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

FDZ-DATENREPORT

Documentation of labour market data

08|2022 EN The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, waves 2000 to 2019 and subsequent quarters 2006 to 2021

Mario Bossler, Nicole Gürtzgen, Alexander Kubis, Benjamin Küfner, Martin Popp

Bundesagentur für Arbeit

The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, waves 2000 to 2019 and subsequent quarters 2006 to 2021

Mario Bossler (IAB), Nicole Gürtzgen (IAB), Alexander Kubis (IAB), Benjamin Küfner (IAB), Martin Popp (IAB)

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

Table of Contents

List of Tables

Abstract

The IAB Job Vacancy Survey is a quarterly and representative establishment survey on labour demand and recruitment processes in Germany. The survey identifies the overall stock of vacancies in the German labour market, including those vacancies that are not reported to the Federal Employment Agency (FEA). The first module of the questionnaire collects information about the number and structure of vacancies, future personnel requirements, about the current economic situation and the expected development of participating establishments. The second module enquires employer attitudes and firm use of current labour market instruments as well as the employer handling of persons disadvantaged in the labour market. The third module asks for information about the last new hiring and the last case of a failed recruitment effort. The Research Data Centre of the Federal Employment Agency offers the data sets of the survey waves from 2000 onwards.

Zusammenfassung

Die IAB-Stellenerhebung ist eine quartalsweise durchgeführte und repräsentative Betriebsbefragung über das gesamtwirtschaftliche Stellenangebot sowie Einstellungsprozesse in Deutschland. Die Erhebung ermittelt die Gesamtzahl aller offenen Stellen am Arbeitsmarkt, einschließlich jener Stellen, die nicht der Bundesagentur für Arbeit (BA) gemeldet werden. Das erste Modul des Fragebogens enthält Informationen zu Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. Das zweite Modul erfragt die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente sowie den betrieblichen Umgang mit am Arbeitsmarkt benachteiligten Personen. Das dritte Modul enthält Fragen zum letzten Fall einer Neueinstellung und zum letzten Fall eines gescheiterten Rekrutierungsversuchs. Das Forschungsdatenzentrum der Bundesagentur für Arbeit stellt die Datensätze der Befragungswellen ab 2000 mit allen Fragebogenteilen einschließlich der Quartalsbefragungen für externe Wissenschaftlerinnen und Wissenschaftler bereit.

Keywords

Establishment survey, job vacancies, recruitment processes

1 Differences to previous versions

Compared to previous versions of the data (see Bossler et al., 2021), the 4th quarter of 2019 and the three subsequent quarters are included. In the 4th quarter of 2019, data on employment opportunities of unemployed persons were collected within Module 2 (see Gürtzgen/Popp, 2022) Starting in the 2nd quarter of 2020, establishments were surveyed about the effects of the COVID-19 crisis on establishments (see also Bossler et al., 2020a, Gürtzgen/Kubis/Küfner 2020, Gürtzgen/Kubis, 2021).

The 2019 survey wave has three special features that need to be considered when using the data. These are explained in detail in Section 3. First, in response to increased information needs during the COVID-19 crisis, the 2019 wave was extended by two additional quarterly follow-up interviews in the 4th quarter of 2020 and in the 1st quarter of 2021 and made available to users at the FDZ. Second, a survey experiment took place in parallel to the main survey in the 4th quarter of 2019. The resulting data are part of the dataset and should be used without weights and for exploratory purposes only. Third, in the process of quality assurance measures, deviant interviewer behaviour was identified for some observations in the 2019 survey wave (Bossler et al., 2022). The data provided at the FDZ were cleaned of these observations as part of the revision process.

In addition, inconsistencies in previous waves regarding the handling of questions with multiple responses were corrected (see also 4.3). All unselected answer options respective variables are set to missing.

2 Description of the data set

2.1 Short description

Table 1: Content characteristics

Category	Description
Topic/variable groups	Employment, personnel requirements, activity impediments, stopped search effort, occupations with increasing demand and with expected bottlenecks in the next 3 years, marginal employment, part-time employment, labour market reform, employment opportunities of (long-term) unemployed persons, further education of workforce, One-Euro-Jobs, labour market participation of older employees, establishment-level integration of refugees, handling of the COVID-19 crisis, detailed information on the last hiring and to the most recent stopped search effort
Unit of observation	Establishments
Number of observations	7,500 to 15,000 establishments
Time period	IV.2000 to I.2021
Frequency	Quarterly
Regional subdivision	East/West Germany, federal state; see “sensitive attributes” for details

Table 2: Methodological characteristics

Category	Description
Survey design	Representative sample of establishments, stratified by size, industry and West/East Germany
Participating institutions	**Client:** Research unit AMPI of the IAB **Implementation:** Economix Research & Consulting, Munich
Frequency of data collection	Yearly self-administered (paper or web) main survey and quarterly follow-up interviews by phone
File format/size	Stata; all data sets together ca. 264 MB
File organization	By wave

Table 3: Data access

Category	Description
Data access	Remote data access, on-site use
Degree of anonymization	Weakly anonymized
Sensitive attributes	Industry classifications: WZ73 3 digits (2000-2003), WZ03 3 digits (2004)/5 digits (2005-2009) WZ08 3/5 digits (since 2010), districts (since 2007), counties (since 2005)
Proper citation	Data: "Gürtzgen, Nicole; Kubis, Alexander; Bossler, Mario; Küfner, Benjamin; Popp, Martin (2022): "German Job Vacancy Survey of the IAB – Version 0019 v1". Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.IABSE0019.de.en.v1 The data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and/or remote data execution." Data documentation: - Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin; Popp, Martin (2022): The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, waves 2000 to 2019 and subsequent quarters 2006 to 2021. FDZ-Datenreport, 08/2022 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2208.en.v1 - Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin; Popp, Martin (2022): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Wellen 2000 bis 2019 mit Folgequartalen von 2006 bis 2021. FDZ-Datenreport, 08/2022 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2208.de.v1

Details about the several ways of and the requirements for data access may be found on the website of the Research Data Centre at http://fdz.iab.de.

