RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

FDZ-METHODENREPORT

Methodological aspects of labour market data

07|2022 EN Technical Report on the IAB Establishment Panel — Wave 29 (2021)

Miriam Gensicke, Nikolai Tschersich, Susanne Kohaut, Christina Mielewczyk, Sophie Künzel

Bundesagentur für Arbeit

KANTAR

Technical Report on the IAB Establishment Panel

Wave 29 (2021)

Miriam Gensicke (Kantar)
Nikolai Tschersich (Kantar)
Susanne Kohaut (IAB)
Christina Mielewczyk (Kantar)
Sophie Künzel (Kantar)

Munich, June 2022
Ag 3163301679

Content

1 Population, Sampling Frame, Sample

1.1 Population and Sampling Frame

The population of the IAB Establishment Panel comprises all establishments with at least one employee subject to social insurance contributions as of the reference date 30 June of the previous year.[1] The basis for sampling is the Federal Employment Agency establishment file. This contains all the establishments that in the context of the social security registration process notify the social security agencies of their employees who are subject to social insurance contributions, and are given an establishment number. As of the reference date 30 June 2020 the establishment file contained 2,125,638 establishments belonging to the population (427,588 for eastern Germany and 1,698,050 for western Germany), with a total of 33,255 thousand employees subject to social insurance contributions. Establishments without employees subject to social insurance contributions, for example one-person establishments or establishments only with marginal employment or employing only civil servants, are not covered by the IAB Establishment Panel. For this reason, for example the national accounts exhibit distinctly more people employed than the IAB Establishment Panel.

1.2 The Logic behind the Establishment Number

An establishment denotes a regionally and economically separate unit with employees, which is awarded its own establishment number in accordance with certain rules during the registration process for the social security agencies.[2]

Branches of one employer in different local authority districts strictly receive their own establishment number.

Branches of one employer within one local authority district are merged under one establishment with one establishment number provided if they exercise the same economic activity. Branches with different economic activities are given different establishment numbers.

An establishment itself can have several establishment numbers; this applies in particular to larger establishments with different functional areas that are considered as independent concerning their administration.

Crossholdings between companies make no difference. Every legally independent company is given establishment numbers according to the rules just mentioned.

1 Private households and exterritorial organisations have been excluded since the 2004 survey.
2 For the logic behind establishment numbers and the rules on issuing them and further information on the website https://www.arbeitsagentur.de/web/content/DE/Unternehmen/Sozialversicherung/Betriebsnummernvergabe/index.htm.

Establishment numbers are (re-)allocated when

the establishment previously did not have an establishment number (usually because it is the first time that the establishment has an employee who is subject to social insurance contributions),

the establishment's economic activity has changed or

there is a change of ownership.

1.3 Stratification Matrix and Sub-Samples

The sample is disproportionally stratified by the size of the establishment, sector and federal state. On the one hand the target degree of completion of the individual cells is determined by the scope of the basic sample and that of the extension samples specific to the federal states and sectors. On the other hand, the individual cells are drawn according to the principle of optimal stratification proportional to the number of employees. For these reasons, large establishments, small federal states and small sectors as well as the manufacturing industry in East Germany are overrepresented in the sample. These disproportionalities are corrected with the aid of a weighting procedure afterwards (cf. in detail Chapters 6 and 7). Table 9 and Table 10 in the Appendix provide an overview of the classification of the sectors and establishment sizes. The last major changes to the stratification matrix were undertaken in 2009 in respect of the delineation of sectors, during the changeover from the economic sector classification WZ2003 to WZ2008. Since 2007 East and West Berlin have been combined.[3]

There are three sub-samples in total:

Panel sample: This comprises all the establishments that are willing to participate and have a valid interview from the previous year. The continuer sample reflects the panel nature of the IAB Establishment Panel. It is necessary so that panel evaluations extending beyond pure time-series analyses can be undertaken. Panel analyses examine the developments at establishment level over time. On the other hand, cross-sectional data from at least two points in time are sufficient for time-series analyses.

Follow-up sample: This comprises all the establishments that are willing to participate and have a valid interview from the year before last.[4] This sub-sample increases the number of cases that can be evaluated cross-sectionally.

Refreshment sample: This includes refreshment samples that are specific to federal states, and a sector-specific refreshments for the manufacturing industry in East German federal states. It also comprises establishments with a new establishment number. The aim of this sub-sample is to replace panel attrition, to achieve the sample sizes required in the individual federal states and in the manufacturing industry, and to illustrate economic structural change. The attribute "New Establishment Number" has to be used carefully and can not simply be put on the same level with „New Establishment".[5]

3 At the wish of the Halle Institute for Economic Research (IWH), Berlin was excluded from the extension to the manufacturing industry in East Germany. Since then this extension has only included East German federal states.

4 These are generally establishments which expressed the wish to miss a year, but would then be willing to participate again in the subsequent year.

5 New establishment numbers are characterised by the fact that as of the reference date of the current survey wave (= 30 June of the previous year) they had at least one employee subject to social insurance contributions, whereas a year earlier they had none. This definition is aligned with the system used for the Federal Employment Agency establishment file, and is only suitable to a limited

Up to 2002 unit-non-responses were generally excluded from further sampling processes. In the greater size classes in particular, there were ultimately hardly any new establishments in the sampling frame which might have been included in the sample. Since then, unit-non-responses can be drawn again after a three-year rest period. Establishments of this kind that are newly drawn are reincluded in the sample with a new identification number.

extent for identifying the genuine founding of new establishments. The establishment might have existed before as an establishment without employees subject to social insurance contributions. It also happens that establishments do not continuously have employees; these so-called perforated establishment numbers can appear in the sample of a survey year as new establishment numbers under the above definition (cf. also the overview of how establishment numbers are awarded in Section 1.2). For this reason, this characteristic should be used with care, and only when additional characteristics from the questionnaire are taken into account, such as information on the manner in which they were founded.

2 Questionnaire and Pretest

2.1 Questionnaire Development and Module System

The development of the questionnaire also takes account of the panel character of the IAB Establishment Panel. In every wave so-called basic modules are used with as many unchanged questions as possible. These basic information on establishments are available for every year since 2008. The basic modules are supplemented by additional modules, in which more in-depth questions are asked at defined intervals of time, usually in a two-year cycle.

Every year there are also one or more focus areas for questions which are coordinated with the various clients (cf. Appendix). Here more in-depth questions are asked about particular topics or current labour market trends.[6] Table 1 below lists the questions that were included in the questionnaire besides the module system. In the development of the questionnaire itself, a compromise must be found between comparability over time and the adaptation or modification of existing questions as well as the inclusion of new and current topics.

6 An overview of the individual questions and the questionnaire can be found in the tools for the IAB Establishment Panel at-https://fdz.iab.de/.

