

Lochner, Benjamin; Wolter, Stefanie; Seth, Stefan

**Article**

## AKM effects for German labour market data from 1985 to 2021

Journal of Economics and Statistics

**Provided in Cooperation with:**

De Gruyter Brill

*Suggested Citation:* Lochner, Benjamin; Wolter, Stefanie; Seth, Stefan (2024) : AKM effects for German labour market data from 1985 to 2021, Journal of Economics and Statistics, ISSN 2366-049X, De Gruyter Oldenbourg, Berlin, Vol. 244, Iss. 4, pp. 425-431, <https://doi.org/10.1515/jbnst-2023-0018>

This Version is available at:

<https://hdl.handle.net/10419/333290>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

## Data Observer

Benjamin Lochner, Stefanie Wolter\* and Stefan Seth

# AKM Effects for German Labour Market Data from 1985 to 2021

<https://doi.org/10.1515/jbnst-2023-0018>

Received March 27, 2023; accepted March 28, 2023

**Abstract:** This article describes the processing and accessibility of the person and establishment fixed wage effects in German administrative data. These effects have been estimated following the approach of Abowd, J., Kramarz, F., and Margolis, D. (1999. High wage workers and high wage firms. *Econometrica* 67: 251–333) and Card, D., Heining, J., and Kline, P. (2013. Workplace heterogeneity and the rise of West German wage inequality. *Q. J. Econ.* 128: 967–1015). They can be linked to most of the available administrative datasets provided by the Research Data Center (FDZ) of the German Federal Employment Agency at the Institute for Employment Research (IAB). They are available for different time intervals from 1985 until 2021. These effects have been used in numerous articles that deal with the contributions of workers and establishments to earnings inequality.

**Keywords:** wages, worker and firm fixed effects, AKM, German administrative microdata, labour market data

**JEL Classification:** J31, J62, C23

## 1 Introduction

In an influential article, Card et al. (2013) (henceforth CHK) study the role of establishment-specific wage premiums for wage inequality in Germany. Their approach builds upon a framework initially proposed by Abowd et al. (1999)

---

**\*Corresponding author: Stefanie Wolter**, Institut für Arbeitsmarkt-und Berufsforschung, Nürnberg, Germany, E-mail: Stefanie.Wolter@iab.de

**Benjamin Lochner**, Institut für Arbeitsmarkt-und Berufsforschung, Nürnberg, Germany; Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg, Germany; Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA), Bonn, Germany; and The Labor and Socio-Economic Research Center (LASER), Nuremberg, Germany

**Stefan Seth**, Institut für Arbeitsmarkt-und Berufsforschung, Nürnberg, Germany

(henceforth AKM). It uses matched administrative employer–employee data to estimate a statistical model of wage determination, where wages are expressed as the sum of worker effects, establishment effects, time-varying covariates, and idiosyncratic error terms. Since then, numerous articles<sup>1</sup> have used this technique to answer many related research questions.

The Research Data Center (IAB-FDZ) of the German Federal Employment Agency at the Institute for Employment Research (IAB) has applied the CHK/AKM approach to the universe of German employment data and estimated the person and establishment fixed effects (henceforth AKM effects). The IAB-FDZ has offered these effects to external data users since 2015 and has recently updated the procedure and the corresponding data files. They are available for the most common administrative data products, such as the Sample of Integrated Employment Biographies (SIAB; see Frodermann et al. 2021) or the Linked Employer–Employee Data from the IAB (LIAB; see Ruf et al. 2021). Currently, the AKM effects are available for different time windows from 1985 until 2021. Users with an unexpired data usage agreement can access the effects in addition to their primary data product.

This article describes how these effects are estimated and processed, how they can be linked to the existing available data, and for which research topics the data might be relevant. It is organized as follows. Section 2 explains the data sources. Section 3 explains the initial data preparation that has been applied in order to estimate the AKM effects. Section 4 explains the details of the estimation procedure. Section 5 provides descriptive results about the estimated effects. Section 6 explains the data structure and how to access them. Section 7 summarizes and gives an overview of potential research questions.

## 2 Data Source

The data source that is used to estimate the AKM effects is the IAB Employment History File (BEH) which originates from the Integrated Employment Biographies (IEB) which are processed by the IAB.