2.2 Overview

The IAB Job Vacancy Survey is a quarterly and representative establishment survey on labour demand and recruitment processes in Germany (see also Bossler et al., 2020b). The survey identifies the overall stock of vacancies in the German labour market, including those vacancies that are not reported to the FEA, Germany's public employment agency. This number alone allows for drawing valid and unbiased conclusions about labour demand in the economy as a whole. This is because the reporting quota, which is the share of vacancies reported to the FEA among all vacancies, greatly fluctuates over time and exhibits systematic differences between industries, regions and occupations.

Since 1989, the survey has been conducted in the 4th quarter of every year using a multi-part paper & pencil questionnaire. Since 2002 (with interruptions from 2005 to 2010) establishments may participate online via internet as well. This web questionnaire is identical to the paper & pencil questionnaire. Module 1, formerly called main questionnaire, enquires information about the number

and structure of vacancies, about future personnel requirements, about the current economic situation and the expected development of participating establishments. Module 2, formerly called special questionnaire, has been used since 2000 in the context of SGBII to survey employers' assessment and use of current labour market policy instruments. In addition, data are collected on how establishments deal with people who are disadvantaged in the labour market. Module 3, formerly called additional questionnaire, collects information about the last new hire and the last case of a failed recruitment effort. This regular and detailed enquiry into hiring processes is a unique characteristic of the IAB Job Vacancy Survey, providing information on characteristics of the job, the person hired, search and recruitment channels including the use of job placement services, search and recruitment durations, the number of applicants, hiring difficulties as well as any compromises made. Module 3 is only answered by establishments who have hired a new employee in the past year or have tried to fill an open position.

Module 1 and Module 3 are basically the same from wave to wave except for editing changes and possible shifts in emphasis. By contrast, Modul 2 changes from wave to wave. Since 2000, it enquires employer attitudes and firms' use of current labour market instruments. From 2005 to 2013, it mainly focused on the workfare scheme One-Euro-Jobs. During this period, therefore, it targeted only those industries in the public sector in which the vast majority of one-euro jobs were offered and carried out.

Since 2005, there are short follow-up interviews by phone on the core questions among participants of the main survey in the first three quarters of the following year. These quarterly telephone follow-up interviews complement the main survey in the respective 4th quarter. They serve to update the most important data points in module 1, including the number of employees, the assessment of current workforce developments and the number of vacancies. In the 2019 survey wave, additional quarterly follow-up interviews took place in the 4th quarter of 2020 and in the 1st quarter 2021.

The Research Data Centre of the Federal Employment Agency makes the survey waves from 2000 onwards available to external researchers. The survey waves include the data from all modules including the quarterly follow-up interviews. Since September 2015, all documentation including variable and value labels, data reports and code books, has been available in English as well.

2.3 Population and sample of the main survey

The population of the main survey in the 4th quarter of every year are all establishments in Germany with at least one employee subject to social security contributions in June (until wave 2004) or December (waves 2005 and later) of the preceding year, excluding private households. "Establishment" is an economic unit according to the establishment identifier concept of the establishment identifier service of the Federal Employment Agency (Bundesagentur für Arbeit, 2020).

A new stratified random sample is drawn every year from this population. It is stratified by region, establishment size category as well as industry, creating a three-dimensional sampling matrix. Table 4 summarizes the stratification variables. During the waves 2005 to 2013, an additional sample was drawn for the public sector in order to collect information on One-Euro-Jobs. Only in the year 2005, another additional sample was drawn to increase the number of observations due to the low

response rate. In both cases, the additional samples were drawn only after removing already-drawn establishments from the population.

Table 4: Stratification variables of the gross samples 2000–2019

Variable	Divisions
Region	**2000–2003:** West Germany including West Berlin, East Germany including East Berlin **since 2004:** West Germany, East Germany including all of Berlin
Establishment size (Total number of employees)	**2000–2003:** 1–9, 10–19, 20–49, 50–199, 200–499, 500+ **2004:** 1–9, 10–19, 20–49, 50–99, 100–199, 200–499, 500–999, 1000+ **2005–2013:** 1–9, 10–19, 20–49, 50–199, 200–499, 500–999, 1000+ **since 2014:** 1–9, 10–19, 20–49, 50–249, 250–499, 500–999, 1000+
Industry	**2000–2002:** 22 Industries based on German classification 1973 **2003:** 14 Industries based on German classification 1973 **2004–2009:** 28 Industries based on German classification 2003 **2010–2014:** 23 Industries based on German classification 2008 **since 2015:** 24 Industries based on German classification 2008

From this gross sample, all establishments that do not wish to be contacted in future waves are removed from the distribution sample, which forms the basis for the response rate stated in Table 5.

All establishments of the sample are contacted by mail at the beginning of October and are asked to participate in the survey. The mailing includes a cover letter, a privacy/data protection statement, the questionnaire as well as a sheet with instructions on how to fill out the questionnaire and answers to frequently asked questions. A few weeks later, a second mailing with a cover letter reminding recipients of participating is sent out that otherwise includes the same content. Both cover letters include a telephone number to a helpline that takes comments and answers any questions that employers may have about the survey. The field phase usually lasts until the beginning of January, with most responses coming back between the end of October and the beginning of December.