Table 1: Questions (re)included in the questionnaire in 2021[7]

Question	Section	Question text	Response options	Last surveyed (question)	Changes compared to last survey
-	One request to all establishments/office at the beginning	The informative value of this multi-year survey can be significantly increased if we link your information of each survey wave for research purposes with data available at the Institute for Employment Research. Of course, all data protection rules will be observed here as well. Do you agree with this?	yes/no	2020	chapter heading/question wording
1a	Impact of the Corona pandemic	Had or has the Corona pandemic had an economic impact on your establishment/ office? Both negative and positive effects are meant.	yes/no/difficult to tell	first raised	-
1b	Impact of the Corona pandemic	Were or are the economic impacts on your establishment/ office predominantly negative, predominantly positive, or both to the same degree?	single choice	first raised	-
1c	Impact of the Corona pandemic	How much was or is your establishment/ office negatively affected economically by the Corona pandemic? Please say it using this scale!	Scale of 5 (1 = slight/5 = very strong)	2020 (q. 1c)	question wording
2	Impact of the Corona pandemic	Is your establishment/ office threatened in its existence by the Corona pandemic?	yes/no	2020 (q. 3a)	new response option 'Does not apply'.
15a	Staff structure	Did you have short-time work in the 1st half of 2021?	yes/no	2010 (q. 77a)	-
15b	Staff structure	How many of your employees were on short-time work at the peak?	Number (numeric)	first raised	-
15c	Staff structure	In which month(s) did you have this peak or peaks?	Multiple answers	first raised	-
16a	Staff structure	Did employees on short-time work participate in further training measures in the 1st half of 2021?	yes/no	first raised	-
16b	Staff structure	What spoke against further training measures during short-time work?	Multiple answers	first raised	-
42a	Mobile working	Does your company/office offer employees the opportunity to work from home using digital devices, whether in a home office, teleworking or mobile?	yes/no	first raised	-
42b	Mobile working	Please indicate either the number or the percentage of employees who can make use of this option. If exact information is not possible, please estimate!	Number (numeric)/share (%)/no answer possible	first raised	-
42c	Mobile working	In this context: Does your company/office have regulations in place to protect employees from overwork due to constant accessibility or extended working hours, e.g., shutting down servers in the evening?	yes/no	2018 (q. 81)	-

7 Questions included or dropped due to the module system are not shown in the table, but only questions are shown that are included in individual waves (or for short periods of time).

43	Mobile working	When you think about the time after the Corona pandemic: To what extent do you want to enable your employees to work from home using digital devices in the future, be it in a home office, teleworking or mobile?	Single choice	first raised	-
62	Vocational training and apprenticeships	As a result of the Corona pandemic, did you offer more apprenticeships, fewer apprenticeships, or no apprenticeships for the 2021/2022 training year compared to your original plan, or was there no change?	Single choice	first raised	-
63	Vocational training and apprenticeships	Thinking back to the time before the Corona pandemic, did you get more applications or fewer applications for the 2021/2022 training year, or was there no change?	Single choice	first raised	-
64a	Vocational training and apprenticeships	In the wake of the Corona crisis, the German government launched the "Securing Apprenticeships" program. Under this program, employers can apply for grants for training. Are you familiar with this program?	yes/no	first raised	-
64b	Vocational training and apprenticeships	Utilization is tied to certain requirements. Does your company/office meet these requirements?	yes/no/don’t know, requirement not known	first raised	-
64c	Vocational training and apprenticeships	Have you received grants for training from this program?	yes/no	first raised	-
65	Vocational training and apprenticeships	Have you received other financial benefits to support in-company vocational training from the federal or state governments?	yes/no	first raised	-
73	Continuing education in the workplace	In addition to the classic forms of continuing education, there are increasingly digital continuing education formats. Have the following digital continuing education formats been used in your company/office?	yes/no	first raised	
74	Continuing education in the workplace	What experiences have you had with digital formats of continuing education? To what extent do you agree with the following statements for your establishment/office?	Scale of 4 (1 = fully agree/ 2 = rather agree/ 3 = rather disagree/ 4 = do not agree at all)	first raised	-

2.2 Cognitive Pretest

For quality assurance, new questions are first tested in a cognitive pretest to determine whether they appear suitable in principle for an establishment survey, and whether the establishments are likely to be capable of providing information on the content being asked, to the effect that the desired information is available at establishment level. For this purpose, interviews are carried out in around 90 establishments from different sectors and establishment size classes. The pretest interview consists of two parts. Firstly the test person is asked to answer the questionnaire, then the second stage comprises the actual cognitive test. The establishments are requested to comment on each question, and are asked whether they had any difficulties in answering. The establishments' ability to provide information and their difficulties in answering as well as their problems in interpreting the questions are of particular interest. The results of the cognitive pretest flow into the further development of the questions: The pretest is undertaken by specially trained project staff from the IAB.

2.3 Computer-assisted questionnaire

Since 2018, the questionnaire is also available as a computer-assisted questionnaire that reflects the functionality and flexibility of the paper questionnaire. The core functionalities include the following:

The computer-assisted questionnaire can be used both in CAPI mode and in CAWI mode.

The interviewers are thereby able to conduct the interview on site with their CAPI laptop.

Furthermore, the questionnaire can be left to the establishment for partial or complete completion via internet (CAWI).

The computer-assisted questionnaire makes it possible to jump to specific questions or question modules with pinpoint accuracy. The target person is able to browse back and forth like in a paper questionnaire.

Several thematically related questions are displayed simultaneously on one screen. Filtered questions are not hidden but only deactivated. This is to avoid mode effects compared to the paper questionnaire.

The establishments are able to archive the completed questionnaire as a PDF or printed document for internal documentation purposes.

3 Study Design and Field Organisation

3.1 Study Design

Due to the Corona pandemic, the interviews should preferably be conducted as a computer assisted telephone interview.Only at the explicit request of the target person and the interviewer, the interview could be conducted face-to-face. The questionnaire asks for a series of items of operational information (e.g. turnover, investments, total wages and salaries), about which in principle the establishment ought to be capable of providing, but which cannot always be spontaneously recalled by the person being interviewed. In such cases the interviewer can leave the questionnaire at the establishment or as a CAWI questionaire for self-completion. The person being interviewed can look up the missing information and then complete the questionnaire. At the request of the establishment the questionnaire can also be left for total self-completion. The majority of establishments are interviewed by telephone (cf. also Table 5 in section 5.2). The option of self-completion (partly or fully) is taken up most frequently by larger establishments.

The use of trained interviewers leads to fewer errors in the completed questionnaires in principle compared to the ones optained in the self-completion mode, and the proportion of missing information is lower.

The interviews are undertaken exclusively by interviewers from the in-house interviewer staff at Kantar.

The data collection takes place annually from the end of June to the end of October. In parallel the data that has been gathered undergo checks and errors are eliminated (for this cf. Chapter 5). In advance the establishments receive an announcement letter from the Federal Employment Agency (BA), a letter of recommendation from the Confederation of German Employers' Associations (BDA), a data privacy declaration and a summary of the most important results from the previous year – in the form of broschure.