The BEH covers all employees who are subject to social security at least one day in a given year. Since 1999, the data also cover employees in marginal part-time employment and unpaid family workers. Workers' employment spells are supplemented by establishment information from the Establishment History Panel ("Betriebs-Historik-Panel", see Ganzer et al. 2022), which aggregates establishment-level variables (with June 30 each year as the record day). The IEB unites data from different data sources, each of which may contain information from different

---

1 Up until March 2023, according to Google Scholar Abowd et al. (1999) has been cited over 3600 times.

administrative procedures. The main data source is the integrated notification procedure for health, pension, and unemployment insurance.

The BeH covers the majority of the German workforce, only excluding civil servants and the self-employed.

The main ingredients that the CHK/AKM approach builds upon are the BeH information on employment and total earnings. Since the identification of the effects relies on worker mobility, being able to track workers is crucial. The BeH offers an ideal laboratory because each worker and establishment has distinct identification numbers.

### 3 Data Preparation

The data preparation mostly follows CHK, with a few differences, which we explain in detail below. From the universe of BeH employment spells, we restrict the sample to full-time aged 20 to 60. For each of these workers, we identify the main job in a given year as the job with the highest total earnings (including any bonus payments).

In the BeH data, earnings are right-censored at the contribution assessment ceiling.

(“Beitragsbemessungsgrenze”).

We define a wage observation as censored whenever the reported wage is higher than 98 % of the censoring thresholds. Following CHK, we fit a series of Tobit regressions to impute the right tail of the wage distribution. We estimate the Tobit regressions by year, sex, education, and age group.<sup>2</sup> As opposed to CHK, we do not restrict our sample to West Germany and males.

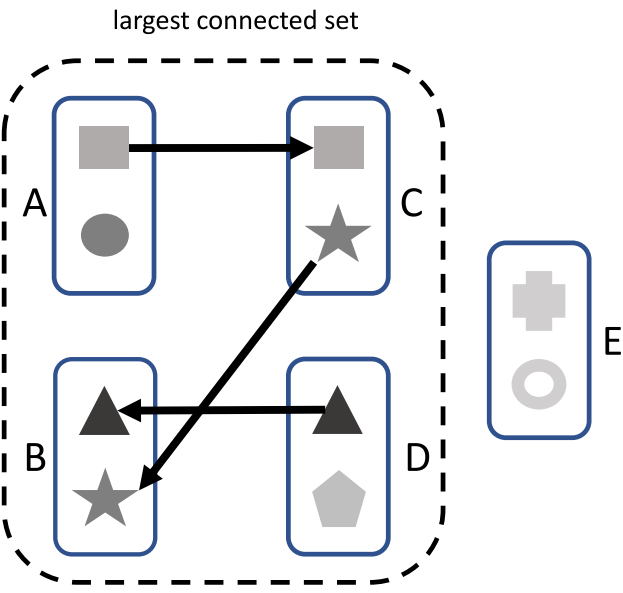
The estimation is performed on the largest connected set of movers that contains the group of establishments that are connected by worker mobility over each sample interval.

Figure 1 illustrates the concept of the largest connected set with four connected (A, B, C, D) and one unconnected establishment (E). We estimate the AKM model for five non-overlapping time intervals: 1985–1992, 1993 – 1999, 2000 – 2006, 2007 – 2013, and 2014 – 2021.

Table 1 shows the number of workers and establishments that are in the largest connected set.

---

<sup>2</sup> We also account for age, the mean log wage of the worker in other years, the fraction of censored wages in other years, the number of full-time employees at the current establishment and its square, a dummy for eleven or more employees, mean years of schooling and fraction of university graduates at the current establishment, mean log wage of co-workers and fraction of co-workers with censored wages, a dummy for individuals observed only one year between 1985 and 2021, a dummy for employees of one-worker establishments, and a dummy for East Germany.



**Figure 1:** Illustration of the largest connected set. A, B, C, and D illustrate different establishments, the symbols illustrate different workers. The arrows represent worker transitions between establishments. Source: Own representation.

