

Heiko Stüber\* and Stefan Seth

# The Public Release Data of the Administrative Wage and Labor Market Flow Panel

<https://doi.org/10.1515/jbnst-2017-1002>

## 1 Introduction


The Administrative Wage and Labor Market Flow Panel (AWFP) is a dataset on labor market flows and stocks for the universe of German establishments (*Betriebe*). It contains data on job flows (changes in the number of employees per establishment), worker flows (information about the hiring and firing activity), and wages for each establishment. The AWFP was generated within the framework of the projects “Custom Shaped Administrative Data for the Analysis of Labour Market” and “Wages, Heterogeneities, and Labor Market Dynamics”. Both projects are part of the German Science Foundation (DFG) priority program “The German Labor Market in a Globalized World” (SPP 1764).

This article introduces the aggregated public release data of the AWFP. The datasets include among other things information on job and worker flows for groups of firms and thus can be used to study the cyclical dynamics (of fractions) of the labor market in terms of turnover and churning.

So far the (aggregated) AWFP data have been used in three papers: First, Merkl and Stüber (2016) analyze the link between establishment-specific wage cyclicality and establishment-specific labor market flow dynamics. Second, Merkl and Stüber (2017) link the AWFP data with the IAB Establishment Panel and test whether the comovements between establishments’ value added /wage and hires /separations are in line with a simple theoretical framework. Third, Bachmann et al. (2017) use aggregated AWFP data – similar to the public release

---

**\*Corresponding author: Heiko Stüber**, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Lange Gasse 20, 90403 Nürnberg, Germany; and Institute for Employment Research (IAB), Regensburger Straße 104, 90478 Nürnberg, Germany, E-mail: Heiko.Stueber@iab.de  
**Stefan Seth**, Institute for Employment Research (IAB), Regensburger Straße 104, 90478 Nürnberg, Germany

 Stüber and Seth, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.

data – to establish stylized facts on the cyclical nature of worker and job flows in Germany.

The article is organized as follows. Section 2 introduces the AFWP and gives important definitions. Section 3 introduces the aggregated public release datasets of the AFWP and Section 4 gives an outlook for scheduled and intended future datasets based on the AFWP.

## 2 The administrative wage and labor market flow panel

The public release data of the AFWP are generated by aggregating information from the AFWP. In this section we therefore briefly introduce the AFWP and give some important definitions. We focus on aspects that are relevant for the public release data. For a detailed description of the AFWP refer to the data report (Seth and Stüber 2017).

### 2.1 Data basis and data preparation

The AFWP measures employment, labor flows, and wage data for the universe of German establishments for the years 1975–2014. Its main data source is the Employment History (*Beschäftigten-Historik*, BeH) of the IAB.

#### 2.1.1 The employment history (BeH)

The BeH is an individual-level dataset covering all workers in Germany subject to social security from 1 January 1975 to 31 December 2014.<sup>1</sup> The information in the BeH originates from the German notification procedure for social security. This mandatory procedure requires employers to keep the social security agencies informed about their employees by reporting any start or end of employment and by annually confirming existing employment relationships. The main types of employees not liable to social security and thus not covered by the BeH are civil servants (*Beamte*), military personnel, and the self-employed.

---

<sup>1</sup> East Germany is included since 1991, marginal part-time workers (*geringfügig Beschäftigte*) since 1999.

From the BeH, the AWFPP aggregates worker flows, job flows, and wage information to the establishment level such that an establishment becomes the observational unit. Before this aggregation, the BeH data are subject to validation procedures briefly described below.

### 2.1.2 Data preparation and data quality

The validation procedures applied to the BeH data are identical to the ones used for generating the IAB Establishment History Panel 1975-2014 (*Betriebs-Historik-Panel*, BHP). The BHP data report (Schmucker et al. 2016) provides a detailed presentation of all procedures; here we only provide a brief overview.<sup>2</sup>

The BeH includes information for employees from West Germany from 1975 onwards. Social security notifications for East Germany can only be assumed to be complete from 1993 onwards, therefore analyses of East German establishments should not begin before 1993.