3.2 Field organization

In view of the Corona pandemic, the IAB and Kantar have decided to redesign the field organization in such way that the IAB Establishment Panel could be conducted even under very restrictive conditions. For example, at that time it was not foreseeable whether a contact restriction due to legal requirements would prevent the IAB Establishment Panel from being conducted face-to-face during the period scheduled for the field work. In addition, it was unclear whether a face-to-face interview situation was at all justifiable for the interviewers and the target person against the background of the infection process. Therefore, the following basic procedure in the field organization were also agreed upon for the 2021 wave:

Interviewer should perform the contacting of the establishment and, as far as possible, the interview by telephone. The interviewers should possibly transfer the information provided by telephone directly into the computer assisted instrument.

If the interview was conducted by telephone, the interviewers should ensure that the target person had access to the questionnaire. This could be done by sending the paper questionnaire with the invitiation letter, by accssing the website iab-betriebspanel.kantar.com or by a questionnaire sent ad hoc by the interviewer (by post or by e-mail as a PDF).

Alternatively, the interviewers could also motivate the target person to complete the questionnaire themselves, either online in the CAWI instrument or written via paper questionnaire.

Interviews were also allowed to be conducted face-to-face in case of an explicit request of the target person (and with the consent of the interviewer). This option was especially kept for establishments of the continuer sample, which would otherwise possibly refuse to participate in the IAB Establishment Panel.

3.3 Identification of the Correct Establishment Unit

The interviewer is responsible for ensuring that the correct establishment unit is surveyed. For establishments being surveyed for the first time, the establishment number defines the survey unit. The interviewer has to decide on the spot whether the information requested in the questionnaire is actually available for this unit. The establishment number itself is a characteristic that is difficult to handle when identifying the correct establishment unit. For that reason the interviewers – aside from the name of the establishment – rely on the sector and the number of employees subject to social insurance contributions as of the reference date of the previous year, according to the details held by the Federal Employment Agency (BA).

In the case of establishments periodically interviewed, the same unit as the unit in the previous year should be interviewed. This applies irrespective of the relationship between the unit surveyed the previous year and the originally stipulated establishment number. The connection to the previous year is established using the details of the total number of employees and sector from the previous year. Only when the provided information in the current year on the reference date of the previous year are identical with the corresponding information given in the previous year, it can be assumed that the interview is taking place in the same establishment as last year (in terms of the employee numbers within a defined range of tolerance).[8] This information forms the basis of the panel case definition provided, and thus of the individual longitudinal sections (cf. Chapter 7 on this in detail).

8 For establishments with more than 20 employees the range of tolerance is +/– 5 %, for establishments with up to 20 employees +/– 1.

3.4 Training and Monitoring Interviewers

Kantar continuously recruits interviewers for its in-house f2f interviewer staff. Interviewers are selected in an extensive, multistage selection and assessment process, within the prospective interviewers receive training face-to-face and in writing. They receive detailed basic information about statistical selection procedures, data protection and interviewing behaviour. The so called contact interviewers, experienced interviewers who take over the local supervision of the new interviewers, practise the later interview situation with the new interviewers. Since contact interviewers also work as interviewers, they are able to pass on their experiences and practical tips to the new interviewers. The new interviewers are intensively supervised by the contact interviewers during their first projects. The interviewers at Kantar receive follow-up training as standard at regular intervals.

As well as the intensive interviewer training for quality assurance purposes the interviewers' work is continuously monitored to ensure "that no significant falsification of the research results takes place through interviews – consciously or unconsciously – not being conducted correctly"[9]:

For the IAB Establishment Panel a project-specific control of interviews that have been conducted takes place through the extensive ex-post data validation and follow-up telephone surveys (cf. in detail Chapter 5).

Beside this project specific control in the IAB Establishment Panel, the interviewers' work is also randomly checked in other projects using random sampling procedure. The interviews of those interviewers that were conspicous in the context of these control measures were included in the project-specific control of the IAB Establishment Panel.

In addition, all interviews of the IAB Establishment Panel are checked in the so-called Similarity Check. This check is based on the hypothesis, which is supported by experience, that interviewers who 'fill out questionnaires themselves' generate data with less variance than in real interviews. Statistical methods are used to test whether there are greater similarities between the interviews of one and the same interviewer than between the interviews of other interviewers. A low variance is only a first indication that the interviewer has falsified data. Therefore, such interviews are given to the telephone follow-up check described above.

[9] Sommer, Rudolf; Unholzer, Gerhard, and Erich Wiegand (1999): Standards zur Qualitätssicherung in der Markt- und Sozialforschung. ADM: Frankfurt a. M., p. 414

4 Result of Field Work

4.1 Response: Unit-Non-Response

The fieldwork started on July 01, 2021, the last interview was realized on November 19, 2021. From the gross sample of 36,306 establishments used in the fieldwork, a total of 15,799 evaluable interviews were realized. The difference between the total sample (37,230) and the gross number of interviews used is rooted in the processing steps prior to the fieldwork start, in which some panel establishments are identified as not being able to be used again. This is largely due to an active refusal to participate in further waves of the survey. In terms of gross data, a utilization rate of 43.5% was achieved (see Table 2).

Table 2: Number of evaluable interviews and response rate

Subsamples	Region	Sample (absolute)	Used gross (absolute)	Evaluable interviews	
				absolute	in % of gross
Repeater sample	West	10,881	10,673	7,593	71.1%
	Ost	6,497	6,379	4,783	75.0%
	total	**17,378**	**17,052**	**12,376**	**72.6%**
Follow-up sample	West	699	343	130	37.9%
	Ost	657	415	162	39.0%
	total	**1,356**	**758**	**292**	**38.5%**
First survey sample	West	10,366	10,366	1,732	16.7%
	Ost	8,130	8,130	1,399	17.2%
	total	**18,496**	**18,496**	**3,131**	**16.9%**
total	West	21,946	21,382	9,455	44.2%
	Ost	15,284	14,924	6,344	42.5%
	total	**37,230**	**36,306**	**15,799**	**43.5%**

The various sub-samples differ significantly in terms of their response rates.

In the case of the sub-sample of 17,052 establishments already surveyed in the previous year (the so-called repeat sample), a response rate of 72.6% was achieved with 12,376 evaluable interviews. This is lower than the previous year's figure (77.0%). In this sub-sample, which is particularly important for the project, a lower response rate was again achieved after it had already declined in the previous wave. In addition to the effects of the Corona pandemic, such as poorer accessibility, this can also be attributed to the fact that establishments that participated for the first time in the previous year as self-

completers have a lower level of commitment to the project, of which there were many cases in the 2021 wave.

The supplementary sample of 18,496 establishments to be surveyed yielded 3,131 evaluable interviews, which corresponds to an exhaustion rate of 16.9%. This figure is slightly higher than the corresponding figure for the previous year (15.5%).

In the follow-up sample - which is very small in comparison to the other two sub-samples (respondents from the previous year who were not able to be interviewed again) - a total of 292 evaluable interviews were realized from a gross of 758 cases; this corresponds to an exhaustion rate of 38.5%. This value is also higher than the previous year's figure (30.7%).

