FORSCHUNGSDATENZENTRUM
der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

14|2020 DE Stichprobe Integrierter Employer-Employee Daten (SIEED) 1975-2018

Lisa Schmidtlein, Stefan Seth, Philipp vom Berge

Bundesagentur für Arbeit

Stichprobe Integrierter Employer-Employee Daten (SIEED) 1975-2018

Lisa Schmidtlein (IAB), Stefan Seth (IAB), Philipp vom Berge (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.

Inhaltsverzeichnis

1 Einleitung und Kurzbeschreibung 6

1.1 Einleitung 6

1.2 Datennutzung 6

1.2.1 Datenzugang 6

1.2.2 Datenhaltung 7

1.3 Steckbrief 8

1.4 Merkmalsliste 9

1.5 Mengengerüst 13

2 Datenquelle 13

3 Datenaufbereitung und Stichproben-ziehung 14

3.1 Bereinigungen und Aufbereitungen 14

3.2 Episodensplitting 15

3.3 Stichprobenziehung 16

3.4 Fehlende Werte 17

4 Datenqualität und Probleme 17

4.1 Gesamte IEB 17

4.2 Beschäftigtenhistorik 18

5 Variablenbeschreibung 21

5.1 Identifikatoren 21

5.1.1 Systemfreie Personennummer (persnr) 21

5.1.2 Systemfreie Betriebsnummer (betnr) 21

5.2 Generierte technische Merkmale 22

5.2.1 Identifikator Panelbetrieb (betnr_pan) 22

5.2.2 Satzzähler pro Konto (spell) 22

5.2.3 Jahr (jahr) 23

5.3 Gültigkeitszeitraum 23

5.3.1 Beginndatum Originalbeobachtung (begorig) 23

5.3.2 Enddatum Originalbeobachtung (endorig) 23

5.3.3 Beginndatum Episode (begepi) 24

5.3.4 Enddatum Episode (endepi) 24

5.4 Informationen zur Person 24

5.4.1 Geschlecht (frau) 24

5.4.2 Geburtsjahr (gebjahr) 24

5.4.3 Geburtsmonat (gebmon) 25

5.4.4 Staatsangehörigkeit (nation) 25

5.4.5 Staatsangehörigkeit vergröbert (nation_gr) 25
5.4.6 Ausbildung (ausbildung) 25
5.4.7 Ausbildung (imputiert) (ausbildung_imp) 26
5.4.8 Schulausbildung (schule) 27
5.5 Informationen zu Beschäftigung 27
5.5.1 Tagesentgelt/ täglicher Leistungssatz (tentgelt) 27
5.5.2 Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf) 28
5.5.3 Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3) 29
5.5.4 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4) 29
5.5.5 Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau) 30
5.5.6 Teilzeit (teilzeit) 31
5.5.7 Stellung im Beruf und Arbeitszeit (stib) 31
5.5.8 Erwerbsstatus (erwstat) 32
5.5.9 Gleitzone (gleitz) 32
5.5.10 Leiharbeit (leih) 33
5.5.11 Befristung (befrist) 33
5.5.12 Abmeldegrund / Abgabegrund / Beendigungsgrund (grund) 33
5.5.13 Dauer der Arbeitslosigkeit (alo_dau) 34
5.6 Ortsangaben 34
5.6.1 Wohnort - Kreis (wo_kreis) 34
5.6.2 Wohnort - Bundesland (wo_bula) 35
5.6.3 Wohnort - Arbeitsagentur (wo_aa) 35
5.6.4 Wohnort - Regionaldirektion (wo_rd) 36
5.7 Betriebsmerkmale 36
5.7.1 WZ73 3-Steller (w73_3) 36
5.7.2 WZ93 5-Steller (w93_5) 36
5.7.3 WZ93 3-Steller (w93_3) 37
5.7.4 WZ03 5-Steller (w03_5) 37
5.7.5 WZ03 3-Steller (w03_3) 38
5.7.6 WZ08 5-Steller (w08_5) 38
5.7.7 WZ08 3-Steller (w08_3) 39
5.7.8 w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen) 39
5.7.9 Art der Vervollständigung w73_3 (group_w73_3) 39
5.7.10 w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen) 40
5.7.11 Art der Vervollständigung w93_3 (group_w93_3) 40
5.7.12 w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen) 40
5.7.13 Art der Vervollständigung w08_3 (group_w08_3) 41
5.7.14 Jahr des 1. Auftretens der Betriebsnummer (grd_jahr) 41
5.7.15 Erstes Auftreten Betriebsnummer (grd_dat) 41
5.7.16 Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) 42
5.7.17 Letztes Auftreten Betriebsnummer (lzt_dat) 42
5.7.18 Anzahl Beschäftigte gesamt (az_ges) 43
5.7.19 Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz) 43

5.7.20 Anzahl geringfügig Beschäftigte (az_gf) 43
5.7.21 Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw) 43
5.7.22 Arbeitsort - Kreis (ao_kreis) 44
5.7.23 Arbeitsort - Bundesland (ao_bula) 44

6 Literatur 45

7 Anhang 47

7.1 Auszählungen 47

7.2 Abkürzungsverzeichnis 47

Zusammenfassung

Dieser Datenreport beschreibt die Stichprobe Integrierter Employer-Employee Daten (SIEED) 1975 - 2018.

Abstract

This data report describes the Sample of Integrated Employer-Employee Data (SIEED) 1975 - 2018.

Keywords

German administrative micro data, labour market data, data manual

Danksagung

Für ihre Mitarbeit und Unterstützung danken wir den Kollegen des Forschungsdatenzentrums (FDZ) sowie des Geschäftsbereichs Daten- und IT-Management (DIM) des Instituts für Arbeitsmarkt- und Berufsforschung. Einzelne Passagen wurden aus IAB-internen Datendokumentationen von IAB-DIM übernommen.

Das vorliegende Datenprodukt wurde als Teil des Projekts „Custom Shaped Administrative Data for the Analysis of Labor Markets (CADAL)“ erstellt. Das Projekt wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Schwerpunktprogramms 1764 „The German Labor Market in a Globalized World: Challenges through Trade, Technology, and Demographics“ finanziell unterstützt (BE 6283/4-1; MO 523/9-1).

Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich. Nähere Informationen dazu auf der Internetseite https://fdz.iab.de/.

1 Einleitung und Kurzbeschreibung

1.1 Einleitung

Die Stichprobe Integrierter Employer-Employee Daten (SIEED) ist eine Stichprobe aus der Grundgesamtheit der Integrierten Erwerbsbiografien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Die IEB enthalten Daten zu allen Personen, die im Beobachtungszeitraum mindestens einmal einen der folgenden Zustände aufweisen:

- sozialversicherungspflichtige Beschäftigung in Deutschland (erfasst ab 1975),
- geringfügige Beschäftigung in Deutschland (erfasst ab 1999),
- Bezug von Leistungen nach dem Rechtskreis SGB III (erfasst ab 1975) oder SGB II (erfasst ab 2005),
- bei der Bundesagentur für Arbeit (BA) bzw. den Grundsicherungsträgern als arbeitsuchend gemeldet (erfasst ab 1997),
- Teilnahme an arbeitsmarktpolitischer Maßnahme (erfasst ab 2000).

Diese aus unterschiedlichen Datenquellen stammenden Informationen werden in den IEB zusammengeführt und tagesgenau abgebildet. Die SIEED enthält ausschließlich Daten aus der Beschäftigtenhistorik (BeH), welche Ursprung der Informationen zu sozialversicherungspflichtiger und geringfügiger Beschäftigung ist. Weitere in der IEB enthaltene Datenquellen sind in der SIEED nicht enthalten.

Die SIEED wird im Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB produziert. Der Datensatz umfasst die Erwerbsverläufe von 5.567.883 Individuen in 5.248.850 Betrieben, deren Erwerbsbiografien in insgesamt 172.935.233 Datenzeilen dokumentiert sind. Der vorliegende Datenreport beschreibt die Merkmale der schwach anonymisierten SIEED, welche weitgehend die Originaldaten, d.h. nicht vergröberte Daten, beinhaltet. Einige Merkmale sind allerdings als besonders sensibel eingestuft und werden nur auf gesonderten Antrag (siehe Kapitel 1.2) weitergegeben.

Dieser Datenreport gliedert sich wie folgt: Neben der Einleitung enthält Kapitel 1 unter anderem Informationen zum Datenzugang sowie eine Kurzbeschreibung der Daten, das Mengengerüst und eine Merkmalsliste. Eine Beschreibung der einzelnen Datenquellen findet sich in Kapitel 2. Die Datenaufbereitung sowie die Datenqualität werden in den Kapiteln 3 und 4 diskutiert, während die einzelnen Merkmale in Kapitel 5 dargestellt werden.

1.2 Datennutzung

1.2.1 Datenzugang

Die SIEED-Daten können ausschließlich im Rahmen von Gastaufenthalten am FDZ und anschließender Datenfernverarbeitung ausgewertet werden.

Für die Datennutzung ist zunächst ein Antrag beim FDZ einzureichen. Über die Genehmigung des Forschungsprojekts entscheidet das FDZ im Auftrag und ggf. unter Abstimmung mit dem Bundesministerium für Arbeit und Soziales (BMAS). Nach erfolgter Genehmigung wird ein Nutzungsver-

trag mit der wissenschaftlichen Einrichtung der Forschenden geschlossen. Details zur Beantragung des Datensatzes und Datenverarbeitungsmöglichkeiten können der FDZ-Website entnommen werden.

1.2.2 Datenhaltung

Die SIEED-Daten, die sowohl deutsche als auch englische Labels enthalten,[1] sind modular gegliedert und in mehreren Dateien abgelegt. Ein Modul, das im weiteren Verlauf Personendatei genannt wird, enthält Identifikatoren (systemfreie Personen- und Betriebsnummer), persönliche Merkmale, Informationen zu Beschäftigung, Merkmale zum Wohnort sowie technische Merkmale. Ein zweites Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr und Merkmale zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni. Für Ostdeutschland liegen die Betriebsdaten erst ab 1992 vor.

Somit sind Betriebs- und Personenmerkmale organisatorisch voneinander getrennt, was die Struktur der Daten deutlich macht und Speicherplatz spart (vgl. Abbildung 1). Die Verknüpfung von Basisbetriebs- und Personendatei erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Die in der Merkmalsliste (vgl. S. 9ff) mit einem „*“ gekennzeichneten Variablen sind in der Basisbetriebsdatei enthalten. Zu beachten ist, dass das für die Verknüpfung beider Module notwendige Merkmal „Betriebsnummer“ sowohl in der Personen- als auch in der Betriebsdatei vorhanden ist. Alle Dateinamen in der SIEED 7518 sind ebenfalls Abbildung 1 zu entnehmen.