Table 5: Sample development 2000–2019

Wave	Address file date	Average sampling ratio	Gross sample	Distribution sample	Net sample	Response rate
2000	31-12-1999	1.3%	28,486	28,266	7,578	26.8%
2001	30-06-2000	1.3%	27,994	27,827	7,347	26.4%
2002	30-06-2001	1.3%	27,418	27,147	5,773	21.3%
2003	30-06-2002	1.8%	37,789	37,421	7,310	19.5%
2004	30-06-2003	2.7%	56,926	56,699	11,707	20.6%
2005	31-12-2004	3.9%	78,032	69,702	11,742	16.8%
2006	31-12-2005	3.9%	75,290	69,231	13,537	19.6%
2007	31-12-2006	3.8%	75,128	73,635	14,381	19.5%
2008	31-12-2007	3.8%	77,543	75,035	13,652	18.2%
2009	31-12-2008	3.8%	77,537	74,998	15,288	20.4%
2010	31-12-2009	3.7%	77,739	75,000	15,124	20.2%
2011	31-12-2010	3.6%	77,685	74,660	15,139	20.3%
2012	31-12-2011	3.6%	79,181	75,006	13,807	18.4%
2013	31-12-2012	3.6%	77,214	75,486	14,019	18.6%

Wave	Address file date	Average sampling ratio	Gross sample	Distribution sample	Net sample	Response rate
2014	31-12-2013	3.6%	76,767	75,073	12,750	17.0%
2015	31-12-2014	3.9%	85,556	83,639	12,852	15.4%
2016	31-12-2015	4.0%	110,140	85,000	11,588	13.6%
2017	31-12-2016	5.1%	120,102	109,988	14,614	13.3%
2018	31-12-2017	5.2%	134,848	109,988	14,528	13.2%
2019	31-12-2018	5.1 %	114,613	109,986	13,906	12.6 %

2.4 Extrapolation procedure of the main survey

Details and the background of the current extrapolation procedure may be found in the IAB Research Report No. 4/2016 (Brenzel et al., 2016). To compute the establishment weights, the following steps are taken:

1. Computing the **design weights** for every cell of the sampling matrix as the inverse of the respective sampling ratio.

2. Multiplication of the design weights with a weight computed from **non-response modelling**. The latter is based on a logistic regression model using the establishment size category, industry, average daily wage and average employee age from administrative data of the Federal Employment Agency as auxiliary variables. The non-response weight is the inverse of the estimated response propensity.

3. **Calibration** of the weights from step 2 as the starting weights of a generalized regression estimator (GREG) with the aim of exactly matching the anchor variables “number of establishments” and “number of employees subject to social security contributions” in all cells of the sampling matrix.

As the anchor variable “number of employees subject to social security contributions” has to be estimated for the most recent survey wave (see Hutter, 2020), the calibration procedure is repeated after 18 months with the then-available actual number of employees, resulting in a revised set of extrapolation weights. The data sets at the Research Data Centre will only contain revised extrapolation weights; as a consequence, every new wave of the survey is made available in the Research Data Centre no earlier than 18 months after the end of the field phase. Previously published results by the IAB in the form of press releases and policy reports usually relied on preliminary weights, which is why they cannot be replicated exactly using Research Data Centre data based on revised weights; differences typically lie within the respective 95% confidence interval.

The retrospective application of the new extrapolation procedure on the waves 2000 to 2014 required recoding the different establishment size categories of the original sampling matrices (see Table 4) into a consistent division of six size classes (1–9, 10–19, 20–49, 50–249, 250–499, 500+). The industry classification in the waves 2000 to 2009 have been recoded into consistent 16 classifications of the German classification edition 2003. Between 2010 and 2014, the original sampling matrices already contained the still-current 23 classifications of the German Classification edition 2008. Since 2015, 24 classifications of the German Classification edition 2008 have been used for extrapolation in analogy to the sample matrix.

2.5 Sampling and extrapolation of the quarterly follow-up interviews

In each subsequent quarter, establishments are drawn from the participants of the main survey of the 4th quarter. The target number of establishments is 9,000 per quarter. The net sample of the main survey therefore is the gross sample of follow-up interviews of the three subsequent quarters.

For each of the three subsequent quarterly surveys a separate establishment weight is available, comprised of the establishment weight of the main survey in the 4th quarter of the original year and the result of another non-response model for the respective subsequent quarter.

3 Special features of the survey waves

3.1 Survey wave 2019

3.1.1 Additional quarterly follow-up interviews

The COVID-19 crisis has substantially increased the need for short-term and valid information on job openings and operational management in general. To address this need for information, the 2019 survey wave was extended to include additional quarterly follow-up interviews in the 4th quarter of 2020 and in the 1st quarter of 2021. In these two quarters, a subsample of the surveyed establishments from the 4th quarter of 2019 was surveyed again, as in the regular quarterly follow-up interviews. Thus, the 2019 survey wave represents a panel survey with a measurement prior to the outbreak of the pandemic in the 4th quarter of 2019 and further measurements during the course of the pandemic. Beginning in the 2nd quarter of 2020, all quarterly follow-up interviews were expanded to include additional questions about the impact of the COVID-19 pandemic on establishments. The 2019 survey wave, supplemented by the two additional quarterly follow-up interviews, is also referred to as the "COVID Panel".

Note that independent extrapolation weights are provided for the COVID Panel, analogous to the regular weighting procedure (see Section 4.1). The extrapolation weights for the COVID Panel are specifically labelled. The weights carry the suffix `_covid` and accordingly have the variable names `q419_greggew_covid`, `q120_greggew_covid`, `q220_greggew_covid`, `q320_greggew_covid`, `q420_greggew_covid`, and `q121_greggew_covid`, respectively. In 1st, 2nd, and 3rd quarter of 2020, the weights with the `_covid` suffix correspond to the regular extrapolation weights without the suffix. Since the quarterly follow-up interviews in the 4th quarter of 2020 and in the 1st quarter of 2021 are exceptional and supplementary, no regular weights are available for these two quarters of the 2019 wave, but only the weights with suffix `_covid`.