As already described in Section 1.3, the IAB Establishment Panel is disproportionately stratified according to different characteristics. Table 3 provides an overview of the extent of the gross sample used and the numbers of cases actually realised for each federal state (federal state-specific extension samples) and for the manufacturing industry in East Germany (sector-specific extension sample).

Table 3: Overview of gross and net sample, by federal state and manufacturing industry in East Germany[10]

	Gross	Net (actual)
Schleswig-Holstein	1.197	719
Hamburg	663	246
Lower Saxony	1.505	933
Bremen	1.580	836
North Rhine-Westphalia	2.615	1.392
Hesse	4.383	1.204
Rhineland-Palatinate	2.365	899
Baden-Württemberg	2.747	1.126
Bavaria	2.821	1.202
Saarland	1.506	898
Berlin	3.270	978
Brandenburg	2.145	1.074
Mecklenburg-West Pomerania	2.325	952
Saxony	2.562	1.192
Saxony-Anhalt	2.078	1.019
Thuringia	2.544	1.129
Total	**36.306**	**15.799**
Manuf. ind. East Germany (excl. Berlin)	3.580	1.615

Table 4 shows an overview over the different types the interviews where conducted. This information was noted down in the address protocol for each respective establishment. The percentage of interviews conducted entirely by interviewers remained the slight majority of all interviews with a share of 55.5% (48.1% + 7.3%) compared to 54.2% in 2020.[11]

10 The manufacturing industry in East Germany is shown separately as these cases are already contained in the case numbers differentiated by East German federal states (excluding Berlin).

Table 4: Overview of the form of the interview

	2020	2021		
	total	Total	repeater	first-time respondent (incl. temp. failures)
conducted entirely by telephone	44,7%	48,1%	45,0%	59,3%
conducted entirely face-to-face	9,5%	7,3%	9,1%	1,1%
conducted partly by telephone	0,6%	1,0%	1,0%	1,1%
conducted partly face-to-face	0,2%	0,1%	0,1%	0,0%
entirely completed by the respondent himself	45,0%	43,5%	44,8%	38,6%

As previously, the form the interview takes is substantially associated with the size of the establishment: the proportion of establishments surveyed entirely face-to-face falls linearly from 68 % of the smallest establishments (with 1 to 4 employees) to 39 % of the large establishments with 5,000 or more employees. This is because the larger the establishments the more frequently the interviewee is able to provide the complex quantitative details only with extensive preliminary work; this applies in particular to the employee structure, personnel recruitment, appointments and resignations, training, business volume and investments. In these cases it is helpful for the interviewee and the interviewer if the questionnaire can be left at the establishment for further processing or is made available to the respondent in advance (cf. also Section 3).

4.2 Dealing with Missing Information: Item-Non-Response

Missing details arise among others because the corresponding question is difficult to understand, difficult to answer for factual reasons or the interviewee refuses to answer (e. g. for reasons of confidentiality or the time required to look up the answer). To some extent to avoid missing values the interviewees are given the option of stating estimated values (“If it is not possible to answer precisely, please estimate”).

A particulary important factor in minimising missing information is the use of interviewers. Thus, as in previous years, the frequency of missing information has shown clear differences between face-to-face interviews and interviews conducted via mail. In general, the quality of face-to-face interviews (and also of those cases in which the questionnaire was left at the establishment for (partial) self-completion) was significantly better than interviews via mail.[12]

Questions that do not apply to an establishment (e.g. follow-up questions on further vocational training in establishments that have given negative answers to the initial filter question) are simply not asked at all in the IAB Establishment Panel. If a question does not apply to the establishment, the corresponding variable in the data set is empty (system-missing). If however the use of filters means that for example certain groups of employees do not exist in the establishment (e.g. number of employees who have received further training), then the missing information can/must be replaced by the user with a zero.

Generally in literature, when designing questions for item non-response a differentiation is required between the categories “No answer/declined to answer” and “Don’t know”. In our

12 Thus, the average proportion of missing values for interviews conducted entirely face-to-face was at one percent, for those completed by mail at five percent.

view these categories cannot be – with few exceptions – reliably differentiated. Many items of information that are required – even if they are available in the establishment – are associated with a certain research effort, or passing them on is regarded as sensitive.

Against this background, apart from a small number of exceptions the explicit provision of answer categories for “Don’t know” or “No answer/declined to answer” has been avoided in the questionnaire. If such answers are explicitly provided, this thends to attract such responses, as it is easier to tick “Don’t know” than to search for a particular business figure.[13]

In a small number of cases the “Don’t know” category contains utilisable information and can be used for analysis purposes. Thus, for example, the question on the development of the volume of business expected in the current financial year (in comparison to the previous year) includes the category “Don’t know yet”.

Overall in the 2021 survey, 24 % of all questions/variables had less than 0.5 % missing values, 73 % had less than two percent and 93 % had less than five percent missing values. Table 5 provides an overview of the questions with a high proportion (10 % and more) of missing values. When evaluating variables with a lot of missing information, possible distortions as a result of this should of course always be borne in mind.

Table 5: Questions with a high proportion of missing values

Question/ Variable	Content	Unit	Share of missing values
bb36	Share of intermediate inputs/external costs in sales 2019	(%)	23 %
bb33	Business volume 2020 (€)	(EUR)	23 %
bb69	Gross wage/salary total June 2021 (€)	(EUR)	22 %
bb13_1eu	Number of 1-euro jobbers 06/21	(number)	11 %
bb03b	Projected number of employees for the coming year in-total	(number)	11 %
bb46	Share of expansion investments (%)	(%)	11 %
bb53ca	Number of employees exclusively involved in R & D tasks	(number)	10 %
bb73e	Use of digital WB formats: other digital formats	-	10 %

13 This phenomenon comes under the problem of satisficing. Detailed explanations can be found e.g. in Krosnick et al. (1996).

5 Data Verification and Editing

In parallel with the field work, the data that has already been collected is checked both cross-sectionally and longitudinally for its completeness, consistency and plausibility. In the 2021 wave, in total 279 cross-sectional checks, 46 longitudinal checks and 76 filter checks were carried out. A distinction is made between four essentially different checks:

Filter errors: Were the filter instructions obeyed in the intended way, or have establishments mistakenly answered resp. not answered a question?

Checks for completeness: For selected questions checks are undertaken as to whether the question was answered. This applies in particular to questions that are relevant for weighting, such as the question on the number of employees subject to social insurance contributions.

Plausibility checks: These checks assess information that are generally unlikely, but can nonetheless occur in practice. One example of this type of check involves reviewing the per capita incomes. The check indicates an error for values which are comparatively high or low.

Consistency checks: These checks refer to logical connections between different answers. Consistency criteria are infringed, for example, if the questionnaire contains contradictory information. The consistency checks include for instance checking the total amount stated in one question against the total of the individual values.