Bestimmte Merkmale, die eine Deanonymisierung von Personen oder Betrieben erleichtern, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf den Datenzugang explizit begründet wird. Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale sind:

Personendatei:

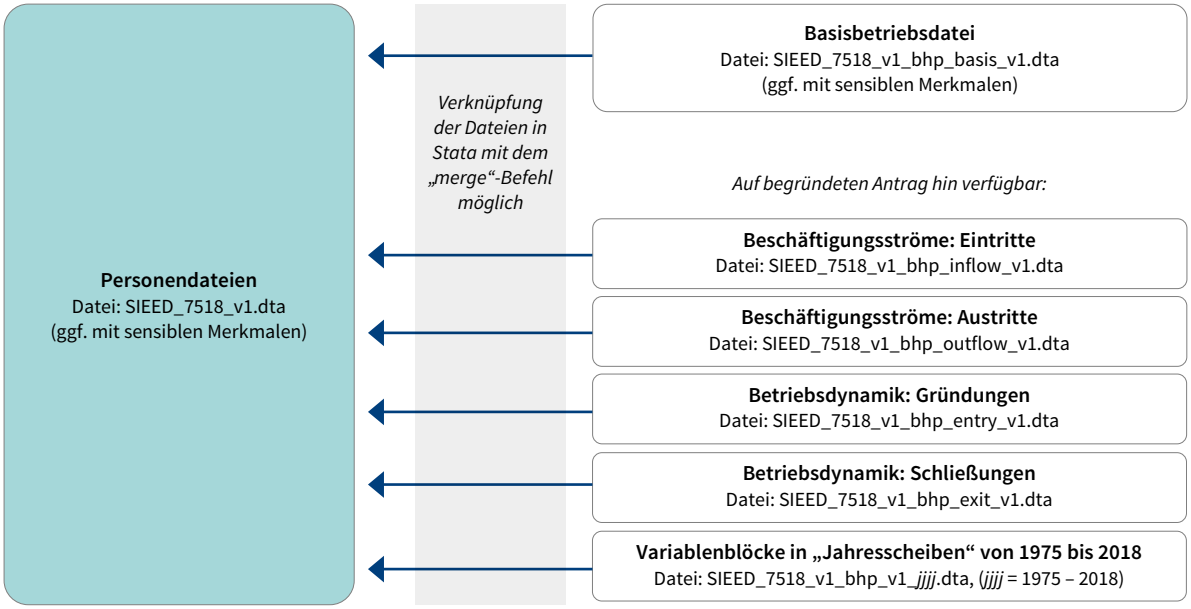
- Staatsangehörigkeit (nation)
- Geburtsmonat (gebmon)
- Berufsuntergruppe (beruf2010_4)
- Wohnort - Arbeitsagentur (wo_aa)
- Wohnort - Kreis (wo_kreis)

Basisbetriebsdatei:

- Arbeitsort - Kreis (ao_kreis)
- Wirtschaftszweig - Wirtschaftsunterklasse (w93_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w03_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w08_5)
- Erstes Auftreten Betriebsnummer (grd_dat)
- Letztes Auftreten Betriebsnummer (lzt_dat)

[1] Mit den Stata-Befehlen *label language en* oder *label language de* kann auf englisch- bzw. deutschsprachige Labels umgestellt werden.

Abbildung 1: Datenhaltung der SIEED-Daten



Auf begründeten Antrag können aus dem Betriebs-Historik-Panel (BHP) weitere Betriebsmerkmale in Variablenblöcken bereitgestellt werden (vgl. Abbildung 1). Zudem können die Erweiterungsmodule Beschäftigungsströme (Eintritte/Austritte) und Betriebsdynamik (Gründungen/Schließungen) mit gesonderter Begründung beantragt werden. Nähere Informationen zum BHP finden sich unter https://fdz.iab.de.

1.3 Steckbrief

Tabelle 1: Steckbrief

Themen/ Merkmalsgruppen	**Beschäftigtenhistorik (BeH):** Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte
Untersuchungseinheit	Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Beschäftigungsbetriebe
Fallzahlen	5.567.883 Personen 5.248.850 Betriebe 157.271.644 Originalsätze 172.935.233 überschneidungsfreie Sätze (nach Episodensplitting)
Zeitraum	BeH: 01.01.1975 - 31.12.2018 (2015: 36-Monats-Datei 2016 und 2017: 18-Monats-Datei, 2018: 6-Monats-Datei)
Zeitbezug	Tagesgenau
Regionale Gliederung	Bundesländer, Kreise

Gebietsstand	Gebietsstandsbereinigt zum 31.12.2018
Erhebungsdesign	Linked-Employer-Employee-Daten; Auswahl einer 1,5% Zufallsstichprobe von Betrieben und Zuspielen der Erwerbsbiographien ihrer Beschäftigten
Beteiligte Institutionen	Sozialversicherungsträger, Bundesagentur für Arbeit, kommunale Träger
Frequenz der Datensammlung	Laufend
Dateiformat/ -größe	Stata; Personendatensatz: 10,3 GB; Basisbetriebsdatei: 1,85 GB
Dateiorganisation	Die Daten werden in zwei Dateien abgelegt. Die eine enthält personenbezogene und die andere betriebsbezogene Informationen. Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen bereitgestellt.
Datenzugang	Gastaufenthalt und anschließende Datenfernverarbeitung
Anonymisierungsgrad	Schwach anonymisiert
Sensible Merkmale	Geburtsmonat (gebmon), Staatsangehörigkeit (nation), Berufsuntergruppe (beruf2010_4), Wohnort – Kreis (wo_kreis), Wohnort – Arbeitsagentur (wo_aa), Arbeitsort - Kreis (ao_kreis), Wirtschaftsunterklasse (w93_5), Wirtschaftsunterklasse (w03_5), Wirtschaftsunterklasse (w08_5), Erstes Auftreten der Betriebsnummer (grd_dat), Letztes Auftreten der Betriebsnummer (lzt_dat)
Zitierung der Daten und Datendokumentation	Daten: „Die Datengrundlage dieses Beitrags bildet die Stichprobe Integrierter Employer-Employee Daten (SIEED) 1975 - 2018. Der Datenzugang erfolgte über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und/oder mittels kontrollierter Datenfernverarbeitung beim FDZ." 10.5164/IAB.SIEED7518.de.en.v1 Datendokumentation: Schmidtlein, Lisa; Seth, Stefan; vom Berge, Philipp (2019): Stichprobe Integrierter Employer-Employee Daten (SIEED) 1975 - 2018. FDZ-Datenreport, 14/2020 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2014.de.v1
Datensatzversion	Stichprobe Integrierter Employer-Employee Daten (SIEED) – Version 7518 v1; DOI: 10.5164/IAB.SIEED7518.de.en.v1

1.4 Merkmalsliste

Die Merkmalsübersicht in Tabelle 3 listet die Variablennamen und die längeren Merkmalsbezeichnungen auf. Weiterhin bietet sie eine Übersicht darüber, ob und wie gut Merkmale in der Datenquelle gefüllt sind. Tabelle 2 macht die Bedeutung der Schattierungen deutlich, die in Tabelle 3 den Füllgrad pro Merkmal und Quelle angeben.

Tabelle 2: Füllgrad der Merkmale

h	Merkmal ist für die Quelle gefüllt. Füllgrad > 0,85
m	Merkmal ist für die Quelle gefüllt. Niedriger bzw. wechselnder Füllgrad, siehe Merkmalsbeschreibung und Häufigkeitsauszählung
l	Merkmal ist für diese Quelle nicht gefüllt. Füllgrad < 0,05

Tabelle 3: Merkmalsliste mit Füllgrad

Merkmalsliste	Modul	Seite	BeH
Identifikatoren		**21**	**l**
Systemfreie Personennummer (persnr)		21	**h**
Systemfreie Betriebsnummer (betnr)	*	21	**h**
Generierte technische Merkmale		**22**	**l**
Satzzähler pro Konto (spell)		22	**h**
Jahr (jahr)	*	23	**l**
Gültigkeitszeitraum		**23**	**l**
Beginndatum Originalbeobachtung (begorig)		23	**h**
Enddatum Originalbeobachtung (endorig)		23	**h**
Beginndatum Episode (begepi)		24	**h**
Enddatum Episode (endepi)		24	**h**
Informationen zur Person		**24**	**l**
Geschlecht (frau)		24	**h**
Geburtsjahr (gebjahr)		24	**h**
Geburtsmonat (gebmon)		25	**h**
Staatsangehörigkeit (nation)		25	**h**
Staatsangehörigkeit vergröbert (nation_gr)		25	**h**
Ausbildung (ausbildung)		25	**m**

Merkmalsliste	Modul	Seite	BeH
Ausbildung (imputiert) (ausbildung_imp)		26	**h**
Schulausbildung (schule)		27	**m**
Informationen zu Beschäftigung		**27**	**l**
Tagesentgelt/ täglicher Leistungssatz (tentgelt)		27	**h**
Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)		28	**h**
Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)		29	**h**
Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)		29	**h**
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)		30	**h**
Teilzeit (teilzeit)		31	**h**
Stellung im Beruf und Arbeitszeit (stib)		31	**h**
Erwerbsstatus (erwstat)		32	**h**
Gleitzone (gleitz)		32	**h**
Leiharbeit (leih)		33	**h**
Befristung (befrist)		33	**h**
Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)		33	**h**
Dauer der Arbeitslosigkeit (alo_dau)		34	**h**
Ortsangaben		**34**	**l**
Wohnort - Kreis (wo_kreis)		34	**h**
Wohnort - Bundesland (wo_bula)		35	**h**
Wohnort - Arbeitsagentur (wo_aa)		35	**h**
Wohnort - Regionaldirektion (wo_rd)		36	**h**
Betriebsmerkmale		**36**	**l**
WZ73 3-Steller (w73_3)	*	36	**h**
WZ93 5-Steller (w93_5)	*	36	**h**
WZ93 3-Steller (w93_3)	*	37	**h**

Merkmalsliste	Modul	Seite	BeH
WZ03 5-Steller (w03_5)	*	37	**h**
WZ03 3-Steller (w03_3)	*	38	**h**
WZ08 5-Steller (w08_5)	*	38	**h**
WZ08 3-Steller (w08_3)	*	39	**h**
w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	*	39	**h**
Art der Vervollständigung w73_3 (group_w73_3)	*	39	**h**
w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	*	40	**h**
Art der Vervollständigung w93_3 (group_w93_3)	*	40	**h**
w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)	*	40	**h**
Art der Vervollständigung w08_3 (group_w08_3)	*	41	**h**
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)	*	41	**h**
Erstes Auftreten Betriebsnummer (grd_dat)	*	41	**h**
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)	*	42	**h**
Letztes Auftreten Betriebsnummer (lzt_dat)	*	42	**h**
Anzahl Beschäftigte gesamt (az_ges)	*	43	**h**
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	*	43	**h**
Anzahl geringfügig Beschäftigte (az_gf)	*	43	**h**
Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	*	43	**h**
Arbeitsort - Kreis (ao_kreis)	*	44	**h**
Arbeitsort - Bundesland (ao_bula)	*	44	**h**

1.5 Mengengerüst

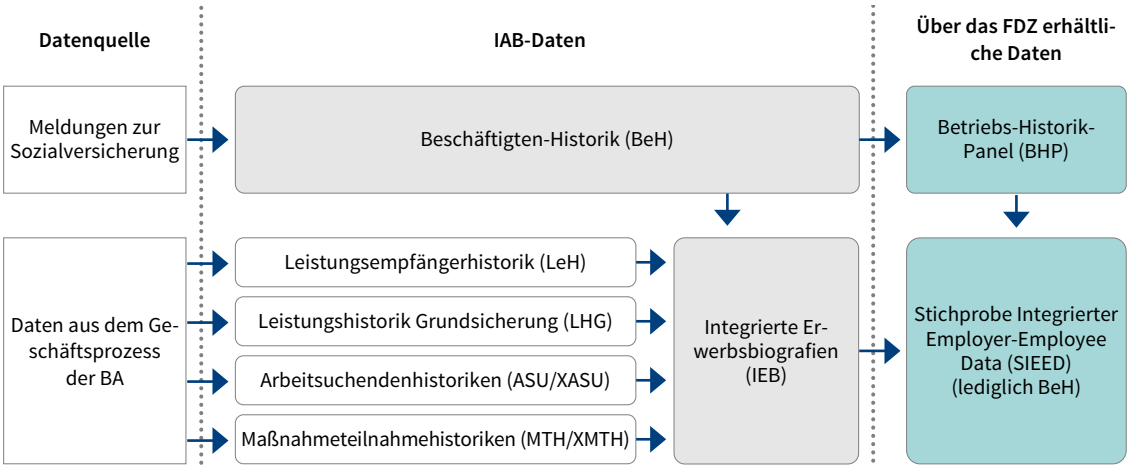
Tabelle 4: Mengengerüst

Fallzahlen	**vor Episodensplitting**	**nach Episodensplitting**
Sätze insgesamt	157.271.644	172.935.233
Personen	5.567.883	

2 Datenquelle

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB) des IAB gezogen. Diese vereinen Daten aus verschiedenen Datenquellen, die jeweils Informationen aus unterschiedlichen EDV-Verfahren enthalten können. Zusätzlich werden einige ergänzende Merkmale aus diesen Datenquellen in die administrativen Personendaten aufgenommen, die nicht Teil der IEB sind. Abbildung 2 veranschaulicht die Datenflüsse, die zur SIEED führen und stellt deren Beziehung zu anderen Datenprodukten des FDZ dar.

