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

09|2021 EN The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, waves 2000 to 2018 and subsequent quarters 2006 to 2019

Mario Bossler, 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 2018 and subsequent quarters 2006 to 2019

Mario Bossler (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.

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), Germany's public employment service. The main questionnaire enquires information about the number and structure of vacancies, future labour demand, the current economic situation and the expected development of participating establishments. The additional questionnaire enquires information about the last new hiring and the last case of a failed recruitment effort. The special questionnaire enquires employer attitudes and firm use of current labour market instruments. 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 und Einstellungsprozessen in Deutschland. Sie ermittelt die Gesamtzahl aller offenen Stellen am Arbeitsmarkt, einschließlich jener Stellen, die nicht der Bundesagentur für Arbeit (BA) gemeldet werden. Der Hauptfragebogen enthält Informationen zu Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. Der Zusatzfragebogen enthält Fragen zum letzten Fall einer Neueinstellung und zum letzten Fall eines gescheiterten Rekrutierungsversuchs. Der Sonderfragebogen erfragt die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente. 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., 2020), the fourth quarter of 2018 and the three subsequent quarters of 2019 are included. In the fourth quarter of 2018, special questions collected data on employment opportunities of unemployed persons (see Pohlan/Rothe, 2020) and on retirement and labour market (see Westermeier/Wolf, 2020). In the second quarter of 2019, establishments were surveyed on Brexit (see Bossler et al., 2019) and the "Participation Opportunities Act".

In comparison to previous versions, a correction was made to the variables *f4i* and *f6i*. In the quarterly telephone surveys, the variables *f4i* and *f6i*, which record the imputed number of employees (subject to social security contributions), were renamed to *f4imp* and *f6imp*, respectively. For the written survey in the fourth quarter, the variable *f4i* was not renamed, as this variable records the number of employees with master craftsmen and technician qualifications.

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, 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 III.2019
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 written and quarterly survey by phone
File format/size	STATA; all data sets together ca. 241 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: "This study uses the German Job Vacancy Survey of the IAB, Wave(s) YYYY - YYYY). 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 subsequently remote data access." DOI: 10.5164/IAB.IABSE0018.de.en.v1 Data documentation: - Bossler, Mario; Kubis, Alexander; Küfner, Benjamin; Popp, Martin (2021): The IAB Job Vacancy Survey. Establishment survey on labour demand and recruitment processes Waves 2000 to 2018 and subsequent quarters 2006 to 2019. FDZ-Datenreport, 09/2021 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2109.en.v1 - Bossler, Mario; Kubis, Alexander; Küfner, Benjamin; Popp, Martin (2021): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Wellen 2000 bis 2018 mit Folgequartalen 2006 bis 2019. FDZ-Datenreport, 09/2021 (de), Nürnberg DOI: 10.5164/IAB.FDZD.2709.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 service. 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.

The survey has been conducted since 1989 in the fourth quarter of every year using a written multi-part questionnaire. Since 2002 (with interruptions), establishments may participate online via internet as well. This web questionnaire is identical to the written questionnaire. The **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. The **additional questionnaire** enquires 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 engagement of job placement services, search and recruitment durations, the number of applicants, hiring difficulties as well as any compromises made. The additional questionnaire is only answered by establishments who have hired a new employee in the past year or have tried to hire for an open position.

Main and additional questionnaire are basically the same from wave to wave except for editing changes and possible shifts in emphasis. By contrast, the **special questionnaire** changes from wave to wave. Since 2000, it enquires employer attitudes and firm use of current labour market instruments. From 2005 to 2013, it mostly focused on the workfare scheme One-Euro-Jobs and for that reason only was provided to establishments in the public sector which provided the vast majority of One-Euro-Jobs.

Since 2005, there are short follow-up surveys by phone on the core questions among participants of the main survey. These **quarterly surveys** complement the written main survey in the respective fourth quarter. They enquire and update the most important data points in the main questionnaire, including the number of employees, the assessment of current workforce developments and the number of vacancies.

In 2017, the survey in the second quarter was used for a vignette survey to study firm's reaction on an increase or decrease of the mandatory national minimum wage. Both variations of the questionnaire are documented in the corresponding codebooks.

The Research Data Centre of the Federal Employment Agency offers the survey waves from 2000 onwards to external researchers. It includes the data from all questionnaire parts including the quarterly surveys. Since September 2015, all documentation including variable and value labels, data reports and code books, is available in English as well.