If missing or incorrect information cannot be supplemented or corrected by means of the questionnaire, an attempt is made to obtain clarity together with the interviewee during a follow-up telephone interview. As a result, missing information can be supplemented and incorrect information corrected. In plausibility checks, implausible values are released with corresponding justification after consultation with the establishment – so despite infringing the test conditions, the value will be accepted.

In the 2021 wave, 32 % of all interviews were completely error-free right from the start, and therefore did not require any further processing. Corrections of the remaining 68 % not completely error-free interviews were undertaken during the editing process, and in the majority of these cases (6,727 respectively 63 %) the establishment was also contacted for a follow-up telephone interview.[14]

14 This corresponds to 43 % of all 15,799 evaluable interviews.

6 Definition of Cross-Sectional Cases and Projection

6.1 Definition of Cross-Sectional Cases

All cases for which a valid questionnaire has been completed and which had at least one employee subject to social insurance contributions as of the reference date of 30 June of the previous year are referred to as cross-sectional cases. Due to the disproportionate structure of the sample the data has to be weighted before descriptive evaluations. With a disproportionate sample structure analyses of unweighted data lead to non-representative results.

For 2021 in total 15,217 cases are available for cross-sectional analyses, of which 9,671 are from West German federal states and 6,093 from East German federal states. 919 further interviews were conducted with establishments which had no employees subject to social insurance contributions as of the reference date of 30 June of the previous year, and therefore did not belong to the population for cross-sectional evaluations; these cases are only used for longitudinal evaluations (see also table 8).

6.2 Cross-Sectional Weighting Process

Weighting is necessary because the sample structure is disproportionate in terms of the establishment size, sector and federal state, and is also needed in order to compensate for any possible differences between the actual and target size of the individual stratification cells (cf. also Section 0).

The weighting of the IAB Establishment Panel essentially takes place in the form of a projection onto the population. This applies for the establishments in the population, in other words for the whole of Germany but also for East and West Germany, for the individual federal states, and for manufacturing industry establishments in East Germany. The target structures are taken from the establishment file of the Federal Employment Agency. The determining factor is the structure (distribution of the establishments) at the time of drawing the sample for the respective wave (in other words as per 30 June of the respective previous year). This projection compensates the disproportionalities as well as the different response rates along the stratification cells in one step.

The weighted sample of the IAB Establishment Panel is proportional to the number of establishments, and thus reflects the distribution of the establishments across the cells of the stratification matrix. Its structure therefore differs from numerous other establishment surveys, in which the over-representation of large establishments is not corrected (and which thus deliver results that are proportional to the number of employees or the turnover, but not to the number of establishments). However, the weighted data of the IAB

Establishment Panel enables analyses that are not only proportional to the establishments, but also proportional to the employees. Moreover, during the cross-sectional weighting, attention is paid to ensuring that at federal state level the employee figures projected from the weighted sample (employees subject to social insurance contributions as of the reference date of 30 June the previous year) correspond to the targets of the Federal Employment Agency. For analyses that are proportional to the number of employees, the weighted number of employees from the establishments to which the characteristic in question applies has to be set in proportion to the total number of employees.

The result of the weighting is an integrated weighting factor that is proportional to the numbers of establishments and employees:

A weighting that is proportional to the number of establishments reflects the distribution of the establishments across the cells of the stratification matrix. This enables representative statements to be made on the percentage of the establishments in Germany to which a particular statement (e.g. establishment has a works council) is applicable.

The weighting that is proportional to the number of employees reflects the distribution of the employees across the federal states in Germany. This enables representative statements to be made on the percentage of the employees that work in establishments to which a particular statement is applicable (e. g. employees work in an establishment with a works council).

For drawing the sample and for projections proportional to the number of establishments the population is subdivided into 19 sectors and 10 establishment size classes per federal state (see Table 9 or Table 10 in the annex). This creates a stratification matrix consisting of 190 cells per federal state. This subdivision has been in place since 2010.

For the IAB Establishment Panel the weighting factors of the cross-section are calculated using generalized regression models, GREG for short.

The aim of a random sampling, i.e. design-based inference, is to estimate certain parameters of the population for a target characteristic y of interest.[15] Important parameters are the sum or the average of this target characteristic in the population. If, in order to estimate such parameters from the population U (= 1, ..., k, ..., N), a sample s (= 1, ..., k, ... , n) with strictly positive selection probabilities for each element ($\pi k = Pr(k \in s) > 0$, $\pi kl = Pr(k \& l \in s) > 0$) based on the sample design,[16] then the design weight dk of a sample element k is the inverse of its selection probability πk, i.e., $dk = \pi k^{-1}$. The Horvitz-Thompson estimator then represents the design-weighted estimate of the parameter, so, for example, for the sum $\hat{y}$ of a feature y, the value $\hat{y} = \sum_s dk\ yk$

In order to take into account not only design-related differences in selection probabilities but also the failure event and to reduce the variance of the estimators, the IAB Establishment Panel uses generalized regression models (GREG) to further adjust the weighting factors by adding certain auxiliary characteristics. Starting from a Horvitz-Thompson estimator, the aim of a calibration by GREG is to adjust the design weights against the background of additionally available information on the sum (or average) of auxiliary characteristics x and to convert them into new weighting factors wk in such a way that the sample after weights represents the sum (or average) of these auxiliary characteristics x, i.e. $\sum_s wk\ xk = \sum_U xk$.

[15] The following statements are based on Deville, J.-C., Särndal, C.-E., Sautory, O., 1993: Generalized Raking Procedures in Survey Sampling. Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1013–1020.

[16] The second requirements of the strictly positive selection probability is necessary to be able to determine the variance of the estimators (cf. e.g. Cassel, C.-M., Särndal, C.E., Wretman J.H., 1977: Foundations of Inference in Survey Sampling. New York: John Wiley & Sons.)

At the same time, the original design weights dk should be changed as little as possible: "Our objective is to derive new weights that modify as little as possible the original sampling weights (dk = πk-1), which have the desirable property of yielding unbiased estimates" .[17]

The weights wk are the solution of a minimization problem under constraints: If G(wk/dk) denotes a function that maps the distance between dk and wk, then the optimization problem is to minimize the function shown below, with respect to wk, where λ represents the vector of Langrange multipliers.

Optimization problem in the context of a generalized regression[18]

$$\sum_{S} d_k \, G\left(\frac{w_k}{d_k}\right) - \boldsymbol{\lambda}'\left(\sum_{S} w_k \boldsymbol{x_k} - \sum_{U} \boldsymbol{x_k}\right)$$

with:

w_k = final weighting factor

x_k = auxiliary characteristics of the elements of the

d_k = design weight

S = sample

U = population

G = distance function

λ = langrange multiplier

Deville et al. (1993) describe several distance functions. The variant described by Deville et al. (1993: 1014) as linear method leads in the application to the establishment panel sample to the best adjustments with slightly larger factor ranges in comparison to the procedure described by them as logit method. Because of the better adjustments, the linear method was therefore chosen. If one uses this method, then for the estimation of the sum of a characteristic the "generalized regression estimator" (GREG, see Deville, Särndal 1993: 1014) results:

17 Deville, J.-C., Särndal, C.-E., 1992: Calibration Estimators in Survey Sampling, Journal of the American Statistical Association, Vol. 87, No. 418, pp.376–382.