Abbildung 2: Quellen der SIEED



Die Quelle für Informationen über Beschäftigung ist die Beschäftigtenhistorik (BeH) des IAB, welche die Hauptdatenquelle für die SIEED darstellt. Die Grundlage der Daten ist das mit Wirkung vom 01.01.1973 eingeführte (und seit dem 01.01.1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung (DEÜV; früher DEVO/DÜVO; vgl. für weitere Details: Bender et al., 1996, S. 4ff.; Wermter/Cramer, 1988). Es verlangt von den Arbeitgebern, mindestens einmal jährlich Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmer an die zuständigen Sozialversicherungsträger abzugeben. Die BeH erfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungs-

pflicht befreit sind. Dies bedeutet, dass verbeamtete und selbstständige Tätigkeiten sowie Studierende[2] (vgl. Cramer, 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 01.01.1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (enthalten erst ab 01.04.1999). Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit gesammelt und aufbereitet und anschließend vom IAB in dessen Historikdatei integriert.

Die administrativen Personendaten werden durch Betriebsdaten (Basisbetriebsdaten und BHP-Erweiterungsmodule) ergänzt. Diese werden dem Betriebs-Historik-Panel (BHP) entnommen, das ebenso auf der BeH basiert.

Bei der Verknüpfung von Personen- mit Betriebsdaten muss berücksichtigt werden, dass die Merkmale in der Basisbetriebsdatei sowie den BHP-Erweiterungsmodulen zum 30.06. eines Jahres aggregiert werden. Die Betriebsmerkmale sind deshalb, im Unterschied zu den Personenangaben, keine zeitraumbezogenen Daten, sondern nur am 30.06. jedes Jahres exakt gültig.[3]

Die Verknüpfung der Betriebs- und Personendatei erfolgt über die programmspezifischen Kommandos der zur Datenaufbereitung und -analyse genutzten Softwarepakete. In Stata können die Daten beispielsweise unter Verwendung des „`merge`“-Befehls und der Jahresangabe sowie der Betriebsnummer in Verbindung mit der entsprechenden Pfadangabe verknüpft werden (siehe Kasten 1).

Kasten 1: Beispielcode zur Verbindung von Personen- und Betriebsdaten in Stata 14

```
use SIEED_7518_v1.dta
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using SIEED_7518_v1_bhp_basis_v1.dta
```

3 Datenaufbereitung und Stichprobenziehung

3.1 Bereinigungen und Aufbereitungen

Die Gesamt-IEB erfahren die folgenden Bereinigungen:

- Sätze, bei denen das Alter an ihrem Ende kleiner 13 oder zu Beginn größer 75 ist, werden gelöscht.
- Datensätze, bei denen das Enddatum kleiner als das Beginndatum ist, werden gelöscht.

[2] Studierende können dennoch in der BeH erscheinen, wenn sie parallel zum Studium z.B. geringfügig beschäftigt waren.

[3] Extrembeispiel: Es liegen Beschäftigungsmeldungen vom 1.1. bis zum 30.05.2006 vor, der Betrieb meldet jedoch Konkurs an, weshalb alle Beschäftigten entlassen werden. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.

- Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum innerhalb eines Kontos werden bereinigt.
- Datensätze ohne Angabe von Geburtsdatum und Geschlecht nach Bereinigung werden gelöscht.

Weitergehende Bereinigungen, wie z.B. das Ergänzen von vermutlich fehlenden Meldungen, Streikbereinigungen, etc., erfolgen nicht.

Die Daten der BeH werden wie folgt aufbereitet und bereinigt:

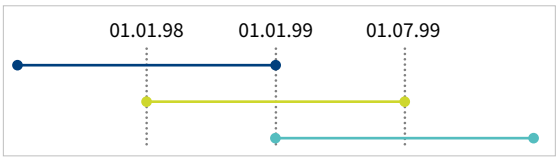
- Um einen im Zeitablauf möglichst konstanten Personenkreis zu erfassen, werden einige Personengruppen ausgeschlossen, deren Daten nicht für den gesamten Beobachtungszeitraum zur Verfügung stehen. Ab dem Meldejahr 2011 stammen die Daten der BeH aus neu konzipierten Quelldaten. Dadurch wurde eine Reihe von Personengruppen reaktiviert bzw. neu eingeführt, da sie von der BA-Statistik als sozialversicherungspflichtig eingestuft werden. Damit sind fortan die Personengruppen 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 und 203 - 205 enthalten sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen. Nicht enthalten sind unter anderem kurzfristig Beschäftigte, d. h. Personengruppen 110, 202 und 210.
- Die Personengruppen 123, 124 und 127 wurden ab 2011 neu eingeführt.
- Aus datenschutzrechtlichen Gründen werden die Personengruppen 107, 111, 113, 114, 127 und 204 zur Personengruppe "Sonstige Erwerbstätige“ (599) zusammengefasst.
- Ab dem Meldejahr 2012 wurden Auszubildende mit den Personengruppen 121 und 122 neu aufgenommen.
- Sätze mit fehlendem Entgelt sowie der Ausprägung 101 beim Merkmal Erwerbsstatus und dem Wert 50 beim Abgabegrund (Jahresmeldung) werden nicht in die IEB übernommen.
- Geschlecht und Geburtsdatum werden aus dem Data Warehouse (DWH) der Statistik der BA übernommen. Dort liegen diese Angaben quellenübergreifend harmonisiert vor.
- Der Gebietsstand für Arbeitsort und Wohnort entspricht dem 31.12.2017.

3.2 Episodensplitting

Die administrativen Personendaten liegen „gesplittet“ vor. Falls sich innerhalb eines Kontos Sätze überlappen, werden diese durch künstliche Sätze mit neuen Datumsangaben derart ersetzt, dass vollständig parallele oder überschneidungsfreie Zeiträume entstehen (Episodensplitting). Die Anzahl der Sätze erhöht sich dadurch (siehe Abbildung 3).

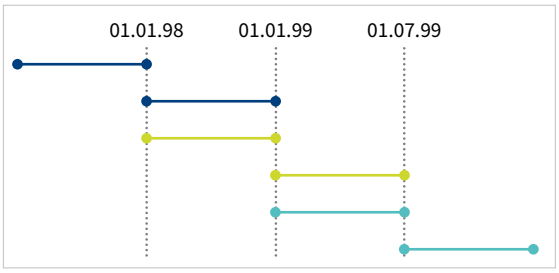
Die ursprünglichen Datumsvariablen zum Beginn und Ende des Originalsatzes (begorig und endorig) bleiben erhalten, die Variablen Beginn und Ende der Episode (begepi und endepi) kennzeichnen Anfang und Ende der gesplitteten Episoden. Ob Sätze gesplittet wurden, kann an einem Vergleich des ursprünglichen Zeitraums (begorig und endorig) mit dem Episodenzeitraum (begepi und endepi) festgestellt werden. Um die Originaldaten ohne Episodensplitting wiederherzustellen bzw. die durch das Episodensplitting künstlich erzeugten Datensätze zu löschen, selektiert man alle Sätze, bei denen der Beginn des Originalsatzes gleich dem Beginn der gesplitteten Episode ist (`begepi == begorig`).

Abbildung 3: Episodensplitting



persnr	*begorig*	*endorig*	*level*
4711	01.01.98	31.12.98	-
4711	01.10.98	30.06.99	-
4711	01.01.99	31.12.99	-

Episodensplitting
Aufteilung der Zeitintervalle in sich nicht überschneidende Intervalle



persnr	*begepi*	*endepi*	*level*
4711	01.01.98	30.09.98	0
4711	01.10.98	31.12.98	0
4711	01.10.98	31.12.98	1
4711	01.01.99	30.06.99	0
4711	01.01.99	30.06.99	1
4711	01.07.99	31.12.99	0

Es ist ratsam, die durch das Splitting erzeugten, vollständig parallelen Sätze einheitlich zu sortieren. Die früher in anderen Datenprodukten, wie z.B. der SIAB, enthaltenen Variablen Satzzähler pro Episode (level2) und Satzzähler pro Episode und Quelle (level1) können bei Bedarf mit folgenden Stata-Befehlen selbst erzeugt werden:

Kasten 2: Beispielcode zum Erstellen weiterer Satzzähler in Stata 14

```
bysort persnr begepi quelle (spell): gen byte level1 = _n-1
bysort persnr begepi (spell): gen byte level2 = _n-1
```

3.3 Stichprobenziehung

Grundlage der Stichprobenziehung in der SIEED sind die Betriebe der Beschäftigtenhistorik (BeH). Die BeH enthält Informationen zu allen Betrieben in Deutschland, die im Zeitraum 1975 bis 2018 mindestens einen sozialversicherungspflichtig Beschäftigten (erfasst ab 1975) bzw. geringfügig Beschäftigten (erfasst ab 1999) hatten. Aus der Grundgesamtheit der Betriebe wird im ersten Schritt eine 1,5% Stichprobe aller Betriebe gezogen. Im zweiten Schritt werden aus den IEB alle Personen gezogen, die im Zeitraum 1975 bis 2018 mindestens einen Tag in einem dieser ausgewählten Betriebe beschäftigt waren. Für diese Personen werden im dritten Schritt die kompletten Erwerbsbiographien für die Quelle BeH aus den IEB gezogen. Die Erwerbsbiographien umfassen dabei den Zeitraum 1975 bis 2018 und enthalten somit Beschäftigungszeiten bei den Stichprobenbetrieben selbst als auch bei anderen Betrieben, die nicht Teil der Stichprobe sind.

Bei Auswertungen auf Basis der SIEED ist zu beachten, dass der Datensatz aufgrund der Ziehungslogik repräsentativ für Betriebe in Deutschland ist, nicht jedoch für Personen. Repräsentative Aussagen z.B. zur Grundgesamtheit der Beschäftigten in Deutschland lassen sich evtl. besser auf Basis der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) treffen.

3.4 Fehlende Werte

In den Daten gilt die folgende Codierung von fehlenden Werten:

Bezeichnung	Wert	Beschreibung
Keine (valide) Angabe	.z	Nicht systematisch fehlender Wert. Das Merkmal ist prinzipiell für die Quelle gefüllt, jedoch ist für den betrachteten Satz ein Wert nicht vorhanden bzw. nicht sinnvoll zu interpretieren.
Fehlt systematisch	.n	Ein Merkmal ist grundsätzlich für eine Quelle (weiße Felder der Merkmalsübersicht im Abschnitt 1.4) oder für einen gewissen Zeitraum nicht gefüllt.

4 Datenqualität und Probleme

4.1 Gesamte IEB

Die IEB beinhalten Erwerbsbiografien. Allerdings ist in den administrativen Daten nicht jede Art von Erwerbstätigkeit erfasst. Bei entsprechenden Lebensverläufen sind bestimmte Personen überhaupt nicht in den IEB abgebildet.

Für Auswertungszwecke ist es oft relevant, Lücken in den abgebildeten Biografien zu kennen (z.B. für Kontrollgruppenbildung, Lebensverlaufsanalysen etc.). Die im Folgenden aufgeführten Lücken sind definiert als Zeiträume nach Ende der Schulausbildung, für die kein Satz in den IEB enthalten ist. Diese Lücken können unterschieden werden in

- Lücken, für die keinerlei Information vorliegt und
- Lücken, für die anhand des Merkmals Abgabegrund/Beendigungsgrund/SGB-II-Einstellungsgrund/Abmeldegrund des direkten Vorgängersatzes eine Information vorliegen kann (sofern ein entsprechender Satz vorhanden ist).

Diese Lücken wurden identifiziert anhand der Ausprägungen der Merkmale Abmeldegrund/Abgabegrund/Beendigungsgrund und Erwerbsstatus in den verschiedenen Quellen. Die Auflistung in Tabelle 6 erhebt keinen Anspruch auf Vollständigkeit. Da nicht alle Quellen in die SIEED übernommen werden, ist die Identifikation von Lücken dadurch weiter eingeschränkt.