2.3 Population and sample of the main questionnaire

The population of the main survey in the fourth 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 2 summarizes the stratification variables. During the waves 2005 to 2013, an additional sample was drawn for the public sector in order the better survey about 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 2000–2018

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 3.

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 2000–2018

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%
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%

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, 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 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 2) 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 survey

In each subsequent quarter, establishments are drawn from the participants of the written main questionnaire of the fourth quarter. The intended number of establishments is 9,000 per quarter. The net sample of the written main survey therefore is the gross sample of follow-up surveys of the three subsequent quarters.

Every one of the three subsequent quarters has its own establishment weight comprised of the establishment weight of the main survey in the fourth quarter of the original year and the result of another non-response model for the respective subsequent quarter.

3 Usage notes for users of the data set

3.1 Structure of the data set

Every wave's data are stored in a separate STATA format data file. It contains the variables of the main, additional and special questionnaires and, since 2005, the data from the follow-up telephone surveys in the three subsequent quarters. Hence, the files do not contain the four quarters of a calendar year, but instead the data of the fourth quarter plus the three quarters of the subsequent year. This division is useful as the participants of the follow-up surveys in the first to third quarter of the following year are recruited from the participants of the main survey in the fourth quarter of the respective original year. This creates an unbalanced panel with four data points in every wave of the survey.

Except for a few exceptions (Table 4), 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 3.2) is therefore called `q312_greggew` in the third quarter of 2012.

Table 5 shows how the quarters are divided among 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 of every 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	Main questionnaire was answered
zf	Additional questionnaire was answered
sf	Special questionnaire was answered
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)]
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 (written or telephone survey)
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** (written main questionnaire)	q45_
	1st quarter 2006 (subsequent phone survey)	q16_
	2nd quarter 2006 (subsequent phone survey)	q26_
	3rd quarter 2006 (subsequent phone survey)	q36_
	(…)	
iabse_18	**4th quarter 2018** (written main questionnaire)	q418_
	1st quarter 2019 (subsequent phone survey)	q119_
	2nd quarter 2019 (subsequent phone survey)	q219_
	3rd quarter 2019 (subsequent phone survey)	q319_

3.2 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 2) are included in the model specification (see Bossler/Geis/Stegmaier, 2019; Winship/Radbill, 1994).

The establishment weight `greggew` must be used for variables from the main questionnaire as well as the special questionnaire, except for questions on the last person hired into a One-Euro-Job. For questions from the additional questionnaire 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 6 summarizes the proper weights for each questionnaire. The data from the follow-up telephone surveys in the three subsequent quarters each use their own establishment weight `greggew`, which differs from the main survey's establishment weight by having a different quarter-specific variable name prefix.

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 questionnaire

Questionnaire	Correct weight variable	
Main questionnaire, follow-up telephone surveys		`Greggew`
Last new hire	**2000-2003:** **Since 2004:**	`greggew*f10` `greggew*f11_04`
Stopped search effort		`greggew*f422`
Special questionnaire (except for Last person hired into a One-Euro-job)		`Greggew`
Last person hired into a One-Euro-Job		`greggew*p29`

[1] When using weights in descriptive analyses, analogue unweighted results must always be specified 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.

3.3 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 7.

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

3.4 Classifications of industries and occupations

An establishment’s industry is not enquired in the questionnaire but instead is taken from the address file of the Federal Employment Agency during the drawing of the sample. As Table 2 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 institute that conducts the survey, as Table 8 shows. In a few cases, the responses were too general to find a proper code, requiring the definition of custom codes outside the official classification, which are listed in Table 9.

Table 10: Occupational classifications by wave

Wave	Occupational classification	Variables
Classification 1 (German system)		
2000-2011	*Klassifikation der Berufe, Ausgabe 1992 (KldB92), 3-Steller*	`c[1-5], bkz[1-3], f418_c[1-5], f423_c[1-5], f424c_[1-5], code, codeab, c430-c432`
Since 2012	*Klassifikation der Berufe, Ausgabe 2010 (KldB10), 5-Steller*	`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.

