentioned above, this takes place in the large federal states with sufficient numbers of cases on the basis of the stratification matrix, and in the smaller federal states on the basis of marginal distributions of industries and establishment size. Projecting the establishments onto the target structure according to the BA can present a problem if very few or no cases at all are present in the individual cells in the sample. In order to have the correct number of establishments nevertheless, cells are combined in such cases.

A separate weighting process is carried out for the extension sample in the manufacturing industry in East Germany. On this case, the weighting factors are calculated across 10 establishment size classes and 15 branches of the manufacturing industry, without differentiation by federal state.

Next, the marginal distribution of establishments via industry and establishment size class is checked again and adapted as necessary.

The final stage is to adjust the weighting factors to the number of employees liable to social security as of the reference date. In this step, care is taken to preserve the number of establishments across establishment size classes, industries and federal states, despite the adjustment to the number of employees liable to social security. For this reason, the adjustment of the employees liable to social security takes place within the stratification cells. The weighting steps are repeated on an iterative basis. This process may produce very large weighting factors. This applies particularly to small establishments, for which the numbers in the population are relatively high and the selected sets in the sample are relatively small. In order to limit the resulting problems of outliers, the level of the weighting factor is checked and limited to a maximum of 3,000.

As a result, an individual weighting factor is given to every establishment. Subsequent to the weighting process, the weighted data are intensively checked

¹³ This applies to Bremen and Schleswig-Holstein, for example.

and compared with external sources in cooperation with the IAB and other institutions¹⁴ involved in evaluating the IAB Establishment Panel, in order to achieve maximum validity. Due to the nature of the weighting process, it is generally the case that an establishment is allocated different weighting factors in two subsequent waves. This applies to both the cross-sectional weighting process and the longitudinal process. There are various reasons for this, e.g. because the target structure (the number of establishments or employees in the target structure of the respective weighting cell) changes, because the number of surveyed establishments in the weighting cell changes due to extensions or non-responses, and / or because an establishment changes size, industry or federal state between two waves.

5. Longitudinal Analyses

Unlike cross-section analyses, longitudinal evaluations offer the possibility of analysing developments and links between establishment characteristics across time, on the individual establishment level. The definition of panel cases as well as the longitudinal weighting process are designed significantly differently to the respective processes in cross-section. The longitudinal processes have to take into account the dynamics of foundations and closures, as well as changes in the individual establishments over time, such as growth and shrinkage.

5.1 Definition of Longitudinal Cases

The starting point for forming a longitudinal section is formed by all cross-sectional cases existing in the starting year. These represent the stock of all existing establishments in the starting year, the development of which shall be observed in the subsequent years. In the subsequent waves, panel cases are then defined as follows:

1. Panel cases from the previous wave that still exist: This group represents the “surviving establishments”. In order to avoid distortion in the analyses, however, only those surviving establishments are continued as panel cases for which the interview could be carried out with the same establishment unit as in the previous wave. If it is not possible to hold an interview with

¹⁴ Alongside the IAB, the IAB Establishment Panel is evaluated by further research bodies on behalf of the participating federal states and institutions. In 2006, these bodies were the BAW Regional Economic Research Institute, the Institute for Applied Economic Research (IAW) Tübingen, the Institute for Socio-economic Structural Analysis (SÖSTRA), the Institute for Economics, Labour and Culture Frankfurt (IWAK), the Halle Institute for Economic Research (IWH) and the International Institute for Empirical Socio-Economics (INIFES).

the same unit as in the previous year, this establishment is still available for the cross-section but not for longitudinal analysis.

2. Panel cases from the previous wave that went out of operation in the previous wave or earlier waves: These establishments depict establishment closures. That means that from the year in which they went out of operation, they are only contained with the information that these establishments were no longer in operation during the relevant period for the panel case definition. No interview can be carried out for these establishments.
3. “New” establishment numbers from the new entries samples of the subsequent years: Alongside existing establishments and those no longer in operation, “new” establishments make up the third part of the establishment dynamic. The “new” establishment numbers, which are adopted into the sample in each wave (see section 2), are thus also a component of the panel case definition, provided an interview was carried out with the establishment unit drawn by the BA.

The establishments that are surveyed in each wave as replacements for non-responders are not included in the panel definition, as the development of the stock establishments is only observed from the respective starting year on. In the IAB Establishment Panel, various longitudinal sections¹⁵ are defined, beginning with five different starting years:

- Longitudinal section 1 (West Germany only) in the period from 1993 to 2006,
- Longitudinal section 2 in the period from 1996 to 2006,
- Longitudinal section 3 with the starting year 2000 up to the latest available data,
- Longitudinal section 4 with the starting year 2003 up to the latest available data,
- Longitudinal section 5 with the starting year 2007 up to the latest available data.

Weighting factors also exist for these defined panel cases. For the analysis, the respective periods for study do not necessarily have to be identical with the longitudinal periods. The analysis period must, however, be completely contained in the respective longitudinal period.¹⁶

¹⁵ Researcher also have the option of creating their own panel case definitions for other time periods. Detailed information on the subject is contained in Fischer et al. (2008).

¹⁶ For example, the longitudinal panel from 1996 to 2006 can also be used to evaluate the period from 1998 to 2001.