The analyses reported by the IAB on job vacancies in the 4th quarter of 2020 and in the 1st quarter of 2021 are based on the 2020 survey wave. The additional quarterly follow-up interviews of the 2019 survey wave were not used for this purpose.

3.1.2 Experimental data from the 2019 survey wave.

In the 4th quarter of 2019, a survey experiment on the effects of explanatory information on response behaviour took place in addition to the regular main survey (for more information on the experimental design and results, see Küfner/Sakshaug/Zins, 2021). For the experiment, a separate sample was drawn and - in contrast to the main survey – a survey using single online mode was conducted. The regular main survey in the 4th quarter of 2019 and all subsequent quarterly follow-up interviews were not part of the experiment and were not affected or influenced by the experiment. The main survey stands on its own and forms the central basis for substantive analyses.

Data from this experimental survey are part of the 2019 annual wave provided to the FDZ. These observations can be uniquely identified via the `experiment2019` indicator variable. It should be noted that no extrapolation weights are available for these experimental observations. Hence, using these observations for weighted analyses is not possible (`q419_greggew == .`). With the exception of the analyses conducted as part of the survey experiment, no additional quality testing of the experimental data took place. Also, it was not investigated to what extent deviations in the survey design have an influence on possible selection effects or measurement effects and thus on the comparability with the main survey. Record linkage with administrative data is not possible (see Section 4.3). Therefore, by default, we recommend excluding these experimental data (`drop if experiment2019 == 1`) for substantive analyses. Nevertheless, the use of the experimental data for exploratory and unweighted purposes can add value and is therefore available to users at the FDZ.

3.1.3 Revision due to deviant behaviour of interviewers

In the 2nd quarter of 2021, routine data quality checks identified conspicuous data patterns and time trends. A detailed investigation substantiated the suspicion that one interviewer had systematically manipulated interviews. The interviewer was found to manipulate responses to a filter question, as (s)he conducted a strikingly large fraction of interviews stating that the surveyed establishments had no vacancies. As a result of this falsified information, no further questions had to be asked about the structure of the vacancies, which shortened the survey duration considerably. During the investigation, another interviewer was identified who showed deviating behaviour in the 1st quarter of 2020. All interviews of these two interviewers were - also retrospectively - excluded from the entire 2019 survey wave. The two interviewers had not conducted any interviews in earlier survey waves.

New extrapolation weights were calculated as part of a revision. In addition, all IAB publications and figures published on the website that contained observations of the falsifiers were revised. The data now published in the FDZ exclude the observations of the two interviewers and include the revised extrapolation weights. Data previously published in the FDZ are not affected by the interview falsification. Further information on the detection of the deviant interviewer behaviour, the revision of data, and about the improved quality assurance of the telephone interviews are recorded in the FDZ method report by Bossler et al. (2022).

3.2 Survey wave 2016

In the 2nd quarter of 2017, the survey was used for a vignette study. This examined the establishment response to an increase or decrease in the national minimum wage (see Bossler/Oberfichtner/Schnabel, 2020). The two associated questionnaire variants are documented in the respective code plans. The corresponding data are part of the data set and the level of the specified minimum wage, which varies across the questionnaires, can be taken from the variable `q217_mindestlohn`.

4 Usage notes for users of the data set

4.1 Structure of the data set

Every wave's data are stored in a separate Stata format data file. It contains the variables of module 1, module 2 and module 3 and, since 2005, the data from the follow-up telephone interviews in the three subsequent quarters. Hence, the files do not contain the four quarters of a calendar year, but instead the data of the 4th quarter plus the subsequent follow-up interviews. This division is useful as the participants of the follow-up interviews are recruited from the participants of the main survey in the 4th quarter of the respective original year. This creates an unbalanced panel with four data points respectively six data points in the 2019 survey wave.

With some exceptions (Table 6), all variable names carry a prefix consisting of the number of the respective quarter as well as the year minus 2000. The establishment weight `greggew` (see Section 4.2) is therefore called `q312_greggew` in the 3rd quarter of 2012.

Table 7 shows how the quarterly data are allocated to the data files as well as the respective variable name prefixes. The meaning of the variables and their possible values can be found in the variable table as well as the codebooks for each wave on the website of the Research Data Centre under Establishment Data: IAB Job Vacancy Survey: Working Tools. To save space, the variable names in the codebooks do not carry their respective prefixes.

Table 6: Control variables that carry no prefix

Variable name	Meaning
key	Anonymized establishment identifier
jahr	Survey wave
hf	Module 1 (Main questionnaire) filled out
sf	Module 2 (Special questionnaire) filled out
zf	Module 3 (Additional questionnaire) filled out
fragebogen	Combination of `hf`, `zf` and `sf`
gebiet04	Region (West/East Germany including all of Berlin)
bl, bl_agg	Federal state (NUTS 1), federal state (aggregated)
wz03_16/wz03_28	**2000–2009:** 16/28 industries based on German classification 2003
wz08_23	**Since 2010:** 23 industries based on German classification 2008
wz08_24	**Since 2015:** 24 industries based on German classification 2008
wz*_3st/wz*_5st	Original industry classification from FEA address file (3/5 digits)[1)]

Variable name	Meaning
kreis	**Since 2007:** district (NUTS 3)[1)]
rbezirk	**Since 2005:** county (NUTS 2)[1)]
gesamtausfueller	**Since 2013:** respondent is superior unit, not single establishment
erh_form	Survey mode (paper &pencil/web or telephone interviews)
antwort	**Since 2011:** Response received by mail or online