18 Here and in the following equations letters in bold denote vectors, in normal letters scalars.

Estimator of the sum of the feature y based on generalized regression (GREG)

$$\hat{y}_{reg} = \sum_{S} w_k \, y_k = \hat{y}_\pi + (\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_{\boldsymbol{\pi}})' \hat{\boldsymbol{B}}_{\boldsymbol{s}}$$

Here, the estimators y and x denoted by the subscript π denote the Horvitz-Thompson estimators of the sums of the characteristic y and the x vector of the auxiliary characteristics, $\hat{\boldsymbol{x}}$ the vector of the sums of the x characteristics known for the population, and $\hat{\boldsymbol{B}}_{\boldsymbol{s}}$ the vector of the regression parameters of y on the x characteristics estimated on the basis of the sample.

The GREG weights can be directly fitted to continuous variables or to their sums. This means for the weighting of the cross-sectional sample that simultaneously the sample can be adjusted to the distribution of establishments and to the distribution of employees.

During the weighting, there are checks of whether the factors in the individual cells are becoming too high or too low, or whether there is no case at all in a cell. In these cases the cell is amalgamated with an appropriate neighbouring cell.

For cross-sectional evaluations, the cross-sectional weighting factor HR2021Q must be used.

7 Definition of Longitudinal Cases and Projection

Longitudinal or panel analyses allow researchers to trace developments in individual establishments over a longer period. Due to its large net sample and its long duration of meanwile 29 survey waves, the IAB Establishment Panel offers a wide range of options for such analysis. Table 6 gives an overview of the number of valid interviews for different starting years in order to produce a balanced panel.

Table 6: Overview of the number of evaluable interviews for different starting years (balanced panel)

	Starting year							
wave	1996	2000	2003	2007	2009	2012	2016	2020
1996	–	–	–	–	–	–	–	–
1997	6,822	–	–	–	–	–	–	–
1998	5,597	–	–	–	–	–	–	–
1999	4,654	–	–	–	–	–	–	–
2000	4,004	–	–	–	–	–	–	–
2001	3,479	10,840	–	–	–	–	–	–
2002	2,999	8,762	–	–	–	–	–	–
2003	2,610	7,295	–	–	–	–	–	–
2004	2,286	6,363	12,775	–	–	–	–	–
2005	2,011	5,542	10,771	–	–	–	–	–
2006	1,761	4,854	9,118	–	–	–	–	–
2007	1,524	4,242	7,845	–	–	–	–	–
2008	1,382	3,774	6,889	12,567	–	–	–	–
2009	1,233	3,338	6,037	10,598	–	–	–	–
2010	1,095	2,973	5,324	9,118	12,524	–	–	–
2011	990	2,666	4,759	7,991	10,653	–	–	–
2012	868	2,361	4,199	6,983	9,132	–	–	–
2013	781	2,108	3,761	6,179	8,019	12,622	–	–
2014	682	1,883	3,352	5,430	6,970	10,515	–	–
2015	603	1,674	2,951	4,800	6,116	9,010	–	–
2016	537	1,520	2,667	4,287	5,447	7,920	–	–
2017	474	1,348	2,363	3,792	4,813	6,950	12,614	–
2018	412	1,173	2,060	3,343	4,230	6,071	10,505	–
2019	337	971	1,708	2,754	3,538	5,047	8,476	–
2020	261	735	1,300	2,088	2,680	3,824	6,267	–
2021	202	568	1,004	1,566	2,049	2,942	4,740	10,256

7.1 Panel Case definition for Longitudinal Cases

For longitudinal weighting, all the establishments which were a panel case the previous year and all new establishment numbers of the supplementary and extension sample from the subsequent years are generally considered as panel cases. To count as a panel case, information must be available for every survey year from the respective first interview to the latest survey, in the form either of a valid interview or the information that the establishment (or the establishment number as applicable) has expired. We therefore differentiate between "panel cases with interview" (panel cases for which valid questionnaires are held for every individual year up to the current survey) and "no longer operational panel cases" (in which from the date on which the establishment ceased operations onwards the only information held is that the establishment no longer exists). For the panel cases with interview, it is also necessary to ensure that the information obtained applies to the same establishment unit every year.[19]

The subgroups comprising the respective panel cases can be shown using the longitudinal section for 2016 – 2021:

All cross-sectional cases from the 2016 wave for which information is available from the 2016 wave onwards.

Establishments from the supplementary and extension samples 2017 (only "new establishment numbers") for which information from the 2017 wave onwards is available.

Establishments from the supplementary and extension samples 2018 (only "new establishment numbers") for which information from the 2018 wave are available.

Establishments from the supplementary and extension samples 2019 (only "new establishment numbers") for which information from wave 2019 are available.

Establishments from the supplementary and extension samples 2020 (only "new establishment numbers"), for which information from wave 2020 are available.

Establishments from the supplementary and extension samples 2021 (only "new establishment numbers") which participated in 2021 wave.

To make it easier for the users of the IAB Establishment Panel to identify different subgroups for cross-sectional and in particular for longitudinal analyses, the relevant subgroups have been marked. For this the following information has to be appropriately combined:

Field result of the ongoing wave (valid interview with/without employees subject to social insurance contributions as of the reference date, establishment no longer operational, cases that did not respond the previous year but are willing to be surveyed again/final non-responses)

Was the establishment surveyed in the previous wave (respondents from the previous wave, non-responses from the previous year that are willing to be surveyed again, supplementary and extension sample)?

19 In companies with several establishments in particular, it can be the case that the interviewees provide details of different units in different years (e.g. once about the local establishment, once about the entire company). This information (from the address protocol or as the result of the validation and editing process) is taken into account when forming the wave code (WELLwxyz) and is thus available for defining the panel cases.

Was the same establishment surveyed as last time?

In consultation with the IAB we have developed the following concept for this. Every case is given a unique identifier in each wave which takes account of the above criteria. This so-called wave code is stored in the variables WELLwxyz, where wxyz stands for the year in which the survey took place (thus WELL1993 for wave 1 in 1993, WELL1994 for wave 2 in 1994 etc.). This labelling takes place using a letter of the alphabet (see Table 7).

Table 7: Group identifiers in the variable WELLwxyz

		Labelling letter *with* employees subject to social insurance contributions as per the respective reference date[20]	Labelling letter *without* employees subject to social insurance contributions as per the respective reference date[20]
1.	Cases **with** interview in the ongoing wave		
1.1	Cases surveyed for the **first time** (= at the date of drawing) from the basic sample and the respective supplementary or top-up samples[21]	**A**	Not permitted
1.2	Cases surveyed **repeatedly**		
1.2.1	**with** interview the previous year		
1.2.1.1	**same** unit interviewed as previous year	**B**	**C**
1.2.1.2	**different** unit interviewed to previous year	**D**	Not permitted
1.2.2	**without** interview in previous year[22]	**E**	Not permitted
2.	Cases **without** interview in the ongoing wave		
2.1	Non-response that can be surveyed again in future		**H**
2.2	Cases from earlier extensions that can no longer be surveyed[23]		**W**
2.3	Non-responses that can no longer be surveyed[24]		**X**
2.4	No longer operational establishments (according to field result, editing or BA file respectively)		
2.4.1	in the ongoing wave		**Y**
2.4.2	earlier than this		**Z**

In principle, other longitudinal sections can also be defined, however there are no weighting factors available for this.