The education and training variable, which exhibits a high level of missing values in the BeH, is improved by means of an imputation procedure using a deterministic replacement rule. After applying this procedure, hardly any values are still missing.

Before 1984 establishments could decide whether or not they included one-off payments (such as holiday pay, 13th monthly salary etc.) in their annual notifications. In 1984, it was made mandatory to report one-off payments as part of the annual earnings subject to social security contributions. This led to an increase in the average daily wage. In particular, the proportion of wages above the upper earnings limit increases considerably in 1984 (cf. Bender et al. 1996).

In some cases, however, the special payment is reported separately even after 1984 (notification reason 54). These payments, too, have to be taken into account when calculating the earnings data of an establishment; for this, the earnings of the extra notifications are added to the earnings of the regular notifications.

In social security notifications, earnings are only reported up to the upper earnings limit for statutory pension insurance contributions. This means that approximately 10% of the information of full-time employees' earnings are censored. This would lead to biased results due to aggregation because means of earnings are biased if the censored observations are not included in the calculation or if censored values are replaced by the censoring limit.

---

<sup>2</sup> The AWFPP data report (Seth and Stüber 2017) also provides a detailed discussion of the procedures, but not as detailed as Schmucker et al. (2016).

Therefore the information on earnings (average daily wages) were imputed before the wage statistics are calculated.

Especially in 1999, a significant increase in notifications of part-time employment can be observed. This is caused both by an actual increase in part-time work as well as by the fact that since 1999 employment notifications have generally been filed more correctly. Also, in about 10% of the notifications submitted between December 1st 2011 and May 31st 2012 the information regarding working hours is missing. We used a logit model to impute the missing information. In the spring of 1984, strike-related lockouts occurred in establishment in Hessen and Baden-Wuerttemberg in the two industries “manufacture of motor vehicles, motor vehicle engines” and “manufacture of parts and accessories for motor vehicles”. The lockouts are reflected in the data in the form of gaps in employment. These gaps frequently overlap one reference date for the quarterly frequency datasets (June 30, 1984); ignoring them would result in considerable distortions for the two industries in the federal states affected. Therefore the gaps were filled using a heuristic.

## 2.2 The aggregation of the employment history data

After applying the validation procedures to the BeH data on individuals, we aggregated the BeH data at establishment level using the establishment ID. Therefore an establishment becomes the observational unit in the AWF. P.

The stocks and flows in the AWF. P. are generally calculated on a “regular worker” basis. In the next subsection we will define the notion “regular worker” and give the AWF. P. standard definition of how stocks and flows are calculated.

## 2.3 Definitions

The stocks and flows in the AWF. P. are calculated on a “regular worker” basis. A **regular worker** (henceforth also “worker” or “employee”) is a full-time employee without specific characteristics. Thus all (marginal) part-time employees, apprentices, employees in partial retirement, interns, etc. are not accounted for as regular workers.

The **stock** of employees of an establishment in some period  $t$  equals the number of employees on the last day of period  $t$ . Inflows and outflows are also calculated using the “end-of-period” definitions.

The number of **inflows** of employees of an establishment for period  $t$  equals the number of employees who were regularly employed on the last day of period  $t$  but were not on the last day of the preceding period,  $t-1$ .

The number of **outflows** of employees of an establishment for period  $t$  equals the number of employees who were regularly employed on the last day of the preceding period ( $t-1$ ) but were not on the last day of period  $t$ .

Note that a worker counted as an inflow is not necessarily a new hire. For instance, an apprentice who becomes a regular worker represents an inflow because an apprentice is not a regular worker. Analogously, a worker counted as an outflow might remain employed in the same establishment. A regular worker who, for instance, reduces hours and changes to a part-time job represents an outflow. Further note that employees who join an establishment and leave it again between two reference dates are not recorded by this flow concept.

## 3 The public release data of the administrative wage and labor market flow panel

The public release datasets of the AWFPP are generated by aggregating the establishment information from the AWFPP. Thus far, six public release datasets are available. Each dataset is available at the quarterly and annual frequency.

### 3.1 Data preparation and data generation

#### 3.1.1 Data preparation

The data preparation procedures described below are applied to the AWFPP establishment data before aggregating it to obtain the public release datasets.