1) Sensitive attribute: provided to external users only upon request

Table 7: File names and variable name prefixes

File name	Quarter	Variable name prefix
iabse_00	**4th quarter 2000**	q40_
iabse_01	**4th quarter 2001**	q41_
iabse_02	**4th quarter 2002**	q42_
iabse_03	**4th quarter 2003**	q43_
iabse_04	**4th quarter 2004**	q44_
iabse_05	**4th quarter 2005** (main survey)	q45_
	1st quarter 2006 (subsequent follow-up interviews)	q16_
	2nd quarter 2006 (subsequent follow-up interviews)	q26_
	3rd quarter 2006 (subsequent follow-up interviews)	q36_
	(…)	
iabse_18	**4th quarter 2018** (main survey)	q418_
	1st quarter 2019 (subsequent follow-up interviews)	q119_
	2nd quarter 2019 (subsequent follow-up interviews)	q219_
	3rd quarter 2019 (subsequent follow-up interviews)	q319_
iabse_19	**4th quarter 2019** (main survey)	q419_
	1st quarter 2020 (subsequent follow-up interviews)	q120_
	2nd quarter 2020 (subsequent follow-up interviews)	q220_
	3rd quarter 2020 (subsequent follow-up interviews)	q320_
	4th quarter 2020 (subsequent follow-up interviews – Additional)	q420_
	1st quarter 2021 (subsequent follow-up interviews – Additional)	q121_

4.2 Using the sampling weights

For descriptive analyses, the sampling weights must be used to infer from the net sample to the underlying population.[1] For regression analyses, sampling weights usually are not used when all stratification variables (see Table 4) are included in the model specification (see Bossler/Geis/Stegmaier, 2019; Winship/Radbill, 1994).

The establishment weight greggew must be used for variables from module 1 as well as module 2, except for questions on the last person hired into a One-Euro-Job. For questions from module 3 on the last new hire, the hiring weight (Establishment weight multiplied with the number of new hires, since 2004 with the number of new hires subject to social security contributions) must be used. For questions on the last stopped search effort, the stopped search weight (establishment weight multiplied with the number of stopped search efforts) must be used. Table 8 summarizes

[1] When using weights in descriptive analyses, analogue unweighted results must always be provided as well. The weighted and the corresponding unweighted result tables must always be listed one directly below the other. This will speed up the check for compliance with data protection laws.

the proper weights for each module. The data from the follow-up telephone interviews in the three subsequent quarters each have their own establishment weight `greggew`, which differs from the main survey's establishment weight by carrying a different quarter-specific variable name prefix. As described in Section 3.1.1, separate extrapolation weights are available for the COVID panel.

These weights are sampling weights by nature, so that in Stata, the weight type `pweights` must be specified, requiring the use of commands that support this type (`svyset`, `svy: tabulate`, `regress`). When using the commands tabulate (without `svy:`) and `summarize`, the weight type `iweight` may be used as well; this however precludes the interpretation of standard errors, for example after chi² tests. If it becomes necessary to test for statistically significant differences between groups during descriptive analyses, the `ttest` command must not be used as it does not accept any sampling weights. Instead, a univariate regression with the group variable as the sole regressor should be used with the proper weight variable specified as a `pweights`, consulting the p-value of the group variable (see Stata example 1) for the significance of the difference between the groups.

Table 8: Correct sampling weight variable by wave and module

Modules	Correct weight variable	
Module 1 and follow-up telephone interviews		`greggew`
Telephone interviews – COVID Panel		`greggew_covid`
Last new hire	**2000-2003:** **Since 2004:**	`greggew*f10` `greggew*f11_04`
Stopped search effort		`greggew*f422`
Module 2 (except for Last person hired into a One-Euro-job)		`greggew`
Last person hired into a One-Euro-Job		`greggew*p29`

4.3 Questions with multiple responses

At several points in the questionnaire, multiple responses are possible. As it is not clear a priori whether a non-checked box means "no" or is a missing value, checked boxes are encoded as "1" while non-checked boxes are encoded as a missing value (.). In practice, all non-checked boxes of a particular question will usually be recoded into "0" ("no") if at least one box within that item battery has been checked (see Stata example 2).

A few questions allow respondents to check "Other" and fill out a free text field. The content of these free text fields is not provided to external researchers to prevent identifying any single respondent. Common answers are however recoded into a variable with the suffix `u` (German for "Umbuchung", recode), which is provided, and the meaning of which is disclosed in the respective value labels in the (test) data sets.

In case a free text answer already exists as a possible answer in the questionnaire, it is appropriately recoded. That is why for every question that allows responding with "other", each answer's variable exists twice, once with the original content, and once including such recodes. Variables "including recodes" receive the suffix `iu` (German for "inklusive Umbuchungen", including recodes). For example, if in the questionnaire on the last new hire, the only search channel chosen is "other search

channel", and the free text field contains "was recommended to us", the variables receive the values as shown in Table 9.

Table 9: Example of recoding a free text answer

Variable name	Label	Value
zf97	Search channel: via own employees/personal contacts	. (not checked)
zf9n	Other search channel	1 (checked)
zf9txt*	Other search channel: free text	„was recommended to us"
zf9u	Recoded into…	11 (personal contacts)
zf97iu	Search channel: via own employees/personal contacts (including recodes)	1 (checked or recoded into)
zf9niu	Other search channel	0 (not checked, or recoded from)

* not included in data set at the Research Data Centre

4.4 Classifications of industries and occupations

An establishment's industry is not asked for in the questionnaire but instead is taken from the address file of the Federal Employment Agency during the drawing of the sample. As Table 4 shows, this was done until 2003 according to the German Classification of Economic Activities, 1973 edition 1973 (WZ73), from 2004 to 2009 according to the 2003 edition (WZ03), and from 2010 on according to the 2008 edition (WZ08). The provided data set however contains the classification WZ03 from 2000 to 2009 with 16 and 28 classifications (from 2000 to 2003 in recoded form) and classification WZ08 from 2010 on. The Research Data Centre provides the original three/five digit codes from the address file of the Federal Employment Agency to external users only upon request.