5.2 Longitudinal Weighting

Similarly to descriptive evaluations of cross-sectional data, descriptive analyses of longitudinal data also lead to distorted results without weighting, due to the disproportionate sample. When carrying out longitudinal weighting for an establishment survey with a disproportionate sample, various requirements must be kept in mind:

The establishments in the stock data in the respective waves must reproduce the industry and establishment site structures as of the reference date for the respective wave within the longitudinal weighting. That means that these establishments have to be weighted for each individual wave in the longitudinal panel to match the target matrix of the establishment file.¹⁷ As many establishments are contained in the stock data in several waves, this means that these establishments have to be weighted in various waves to match differing populations, but only one factor per establishment is possible for the entire longitudinal period. Establishments that grow or shrink during the period present a particular challenge for longitudinal weighting, as far as the changes lead to an alteration in the size class relevant for the weighting process, as the average weighting factors vary strongly depending on the establishment size, due to the disproportionate sample. For new definitions of a longitudinal section, the cross-sectional factor of the basis wave serves as an entry factor for the longitudinal weighting, and for existing longitudinal sections the longitudinal factor of the previous wave is taken. If an establishment switches, e.g. from one size class to the next higher class between two waves, it initially receives an entry factor that is too high – due to the fact that the factor of the previous wave serves as the entry factor – meaning that this establishment would be too highly weighted. Vice versa, if an establishment shrinks from a larger size class to the next lower class in the stratification matrix, this establishment initially receives a factor that is too small, as it had received a smaller factor as a large establishment in the previous year, to balance out the disproportionate sample. These effects are checked, so as to minimise distortions in the evaluation of the respective wave.

The longitudinal weighting process is designed in such a way that an analysis of the panel cases with the longitudinal weighting factors – for comparably defined sub-groups – results in approximately the same distributions for all questions from the previous surveys as in the corresponding cross-section analyses of the wave in question. Even with such a large sample as in the IAB Establishment Panel, longitudinal weighting theoretically requires considerably more parameters to be checked than is practically possible and sensible. Longitudinal weighting therefore has to concentrate on the key parameters.

¹⁷ Due to the low number of longitudinal cases, the industry structure is compounded into seven industries for the longitudinal weighting process.

Longitudinal weighting takes place in eight subsequent steps, which are explained in detail in Fischer et al. (2008). With the exception of the longitudinal panel from 1993, which contains only West German establishments, all steps are carried out separately for East and West Germany.¹⁸ The individual steps towards calculating the longitudinal weighting factors take the following aspects into account:

- The structure of the establishments by industry, size class and federal state contained in the latest available data,
- the number of existing, no longer operational and newly founded establishments in every year of the longitudinal period,
- survival and removal of the establishments from the stock data of the starting year and the “new” establishment numbers added across the longitudinal period,
- growth and shrinkage of the establishments from the stock data of the starting year and of the “new” establishment numbers added across the longitudinal period,
- disproportionate non-responses, on the basis of the questionnaires from the previous year, and
- the industry and size structure of the population in each of the waves contained in the longitudinal period.

6. Data Access

As the surveyed establishments have been assured that their data will only be published in anonymous form and not passed on to third parties, external researchers only have access to the data via the Research Data Centre (FDZ) of the Federal Employment Agency at the Institute for Employment Research. Researchers have the choice between a research visit at the FDZ or controlled remote data access. There is a feasibility study about the data access of the IAB Establishmentpanel via scientific use files in the project “Business Statistical Panel Data and Factual Anonymisation” (FAWE Panel) Drechsler et al., 2007.

The FDZ offers a number of aids for preparing analyses with the IAB Establishment Panel online, e.g. dataset descriptions, test data and questionnaires. The data access routes are also explained on the FDZ homepage, and all application forms are available for download (<http://fdz.iab.de>). For further infor-

¹⁸ As detailed in section 2.2, there has been a new definition of West and East Germany from 2007 on. After this date, West Germany consists only of the West German federal states, not including Berlin. East Germany consists of all East German federal states and Berlin.

mation on the current survey wave, content-based results and a wide range of literature, please see the IAB Establishment Panel homepage (<http://betriebs-panel.iab.de>, only in German).

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.
- Alda, H./Dundler, A./Müller, D./Spengler, A.* (2006): Aufbereitung eines Paneldatensatzes aus den Querschnittsdaten des IAB-Betriebspanels, *FDZ-Datenreport*, 02/2006.
- Bellmann, L.* (2002): Das IAB-Betriebspanel. Konzeption und Anwendungsbereiche, *Allgemeines Statistisches Archiv* 86 (2), 177–188.
- Brixy, U./Fritsch, M.* (2004): The Establishment Files of the German Social Insurance Statistics, *Schmollers Jahrbuch* 124 (1), 183–190.
- Bundesagentur für Arbeit* (1997): Handbuch für die Betriebsnummernstellen der Arbeitsämter zur Aktualisierung der Betriebsdatei mit dem IT-Verfahren coBer-Betriebe / coBer-coStat, unpublished document.
- Bundesagentur für Arbeit* (2005): Statistik der sozialversicherungspflichtigen Beschäftigung und geringfügig entlohnten Beschäftigung. <http://www.pub.arbeitsamt.de/hst/services/statistik/000200/html/qualitaetsberichte/qualitaetsbericht-bst.pdf> (07/08/2007).
- Buttler, G./Fickel, N.* (2002): Statistik mit Stichproben, Reinbek bei Hamburg.
- Drechsler, J./Dundler, A./Bender, S./Rässler, S./Zwick, T.* (2007): A new approach for disclosure control in the IAB Establishment Panel. Multiple imputation for a better data access, *IAB Discussion Paper*, 11/2007.
- Fischer, G./Janik, F./Müller, D./Schmucker, A.* (2008): The IAB Establishment Panel – from Sample to Survey to Survey Projection, *FDZ Methodenreport*, 01.
- Gerlach, K./Hübler, O./Meyer, W.* (2003): The Hannover Firm Panel (HFP), *Schmollers Jahrbuch* 123 (3), 463–470.
- Kölling, A.* (2000): The IAB-Establishment Panel, *Schmollers Jahrbuch* 120 (2), 291–300.