Tabelle 6: Biographische Lücken und Identifikationsmöglichkeiten

Biografische Lücke	**Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle**
Beamte, Berufssoldaten, Richter, Angestellte von Körperschaften oder Stiftungen des öffentlichen Rechts	XASU
Ungeförderte Selbstständige	LeH, ASU
Studierende, Personen in schulischer Weiterbildung	LeH, LHG, ASU, XASU

Personen, die länger als 6 Wochen krank / arbeitsunfähig sind (Krankheit während Arbeitslosigkeit ist aber unter bestimmten Bedingungen in der Quelle ASU abgebildet, vgl. Abschnitt 3.4)	BeH, LeH, ASU
Personen in Altersrente ohne Beschäftigung, wenn nicht Mitglied in einer BG	LeH, LHG, ASU
Personen in Mutterschutz / Elternzeit	XASU
Bezieher von Vorruhestandsgeld	LeH, ASU
Hausgewerbetreibende	
Kurzfristig Beschäftigte	ASU
Personen in Einrichtungen der Jugendhilfe, Berufsbildungswerken, anerkannten Werkstätten oder ähnlichen Einrichtungen für behinderte Menschen	ASU
Teilnehmende an Leistungen zur Teilhabe am Arbeitsleben (Rehabilitanden)	ASU
(Nebenerwerbs-)Landwirte	
Pflegepersonen i.S. von § 19 SBG XI	
Grundwehrdienstleistende	BeH, LeH, LHG, ASU, XASU
Wehrübungsleistende	BeH, LeH, LHG, ASU, XASU
Zivildienstleistende	BeH, LeH, LHG, ASU, XASU
Ableistende eines freiwilligen sozialen bzw. ökologischen Jahres anstelle des Zivildienstes	
Sonstige Personen, die nicht bei der Rentenversicherung oder der BA registriert sind (z.B. Sabbatjahr, Finanzierung aus privatem Vermögen oder Rentenbezügen, Emigration, Beschäftigung im Ausland, ehrenamtliche Tätigkeit, etc.)	BeH, LeH, ASU
Streikende bei Streikdauer über einem Monat	LeH
Sozialhilfe-Bezieher (vor Einführung des SGB II 2005), Sozialgeld-Bezieher (im Sinne des SGB II)	
SGB-II-Leistungsberechtigte, bei deren Trägern Lieferausfälle vorkamen	
Ausgleichsempfänger nach dem FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit)	

4.2 Beschäftigtenhistorik

- Die Übermittlung von Informationen zur Ausbildung, ausgeübten Tätigkeit und Stellung im Beruf erfolgt durch Meldungen der Arbeitgeber nach DEÜV (siehe Abschnitt 2) über einen sogenannten Tätigkeitsschlüssel. Für Zeiträume mit einem Enddatum nach dem 30.11.2011 erfolgte der Wechsel zum neuen Tätigkeitsschlüssel 2010 (siehe Bertat et al., 2013, für mehr Details). Dieser Wechsel wurde von den Spitzenorganisationen der Sozialversicherungsträger beschlossen, da der Tätigkeitsschlüssel 2003 eine Reihe von Sachverhalten nicht mehr zeitgemäß und nicht mehr realitätsnah erfassen konnte. Da die Meldungen der Arbeitgeber nach DEÜV in die BeH einfließen, hat die Umstellung des Tätigkeitsschlüssels ausschließlich Auswirkungen für Beobachtungen aus dieser Quelle. Die Messung folgender bisher enthaltener An-

gaben ist von dieser Umstellung betroffen: Arbeitszeit, Beruf, Stellung im Beruf sowie schulisches und berufliches Qualifikationsniveau. Darüber hinaus gibt es seit der Umstellung auch Angaben zur Befristung eines Beschäftigungsverhältnisses sowie dazu, ob eine Person als Verleihpersonal in einem Zeitarbeitsunternehmen angestellt ist. Die wichtigste Auswirkung ist der Umstieg auf eine neue Berufsklassifikation. Statt der bisherigen Klassifikation der Berufe 1988 (KldB 1988) wird mit dem neuen Tätigkeitsschlüssel die differenziertere KldB 2010 gemeldet.[4]

- Die Variable „Stellung im Beruf und Arbeitszeit“ (stib) ist nur für Meldungen gefüllt, die vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden, und das FDZ nimmt keine Extrapolation oder Imputation für spätere Meldungen vor. Die Kategorien der Variable stib, die über den gesamten Beobachtungszeitraum konsistent beobachtet werden können (Erwerbsstatus als Auszubildender, Unterscheidung von Teil- und Vollzeit), sind in den Variablen erwstat und teilzeit nachzuvollziehen.
- Die Einführung des neuen Tätigkeitsschlüssels hat zu einer Reihe von Problemen geführt. So kam es während des Übergangszeitraums, der Arbeitgebern im Meldeverfahren zur Sozialversicherung eingeräumt wurde,[5] zu einer temporären Zunahme fehlender Angaben. Analysen der Statistik der BA (Bertat et al., 2013, S. 10) zeigen, dass die neuen bzw. umgestellten Merkmale „ausgeübte Tätigkeit“, „Arbeitszeit“ und „Ausbildungsabschluss“ nach der Umstellung zu etwa 20 bis 30 Prozent keine Angaben enthielten. Eine deutliche Verbesserung stellt sich im ersten Halbjahr 2013 ein. Um die Qualität der Arbeitszeitangabe im Übergangszeitraum zu verbessern, haben Ludsteck/Thomsen (2016) ein Imputationsverfahren entwickelt, das die fehlenden durch imputierte Werte ersetzt. Die imputierten Daten werden in der SIEED 7518 verwendet. Für die Lücken in den anderen Merkmalen werden keine Imputationen vorgenommen.
- Aufgrund der Einführung des Meldeverfahrens in den neuen Bundesländern kann erst ab 1993 von einer hinreichenden Vollständigkeit der Meldungen ostdeutscher Betriebe ausgegangen werden. Aus demselben Grund weisen sehr viele Sätze des Jahrs 1991 für einige Merkmale (u.a. Ausbildung, Erwerbsstatus und Tagesentgelt) fehlende Werte auf.
- Der Anstieg der Zahl der BeH-Sätze ab 1999 folgt aus der Einführung der Meldepflicht für geringfügig Beschäftigte ab dem 01.04.1999.
- Es lässt sich insbesondere 1999 ein deutlicher Anstieg bei den Meldungen zu Teilzeitbeschäftigung feststellen. Verursacht wird dieser Befund sowohl vom tatsächlich zu beobachtenden Anstieg der Teilzeitbeschäftigung als auch von der Tatsache, dass seit 1999 die Beschäftigtenmeldungen generell besser ausgefüllt werden.
- Innerhalb des Meldeverfahrens ist eine gewisse Zeitverzögerung nicht vermeidbar. Zwar müssen Änderungen von Beschäftigungsverhältnissen sofort gemeldet und bestehende Beschäftigungsverhältnisse jährlich bis zum April (bis Jahresmeldung 2012) bzw. Mitte Februar (ab Jahresmeldung 2013) des folgenden Jahres bestätigt werden. Faktisch treffen Meldungen jedoch zum Teil um Jahre verspätet ein. Die Historikdatei des IAB wird aber nicht kontinuierlich

[4] Siehe Paulus/Matthes, 2013, für Details zur Klassifikation der Berufe 2010.

[5] Die im Rahmen des Meldeverfahrens eingesetzten Prüfprogramme ließen bis einschließlich Mai 2012 leere Angaben beim Tätigkeitsschlüssel 2010 zu.

fortgeschrieben, sondern in bestimmten Abständen aktualisiert. Dies geschieht unter Verwendung von Dateien mit den Beschäftigten-Meldungen jeweils eines Jahres, die 36, 18, 12 oder 6 Monate nach Ablauf des Berichtsjahres eingetroffen sind (frühestens im Juli 2015 z.B. kann die 18-Monats-Datei für das Jahr 2013 erstellt werden). Mehr als 3 Jahre verspätete Meldungen werden im IAB nicht berücksichtigt, so dass eine 36-Monats-Datei per definitionem einen Füllgrad von 100% aufweist.

- In der zugrundeliegenden Version der IEB ist 2018 das letzte Jahr mit einem Füllgrad der BeH-Meldungen von 100%. Für die Jahre 2016 und 2017 wurden 18-Monats-Dateien verwendet, für 2018 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass Beschäftigungsmeldungen für die Jahre 2016 bis 2017 leicht und für 2018 etwas stärker untererfasst sind. Die Auswertbarkeit der Daten auf Personenebene sollte dadurch aber nicht eingeschränkt sein. Allerdings häufen sich die fehlenden Meldungen in wenigen Betrieben. Dies bedeutet, dass in Einzelfällen die Betriebsangaben, z.B. die Betriebsgröße, grob falsch sind und sich in Folgeversionen stark verändern werden.
- Im Jahr 1984 erfolgte eine Änderung des Meldeverfahrens. Einmalig gezahltes Bruttoarbeitsentgelt wurde ab diesem Zeitpunkt als Teil des beitragspflichtigen Jahresentgelts gemeldet, was zu einer Erhöhung des durchschnittlichen Tagesentgelts führt. Insbesondere steigt ab diesem Jahr der Anteil der Entgelte oberhalb der Beitragsbemessungsgrenze stark an (vgl. Bender et al., 1996).
- Für die Jahre 1992 bis 2000 wurden in den IEB auffällige Einbrüche und Anstiege in der Anzahl der Meldungen festgestellt. Es lassen sich v.a. Einbrüche bei folgenden zehn Kreisen beobachten: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) und Kempten (Allgäu) (09763). Grund hierfür sind Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen.
- Nennenswerte Einbrüche verzeichnen auch die Kreise Salzgitter (03102) und Hoyerswerda (14264).
- Bei den Meldungen zu Vollzeitbeschäftigung in den IEB sind besonders die Kreise Main-Taunus-Kreis (06436) und Alzey-Worms (07331) erwähnenswert. Hier sind überdurchschnittliche Anstiege zu beobachten. Als Grund sind ebenfalls Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen anzuführen.
- In den Jahren 1996 bis 1998 kommen die Werte 841-844 (Ärzte und Apotheker) in der Variable Beruf – ausgeübte Tätigkeit (beruf) im Vergleich zu benachbarten Jahren sehr selten vor. Die Gründe dafür sind nicht bekannt.
- In den Jahren 1975 bis 1977 gab es bisher deutlich zu viele Beschäftigte mit einem Arbeitsort in der niedersächsischen Gemeinde Lahn (Kreis Emsland). Statt im zu erwartenden zweistelligen Bereich gab es pro Jahr bis zu 90.000 Beschäftigtenmeldungen mit diesem Arbeitsort in der Grundgesamtheit der BeH. Ursächlich ist eine historische Fehlzuordnung der Beschäftigtenmeldungen aus der Stadt Lahn. Hierbei handelte es sich um einen nach kurzer Zeit wieder aufgelösten Zusammenschluss der hessischen Gemeinden Wetzlar, Gießen, Heuchelheim, Wettenberg und Lahnau (Kreise Gießen und Lahn-Dill-Kreis). Auf Basis dieser Erkenntnisse und unter der Annahme, dass es keine realen Betriebsumzüge zwischen diesen Regionen gab,

wurde auf Ebene der Betriebsnummer folgende Bereinigungsregel umgesetzt: Sobald der Betriebsort in den Jahren 1975 bis 1978 von der Gemeinde Lahn auf eine der aufgeführten hessischen Gemeinden wechselt, wird die frühere Angabe Gemeinde Lahn mit der späteren korrekten Angabe überschrieben. Über diese Regel wird der Überhang deutlich reduziert und die niedersächsische Gemeinde Lahn weist danach in den kritischen Jahren zumindest nur noch unter 2.000 Beschäftigtenmeldungen in der Grundgesamtheit der BeH auf.

5 Variablenbeschreibung

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in den Arbeitshilfen zur SIEED unter https://fdz.iab.de.

5.1 Identifikatoren

5.1.1 Systemfreie Personennummer (persnr)

Variablenlabel	Systemfreie Personennummer
Variablenname	persnr
Kategorie	Identifikatoren
Datentyp	Numerisch
Detailbeschreibung	Die systemfreie Personennummer zeigt an, welche Sätze zur gleichen Person gehören. Systemfrei bedeutet, dass von dieser Personennummer nicht auf Eigenschaften der Person oder Original-Identifikatoren geschlossen werden kann. Da es in den verschiedenen Datenquellen keinen einheitlichen Personen-Identifikator gibt, ist die Zuordnung der Informationen aus den unterschiedlichen Datenquellen (z.B. Beschäftigung und Leistung) zu Personen(-Konten) nicht immer eindeutig. In solchen Fällen kann es zu unplausiblen Erwerbshistorien kommen. Die Bildung des quellenübergreifenden Personen-Identifikators basiert auf einer von der BA entwickelten Heuristik.