Several parts of the questionnaire ask respondents to specify occupations. Because determining the proper occupational codes is too much to ask of respondents, the questionnaire instead inquires occupations in free text, which is coded into two classifications by the survey institute, as Table 10 shows. In a few cases, the responses to the occupation items were too general to find a proper code, requiring the definition of custom codes outside the official classification, which are listed in Table 11.

Table 10: Occupational classifications by wave

Wave	Occupational classification	Variables
Classification 1 (German system)		
2000-2011	*German Classification of Occupations 1992 (KldB92), 3-Digit*	c[1-5], bkz[1-3], f418_c[1-5], f423_c[1-5], f424c_[1-5], code, codeab, c430-c432
Since 2012	*German Classification of Occupations 2010 – First Version (KldB2010), 5-Digit*	kb10_*
Classification 2 (international system)		
2000-2004	None	
2005-2011	*International Standard Classification of Occupations, 1988 Edition (ISCO-88)*	isco*, isc1-3, f418isc*, f423isc*, f424isc*, i430-i432
Since 2012	*International Standard Classification of Occupations, 2008 Edition (ISCO-08)*	isco08_*

Table 11: Supplement of KldB2010 for responses without detailed activity description (n.o.s.)

Category	Code	Label
Technical assistant w/o detailed activity description	**99001**	Production assistant, production hand, aide
	99002	Factory specialist, technical specialist, journeyman
	99003	Technician, foreman, master craftsman
	99004	Engineer, qualified engineer
	99005	Technical assistant n.o.s.
Commercial assistant w/o detailed activity description	**99101**	Comm. aides
	99102	Businessman, fund manager
	99103	Business administrator
	99104	MBA
	99105	Commercial assistant n.o.s.
Service assistants w/o detailed activity description	**99201**	Service staff, supporter, appointment manager
	99202	Service specialist, advisor, inspector, non-teaching staff
	99203	Consultant, temporary employment agency
	99204	Senior consultant
	99205	Service assistants n.o.s.
Managers w/o detailed activity description	**99301**	(does not occur)
	99302	(does not occur)
	99303	Team manager, managing staff, site manager, department head
	99304	Director, manager, managing director
	99305	Managers n.o.s.
IT assistants w/o detailed activity description	**99401**	IT staff (requirement level 1)
	99402	IT professional
	99403	IT specialist
	99404	IT expert
	99405	IT profession, IT assistant n.o.s.
Teachers w/o detailed activity description	**99501**	School assistant
	99502	Qualified teacher
	99503	Teacher
	99504	Lecturer, teacher with a university degree
	99505	Teacher n.o.s.
Workers in training w/o detailed activity description	**99611**	Trainee
	99621	Intern/volunteer
	99691	Other workers in training
Workers w/o detailed activity description	**99901**	Helper, temporary assistant, semi-skilled employee, unskilled worker
	99902	Professional, worker, assistant
	99903	Specialist
	99904	Expert
	99905	Workers n.o.s.

4.5 BHP record linkage

Starting in the 4th quarter of 2010, all variables of the Establishment History Panel, version BHP 7520 v1, are available with the exception of the extension file on bankruptcies. The BHP record linkage data must be requested separately; the same applies to BHP's extension modules. Please refer to the data set and variable descriptions on the Research Data Centre's website. The BHP extension data sets are named as follows:

Table 12: Filenames of the BHP administrative data sets for record linkage

Filename	Content
iabse_0019_v1_bhp_7520_m06_v1_1975 … **iabse_0019_v1_bhp_7520_m06_v1_2019**	BHP core data set
iabse_0019_v1_bhp_7520_m06_inflow_v1 **iabse_0019_v1_bhp_7520_m06_outflow_v1**	BHP extension „worker flows“
iabse_0019_v1_bhp_7520_m06_entry_v1 **iabse_0019_v1_bhp_7520_m06_exit_v1**	BHP extension „foundations/closures“
iabse_0019_v1_bhp_7520_m06_wgen_v1	BHP extension „generated consistent industry codes”

Please note the following points when using the data:

1. The survey and administrative data must be linked using the variable `key`, which is unique across all waves. The original establishment identifier `betnr` is not included in the BHP version provided for linkage with the IAB Job Vacancy Survey.
2. Record linkage was not possible at all for a small number of establishments that have responded for a larger unit than the establishment. For the same reason, a few keys have record-linked data that cannot be interpreted sensibly; they can be identified by the variable `gesamtausfueller` having a value of “1”.
3. Since the 4th quarter of 2010, the privacy policy sheet that was sent to all employers who were asked to participate in the survey has been worded such that participating in the survey implies consenting to record linkage. This means that record linkage is possible for all establishments except those mentioned in Number 2.
4. The Variable `jahr` in the BHP data sets describes the year at which the BHP variables have been measured. The survey wave in which the respective establishment has responded can be identified using the additional variable `SE_jahr`.
5. Large deviations in establishment size between survey and administrative data are expected in a few cases, for example because respondents provided information on the entire company instead of just the establishment, without stating such.
6. If the sensitive variables `kreis`/`ao_kreis` or `wz08_3st`/`wz08_5st`/`ieb_w08` have been requested, please note that when comparing survey and administrative data, those particular variables are measured at the time of sampling, which is one year before the field phase (see Table 5). In addition, the district information in the survey data reflects the district numbers that were valid at the time of sampling, while the district numbers in the administrative data are the ones that are valid as of 31st December 2020.

The information from the BHP data may differ from the data from the IAB job vacancy survey. All characteristics listed below will be provided.