**Dummy for East Germany (east):** For some establishments the location information is sometimes missing. For establishments with at least one valid observation, the missing information is imputed through writing available information forward and backward in time. The establishments that have missing information in all observations are dropped. A few establishments switch between East and West Germany (or vice versa). These establishments are dropped in all datasets as well.

**Establishment size category (size\_cat):** Each establishment in each period is assigned one of eight establishment size categories using the establishment size of the previous period (see Table 1). We cannot assign size categories for the first

observation period, which is why this period is missing in all public release datasets: In yearly datasets the year 1975 (1993) is missing for West (East) Germany, and in quarterly datasets the first quarter of 1975 (1993) is missing for West (East) Germany.

**Table 1:** Establishment size categories.

Establishment size category	Stock of regular workers in period $t-1$ ( $J_{t-1}$ )	Establishment size category	Stock of regular workers in period $t-1$ ( $J_{t-1}$ )
0	0	4	50–99
1	1–4	5	100–249
2	5–9	6	250–499
3	10–49	7	>500

**Industry Classification Category:** a classification code of economic activity is assigned to each establishment. In the public release data, we use the imputed establishment industry classification of 1993 (see Eberle et al. 2011). For some establishments this information is sometimes missing. For establishments with at least one valid observation, missing information is imputed through writing available information forward and backward in time. Establishments that have missing information in all observations are dropped. We aggregate the classification of economic activity to industry classification categories as shown in Table 2. Because we observe only a few establishments in “extra territorial organizations and bodies” (see Table 2), these are dropped in all public release datasets.

**Table 2:** Industry classification category.

Industry classification category	German Classification of Economic Activities, Edition 1993 (WZ 93)
1	A Agriculture, hunting, and related service activities B Fishing
2	C Mining and quarrying
3	D Manufacturing
4	E Electricity, gas, and water supply
5	F Construction
6	G Wholesale and retail trade; repair of motor vehicles, motorcycles, and personal and household goods
7	H Hotels and restaurants
8	I Transport, storage, and communication
9	J Financial intermediation

(continued)

Table 2: (continued)

Industry classification category	German Classification of Economic Activities, Edition 1993 (WZ 93)
10	K Real estate, renting, and business activities
11	L Public administration and defence; compulsory social security
12	M Education
13	N Health and social work
14	O Other community, social, and personal service activities
15	P Private households with employed persons
16 (dropped)	Q Extra-territorial organizations and bodies

### 3.1.2 Data generation

Three time series from the AWFp are used to construct most data for the public release datasets: First, the stock of employees ( $J_{it}$ ) from establishment  $i$  at the end of the period  $t$  (quarter or year). Second, the number of hires ( $H_{it}$ ) at an establishment and third, the number of separations ( $S_{it}$ ).

Using these variables, we compute the net job flow for each establishment in each period,  $JF_{it} = J_{it} - J_{it-1}$ . When an establishment decreases employment ( $JF_{it} < 0$ ) within a period, we count this as job destruction,  $JD_{it}$ . When employment increases ( $JF_{it} > 0$ ), we count this as job creation,  $JC_{it}$ .

From the AWFp we further use the mean daily wage for all regular workers ( $\overline{w_{it}}$ ), the mean daily wage for inflows ( $\overline{w_{it}^{in}}$ ), the stock of regular workers ( $J_{it}$ ), and the number of inflows ( $J_{it}^{in}$ ) to calculate the wage sums ( $w_{it}$  and  $w_{it}^{in}$ ) for each establishment in each period:  $w_{it} = \overline{w_{it}}J_{it}$  and  $w_{it}^{in} = \overline{w_{it}^{in}}J_{it}^{in}$ . We then aggregate data to higher aggregates (e. g., East and West Germany) and calculate the mean daily wage (for all workers and inflows) for the respective aggregation level:  $\overline{W}_t = \sum w_{it} / \sum J_{it}$  and  $\overline{W}_t^{in} = \sum w_{it}^{in} / \sum J_{it}^{in}$ .

