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

FDZ-METHODENREPORT

Methodological aspects of labour market data

03|2025 EN AKM effects for German labour market data 1985-2023

Benjamin Lochner, Stefanie Wolter

Bundesagentur für Arbeit

AKM effects for German labour market data 1985 - 2023

Benjamin Lochner (Institute for Employment Research, Friedrich-Alexander-University Erlangen-Nürnberg, IZA, and LASER)

Stefanie Wolter (Institute for Employment Research)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

Content

Tables

Zusammenfassung

Dieser FDZ-Methodenreport beschreibt die Schätzung und Aufbereitung der personen- und betriebsspezifischen Lohneffekte (AKM_8523_v1) und wie diese zu einigen der über das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB) verfügbaren Datensätze zugespielt werden können. Der Report aktualisiert den Bericht von Lochner et al. 2023.

Abstract

This FDZ Methodenreport describes the estimation procedure and the processing of the person and establishment fixed wage effects (AKM_8523_v1) and how this information may be combined with several of the data products available at the research Data Center (FDZ) of the German Federal Employment Agency at the Institute for Employment Research (IAB). This FDZ Methodenbericht updates Lochner et al. 20230.

Keywords

German Administrative Data, Person and Establishment Fixed Effects, AKM

Acknowledgements

We would like to thank our colleagues in the Data and IT-Management division (DIM) of the Institute for Employment Research for their cooperation, especially Mahmoud Haouet for his exzellent support with the Stata self-service.

1 Introduction

This report describes the estimation procedure and the the processing of Abowd-Kramarz-Margolis (henceforth AKM) fixed effects in the German labor market. The AKM approach decomposes individual wages into contributions from person fixed effects, firm (or in our case establishment) fixed effects, observable worker characteristics, and residual components, providing a rich understanding of wage determination in the labor market.

The IAB Employment History File (BEH) provides the administrative data used to estimate these AKM models. The BEH covers all workers subject to social security contributions and offers daily records of employment and wages, maintained by the Institute for Employment Research (IAB). This data is a core component of the administrative data products provided by the Research Data Center of the German Federal Employment Agency (FDZ).

As an example of the AKM framework's application in the German context, Card et al. (2013) (henceforth CHK) apply the AKM methodology to analyze wage inequality in West Germany. This report describes the estimation strategy that closely follows CHK, but extends and updates their specification. Specifically, we perform the estimation in the following time windows: 1985-1992, 1993 – 2000, 2001 – 2008, 2009 – 2016 and 2017 – 2023. Our specification includes both male and female workers, resulting in gender-pooled establishment effects. The first period focuses exclusively on West Germany, while later periods include both East and West Germany. The FDZ provides AKM effects for several core data products. To find out, wheter there are AKM effects available for a specific data product, please read the documentation on the specific data product.

The remainder of this report is structured as follows: Section 2 summarizes the updates and changes compared to prior version of the AKM effects (1985–2021). Section 3 provides an overview of the BEH dataset and the sampling structure. Section 4 briefly outlines the AKM estimation strategy. Section 5 describes how to link the AKM effects to the SIAB dataset. Section 6 explains procedures for accessing the AKM data files. Section 7 details the variables contained in the AKM data files.

2 Summary of changes

This update of the AKM effects is closely linked to the previous AKM effects (Lochner et al. 2023). We have compared the estimated fixed effects from the previous versions with the one from this version and found only minor deviations. The following improvements and changes have been implemented:

- **Adaption of periods**. The five periods now cover 1985-1992, 1993 – 2000, 2001 – 2008, 2009 – 2016 and 2017 – 2023. The non-overlapping episodes allow to have more clearly structured observation periods.
- **Changing the implementation from SAS/Matlab to Stata.** Starting with AKM 8523_v1 the data preparation and the estimation of the AKM effects is no longer implemented in

SAS and Matlab. We translated the code into Stata. The advantage is that we rely on the most commonly used software at IAB making replication more easy. Our implementation uses the user written packages group2hdfe (v. 1.01, see Guimaraes, 2016) and reghdfe (v. 6.12.3, see Correia, 2017). The first is used to identify the largest connected set of workers who move between establishments. The latter is used to estimate the two-way fixed effects model. A direct result is that both the estimated person and establishment fixed effects now have mean zero, i.e., the fixed effects are deviations from the mean, not differences from a specific base category as in the previous versions. For more technical details on the estimation method visit the reghdfe website under https://scorreia.com/software/reghdfe/[1] .

- **Wage imputation procedure.** To foster comparability across IAB data products., we now use the imputed wages officially provided for the BEH data. The wage imputation largely follows the method implemented by CHK. A description of the imputation procedure can be found in Drechsler et al. 2023 and Drechsler et al. 2025.

3 Data source and preparation

As mentioned above, the BEH comprises information on the universe of workers' employment subject to social security contributions in Germany. The data contain distinctive (system-free) worker and establishment IDs, which allows to track workers when they move between employers. All information is available on a daily basis, which allows the construction of accurate employment biographies.