Basic variables

- Serial Number (key)
- Wave of the IAB Job Vacancy Survey (SE_jahr)
- Year (jahr)
- Year of first appearance (grd_jahr)
- Year of last appearance (lzt_jahr)

- Place of work: federal state (Bundesland) (ao_bula)
- Total number of employees (az_ges)
- Total number of full-time employees (az_ges_vz)
- Number of marginal part-time workers (az_gf)
- Mean imputed wage (gross average daily wage) for all full-time employees (te_imp_mw)

Thematic variable blocks

- General employment structure (az_ges az_f az_reg az_gf az_azubi az_atz az_vz az_tz az_f_vz az_f_tz az_reg_vz)
- Structure of employees by educational and vocational qualifications (az_gq az_mq az_hq az_gq_vz az_mq_vz az_hq_vz)
- Employee age structure (az_15_19 az_15_19_vz az_20_24 az_20_24_vz az_25_29 az_25_29_vz az_30_34 az_30_34_vz az_35_39 az_35_39_vz az_40_44 az_40_44_vz az_45_49 az_45_49_vz az_50_54 az_50_54_vz az_55_59 az_55_59_vz az_60_64 az_60_64_vz az_ab65 az_ab65_vz alter_mw alter_mw_vz)
- Research and development activities (az_ingnat)
- Number of employees with non-standard job types (az_leih az_bfr)
- Structure of employees by nationality (az_d az_d_vz az_eu)
- Structure of employees by Blossfeld occupational group (az_bf_agr az_bf_emb az_bf_edi az_bf_evb az_bf_qmb az_bf_qdi az_bf_qvb az_bf_tec az_bf_semi az_bf_ing az_bf_prof az_bf_man)
- Structure of employees by level of requirement (az_niv1 az_niv2 az_niv3 az_niv4)
- Wage structure of full-time employees (az_zens te_med te_p25 te_p75 te_imp_mw te_imp_med te_imp_p25 te_imp_p75 te_imp_mw_f te_imp_med_f te_imp_med_m te_imp_mw_gq te_imp_med_gq te_imp_mw_mq te_imp_med_mq te_imp_mw_hq te_imp_med_hq te_imp_med_uq te_imp_mw_d te_imp_med_d te_imp_med_a)
- Extension file – Worker flows (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_vz_f ein_tz_f ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_bg_edi ein_15_19 ein_20_24 ein_25_29 ein_30_34 ein_35_39 ein_40_44 ein_45_49 ein_50_54 ein_55_59 ein_60_64 ein_ab65 aus_ges aus_f aus_reg aus_gf aus_azubi aus_vz aus_tz aus_vz_f aus_tz_f aus_reg_vz aus_temp aus_bw aus_bw_f aus_15_19 aus_20_24 aus_25_29 aus_30_34 aus_35_39 aus_40_44 aus_45_49 aus_50_54 aus_55_59 aus_60_64 aus_ab65 aus_senio_1 aus_senio_2 aus_senio_3)
- Extension file – Entry and exit (eintritt besch inflow betnr_vor besch_vor status_vor austritt besch outflow betnr_nach besch_nach status_nach)

Sensitive variables BHP

Additionally, the following sensitive variables from the BHP can be requested individually:

- Place of work: district code (ao_kreis)
- Classification of Economic Activities 1973, 3-digit code (w73_3)
- Classification of Economic Activities 1993, 3-digit code (w93_3)
- Classification of Economic Activities 1993, 5-digit code (w93_5)
- Classification of Economic Activities 2003, 3-digit code (w03_3)
- Classification of Economic Activities 2003, 5-digit code (w03_5)

- Classification of Economic Activities 2008, 3-digit code (w08_3)
- Classification of Economic Activities 2008, 5-digit code (w08_5)
- Time-consistent industry codes, 3-digit (w73_3_gen, w93_3_gen, w08_3_gen)

4.6 Stata examples

4.6.1 Time series of vacancies

```
/* Display a time series of vacancies from the year 2000 until the most recent
       quarter, with confidence intervals */

capture log close
log using ${log}/05a_StataBeispiel1_en,text replace
set more off

// Loop over every survey wave
forvalues Wave=2000/2019 {
       local WaveShort =strofreal(`Wave'-2000,"%02.0f")
       quietly use ${orig}/iabse_`WaveShort',clear
       quietly label language en

       /*Exclusion of experimental data */
       if `Wave' ==2019 {
              drop if experiment2019 ==1 & jahr == 2019
       }

       // Identify number of quartals
       if `Wave' <2005 {
              local Quartale 4
       }
       else {
              local Quartale 4 1 2 3
       }

       // Loop over every quarter within this survey wave
       foreach Quartal of local Quartale {
              // Form quarter-specific variable prefix
              if `Quartal' ==4 {
                     local Year =`Wave'
              }
              else {
                     local Year =`Wave' +1
              }
              local YearShort =`Year' -2000
              local Prefix q`Quartal'`YearShort'

              quietly {
              // Missing value means zero vacancies
              replace `Prefix'_f20 =0 if `Prefix'_f20 ==.
              replace `Prefix'_f40 =0 if `Prefix'_f40 ==.

              // Sum of positions to be filled immediately and later
              generate `Prefix'_Vacancies =`Prefix'_f20 +`Prefix'_f40

              // Compute extrapolated number of vacancies including 95%
              // confidence intervals, by computing the weighted mean
              // multiplied with the weighted number of firms
              mean `Prefix'_Vacancies [pweight =`Prefix'_greggew]
              matrix Result =r(table)
              quietly summarize `Prefix'_greggew
              local Firms =r(sum)
              local b        =Result[1,1] *`Firms'
              local ll =Result[5,1] *`Firms'
              local ul =Result[6,1] *`Firms'
              } // quietly

              // Display one list of results
              display as text "Vacancies in quarter " as result `Quartal' as text "
of " as result `Year' as text ": "_
              display as result %7.0f `b' as text ", 95% confidence interval: [" as
result %7.0f `ll' "; " %7.0f `ul' as text "]"
       }
}
log close
```