## 3.2 Variables of the Public Release Datasets

### Index of year

Variable name	a
Detailed description	Index for the year (a = 1 being the year 1975).

### Index of quarter

---

Variable name	q
Detailed description	Index for the quarter (q = 1 being the first quarter of 1975).

---

### Stock of establishments

---

Variable name	est_counter
Detailed description	The number of establishments.

---

### East Germany dummy

---

Variable name	east
Detailed description	Dummy (1 for East Germany and 0 for West Germany).

---

### Establishment size category

---

Variable name	size_cat
Detailed description	Indicates the establishment size category (see Table 1).

---

### Industry classification category

---

Variable name	w93_imp
Detailed description	Indicates the industry classification category (see Table 2).

---

### Stock of workers

---

Variable name	st_eop
Detailed description	Stock of workers (see Section 2.3).

---

### Inflows using the standard end-of-period definition

---

Variable name	in_eop
Detailed description	Sum of the inflows (see Section 2.3).

---



## Outflows using the standard end-of-period definition

---

Variable name	out_eop
Detailed description	Sum of the outflows (see Section 2.3).

---

## Mean wage of all regular workers

---

Variable name	dw_mean
Detailed description	Mean daily wage of all workers.

---

## Mean wage of inflows

---

Variable name	dw_mean_in
Detailed description	Mean daily wage of inflows.

---

## Job creation

---

Variable name	JC
Detailed description	Aggregated job creation.

---

## Job destruction

---

Variable name	JD
Detailed description	Aggregated job destruction.

---

## 3.3 Available datasets

At the moment, six public release datasets are available. Each dataset is available on the quarterly and annual frequency. The datasets on the quarterly (annual) frequency cover the period 2nd quarter 1975 to 4th quarter 2014 (1976–2014) for West Germany and 2nd quarter 1993 to 4th quarter 2014 (1994–2014) for East Germany. We intend to release further aggregated public release datasets of the AWF<sup>P</sup> if inquired.

All datasets contain the following variables: stock of establishments (est\_counter), stock of workers (st\_eop), inflows of worker (in\_eop), outflows of workers (out\_eop), job creation (JC), job destruction (JD), mean wage of

regular workers (*dw\_mean\_all*), the mean wage of inflows (*dw\_mean\_in*), and the index of year (*a*). The index of quarter (*q*) is only included in datasets on the quarterly frequency.

**Dataset 1 – Germany:** the dataset contains AWFП establishment data aggregated to the level of Germany.

**Dataset 2 – West and East Germany:** the dataset contains AWFП establishment data aggregated to the level of West Germany and East Germany. Note that Berlin is not included in this dataset.

**Dataset 3 – Germany by establishment size category:** the dataset contains AWFП establishment data aggregated to establishment size categories for Germany.

**Dataset 4 – West and East Germany by establishment size category:** the dataset contains AWFП establishment data aggregated to establishment size categories for West Germany and East Germany. Note that Berlin is not included in this dataset.

**Dataset 5 – Germany by establishment industry classification category:** the dataset contains AWFП establishment data aggregated to establishment industry classification categories for Germany.

**Dataset 6 – West and East Germany by establishment industry classification category:** the dataset contains AWFП establishment data aggregated to establishment industry classification categories for West Germany and East Germany. Note that Berlin is not included in this dataset.

### 3.4 Data access

The public release datasets of the AWFП are downloadable in the category Data (*Daten*) from the IAB website (<http://www.iab.de/en/daten.aspx>) and the website of the Chair of Macroeconomics at the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (<http://www.makro.wiso.fau.de/daten>).

## 4 Outlook: AWFП extensions

We will provide AWFП extensions for datasets offered by the Research Data Centre (FDZ) of the German Federal Employment Agency at the IAB. At the moment we evaluate which variables will be contained in the various AWFП extensions. The FDZ data and the corresponding AWFП extensions may be analyzed in the context of a research visit at the FDZ and subsequent remote data access. In order to be able to use the data, it is first necessary to submit an application to the FDZ.