**Table 1:** Number of workers and establishments in each time interval.

|                          | 1985–1992  | 1993–1999  | 2000–2006  | 2007–2013  | 2014–2021  |
|--------------------------|------------|------------|------------|------------|------------|
| Number of workers        | 29,515,424 | 34,234,988 | 31,418,385 | 30,467,636 | 33,048,534 |
| Number of establishments | 2,354,423  | 3,155,622  | 3,134,233  | 2,786,426  | 2,590,989  |

The sample consists of full-time female and male workers in the connected set, aged 20–60, in their main job. Source: BeH.

## 4 Estimation

Person and establishment fixed effects are estimated using the following regression equation:

$$y_{it} = \alpha_i + \psi_{J(i,t)} + x'_{it}\beta + r_{it}.$$

We estimate this equation separately for each of the aforementioned time intervals. The data comprises  $N^*$  person-year observations on  $N$  workers and  $J$

establishments.  $J(i, t)$  gives the identity of the unique establishment that employs worker  $i$  in year  $t$ .

As in CHK, the daily real wage  $y_{it}$  of individual  $i$  in year  $t$  is assumed to be the sum of a worker component  $\alpha_i$ , an establishment component  $\psi_{J(i,t)}$ , an index of time-varying observable characteristics  $x'_{it}\beta$ , and an error component  $r_{it}$ .  $x'_{it}\beta$  includes an unrestricted set of year dummies as well as quadratic and cubic terms in age fully interacted with educational attainment.

The literature interprets the establishment effect  $\psi_j$  as a wage premium (or discount) that is paid by establishment  $j$  to all of its employees, while the person effect  $\alpha_i$  represents time-invariant factors that are attributed to the worker.

## 5 AKM Effects

Table 2 shows the mean, standard deviation, and the distribution of the estimated AKM effects. The results point to the usual findings that both the person effects and the establishment effects become more variable over time (CHK). However, in the most recent years, this trend has stopped (see Lochner et al. 2020).

**Table 2:** Statistics on person and establishment effects for each interval.

|                    | 1985–1992 | 1993–1999 | 2000–2006 | 2007–2013 | 2014–2021 |
|--------------------|-----------|-----------|-----------|-----------|-----------|
| Mean person effect | 4.334     | 4.347     | 4.418     | 4.472     | 4.521     |
| SD person effect   | 0.361     | 0.373     | 0.411     | 0.439     | 0.434     |
| p1 person effect   | 3.256     | 3.256     | 3.202     | 3.215     | 3.435     |
| P10 person effect  | 3.927     | 3.943     | 3.974     | 3.993     | 4.047     |
| P25 person effect  | 4.156     | 4.154     | 4.205     | 4.230     | 4.249     |
| P50 person effect  | 4.346     | 4.341     | 4.407     | 4.457     | 4.485     |
| P75 person effect  | 4.523     | 4.538     | 4.633     | 4.710     | 4.760     |
| P90 person effect  | 4.752     | 4.803     | 4.932     | 5.028     | 5.094     |
| P99 person effect  | 5.214     | 5.264     | 5.420     | 5.546     | 5.639     |
| Mean firm effect   | −0.064    | −0.052    | −0.103    | −0.178    | −0.112    |
| SD firm effect     | 0.369     | 0.362     | 0.414     | 0.414     | 0.322     |
| p1 firm effect     | −1.327    | −1.249    | −1.522    | −1.610    | −1.283    |
| P10 firm effect    | −0.536    | −0.501    | −0.611    | −0.682    | −0.460    |
| P25 firm effect    | −0.211    | −0.223    | −0.284    | −0.344    | −0.241    |
| P50 firm effect    | 0.013     | 0.008     | −0.031    | −0.104    | −0.073    |
| P75 firm effect    | 0.165     | 0.183     | 0.162     | 0.079     | 0.074     |
| P90 firm effect    | 0.284     | 0.312     | 0.302     | 0.224     | 0.203     |
| P99 firm effect    | 0.601     | 0.629     | 0.665     | 0.585     | 0.529     |

The sample consists of full-time female and male workers in the largest connected set, aged 20–60, in their main job.  
Source: BeH.

## 6 Data Structure and Access

### 6.1 Data Structure

The AKM effects are estimated for five different non-overlapping time intervals. Data users can readily link the effects using the provided unique person and establishment identifiers. The person and establishment effects are provided in separate time-specific variables, i.e. the data is in wide format.