5.1.2 Systemfreie Betriebsnummer (betnr)

Variablenlabel	Systemfreie Betriebsnummer
Variablenname	betnr
Kategorie	Identifikatoren
Datentyp	Numerisch
Detailbeschreibung	Die systemfreie Betriebsnummer zeigt an, welche Sätze zum gleichen Betrieb gehören. Sie basiert auf der von der BA vergebenen Betriebsnummer, die durch eine systemfreie Nummer ersetzt wurde (weitere Informationen zur Betriebsnummernvergabe der BA finden sich in Bender et al., (1996, S. 15f. und S. 27-30) sowie direkt bei dem Betriebsnummern-Service der BA unter https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige). Die Betriebsnummer kann zusammen mit der Jahresangabe benutzt werden, um Personen- und Betriebsinformationen zusammenzuspielen. Für die Betriebsnummer sollte Folgendes grundsätzlich beachtet werden:

	Besteht das Unternehmen aus nur einer Niederlassung oder hat das Unternehmen in einer Gemeinde nur eine Niederlassung, so ist diese Niederlassung der Betrieb und erhält eine Betriebsnummer. Hat das Unternehmen **mehrere** Niederlassungen in **einer** Gemeinde, so sind diese Betriebs- und Arbeitsstätten zu einem Betrieb unter einer Betriebsnummer zusammenzufassen, wenn sie derselben Wirtschaftsklasse angehören bzw. ist jede Niederlassung ein Betrieb und erhält je eine Betriebsnummer, wenn sie **nicht** derselben Wirtschaftsklasse angehören. Hat das Unternehmen **mehrere** Niederlassungen in **mehreren** Gemeinden, so ist jede dieser Niederlassungen ein Betrieb und erhält je eine Betriebsnummer. In diesem Zusammenhang sind auch folgende Definitionen im Sinne der Betriebsnummernvergabe im Rahmen des Meldeverfahrens zur Sozialversicherung zu beachten: Ein **Betrieb** ist eine regional und wirtschaftsfachlich abgegrenzte Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen eine Betriebsnummer vergeben wird. Eine **Arbeitsstätte** ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen keine Betriebsnummer zu vergeben ist. Ein **Unternehmen** als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen. Ein **Arbeitgeber** ist jede natürliche oder juristische Person, für die mindestens ein sozialversicherungspflichtig bzw. geringfügig Beschäftigter tätig ist. Betrieb und Betriebsstätten sind synonyme Begriffe; Niederlassung ist dann ein synonymer Begriff für Filialen, Zweigstellen, Auslagerungen, Dependancen, Arbeitsstätten u.ä., wenn sie kein Betrieb ist.
Qualitätshinweis	Die Betriebsnummer fehlt in nur sehr wenigen BeH-Meldungen. Bei diesen Meldungen handelt es sich in den meisten Fällen um Meldungen der Personengruppe „205“ (Entgeltmeldungen für unständig Beschäftigte). Da Betriebsmerkmale (Arbeitsort, Wirtschaftszweig, Betriebsgröße usw.) über die Betriebsnummer zugespielt werden, fehlen sie in diesen Meldungen.

5.2 Generierte technische Merkmale

5.2.1 Identifikator Panelbetrieb (betnr_pan)

Variablenlabel	Identifikator Panelbetrieb
Variablenname	betnr_pan
Kategorie	Generierte technische Merkmale
Datentyp	Numerisch
Detailbeschreibung	Der Identifikator gibt an, ob ein Betrieb zur ursprünglich gezogenen 1,5 % Betriebsstichprobe gehört oder nicht. Im Fall von Null bedeutet das, dass der Betrieb lediglich als Teil der Längsschnittziehung auf Personenebene mitgezogen wurde. Für diesen Betrieb muss also davon ausgegangen werden, dass nicht notwendigerweise alle Beschäftigten im Datensatz enthalten sind.

5.2.2 Satzzähler pro Konto (spell)

Variablenlabel	Satzzähler pro Konto

Variablenname	spell
Kategorie	Generierte technische Merkmale
Datentyp	Numerisch
Detailbeschreibung	Der Satzzähler pro Konto nummeriert aufsteigend die Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal „Satzzähler pro Konto“ kann die ursprüngliche Sortierung auf einfache Art und Weise wiederhergestellt werden. Die Sortierung erfolgt grundsätzlich zunächst nach dem Beginndatum der Episode und dann nach der Quelle. Innerhalb der Beschäftigungsmeldungen werden sozialversicherungspflichtige vor geringfügige Beschäftigungsmeldungen und höhere vor niedrigere Tagesentgelte sortiert. Eine Ausnahme bilden Einmalzahlungen, die nach hinten sortiert werden.

5.2.3 Jahr (jahr)

Variablenlabel	Jahr
Variablenname	jahr
Kategorie	Generierte technische Merkmale
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal ist nur in der Betriebsdatei enthalten. Es gibt das Gültigkeitsjahr der Betriebsinformationen zum Stichtag 30.06 an. Mit Hilfe dieses Merkmals und der Betriebsnummer können Personen- und Betriebsdatei verknüpft werden. Siehe Kasten 1 auf Seite 14 für einen Beispielcode mit dem „merge“-Befehl in Stata 14.

5.3 Gültigkeitszeitraum

5.3.1 Beginndatum Originalbeobachtung (begorig)

Variablenlabel	Beginndatum Originalbeobachtung
Variablenname	begorig
Kategorie	Gültigkeitszeitraum
Datentyp	Datum
Detailbeschreibung	Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Dies kann vom Beginndatum der betreffenden Datenzeile, dem sogenannten Episoden-Beginndatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2). In BeH-Meldungen sind Beginn- und Endjahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen). Ein durchgehendes Beschäftigungsverhältnis kann also auf mehrere Meldungen verteilt sein.

5.3.2 Enddatum Originalbeobachtung (endorig)

Variablenlabel	Enddatum Originalbeobachtung
Variablenname	endorig
Kategorie	Gültigkeitszeitraum
Datentyp	Datum

Detailbeschreibung	Das Enddatum des Originalsatzes entspricht dem ursprünglichen Enddatum der Meldung. Dies kann vom Enddatum der betreffenden Datenzeile, dem sogenannten Episoden-Enddatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2). In BeH-Meldungen sind Beginn- und Endjahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen). Ein durchgehendes Beschäftigungsverhältnis kann also auf mehrere Meldungen verteilt sein.

5.3.3 Beginndatum Episode (begepi)

Variablenlabel	Beginndatum Episode
Variablenname	begepi
Kategorie	Generierter Gültigkeitszeitraum
Datentyp	Datum
Detailbeschreibung	Das Beginndatum der Episode ist immer gleich dem oder größer als das Beginndatum der Originalbeobachtung (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2).

5.3.4 Enddatum Episode (endepi)

Variablenlabel	Enddatum Episode
Variablenname	endepi
Kategorie	Generierter Gültigkeitszeitraum
Datentyp	Datum
Detailbeschreibung	Das Enddatum der Episode ist immer gleich dem oder kleiner als das Enddatum der Originalbeobachtung (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2).

5.4 Informationen zur Person

5.4.1 Geschlecht (frau)

Variablenlabel	Geschlecht
Variablenname	frau
Kategorie	Persönliche Merkmale
Datentyp	Numerisch
Detailbeschreibung	Geschlechter-Dummy (0 - Mann, 1 - Frau). Die Geschlechtsangabe ist innerhalb eines Personenkontos konstant.

5.4.2 Geburtsjahr (gebjahr)

Variablenlabel	Geburtsjahr
Variablenname	gebjahr
Kategorie	Persönliche Merkmale
Datentyp	Numerisch
Detailbeschreibung	Das Geburtsjahr ist innerhalb eines Personenkontos konstant.

Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsdatum zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus dem Leistungsbezug die höchste Priorität zu.

5.4.3 Geburtsmonat (gebmon)

Variablenlabel	Geburtsmonat
Variablenname	gebmon
Kategorie	Persönliche Merkmale
Datentyp	Numerisch

Detailbeschreibung	Der Geburtsmonat ist innerhalb eines Personenkontos konstant.
	Um aus den Angaben zum Geburtsjahr (gebjahr) und -monat (gebmon) eine Variable im Datumsformat JJJJmM (z.B. 1984m6) zu generieren, kann folgender Befehl verwendet werden:
	`gen int gebdat = ym(gebjahr, gebmon)`
	`format gebdat %tm`

Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsdatum zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus dem Leistungsbezug die höchste Priorität zu.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen bereitgestellt. Ansonsten wird lediglich das Geburtsjahr (gebjahr) bereitgestellt.

5.4.4 Staatsangehörigkeit (nation)

Variablenlabel	Staatsangehörigkeit
Variablenname	nation
Kategorie	Persönliche Merkmale
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die Staatenschlüssel des Statistischen Bundesamtes (Statistisches Bundesamt, 2019).
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird als Staatsangehörigkeitsmerkmal lediglich die Staatsangehörigkeit vergröbert (nation_gr) bereitgestellt.

5.4.5 Staatsangehörigkeit vergröbert (nation_gr)

Variablenlabel	Staatsangehörigkeit vergröbert
Variablenname	nation_gr
Kategorie	Persönliche Merkmale
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die zu Gruppen zusammengefassten Staatenschlüssel des Statistischen Bundesamtes (Statistisches Bundesamt, 2019).

5.4.6 Ausbildung (ausbildung)

Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Datentyp	Numerisch

Detailbeschreibung	Das Merkmal enthält für BeH-Sätze die von den Arbeitgebern im Rahmen des Meldeverfahrens anzugebende Berufsausbildung. Folgende Ausprägungen kommen vor: 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 11 FH ohne nähere Spezifikation 12 Hochschulabschluss ohne nähere Spezifikation Bei Meldungen nach dem neuen Tätigkeitsschlüssel (siehe Abschnitt 4.2) ist es nicht mehr möglich, FH-Absolventen wie bisher eindeutig zu identifizieren, da der neue Tätigkeitsschlüssel die separate Erfassung dieses Bildungsabschlusses nicht länger erlaubt. Sie werden der Kategorie 12 zugeordnet.
Qualitätshinweis	„Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammenstellt werden. Wenn ein Beschäftigter z.B. durch berufsbegleitende Weiterbildung einen höheren Abschluss erlangt hat, dann wird dieser Statuswechsel wahrscheinlich erst von einem neuen Beschäftigungsbetrieb erfasst. Es ist in der Regel davon auszugehen, dass bei längeren Beschäftigungszeiten innerhalb eines Betriebs die einmal erhobenen personenbezogenen Daten fortgeschrieben werden.“ (Meinken/Koch, 2004, S. 63). Der Anteil der fehlenden Werte steigt in der BeH im Zeitverlauf fast kontinuierlich an. Durch die Einführung des neuen Tätigkeitsschlüssels gab es in 2011 einen stark erhöhten Anteil von rund 51%. Seit 2014 pendelt sich der Anteil der fehlenden Werte in der BeH bei rund 40% ein. Fehlende Werte treten besonders häufig bei folgenden Personengruppen auf: geringfügig Beschäftigte, Personen in Teilzeit, ausländische Beschäftigte und ostdeutsche Arbeitnehmer. Hintergrund ist, dass das Merkmal versicherungsrechtlich keine besondere Bedeutung hat (siehe Meinken/Koch, 2004, S. 63). Für das Merkmal Ausbildung imputiert (ausbildung_imp, siehe Abschnitt 5.4.7) wurde eine Methode angewendet, um fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben zu bereinigen. Diese Variable ist jedoch nur in der Quelle BeH gefüllt und hat andere Ausprägungen als die Variable ausbildung.