## 4.1 Extension for the establishment history panel

The Establishment History Panel (*Betriebs-Historik-Panel*, BHP) is composed of cross sectional datasets since 1975 for West Germany and 1992 for East Germany. Every cross section contains all the establishments in Germany which are covered by the BeH on June 30th. The BHP contains, for example, information about the industry classification and the location of an establishment. It reports inter alia the numbers of employees subject to social security and marginal part-time employees, both in total and broken down by, e. g., gender, age, and qualification. For an introduction to the BHP see Schmucker et al. (2016).

Researchers may work with a 50% random sample of the BHP. Certain information from the AWFP will be available as extension for the random sample of the BHP, scheduled to be available in 2018.<sup>3</sup>

## 4.2 Extension for the IAB establishment panel /the linked-employer-employee data

The IAB Establishment Panel is an annual representative survey on various topics such as the determinants of labor demand. It has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany. For an introduction to the IAB Establishment Panel see, e. g., Fischer et al. (2009). The IAB Establishment Panel is also available linked with individual administrative data as the Linked Employer/Employee Data (LIAB).<sup>4</sup>

We intend to provide certain information from the AWFP as extensions for the IAB Establishment Panel (*IAB-Betriebspanel*) and hence the LIAB, too.

## 4.3 Extension for the German job vacancy survey

The German Job Vacancy Survey of the IAB (*IAB-Stellenerhebung*) is an annual survey of up to 15,000 establishments.<sup>5</sup> The questionnaire asks establishments how many vacant positions they have by qualification categories. In addition,

---

<sup>3</sup> For a data report on the recent BHP, see Schmucker et al. (2016).

<sup>4</sup> See [http://fdz.iab.de/en/Integrated\\_Establishment\\_and\\_Individual\\_Data/LIAB.aspx](http://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/LIAB.aspx) for more information.

<sup>5</sup> See [http://fdz.iab.de/en/FDZ\\_Establishment\\_Data/IAB\\_Job\\_Vacancy\\_Survey.aspx](http://fdz.iab.de/en/FDZ_Establishment_Data/IAB_Job_Vacancy_Survey.aspx) for more information.

the survey contains detailed information on the last actual hire, such as the beginning and the end of the search process (i. e., search duration), the used search channels (e. g., newspapers, online channels, Federal Employment Agency), the hours spent for filling the position, and the number of suitable applicants. The Job Vacancy Survey contains direct measures of recruiting intensity (such as the number of search channels or the money spent on search).

We intend to provide certain information from the AWFPP as extensions for the Job Vacancy Survey of the IAB from 2010 onwards.

**Acknowledgments:** We gratefully acknowledge support from the German Research Foundation (DFG) under priority program “The German Labor Market in a Globalized World” (SPP 1764). For helping us with the conceptual design of the AWFPP, we thank Christian Bayer, Christian Merkl, and Felix M. Wellschmied.

## References

- Bachmann, R., C. Bayer, C. Merkl, S. Seth, H. Stüber, F. Wellschmied (2017), Job and Worker Flows: New Stylized Facts for Germany. FAU Discussion Papers in Economics 02/2017.
- Bender, S., J. Hilzendege, G. Rohwer, H. Rudolph (1996), Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- Eberle, J., P. Jacobebbinghaus, J. Ludsteck, J. Witter (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.
- Fischer, G., F. Janik, D. Müller, A. Schmucker (2009), The IAB Establishment Panel – Things Users Should Know. Schmollers Jahrbuch 129: 133–148.
- Merkl, C., H. Stüber (2016), Wage Cyclicalities and Labor Market Dynamics at the Establishment Level: Theory and Evidence. FAU Discussion Papers in Economics 12/2016.
- Merkl, C., H. Stüber (2017), Value Added, Wages, and Labor Market Flows at the Establishment Level. FAU Discussion Papers in Economics 03/2017.
- Schmucker, A., S. Seth, J. Ludsteck, J. Eberle, A. Ganzer (2016), Establishment History Panel 1975–2014. FDZ-Datenreport 03/2016 (en), Nürnberg.
- Seth, S., H. Stüber (2017), Administrative Wage and Labor Market Flow Panel (AWFP) 1975–2014. FAU Discussion Papers in Economics 1/2017.