20 The cut-off date is always June 30 of the previous year, so for wave 28 (2020), for example, June 30, 2019.The weighting takes place using the questionnaire information and the targets from the BA establishment file as of this reference date. No newer information from the BA establishment file is yet available as per the respective weighting date.

21 Note: Dropouts from wave 1 who were interviewed for the first time in wave 2 and dropouts from the basic sample East (wave 4) who were interviewed for the first time in wave 5 do not belong to this group; these cases were assigned the code letter H in wave 1 or wave 4, respectively, and the code letter E in the follow-up wave.

22 Since 2002 non-responses from the previous year (H cases) have been treated in the same way as establishments being surveyed for the first time and the employee details from the previous year according to the BA used for identification. For this reason, there are no longer any G cases (establishments repeatedly surveyed without an interview the previous year and different unit to that surveyed the year before last).

23 Thus e.g. the cases from the extension sample in 1997 in agriculture in Mecklenburg-West Pomerania, or the cases from the Halle Institute for Economic Research (IWH) extension sample in 1998 for the construction sector in East Germany.

24 These include (a) establishments which declined to be interviewed, unless they expressly consented to being interviewed again the following year, and (b) non-responses from the previous wave from which no interview has also been obtained in the ongoing wave (i.e. the combination HH is not permitted and becomes HX).

7.2 Overview of the Longitudinal Sections provided 2021

There are three longitudinal sections for which a panel weighting was undertaken for the 2021 wave. Table 8 depicts the case numbers for the respective longitudinal sections from 2012, 2016 and 2020 onwards.

As already mentioned above, the definition of panel cases essentially comprises all the establishments which were a panel case the previous year – either with an interview or as a no longer operational establishment – and all the establishments that were newly founded between the reference date of the year before last and the reference date of the previous year. For this reason, the number of newly-founded establishments (in the sample) is identical for all longitudinal sections. After the longitudinal section has existed for a certain period of time, the number of newly-founded establishments more or less counterbalances the number of non-responses. Hence after 3 – 5 waves, largely stable case numbers are achieved. For this reason, and because the longitudinal sections shown in Table 8 started with relatively similar case numbers, the number of panel cases in the individual longitudinal sections (with exception of the new longitudinal section of 2020) in 2021 is at a similarly high level.

Table 8: Overview of all longitudinal sections

Starting year of the respective longit. section	No. of panel cases in start year	Number 2021			
		No. of panel cases	No. of repeaters	No. of newly founded current wave	No. of no longer operational
2012	14,186	8,949	5,411	1,281	2,257
2016	14,250	8,680	6,300	1,281	1,099
2020	11,753	11,753	10,256	1,281	216

7.3 Longitudinal Weighting Process

Specific panel weighting factors are calculated for every longitudinal section. As for the cross-sectional weighting, the panel weighting takes place in the form of a projection onto the population. The aim of the panel weighting is that panel analyses using the respective longitudinal weighting factor should as far as possible generate the same distributions for all the waves involved as those from cross-sectional analyses of the individual waves.

Panel weighting essentially takes place in eight steps, each forming the basis for the next, with East and West Germany being differentiated in all steps. Unlike the cross-sectional weighting, for the longitudinal weighting a weighting at federal state level takes place only in the first step. In the subsequent steps the federal state level is no longer monitored. In total the eight weighting steps of the longitudinal weighting are repeated iteratively for as long as this is necessary and purposive.

The first step is to adapt the latest available cross-sectional cases to the structure of the population (number of establishments and employees subject to social insurance contributions as of 30 June of the previous year). The entry factor is the previous year's panel weighting factor and for new establishment numbers the cross- sectional weighting factor of the current wave.

The next stage is to adjust the key figure for the stock of establishments for each individual year included in the longitudinal section, and for the establishment numbers that in the meantime have become no longer operational or are new. Each year's stock includes the establishments with at least one employee subject to social insurance contributions as of the reference date. The new establishment numbers consist of those from the stock that had no employees subject to social insurance contributions as of the reference date one year before. The no longer operational establishment numbers are those in the stock that no longer had any employees subject to social insurance contributions as of the reference date one year later. This step depicts the recovery or termination of establishment numbers over the course of time.

The third step is to adjust for the establishments that are surviving or no longer operational from the different entry cohorts.

The fourth and fifth steps take account of the establishment dynamics (growth and shrinkage of establishments). In particular in the longitudinal sections that have been running for a longer period, an extremely large number of combinations is possible in respect of the development of the number of employees. Simply because of the limited case numbers, it is not possible to obtain a finely differentiated picture of the possible development paths (changes between individual establishment size classes from one year to the next year). We have therefore applied a simplified procedure for the panel weighting in consultation with the IAB: a change of establishment size class is only taken into account between the starting wave of the respective longitudinal section (or as applicable for new establishment numbers, the first time they were surveyed) and the latest available data. No check is kept on changes occurring in between. Due to extremely low numbers of cases in certain combinations, the theoretically possible combinations are compounded as follows: as of the latest available data the establishment is in the same establishment size class as it was at the beginning, or has grown, or has shrunk. It is self-evident that when this approach is used the development paths of establishments can only be represented approximately during the panel weighting. In the fourth step, the establishment dynamics are taken into account for the establishments from the stock of the starting wave of the respective longitudinal section. In the fifth step, this is done for the "new" establishment numbers added during the longitudinal period.

The sixth step is to correct any disproportional non-responses depending on the answers to individual questions from the previous year's survey. This is done by undertaking multivariate analyses of non-responses in comparison between the previous wave and the most recent available data.

The two subsequent steps involve adapting the cases belonging to the respective cross-section to the requirements of the stratification matrix for every survey year included in the longitudinal section. To limit the number of weighting cells, only the establishment size classes (seventh step) and a simplified sector structure with six categories (step eight) are taken into account (cf. Appendix Table 11).

In individual cases using the aforementioned parameters during the panel weighting can lead to extremely high weighting factors. To limit the resulting problems of outliers, only weighting factors up to a maximum of 4,000 are permitted for the panel weighting. As a consequence, the theoretically necessary target numbers for individual characteristic values may not be reached due to the weighting process. Thus, for example, new establishment numbers in certain years are substantially underestimated during the panel weighting.

For longitudinal evaluations, the following factors are to be used:

For the longitudinal section 2012 - 2021: HR12_21P

For the longitudinal section 2016 - 2021: HR16_21P

For the longitudinal section 2020 - 2021: HR21_21P

8 General Information about Evaluation

Descriptive results should always be calculated and interpreted on a projected level, so that the disproportional sample structure and selective non-responses are corrected.