5.4.7 Ausbildung (imputiert) (ausbildung_imp)

Variablenlabel	Ausbildung (imputiert)
Variablenname	ausbildung_imp
Kategorie	Persönliche Merkmale
Datentyp	Numerisch

Detailbeschreibung	Das Merkmal Ausbildung (imputiert) ist ein Zusatzmerkmal zum Merkmal Ausbildung (ausbildung) und enthält für BeH-Meldungen ergänzte und harmonisierte Angaben zur Berufsausbildung von Beschäftigten. Das Merkmal bietet damit einen Lösungsansatz für die in Abschnitt 5.4.6 beschriebenen Probleme der Variable Ausbildung (ausbildung) bei BeH-Meldungen. Das Imputationsverfahren wird in Thomsen et al. (2018) beschrieben, welches auf der Arbeit von Fitzenberger et al. (2006) basiert. Da das Merkmal nur Bildungsangaben der BeH-Meldungen nutzt, und weil die Bildungskategorien des alten und des neuen Tätigkeitsschlüssels für das Merkmal Ausbildung (ausbildung) harmonisiert werden mussten, bietet das Merkmal Ausbildung (imputiert) (ausbildung_imp) andere Kategorien als das Merkmal Ausbildung (ausbildung).

5.4.8 Schulausbildung (schule)

Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält den Schulabschluss. Mit der Umstellung des Tätigkeitsschlüssels ändern sich die möglichen Ausprägungen (siehe Abschnitt 4.2). Davor gelten: 5 Volks-/Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung 8 Fachhochschulreife oder Abitur/Hochschulreife 9 Abitur/Hochschulreife Danach gelten: 1 Kein Hauptschulabschluss 4 Hauptschulabschluss/Volksschulabschluss 6 Mittlere Reife 8 Fachhochschulreife oder Abitur/Hochschulreife
Qualitätshinweis	Der Füllgrad nimmt in der BeH über die Jahre kontinuierlich ab und liegt in jüngeren Jahren unter 2/3.

5.5 Informationen zu Beschäftigung

5.5.1 Tagesentgelt/ täglicher Leistungssatz (tentgelt)

Variablenlabel	Tagesentgelt/täglicher Leistungssatz
Variablenname	tentgelt
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	In BeH-Sätzen weist das Merkmal jeweils das Bruttotagesentgelt eines Beschäftigten aus. Es wird berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der (ungesplitteten) Originalmeldung in Kalendertagen. Das Tagesentgelt ist in Euro angegeben.

	Bis 1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt. Entgelte, die unterhalb der Geringfügigkeitsgrenze lagen, wurden nicht gemeldet. Entgelte, welche die Beitragsbemessungsgrenze in der gesetzlichen Rentenversicherung überschritten, werden nur bis zu dieser Grenze gemeldet. In der gesetzlichen Rentenversicherung gibt es zwei Beitragsbemessungsgrenzen. Die Grenze der Knappschaftlichen Rentenversicherung liegt meist höher als die Grenze der Arbeiter- und Angestelltenrentenversicherung. Jedoch ist eine Unterscheidung der beiden Versicherungsträger in den Daten nicht möglich. Seit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren zum 01.04.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze erfasst. Die Beitragsbemessungsgrenze als obere Kappungsgrenze gilt jedoch weiterhin. In einigen Fällen übersteigen die gemeldeten Entgelte die Beitragsbemessungsgrenze allerdings trotzdem. Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnen kann. Dabei ist es dann unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei der Beschäftigungszeit handeln. (Die Entgelt-Informationen hingegen können auf Grund ihrer Versicherungsrelevanz als weniger fehleranfällig gesehen werden.) Die Geringfügigkeits- und Beitragsbemessungsgrenzen unterscheiden sich von Jahr zu Jahr und in West- und Ostdeutschland (maßgeblich ist jeweils der Standort des Betriebs). Eine Übersicht über diese Grenzen findet sich unter https://fdz.iab.de. Eine Angabe von 0 Euro ist auf "Unterbrechungsmeldungen" zurückzuführen. In diesen Zeiten besteht das Beschäftigungsverhältnis zwar rechtlich fort, es wird aber kein Entgelt gezahlt. Dies ist der Fall bei Krankheit nach Ende der Lohnfortzahlung, Mutterschaftszeiten und Sabbaticals. Ab 2013 nimmt die Anzahl der Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) sprunghaft zu (siehe Abschnitt 5.5.12). Es ist davon auszugehen, dass Sonderzahlungen, die vor 2013 mit den Jahresmeldungen angegeben wurden, nun extra gemeldet werden. Daher empfiehlt es sich bei Analysen der Entgelte im Zeitverlauf, die Sonderzahlungen zu den entsprechenden Entgelten von zeitgleichen Beschäftigungsepisoden im gleichen Betrieb dazu zu addieren. Das Tagesentgelt wird mit zwei Nachkommastellen ausgewiesen. Alle Werte größer als 0 und kleiner als 0,01 wurden auf 0,01 aufgerundet. Somit können die o.a. Unterbrechungsmeldungen mit der Bedingung Tagesentgelt = 0 identifiziert werden.

5.5.2 Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)

Variablenlabel	Beruf – ausgeübte/letzte Tätigkeit (KldB 1988)
Variablenname	beruf
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).

	Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifizierung der Berufe 1988“ (Bundesanstalt für Arbeit, 1988), in der ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Ausprägungen. Beschäftigungsmeldungen mit einem Enddatum nach dem 30.11.2011 werden nach dem neuen Tätigkeitsschlüssel 2010 (KldB2010) gemeldet (siehe Abschnitt 4.2). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB1988 zurückgeschlüsselt. Dabei kommt es zu Unschärfen.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.

5.5.3 Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)

Variablenlabel	Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 3-Steller
Variablenname	beruf2010_3
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsgruppe wird durch die ersten drei Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.2). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.

5.5.4 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)

Variablenlabel	Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 4-Steller
Variablenname	beruf2010_4
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch

Detailbeschreibung	Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit". Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010" (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsuntergruppe wird durch die ersten vier Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.2). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird als Berufsangabe gemäß KldB2010 lediglich die Berufsgruppe (beruf2010_3) bereitgestellt.

5.5.5 Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)

Variablenlabel	Anforderungsniveau - ausgeübte/letzte Tätigkeit (KldB 2010)
Variablenname	niveau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit". Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010" (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Das an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.2). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.

Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.

5.5.6 Teilzeit (teilzeit)

Variablenlabel	Teilzeit
Variablenname	teilzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal Teilzeit (teilzeit) unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. Maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Um dieses Problem abzumildern, wurde am IAB eine Imputation der Arbeitszeit für den fraglichen Zeitraum vorgenommen. Nähere Informationen zum Verfahren können Ludsteck/Thomsen (2016) entnommen werden. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf. Hier wird keine Imputation vorgenommen.

5.5.7 Stellung im Beruf und Arbeitszeit (stib)

Variablenlabel	Stellung im Beruf und Arbeitszeit
Variablenname	stib
Kategorie	Informationen zu Beschäftigung
Datentyp	Numerisch
Detailbeschreibung	Die berufliche Stellung eines Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt. Das Merkmal „Stellung im Beruf und Arbeitszeit“ unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. Maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit. Für Teilzeitbeschäftigte wird nur erfasst, ob ihre Stundenzahl eine bestimmte Grenze übersteigt oder nicht. Diese Grenze lag bis 1978 bei 20 Wochenstunden, zwischen 1979 und 1987 bei 15 Wochenstunden und seit 1988 bei 18 Wochenstunden. Nur für Vollzeitbeschäftigte liefert die Variable tatsächlich Angaben zur beruflichen Stellung, wobei unter anderem zwischen vollzeitbeschäftigten Arbeitern und Angestellten sowie Auszubildenden unterschieden wird. Die Unterscheidung zwischen vollzeitbeschäftigten Arbeitern und Angestellten richtet sich ausschließlich nach der Art der Rentenversicherungspflicht (BfA für Angestellte und LVA für Arbeiter). Meister und Poliere werden nur dann in einer eigenen Kategorie erfasst, wenn sie in der Arbeiterrentenversicherung pflichtversichert sind. Die Zuordnung eines Meisters oder Poliers zu den Arbeitern oder Angestellten kann nur über den jeweiligen Rententräger erfolgen. Neben den Auszubildenden, Praktikanten und Volontären zählen auch Anlernlinge, Schüler an Schulen des Gesundheitswesens und Teilnehmer an geförderten Maßnahmen zur beruflichen Fortbildung, Umschulung und betrieblichen Einarbeitung zur Kategorie Beschäftigte in beruflicher Ausbildung. Falls mehrere Schlüsselzahlen in Betracht kommen, ist der Arbeitgeber dazu angehalten, die Einstufung nach der überwiegend ausgeübten Tätigkeit vorzunehmen. Lässt sich diese nicht eindeutig feststellen, ist die Schlüsselzahl der höherwertigen Stellung im Beruf einzutragen (vgl. BA 2005, S. VI).

	Mit der Einführung des neuen Tätigkeitsschlüssels (siehe Abschnitt 4.2) fiel diese Unterscheidung jedoch weg. Die Variable „stib“ ist daher nur für Meldungen gefüllt, die noch vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden.
Qualitätshinweis	In 1991 kommt es aufgrund der Wiedervereinigung zu einer deutlichen Häufung der fehlenden Angaben.

5.5.8 Erwerbsstatus (erwstat)

Variablenlabel	Erwerbsstatus
Variablenname	erwstat
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Für BeH-Sätze entspricht das Merkmal Erwerbsstatus der im neuen Meldeverfahren (DEÜV) ab dem 01.01.1999 erfassten Personengruppe. Sie weist beitrags- oder leistungsrechtliche Besonderheiten des Beschäftigungsverhältnisses aus. Treffen für eine Beschäftigungsmeldung mehrere Schlüssel zu, ist vom meldenden Arbeitgeber der kleinste anzugeben. Mehrheitlich handelt es sich dabei um sozialversicherungspflichtige Beschäftigungsverhältnisse ohne Besonderheiten, die unter der Schlüsselnummer 101 geführt werden. Entsprechend ist es möglich, dass diese Beschäftigungen leicht überschätzt werden. Im Meldeverfahren ist geregelt, dass Wechsel im Erwerbsstatus - z.B. die Übernahme eines Auszubildenden nach dem Ende der Berufsausbildung im Ausbildungsbetrieb - durch eine erneute Meldung erfolgen müssen. Die Personengruppe kann in Beschäftigungsmeldungen, die sich auf die Jahre vor 1999 beziehen, aber erst 1999 oder später eingegangen sind, enthalten sein. Für Meldungen, die vor 1999 eingegangen sind, wird versucht, anhand bestimmter Regeln mit Hilfe der Merkmale „Ausbildung“, „Stellung im Beruf und Arbeitszeit“ und „Beruf“ sowie weiterer Informationen eine Zuordnung zu den Personengruppen durchzuführen. In vielen Fällen ist das jedoch nicht eindeutig möglich. Seit dem 01.04.1999 werden im Meldeverfahren DEÜV auch geringfügig Beschäftigte erfasst. Diese Personengruppe lässt sich über die Ausprägungen 109 und 209 abgrenzen. Für geringfügig Beschäftigte konnten keine Daten vor dem Zeitpunkt der Einführung der Meldepflicht im Jahr 1999 zurückgeschrieben werden.

5.5.9 Gleitzone (gleitz)

Variablenlabel	Gleitzone
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal ist erst ab 2003 gefüllt. Es zeigt an, ob es sich bei der Beschäftigungsmeldung um eine Beschäftigung im Niedriglohnbereich innerhalb der sogenannten Gleitzone mit einem monatlichen Bruttoarbeitsentgelt von 400,01 bis 800,00 € handelt (sog. Midi-Job), für die der Arbeitnehmer nur einen reduzierten Beitrag zum Gesamtsozialversicherungsbeitrag leisten muss. Da Arbeitnehmer mit einem Entgelt in der Gleitzone freiwillig den „normalen“ Sozialversicherungsbeitrag zahlen können, sind nicht automatisch alle Beschäftigungen mit einem entsprechen Entgelt als Gleitzone gekennzeichnet. Die entsprechende gesetzliche Regelung besteht seit dem 01.04.2003.