3.5 BHP record linkage

Starting in the fourth quarter of 2010, all variables of the Establishment History Panel, version BHP 7519 v2, 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_0018_v1_bhp_7519_m06_v2_1975 … **iabse_0018_v1_bhp_7519_m06_v2_2019**	BHP core data set
iabse_0018_v1_bhp_7519_m06_inflow_v2 **iabse_0018_v1_bhp_7519_m06_outflow_v2**	BHP extension „worker flows“
iabse_0018_v1_bhp_7519_m06_entry_v2 **iabse_0018_v1_bhp_7519_m06_exit_v2**	BHP extension „foundations/closures“
iabse_0018_v1_bhp_7519_m06_wgen_v2	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. The privacy policy sheet that was sent to all employers who were asked to participate in the survey has been worded since the fourth quarter of 2010 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 when the survey sample is drawn, which is one year before the field phase (see Table 3). In addition, the district information in the survey data reflects the district numbers that were valid when the sample was drawn, while the district numbers in the administrative data are the ones that are *currently* valid.

3.6 STATA examples

3.6.1 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=2000/2018 {
   local WaveShort =strofreal(`Wave'-2000,"%02.0f")
   quietly use ${orig}/iabse_`WaveShort',clear
   quietly label language en

   // Anzahl der Quartale in dieser Wave ermitteln
   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
```

3.6.2 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_18,clear
label language en

/* Create hiring weight */
generate gregneu = q418_greggew*q418_f11_04

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

/* Share of difficult hirings */
svy: tabulate q418_zf8

/* If a reason was checked, recode all non-checked reasons as “no”,
 Otherwise keep them on “missing” */
generate byte answered_zf8reason = q418_zf8aiu==1 | q418_zf8b1iu==1 | ///
q418_zf8c1iu==1 | q418_zf8eiu==1 | q418_zf8niu==1
foreach var of varlist q418_zf8aiu q418_zf8b1iu q418_zf8c1iu q418_zf8eiu
q417_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 q418_zf8 != 1

/* Print reasons one after the other */
foreach var of varlist q418_zf8aiu q418_zf8b1iu q418_zf8c1iu q418_zf8eiu
q417_zf8niu {
       svy: tabulate `var' if answered_zf8reason
}
log close
```

References

Bossler, Mario; Gartner, Hermann; Kubis, Alexander; Küfner, Benjamin (2020): The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, Waves 2000 to 2017 and subsequent quarters 2006 to 2018. (FDZ-Datenreport, 06/2020 (en)), Nürnberg, 18 S.

Bossler, M., Gürtzgen, N., Kubis, A., Küfner, B., & Lochner, B. (2020b). The IAB Job Vacancy Survey: design and research potential. Journal for Labour Market Research, Vol. 54, No. 1, pp.. 1-12.

Bossler, M., Fuchs, J., Kubis, A., & Schneider, L. (2019). Mögliche Brexit-Folgen für den deutschen Arbeitsmarkt. In: Wirtschaftsdienst, Vol. 99, No. 10, S. 687-692.

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, Vol. 52, No. 2, pp.. 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 of the IAB job vacancy survey * backgrounds, methods and results. (IAB-Forschungsbericht, 04/2016 (en)), Nürnberg, 61 S.

Pohlan, Laura; Rothe, Thomas (2020): Personalrekrutierung von Beschäftigten, Kurz- und Langzeitarbeitslosen: Unterschiede bei Besetzungswegen und Beschäftigungsqualität. (IAB-Kurzbericht, 06/2020), Nürnberg, 12 S.

Westermeier, Christian; Wolf, Mario (2020): IAB-Stellenerhebung: Betriebe wollten 2018 deutlich mehr rentenberechtigte Mitarbeiter halten als 2015. (IAB-Kurzbericht, 18/2020), Nürnberg, 12 S.

Winship, Christopher; Radbill, Larry (1994): Sampling Weights and Regression Analysis. In: Sociological Methods & Research, Vol. 23, No. 2, pp. 230–257.

Imprint

FDZ-Datenreport 09|2021 EN

Date of publication
25 August 2021

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
http://doku.iab.de/fdz/reporte/2021/DR_09-21_EN.pdf

Documentation version
IABSE0018_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2109.en.v1

Dataset version
IAB SE 0018 v1, DOI: 10.5164/IAB.IABSE0018.de.en.v1

All publications in the series "FDZ-Datenreport" can be downloaded from
https://fdz.iab.de/en/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx

Website
http://fdz.iab.de

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