The results should always be interpreted against the background of the underlying (unweighted) case numbers. The lower the number of cases considered, the more uncertain the results. The area of uncertainty for different unweighted case numbers is shown in the error tolerance table (cf. Appendix). The case number of 15,000 establishments can exhibit differences of two percentage points and more as significant differences.

Even when the projection factor contains an adjustment proportional to the numbers of employees, we recommend always showing employee numbers rounded to full thousands.

The panel weighting does indeed take place in the form of a projection, however for methodological reasons it is associated with greater inaccuracies in terms of the adjustment to the different target structures than the respective cross-sectional weightings, because the target structures of the various cross-sections, which fluctuate to some extent, has to be achieved with just one projection factor. This limitation applies to both the projected number of establishments and – to an even greater extent – to projected employee figures from the IAB Establishment panel.

We therefore urgently recommend that when analyses take place using longitudinal factors not to show absolute figures. Projected absolute figures (even if rounded to full thousands) give the reader the impression of an accuracy that cannot be achieved with the IAB Establishment Panel when longitudinal analyses are used.

As with all sample-based data, we recommend to use distribution measures, in other words statements of the type “X % of establishments have grown since 2003, Y % have contracted” or “The number of employees by Year Y developed better/worse in establishments which invested more than average in Year X than was the case for establishments which made no investments in Year X”. It should also be borne in mind that the percentage values obtained can also be afflicted with inaccuracies, so the values should if anything be interpreted as orders of magnitude.

When undertaking analyses, particularly in comparisons across time, account needs to be taken of changes to the delineations of the sectors, the regional classification and the population. Such changes on their own can lead to sometimes substantial changes in the parameters and distributions. Thus when the employee statistics were revised with the 2015 wave, new groups of persons were included; hence the increase at that time in the number of employees subject to social insurance contributions by a good three percent can also be ascribed to the revision of the employee statistics, and not just to changes in the real employee situation.

Moreover, in 2009 the last change of the stratification matrix was made in the course of the changeover from the WZ2003 to the WZ2008 classification of economic sectors. Changes in the shares of establishments by sector can therefore only be attributed to real changes in the economy structure to a limited extent.

Since 2007 East and West Berlin have been aggregated. In the consequence of this, Berlin was excluded from the extention sample in manufacturing industry in eastern Germany at the request of the Leibniz Institute for Economic Research in Halle (IWH). Since then, this extention sample only includes eastern German territorial states. The IWH sample since 2007 is therefore only partially comparable with the IWH sample before 2007.

Generally when undertaking time series and panel analyses, account should be taken of changes to the question or the individual items, so that differences in results are actually ascribable to real changes and not to changes to the question and/or the answer requirements. Account should also be taken of the respective timeframe. Thus establishment sizes refer to the reference date 30 June, flow figures to the first half year, and some questions – e. g. the business volume –relate to the last year or the financial year that has recently expired.

9 Appendix

The study “Employment Trends – Employer Survey 2021” (IAB Establishment Panel 2021) was undertaken on behalf of the Institute for Employment Research (IAB) of the Federal Employment Agency (BA).

Regional and sectoral expansions of the sample were commissioned by:

- The Senate Department for Integration, Labour and Social Affairs of the Federal State of Berlin
- The Ministry for Labour, Social Affairs, Health, Women and Family in the Federal State of Brandenburg
- The Ministry for Economic Affairs, Labour and Health for the Federal State of Mecklenburg-West Pomerania
- The Saxony State Ministry for Economic Affairs, Labour and Transport
- The Ministry of Labour, Social Affairs and Integration Saxony-Anhalt
- The Thuringian Ministry for Labour, Social Welfare, Health, Women and Family Affairs
- The Halle Institute for Economic Research (IWH)
- The Ministry of Finance and Economic Affairs Baden-Württemberg
- The Bavarian Ministry of Labour and Social Affairs, Family and Integration
- The Senator for Economic Affairs, Labour and Ports of the Federal State of Bremen
- The Hessian Ministry of Economics, Energy, Transport and Regional Development
- The Ministry for Social Affairs, Labour, Health and Demography of Rhineland-Palatinate
- The State Ministry of Saarland for Economic Affairs, Labour, Energy and Traffic

Table 9: Classification of economic activities by 19 sectors for sampling and cross-sectional weighting from the 2010 wave onwards

Variable BR19BAxy	Sector	WZ2008 code	Questionnaire sector
1	Agriculture, forestry and fishing	1 – 3	1
2	Mining and quarrying, electricity, gas and water supply; sewerage and waste management	05 – 09, 35 – 39	2 – 3
3	Food products, beverages and tobacco	10 – 12	4
4	Consumer products (excluding manufacture of timber products)	13 – 18	5 – 6
5	Industrial goods (including manufacture of timber products)	19 – 24	7 – 10
6	Capital and consumer goods	25 – 33	11 – 17
7	Construction	41 – 43	18 – 19
8	Wholesale, sale and repair of motor vehicles	45 – 46	20 – 21
9	Retail	47	22
10	Transport and warehousing	49 – 53	23
11	Information and communication	58 – 63	24
12	Hotels and restaurants	55 – 56	25
13	Financial and insurance services	64 – 66	26
14	Economic, scientific and freelance services	68 – 82	27 – 36
15	Education	85	37
16	Health and social services	86 – 88	38
17	Other services	90 – 93, 95, 96	39 – 41
18	Representations of interests	94	42
19	Public administration, defence, social security	84	43

Table 10: Establishment size class classifications for sampling and cross-sectional and longitudinal weightings

Employees subject to social insurance contributions on 30 June of the previous year
1 – 4
5 – 9
10 – 19
20 – 49
50– 99
100 – 199
200 – 499
500 – 999
1000+

Table 11: Classification of economic activities by 6 sectors for longitudinal weighting from the 2009 wave onwards

Sector	Code from classification by 19 sectors
Agriculture and forestry, fishing	1
Manufacturing industry	3 – 6
Other production industry	2, 7
Retail/transport and warehousing/hotels and restaurants	8 – 12
Financial and insurance services/business services	13, 14
Public and private services	15 – 19

Imprint

FDZ-Methodenreport 7|2022 EN

Date of publication
07 December 2022

Publisher
Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nuremberg

Download
https://doku.iab.de/fdz/reporte/2022/MR_07-22_EN.pdf

Documentation version
IABBP9321_EN_v1_dok1, DOI: 10.5164/IAB.FDZM.2207.en.v1

Dataset version
IABBP 9321 v1, DOI: 10.5164/IAB.IABBP9321.de.en.v1

All publications in the series "FDZ-Methodenreport" can be downloaded from
https://fdz.iab.de/en/research/publications/fdz-methodenreport-series/

Website
https://fdz.iab.de

Corresponding author
Susanne Kohaut
Phone: +49 911 179-3253
Email: susanne.kohaut@iab.de