5.5.10 Leiharbeit (leih)

Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein Leiharbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Im Jahr 2012 beträgt die Missingquote nur noch rund 3 %.

5.5.11 Befristung (befrist)

Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein befristetes oder unbefristetes Arbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Im Jahr 2012 beträgt die Missingquote nur noch rund 3 %.

5.5.12 Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)

Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung übermittelt hat. Allerdings werden in die IEB nicht alle im Rahmen des Meldeverfahrens möglichen Abgabegründe integriert. So beinhalten die IEB nur Meldungen mit Entgeltangaben (also Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen wegen fehlender Entgeltinformationen nicht enthalten sind. Ein Informationsverlust ist damit jedoch nicht verbunden, da die Angaben aus einer Anmeldung mit der darauf folgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden. Die Abgabegründe für Beschäftigungsmeldungen sind nach den Regeln des seit 01.01.1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert.
Qualitätshinweis	Ab 2013 nimmt die Anzahl der Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) sprunghaft zu. Solange ein Beschäftigungsverhältnis besteht, können Sonderzahlungen, die bis März des Folgejahres ausgezahlt werden, mit der normalen Meldung (v.a. Jahresmeldung) für das Vorjahr angegeben werden. Eine gesonderte Meldung mit Abmeldegrund 54 ist dann nicht erforderlich. Bis 2012 konnten die Jahresmeldungen bis Mitte April abgegeben werden, seit 2013 müssen sie nun bis spätestens Mitte Februar erfolgen. Für Sonderzahlungen, die im Februar und März ausgezahlt werden, muss nun eine extra Meldung erfolgen.

5.5.13 Dauer der Arbeitslosigkeit (alo_dau)

Variablenlabel	Dauer der Arbeitslosigkeit
Variablenname	alo_dau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt die Dauer (in Tagen) einer ununterbrochenen Folge von Arbeitslosigkeitsperioden an und ist zu Beginn des Satzes gültig. Es wird auf Basis weiterer Quellen zu Leistungsbezug, Arbeitssuche und Maßnahmenteilnahme in der IEB berechnet (siehe dazu Antoni et al., 2019). DOI: 10.5164/IAB.FDZD.1902.de.v1). Folgende Lücken führen nicht zu einer Unterbrechung der Arbeitslosigkeit: • jede Lücke kleiner oder gleich 7 Tage • Krankheitslücken bis zu 42 Tagen Vor 1997 bedeutet der Wert „0“ nicht, dass keine Arbeitslosigkeit vorliegt, da hier die Quellen ASU/XASU nicht gefüllt sind.

5.6 Ortsangaben

5.6.1 Wohnort - Kreis (wo_kreis)

Variablenlabel	Wohnort - Kreis
Variablenname	wo_kreis
Kategorie	Ortsangaben
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	In BeH- Sätzen liegt der Kreis des Wohnortes erst für die Jahre ab 1999 vor. Darin ist jeweils der Kreis (kreisfreie Stadt oder Landkreis) angegeben, in dem der Versicherte wohnt. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes (NUTS-1) und in der ersten bis dritten Stelle den Regierungsbezirk (NUTS-2). Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. In der Systematik NUTS (Nomenclature des unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3. In der BeH wird der Wohnort jeweils zum Jahresende ermittelt und dann einheitlich für alle Sätze eines Jahres zugespielt. Mit der Dauer des Spells steigt also das Risiko, dass der Wohnort veraltet ist und die Angaben zu späteren Stichtagen falsch sind. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2018 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum

	31.12.2018 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Wohnortes wechselt, ohne dass der/die betreffende Versicherte umgezogen ist.
Qualitätshinweis	Hinsichtlich der Wohnortangaben bestehen für einzelne Beschäftigte Erhebungsungenauigkeiten. Die Meldevorschrift stellt nicht klar, welcher Wohnsitz – Haupt- oder Nebenwohnsitz mit überwiegendem Aufenthaltsort – vom Arbeitgeber zu melden ist. Dies kann in der Beschäftigungsstatistik zum Nachweis von "Fernpendlern" zwischen gemeldetem Hauptwohnsitz und Arbeitsort führen, obwohl die beschäftigte Person ihrer Beschäftigung am Nebenwohnsitz nachgeht, also faktisch nicht pendelt. Im Jahr 2015 kommt es in den Daten zu einer Reduktion der Wohnortwechsel von ca. 10-15% gegenüber dem üblichen Niveau der umliegenden Jahre, wobei regionale Unterschiede auftreten. Der genaue Grund für diese Abweichung konnte bisher nicht ermittelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (wo_bula) bereitgestellt.

5.6.2 Wohnort - Bundesland (wo_bula)

Variablenlabel	Wohnort - Bundesland
Variablenname	wo_bula
Kategorie	Ortsangaben
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort - Kreis auf die 16 deutschen Bundesländer. In BeH- Sätzen liegt das Bundesland des Wohnortes erst für die Jahre ab 1999 vor. Weitere Informationen zum Kreis des Wohnortes finden sich unter Wohnort - Kreis (wo_kreis).

5.6.3 Wohnort - Arbeitsagentur (wo_aa)

Variablenlabel	Wohnort - Arbeitsagentur
Variablenname	wo_aa
Kategorie	Ortsangaben
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Das Merkmal enthält ab 1999 den Agenturbezirk der am Wohnort des Beschäftigten/Kunden zuständigen Arbeitsagentur. Sie wird aus der Wohnadresse ermittelt. Mit der Dauer des Spells steigt also das Risiko, dass der Wohnort veraltet ist und die Angaben zu späteren Stichtagen falsch sind. Es ist für die BeH erst ab 1999 vorhanden. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Angaben für alle Quellen auf den Gebietsstand vom 31.12.2018 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Agenturbezirk nach den Grenzen, welche die Bezirke zum 31.12.2018 hatten. Da sich die Grenzen im Zeitverlauf geändert haben, würden ohne diese Gebietsstandsaktualisierung Fälle auftreten, bei denen der Bezirk einer Person wechselt, ohne dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandsaktualisierung allerdings Berlin dar: Die

	Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt werden konnte. Berlin sollte möglichst als Einheit und nicht getrennt nach Ost und West analysiert werden, da seit der Gebietsreform der Arbeitsämter in Berlin zum 01.07.1997 deren eindeutige Zuordnung zu Ost bzw. West nicht mehr möglich ist. Von der Statistik der BA wird Berlin einheitlich dem Osten zugeordnet.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort des Versicherten liegt, ausgewiesen.

5.6.4 Wohnort - Regionaldirektion (wo_rd)

Variablenlabel	Wohnort - Regionaldirektion
Variablenname	wo_rd
Kategorie	Ortsangaben
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort - Arbeitsagentur auf die Ebene der Regionaldirektionen. Weitere Informationen finden sich dort.

5.7 Betriebsmerkmale

5.7.1 WZ73 3-Steller (w73_3)

Variablenlabel	WZ73 3-Steller
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller) | Wirtschaftsgruppe (2-Steller) | Wirtschaftsklasse (3-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus und ist für die Jahre 1975 bis einschließlich 2002 gefüllt. WS73 steht für die "Klassifikation der Wirtschaftszweige 1973“ (Bundesagentur für Arbeit, 2009). Mittels eines dreistelligen Zahlencodes werden 269 Wirtschaftsklassen unterschieden, wobei die erste Stelle dieses Codes die Wirtschaftsabteilung, insgesamt 10, definiert und die beiden ersten Stellen zusammen die jeweilige Gruppe, insgesamt 95, festlegen. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Die Zuordnung zu der entsprechenden Wirtschaftsklasse erfolgt unter Beachtung der institutionellen Ausrichtung des Betriebes.

5.7.2 WZ93 5-Steller (w93_5)

Variablenlabel	WZ93 5-Steller
Variablenname	w93_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch

Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 5-Steller aus und ist für die Jahre 1999 bis inklusive 2003 gefüllt. WZ93 steht für die „Klassifikation der Wirtschaftszweige, Ausgabe 1993“ (Statistisches Bundesamt, 2002). Die WZ93 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 („Nomenclature génerale des activités économiques dans les communautés européennes“) auf, die vier Gliederungsebenen vorsieht und bezüglich der ersten beiden wiederum auf dem internationalen Standard ISIC Rev.3 („International Standard Industrial Classification of All Economic Activities“) basiert. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w93_3) bereitgestellt.

5.7.3 WZ93 3-Steller (w93_3)

Variablenlabel	WZ93 3-Steller
Variablenname	w93_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 3-Steller aus und ist für die Jahre 1999 bis inklusive 2003 gefüllt. WZ93 steht für die „Klassifikation der Wirtschaftszweige, Ausgabe 1993“ (Statistisches Bundesamt, 2002). Die WZ93 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 („Nomenclature génerale des activités économiques dans les communautés européennes“) auf, die vier Gliederungsebenen vorsieht und bezüglich der ersten beiden wiederum auf dem internationalen Standard ISIC Rev.3 („International Standard Industrial Classification of All Economic Activities“) basiert. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.4 WZ03 5-Steller (w03_5)

Variablenlabel	WZ03 5-Steller
Variablenname	w03_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 5-Steller aus und ist für die Jahre 2003 bis 2008 gefüllt. WZ03 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2003" (Statistisches Bundesamt, 2003). Die

	WZ03 baut ebenso wie die WZ93 auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 auf (siehe Variablenbeschreibung w93_3, w93_5). Die Klassifikationen der Wirtschaftszweige wurden aktualisiert, wobei die Struktur der WZ93 weitgehend beibehalten wurde. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w03_3) bereitgestellt.

5.7.5 WZ03 3-Steller (w03_3)

Variablenlabel	WZ03 3-Steller
Variablenname	w03_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller)| Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 3-Steller aus und ist für die Jahre 2003 bis 2008 gefüllt. WZ03 steht für die "Klassifikation der Wirtschaftszweige Ausgabe 2003" (Statistisches Bundesamt, 2003). Die WZ03 baut ebenso wie die WZ93 auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 auf (siehe Variablenbeschreibung w93_3, w93_5). Die Klassifikationen der Wirtschaftszweige wurden aktualisiert, wobei die Struktur der WZ93 weitgehend beibehalten wurde. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.6 WZ08 5-Steller (w08_5)

Variablenlabel	WZ08 5-Steller
Variablenname	w08_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 5-Steller aus und ist für die Jahre ab 2008 gefüllt. WZ08 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2008" (Statistisches Bundesamt, 2008). Die WZ08 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.2 auf. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w08_3) bereitgestellt.

5.7.7 WZ08 3-Steller (w08_3)

Variablenlabel	WZ08 3-Steller
Variablenname	w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 3-Steller aus und ist für die Jahre ab 2008 gefüllt. WZ08 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2008" (Statistisches Bundesamt, 2008). Die WZ08 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.2 auf. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.8 w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)

Variablenlabel	w73_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w73_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller) | Wirtschaftsgruppe (2-Steller) | Wirtschaftsklasse (3-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus. Von 1975 bis einschließlich 2002 enthält das Merkmal die Originalwerte aus w73_3. Ab 2003 wird die Information entweder fortgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt. Das Merkmal bietet somit eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WS73. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011. Weitere Informationen zur Klassifikation WS73 finden sich bei der Variablenbeschreibung w73_3.

5.7.9 Art der Vervollständigung w73_3 (group_w73_3)

Variablenlabel	Art der Vervollständigung w73_3
Variablenname	group_w73_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w73_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w73_3_gen dem Originalwert von w73_3 entspricht, weiterhin fehlend ist bzw. extrapoliert oder auf Basis von Umschlüsselungstabellen imputiert wurde. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

5.7.10 w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)

Variablenlabel	w93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 3-Steller aus. Von 1998 bis einschließlich 2003 enthält das Merkmal die Originalwerte aus w93_3. Vor 1998 und nach 2003 wird die Information entweder rück- oder fortgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt, sodass das Merkmal eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WZ93 enthält. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011. Weitere Informationen zur Klassifikation WZ93 finden sich bei der Variablenbeschreibung w93_3.