From the universe of employment spells, we construct a person-year data set. To this end, we select the employment spell of each worker with the highest earnings in a given year. We do this beaucse workers can have several jobs at the same time with multiple employers.

We further restrict our sample to full-time workers (based on the person group "101") aged between 20 and 60 and discard jobs with (real) daily wages below 10 Euros.

4 Estimation

For the estimation of person and establishment fixed effects, we largely follow CHK:

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

In a given time interval, the data set contains N^*person-year observations on N workers and J establishments. The function $\boldsymbol{J}(i,t)$ gives the identity of the unique establishment that employs worker i in year t. It is assumed that the log daily real wage y_{it} of individual i in year t is the sum of a worker component α_i, an establishment component $\psi_{J(i,t)}$, an index of time-varying observable characteristics $x'_{it}\beta$, and an error component r_{it}. The person effect α_i is often interpreted as a combination of skills and other time-invariant factors that are rewarded equally across employers.

[1] Last accessed on 05/28/2025.

Likewise, the index $x_{it}'\beta$ is interpreted as a combination of life-cycle and aggregate time-variant factors that affect worker i's productivity at all jobs. An unrestricted set of year dummies as well as quadratic and cubic terms in age fully interacted with educational attainment is included in x_{it}. Finally, the establishment effect ψ_j is often interpreted as a proportional pay premium (or discount) that is paid by establishment j to all employees (i.e., all those with $\boldsymbol{J}(i,t) = j$. Such a premium could represent rent-sharing, an efficiency wage premium, or strategic wage posting behavior.

The above estimation is estimated on the connected set of movers, i.e., individuals that change employers within one observational period. The estimated person effects are transferred to the group of stayers within the same age, education and gender cell (see the CHK online appendix). Table 4-1 shows the number of the estimated person and establishments effects in our final sample in each observation period.

Table 4-1 Number of estimated AKM effects for individuals and establishments

	1985-1992	1993-2000	2001-2008	2009-2016	2017-2023
number of person effects	30,903,562	34,493,240	31,006,815	30,723,828	31,288,177
number of establ. effects	1,723,811	2,487,742	2,312,726	1,944,278	1,727,207
adjusted R squared	0.905	0.914	0.926	0.929	0.920
size of connected set (%)	0.972	0.983	0.977	0.977	0.978
sd log wages	0.451	0.477	0.539	0.540	0.495
sd person effects	0.350	0.350	0.394	0.408	0.404
sd estbl. effects	0.185	0.222	0.263	0.232	0.180
sd xb	0.103	0.072	0.079	0.104	0.092
corr. person/ establ. effects	0.174	0.243	0.261	0.325	0.346

Note: Rows 5-9 are calculated across person-year observations.

5 AKM effects and linkage

The AKM effects can be directly merged to the respective data set using the provided system-free person and the establishment identifier.

Note however that the AKM person effects have been estimated only for regularly employed, full time workers, excluding, for example, marginal workers. Consequently, the AKM effects may only be meaningfully interpreted for code 101 of the *erwstat* variable **and** for the code 0 of the teilzeit variable (`inlist(teilzeit, 0) & inlist(erwstat,101)`).

However, the merging procedure described below links AKM effects via persnr (person identifier) where the *erwstat* and *teilzeit* variables may have different values. It is up to the researcher whether to use the information for these observations or not.

Further note, that the AKM effects are time-specific. Hence, both the person and the establishment effect should be normalized when using data that includes more than one time interval.

The provided establishment and individual level file contain variable labels in German and English. To change the label language you can type `label language de/en` in Stata.

5.1 Exemplary linkage to the Sample of Integrated Labor Market Biographies (SIAB 7523)

Next, we provide exemplary STATA code for mergingthe AKM effects to the SIAB 7523. Other datasets can be linked in the same manner. Note that, if you link the effects to other data the person and establishment identifier might differ. Please read the data reports on the specific data product for further details.

The AKM effects may be linked to the SIAB 7523 data (see Schmucker and vom Berge 2025) by using the *persnr_siab* variable in order to link the person effect, respectively the *betnr_siab* variable to link the establishment effects. The researcher might first want to set up the SIAB data in annual structure and generate five AKM periods in order to restrict to observations that overlap with the different periods in the AKM data. An example how to generate annual data is provided in vom Berge and Schmucker (2025).

```
use SIAB_7523_v1.dta, clear

merge m:1 persnr_siab using SIAB_7523_v1_akm_pers.dta
```

```
use SIAB_7523_v1.dta, clear

merge m:1 betnr_siab using SIAB_7523_v1_akm_estab.dta
```

Table 5-1 provides an overview on the number of linked AKM effects in the SIAB 7523 data. We show the number of individuals and establishments that have a match in the respective interval. The sample is restricted to establishments with fulltime employees and to employees that meet the criteria of the AKM estimation, i.e. regular fulltime employees within the age range of 20 to 60 years.