4.6.2 Tabulating a multi-response question

```
/* Tabulate the share of difficult hirings and the reasons for the difficulty */

capture log close
log using ${log}/05b_StataBeispiel2_en,text replace
set more off

use ${orig}/iabse_19,clear

/*Exclusion of experimental data */
drop if experiment2019 ==1

label language en

/* Create hiring weight */
generate gregneu = q419_greggew*q419_f11_04

/* Weight for all subsequent tabulations */
svyset [pweight=gregneu]

/* Share of difficult hirings */
svy: tabulate q419_zf8

/* If a reason was checked, recode all non-checked reasons as “no”,
 therwise keep them on “missing” */
generate byte answered_zf8reason = q419_zf8aiu==1 | q419_zf8b1iu==1 | ///
q419_zf8c1iu==1 | q419_zf8eiu==1 | q419_zf8niu==1
foreach var of varlist q419_zf8aiu q419_zf8b1iu q419_zf8c1iu q419_zf8eiu q419_zf8niu
{
       replace `var'=0 if `var'==. & answered_zf8reason==1
}

/* Make sure that a reason was checked only for difficult hirings */
assert answered_zf8reason == 0 if q419_zf8 != 1

/* Print reasons one after the other */
foreach var of varlist q419_zf8aiu q419_zf8b1iu q419_zf8c1iu q419_zf8eiu q419_zf8niu
{
       svy: tabulate `var' if answered_zf8reason
}
log close
```

References

Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin; Olbrich, Lukas; Schwanhäuser, Silvia (2022): Revision and new data quality concept due to deviant interviewer behavior in the IAB Job Vacancy Survey. (FDZ-Methodenreport, 05/2022), Nürnberg, 19 S.

Bossler, Mario; Kubis, Alexander; Küfner, Benjamin; Popp, Martin (2021): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Welle 2000 bis 2018 mit Folgequartalen von 2006 bis 2019. (FDZ-Datenreport, 09/2021 (de)), Nürnberg, 19 S.

Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin (2020a): IAB-Stellenerhebung im ersten Quartal 2020: Mit dem Corona-Shutdown ging zuerst die Zahl der offenen Stellen zurück. (IAB-Kurzbericht, 12/2020), Nürnberg, 12 S.

Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin; Lochner, Benjamin (2020b). The IAB Job Vacancy Survey: design and research potential. Journal for Labour Market Research, Jg. 54, H. 1, S. 1-12.

Bossler, Mario; Oberfichtner, Michael; Schnabel, Claus (2020): Employment adjustments following rises and reductions in minimum wages: New insights from a survey experiment. In: Labour, Vol. 34, No. 3, S. 323-346.

Bossler, Mario; Geis, Gregor; Stegmaier, Jens (2018): Comparing survey data with an official administrative population: Assessing sample-selectivity in the IAB Establishment Panel. In: Quality & Quantity, Jg. 52, H. 2, S. 899-920.

Bundesagentur für Arbeit (2020): Steckbrief Betriebsnummervergabe. https://www.informationsportal.de/wp-content/uploads/Steckbrief-Betriebsnummer.pdf [2021-06-08]

Brenzel, Hanna; Czepek, Judith; Kiesl, Hans; Kriechel, Ben; Kubis, Alexander; Moczall, Andreas; Rebien, Martina; Röttger, Christof; Szameitat, Jörg; Warning, Anja; Weber, Enzo (2016): Revision der IAB-Stellenerhebung. Hintergründe, Methode und Ergebnisse. (IAB-Forschungsbericht, 04/2016), Nürnberg

Gürtzgen, Nicole; Popp, Martin (2022): Betriebliche Vorbehalte gegenüber Langzeitarbeitslosen sinken leicht in Krisenzeiten. Unpublished Manuscript. Forthcoming als IAB-Kurzbericht (prospective 17/2022. Nürnberg.

Gürtzgen, Nicole; Kubis, Alexander (2021): Über ein Viertel der mit Corona-Hilfen unterstützten Betriebe sieht dennoch eine Insolvenzgefahr (Serie "Corona-Krise: Folgen für den Arbeitsmarkt"). In: IAB-Forum, 06.04.2021, o. Sz.

Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin (2020): Großbetriebe haben während des Covid-19-Shutdowns seltener als kleine Betriebe Beschäftigte entlassen (Serie "Corona-Krise: Folgen für den Arbeitsmarkt"). In: IAB-Forum, 03.07.2020, o. Sz.

Hutter, Christian (2020): A new indicator for nowcasting employment subject to social security contributions in Germany. In: Journal for labour market research, Vol. 54, No. 1, Art. 4, S. 1-10.

Küfner, Benjamin; Zins, Stefan; Sakshaug, Joseph (2021): More Clarification, Less Item Nonresponse in Establishment Surveys? A Split-Ballot Experiment. In: Survey research methods, Vol. 15, No. 2, S. 195-206.

Winship, Christopher; Radbill, Larry (1994): Sampling Weights and Regression Analysis. In: Sociological Methods & Research, Jg. 23, H. 2, S. 230–257.

Imprint

FDZ-Datenreport 08|2022 EN

Date of publication
15 September 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/DR_08-22_EN.pdf

Documentation version
IABSE0019_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2208.en.v1

Dataset version
IAB SE 0019 v1, DOI: 10.5164/IAB.IABSE0019.de.en.v1

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

Website
https://fdz.iab.de/

Corresponding author
Alexander Kubis
Phone: +49 911 179-8978
Email: alexander.kubis@iab.de