5.7.11 Art der Vervollständigung w93_3 (group_w93_3)

Variablenlabel	Art der Vervollständigung w93_3
Variablenname	group_w93_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w93_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w93_3_gen dem Originalwert von w93_3 entspricht, weiterhin fehlend ist bzw. extrapoliert oder auf Basis von Umschlüsselungstabellen imputiert wurde. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

5.7.12 w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)

Variablenlabel	w08_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w08_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 3-Steller aus. Von 2008 bis 2017 enthält das Merkmal die Originalwerte aus w08_3. Vor 2008 wird die Information rückgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt, sodass das Merkmal eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WZ08 enthält. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011. Weitere Informationen zur Klassifikation WZ08 finden sich bei der Variablenbeschreibung w08_3.

5.7.13 Art der Vervollständigung w08_3 (group_w08_3)

Variablenlabel	Art der Vervollständigung w08_3
Variablenname	group_w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w08_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w08_3_gen dem Originalwert von w08_3 entspricht, weiterhin fehlend ist bzw. extrapoliert oder auf Basis von Umschlüsselungstabellen imputiert wurde. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

5.7.14 Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)

Variablenlabel	Jahr des 1. Auftretens der Betriebsnummer
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Bei dieser Variablen handelt es sich um das erste Auftreten der Nummer des Betriebes im Datensatz. Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde (zur Betriebsnummernvergabe siehe Bundesagentur für Arbeit, 2007, S. 9-11). Ebenfalls könnte es sein, dass der Betrieb bereits vorher existierte, aber keine sozialversicherungspflichtig bzw. ab 1999 auch keine geringfügig Beschäftigten hatte. Ein Betrieb muss nicht zwingend in dem Jahr des ersten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Gründungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.

5.7.15 Erstes Auftreten Betriebsnummer (grd_dat)

Variablenlabel	Erstes Auftreten Betriebsnummer
Variablenname	grd_dat
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Datum
Detailbeschreibung	Bei dieser Variablen handelt es sich um das tagesgenaue erste Auftreten der Nummer des Betriebes in der BeH. Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde (zur Betriebsnummernvergabe siehe Bundesagentur für Arbeit, 2007, S. 9-11). Ebenfalls könnte es sein, dass der Betrieb bereits vorher existierte, aber keine sozialversicherungspflichtig bzw. ab 1999 auch keine geringfügig Beschäftigten hatte.

	Ein Betrieb muss nicht zwingend in dem Jahr des ersten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Gründungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des ersten Auftretens der Betriebsnummer (grd_jahr) bereitgestellt.

5.7.16 Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)

Variablenlabel	Jahr des letzten Auftretens der Betriebsnummer
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Bei dieser Variablen handelt es sich um das letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11). Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.

5.7.17 Letztes Auftreten Betriebsnummer (lzt_dat)

Variablenlabel	Letztes Auftreten Betriebsnummer
Variablenname	lzt_dat
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Datum
Detailbeschreibung	Bei dieser Variablen handelt es sich um das tagesgenaue letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11). Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.

Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) bereitgestellt.

5.7.18 Anzahl Beschäftigte gesamt (az_ges)

Variablenlabel	Anzahl Beschäftigte gesamt
Variablenname	az_ges
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Die Variable enthält die Anzahl aller zum 30.06. eines Jahres zur Sozialversicherung gemeldeten Mitarbeiter eines Betriebes. Seit Einführung der neuen Meldevorschriften im Jahr 1999 sind auch die geringfügig Beschäftigten erfasst. Ruhende Beschäftigungsverhältnisse (Tagesentgelt von Null) sind nicht enthalten.

5.7.19 Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)

Variablenlabel	Anzahl Vollzeit (Normalbeschaeftigte + sonstige)
Variablenname	az_vz
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Die Variable enthält die Anzahl der Personen im Betrieb, die die am 30.06. eines Jahres mit den Personengruppenschlüsseln 101, 140, 143, 105, 106, 112, 113, 114, 118, 119, 120, 149, 201, 203, 205, 999 und YYY und als Vollzeitbeschäftigte gemeldet sind. Azubis, geringfüg Beschäftigte und Personen in Altersteilzeit werden nicht berücksichtigt.

5.7.20 Anzahl geringfügig Beschäftigte (az_gf)

Variablenlabel	Anzahl geringfügig Beschäftigte
Variablenname	az_gf
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Die Anzahl der geringfügig Beschäftigten wird anhand des Personengruppenschlüssels – Ausprägungen 109 und 209 – gebildet. Dieses Merkmal ist erst seit 1999 im Datensatz enthalten, da die geringfügig Beschäftigten erst seit diesem Jahr in das Meldeverfahren zur Sozialversicherung integriert wurden.

5.7.21 Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Variablenlabel	Mittelwert imp. Bruttotagesentgelt Vollzeitbeschaeftigte
Variablenname	te_imp_mw
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch

Detailbeschreibung	Diese Variable enthält den Mittelwert des imputierten Bruttotagesentgeltes der Vollzeitbeschäftigten eines Betriebes. Nicht enthalten sind die Löhne der geringfügig Beschäftigten, der Azubis sowie der Personen in Altersteilzeit. Die Werte sind für alle Jahre in Euro angegeben. Gemäß den Meldevorschriften zur Sozialversicherung müssen Arbeitgeber angeben, wie hoch das beitragspflichtige Bruttoarbeitsentgelt des Beschäftigten für einen bestimmten Zeitraum war (Zeitraumentgelt). Bis Ende 1998 hatten Arbeitgeber nur den sozialversicherungspflichtigen Bruttolohn zu melden. Damit wurden nur Arbeitsentgelte erfasst, die oberhalb der Geringfügigkeits- und unterhalb der Beitragsbemessungsgrenze lagen. Seit 1999 werden im Rahmen des neuen Meldeverfahrens auch Entgelte gemeldet, die unterhalb der Geringfügigkeitsgrenze liegen. Bruttolöhne, die oberhalb der Beitragsbemessungsgrenze liegen, werden jedoch weiterhin gekappt. Zur Ermittlung des Bruttotageslohnes wird das Zeitraumentgelt durch die Anzahl der Kalendertage des Zeitraums geteilt. Zur Berechnung der Mittelwerte wurden diese zensierten Löhne imputiert (vgl. Abschnitt 3.1.3.3 in Schmucker et al. 2018). Diese Daten wurden dann auf Betriebsebene aggregiert. Die Werte sind auf zwei Dezimalstellen gerundet. Aufgrund des "storage types" in Stata werden jedoch weitere Dezimalstellen angezeigt, die nicht korrekt sind.

5.7.22 Arbeitsort - Kreis (ao_kreis)

Variablenlabel	Arbeitsort - Kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Angegeben ist jeweils der Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes (NUTS-1), in der ersten bis dritten Stelle den Regierungsbezirk (NUTS-2). Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. In der Systematik NUTS (Nomenclature des unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2017 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2017 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Betriebsortes wechselt, ohne dass dieser Betrieb seinen Standort verlagert hat.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (ao_bula) bereitgestellt.

5.7.23 Arbeitsort - Bundesland (ao_bula)

Variablenlabel	Arbeitsort - Bundesland
Variablenname	ao_bula
Kategorie	Ortsangaben

Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Angegeben ist jeweils das Bundesland, in dem der Betrieb angesiedelt ist. Dieses Merkmal wird aus der Kreiskennziffer (ao_kreis) generiert. Die ersten beiden Stellen der Kreiskennziffer bezeichnen das Bundesland (NUTS-1).

6 Literatur

Antoni, Manfred; Schmucker, Alexandra; Seth, Stefan; vom Berge, Philipp (2019): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2017. FDZ-Datenreport, 02/2019 (de), Nürnberg. DOI:10.5164/IAB.FDZD.1902.de.v1

Bender, Stefan / Hilzendegen, Jürgen / Rohwer, Götz / Rudolph, Helmut (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.

Bertat, Thomas / Dundler, Agnes / Grimm, Christopher / Kiewitt, Jochen / Schomaker, Christine / Schridde, Dr. Henning / Zemann, Dr. Christian (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, URL: https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erbebungsinhalte-Arbeitszeita-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf (Zugriff: 12.06.2019).

Bundesagentur für Arbeit (Hrsg.) (2005): Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. Ausgabe Januar 2005, Nürnberg.

Bundesagentur für Arbeit (Hrsg.) (2007): Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung. Ausgabe Dezember 2007, Nürnberg.

Bundesagentur für Arbeit (2009): Klassifikation der Wirtschaftszweige 1973, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirtschaftszweige-1973-2003-Nav.html (Zugriff: 12.06.2019).

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html (Zugriff: 12.06.2019).

Bundesanstalt für Arbeit (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html (Zugriff: 12.06.2019).

Cramer, Ulrich (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv 69: S. 56-68.

Eberle, Johanna/ Jacobebbinghaus, Peter/ Ludsteck, Johannes/ Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes * Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport, 05/2011 (en), Nürnberg.

Fitzenberger, Bernd/ Osikominu, Aderonke/ Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.

Ludsteck, Johannes/ Thomsen, Ulrich (2016): Imputation of the Working Time Information for the Employment Register Data. FDZ Methodenreport 01/2016 (en), Nürnberg.

Meinken, Holger/ Koch, Iris (2004): BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.

Paulus, Wiebke/ Matthes, Britta (2013): Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport, 08/2013 (de), Nürnberg.

Schmucker, Alexandra/Eberle, Johanna/ Ganzer, Andreas/Stegmaier, Jens/ Umkehrer, Matthias (2018): Betriebs-Historik-Panel 1975-2016. FDZ-Datenreport, 01/2018 (de), Nürnberg.

Statistisches Bundesamt (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ2003.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2019): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5 (Zugriff: 12.06.2019).

Thomsen, Ulrich/ Ludsteck, Johannes/ Schmucker, Alexandra (2018): Skilled or unskilled - Improving the information on qualification for employee data in the IAB Employee Biography. FDZ-Methodenreport, 09/2018 (en), Nürnberg.

Wermter, Winfried/ Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt – und Berufsforschung 4/88, S. 468-482.

7 Anhang

7.1 Auszählungen

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in gesonderten Dateien unter https://fdz.iab.de.

7.2 Abkürzungsverzeichnis

AA	Agentur für Arbeit / Arbeitsamt
ASU	Arbeitsuchendenhistorik
BA	Bundesagentur für Arbeit
BeH	Beschäftigtenhistorik
BfA	Bundesversicherungsanstalt für Angestellte
BG	Bedarfsgemeinschaft
BHP	Betriebs-Historik-Panel
BMAS	Bundesministerium für Arbeit und Soziales
DEÜV	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Datenerfassungs- und -übermittlungsverordnung –
DEVO	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –
DIM	Daten- und IT-Management
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung
EDV	Elektronische Datenverarbeitung
FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit
IAB	Institut für Arbeitsmarkt- und Berufsforschung
IEB	Integrierte Erwerbsbiografien
ISIC	International Standard Industrial Classification of All Economic Activities
LeH	Leistungsempfängerhistorik
LHG	Leistungshistorik Grundsicherung
LVA	Landesversicherungsanstalt
MTH	Maßnahmeteilnahmehistorik
NACE	Nomenclature génerale des activités économiques dans les communautés européennes
NUTS	Nomenclature des unités territoriales statistiques
SGB	Sozialgesetzbuch

SIEED	Stichprobe der Integrierten Arbeitsmarktbiografien
XASU	Arbeitsuchendenhistorik aus XSozial-BA-SGB II
XMTH	Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II

Impressum

FDZ-Datenreport 14|2020

Veröffentlichungsdatum
18. November 2020

Herausgeber
Forschungsdatenzentrum
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Straße 104
90478 Nürnberg

Rechte
Nachdruck - auch auszugsweise - nur mit Genehmigung des FDZ gestattet

Bezugsmöglichkeit dieses Dokuments
http://doku.iab.de/fdz/reporte/2020/DR_14-20.pdf

Dokumentation Version
SIEED7518_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2014.de.v1

Datensatz Version
SIEED 7518 v1, DOI: 10.5164/IAB.SIEED7518.de.en.v1

Bezugsmöglichkeit aller Veröffentlichungen der Reihe „FDZ-Datenreport“
https://fdz.iab.de/de/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx

Website
https://fdz.iab.de/

Rückfragen zum Inhalt
Philipp vom Berge
Telefon: 0911 179-5020
E-Mail: Philipp.vom-Berge@iab.de