### 6.2 Data Access

The estimated person and establishment effects are offered by the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research (IAB-RDC) in addition to the SIAB and LIAB data. The effects can be accessed during guest visits to one of the locations of the Research Data Centre of the IAB-RDC. Subsequently, remote data processing is possible.

Initially, an application for data use must be submitted to the IAB-RDC. After approval, a contract needs to be signed between the institution of the researcher and the IAB. An unexpired agreement is necessary for using the estimated effects. Details on how to apply for the data set and the data processing options can be found on the website of the IAB-RDC (<https://fdz.iab.de/en/data-access/>). Access to the data can only be granted once the applicant has signed contracts with both research data centers.

## 7 Summary and Outlook

The AKM approach allows researchers to separately identify unobserved, time-invariant worker and establishment heterogeneity in wages. Using the universe of German employment data of full-time workers aged 20–60, we have applied the AKM wage model to estimate person and establishment fixed wage effects. These effects have been estimated for different time intervals from 1985 to 2021 and can be linked to the workhorse data sets provided by the IAB-FDZ.

So far, many articles have used the AKM effects since the IAB-FDZ started to offer them in 2020. The following paragraph summarizes a selective list of them (incomplete and without any specific ordering):

Butschek (2022) investigates whether the minimum wage introduction raised hiring standards, using worker fixed effects as a proxy for worker productivity. Zimmermann (2022) studies the relationship between worker productivity and the

share of female managers. Müller and Neuschaeffer (2021) study the relationship between worker participation, worker sorting, and firm performance and find that establishments with works councils employ higher-quality workers. Niebuhr and Peters (2020) study workforce composition and individual wages, using the estimated worker and establishment fixed effects as a proxy for unobserved factors that might influence the entry wages of new hires. Illing et al. (2021) analyse the gender wage gap after job loss and show, among other things, that women tend to work at establishments with lower AKM establishment effects. Lochner and Merkl (2022) explore the role of gender-specific application behavior as well the selection behavior of firms for the gender earnings gap in Germany. They find that women tend to apply less often at high-wage establishments, but if they do they have about the same likelihood of being selected by these firms.

As the aforementioned examples show, the provided AKM effects offer an ideal laboratory to exploit some sort of worker and establishment heterogeneity in the context of many different research questions.

## References

- Abowd, J., Kramarz, F., and Margolis, D. (1999). High wage workers and high wage firms. *Econometrica* 67: 251–333.
- Butschek, S. (2022). Raising the bar: minimum wages and employers' hiring standards. *Am. Econ. J. Econ. Pol.* 14: 91–124.
- Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Q. J. Econ.* 128: 967–1015.
- Frodermann, C., Schmucker, A., Seth, S., and Berge, P.V. (2021). *Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975–2019*. FDZ-Datenreport 01 (de), Nürnberg.
- Ganzer, A., Schmucker, A., Stegmaier, J., and Stüber, H. (2022). *Betriebs-historik-panel 1975–2020*. FDZ-Datenreport 03 (de), Nürnberg.
- Illing, H., Schmieder, J.F., and Trenkle, S. (2021). The gender gap in earnings losses after job displacement. *National bureau of economic research working paper 29251*.
- Lochner, B. and Merkl, C. (2022). Gender-specific application behavior, matching, and the residual gender earnings gap. *IAB Discussion Paper 22*, Nürnberg.
- Lochner, B., Seth, S., and Wolter, S. (2020). Decomposing the large firm wage premium in Germany. *Econ. Lett.* 194, <https://doi.org/10.1016/j.econlet.2020.109368>.
- Mueller, S. and Neuschaeffer, G. (2021). Worker participation in decision-making, worker sorting, and firm performance. *Ind. Relat.* 60: 436–478.
- Niebuhr, A. and Peters, J.C. (2020). Workforce composition and individual wages — an employer–employee data analysis. *Br. J. Ind. Relat.* 58: 719–742.
- Ruf, K., Schmidtlein, L., Seth, S., Stüber, H., and Umkehrer, M. (2021). *Linked-employer-employee-daten des IAB: LIAB Längsschnittmodell (LIAB LM) 1975 – 2019*. FDZ-Datenreport 06 (de), Nürnberg.
- Zimmermann, F. (2022). Managing the gender wage gap—how female managers influence the gender wage gap among workers. *Eur. Socio Rev.* 38: 355–370.