Table 5-1 Number of linked AKM - Person and Establishment effects in the SIAB 7523 data

	Person effects		Establisment effects	
Interval	Total	Percentage	Total	Percentage
1985 – 1992	617,383	89.1	470,261	88.2
1993 – 2000	687,119	95.4	683,125	94.4
2001 – 2008	618,44	94.1	715,257	91.6
2009 – 2016	613,339	94.5	692,993	90.4
2017 – 2023	625,685	95.6	673,026	89.7

Note: East Germany is included in the AKM estimation from 1993 onwards. There are observations for East Germany in the SIAB in 1992 which decreases the average match rate for the period 1985-1992.

6 Data access

For new projects the AKM effects can be applied for at the field for sensitive variables and extension modules. For exiting projects users can add AKM effects via an informal email to the FDZ (iab.fdz@iab.de). An unexpired use agreement covering the respective data is required. AKM effects are available for data that can be accesed via on-site stay and/ or remote data access, i.e. the effects might not be combined with SUFs.

7 Description of variables

7.1 Identifiers

7.1.1 Individual ID (persnr)

Category	**Description**
Variable label	Indiviudal ID
Variable name	persnr
Category	Identifier
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	The artificial individual ID indicates which observations belong to the same person. Artificial means that it is not possible to infer any of the person's characteristics or any original identifiers from this individual ID. The person identifier may differ in the data product. Please refer to the data report of the data set you want to use.

7.1.2 Establishment ID (betnr)

Category	**Description**
Variable label	Establishment ID

Category	Description
Variable name	betnr
Category	Identifier
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	Artificial establishment ID that is used for all administrative data products provided by FDZ.

7.2 Personal Variables

7.2.1 Person Effect 1985 – 1992 (peff_1985_1992)

Category	Description
Variable label	Person Effect 1985-1992
Variable name	peff_1985-1992
Category	Personal variable
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated person effects for both male and female workers in the period 1985 to 1992

7.2.2 Person Effect 1993 – 2000 (peff_1993_2000)

Category	Description
Variable label	Person effect 1993-2000
Variable name	peff_1993-2000
Category	Personal variable
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated person effects for both male and female workers in the period 1993 to 2000.

7.2.3 Person Effect 2001 – 2008 (peff_2001_2008)

Category	Description
Variable label	Person effect 2001 – 2008
Variable name	peff_2001_2008
Category	Personal variable
Origin	BeH

Category	Description
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated person effects for both male and female workers in the period 2001 to 2008.

7.2.4 Person Effect 2009 – 2016 (peff_2009_2016)

Category	Description
Variable label	Person effect 2009-2016
Variable name	peff_2009_2016
Category	Personal variable
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated person effects for both male and female workers in the period 2009 to 2016.

7.2.5 Person Effect 2017 – 2023 (peff_2017_2023)

Category	Description
Variable label	Person effect 2017-2023
Variable name	peff_2017-2023
Category	Personal variable
Origin	BeH
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated person effects for both male and female workers in the period 2017 to 2023.

7.3 Establishment Variables

7.3.1 Establishment Effect 1985 – 1992 (feff_1985-1992)

Category	Description
Variable label	Establishment Effect 1985-1992
Variable name	feff_1985-1992
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 1985 to 1992.

7.3.2 Establishment Effect 1993 – 2000 (feff_1993_2000)

Category	Description
Variable label	Establishment Effect 1993-2000
Variable name	feff_1993-2000
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 1993 to 2000.

7.3.3 Establishment Effect 2001 – 2008 (feff_2001_2008)

Category	Description
Variable label	Establishment Effect 2001-2008
Variable name	feff_2001_2008
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2001 to 2008.

7.3.4 Establishment Effect 2009 – 2016 (feff_2009_2016)

Category	Description
Variable label	Establishment Effect 2009-2016
Variable name	feff_2009-2016
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2009 to 2016.

7.3.5 Establishment Effect 2017 – 2023 (feff_2017_2023)

Category	Description
Variable label	Establishment Effect 2017-2023
Variable name	feff_2017-2023
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2017 to 2023.

Literature

Abowd, John; Kramarz, Francis; Margolis, David (1999), 'High Wage Workers and High Wage Firms, In: Econometrica, Vol. 67, No. 2, pp. 251–333.

Bonhomme, Stéphane; Holzheu, Kerstin; Lamadon, Thibaut; Manresa, Elena; Mogstad, Magne; Setzler, Bradley (2023): How Much Should We Trust Estimates of Firm Effects and Worker Sorting?, In: Journal of Labor Economics, Vol. 41, No. 1.

Card, David; Heining, Jörg; Kline, Patrick (2013): Workplace heterogeneity and the rise of West German wage inequality. In: The Quarterly Journal of Economics, Vol. 128, No. 3, pp. 967-1015.

Card, David; Heining, Jörg; Kline, Patrick (2015): CHK effects. (FDZ-Methodenreport, 06/2015 (en)), Nürnberg, 35 S.

Correia, Sergio (2017): reghdfe: Stata module for linear and instrumental-variable/gmm regression absorbing multiple levels of fixed effects. Statistical Software Components s457874, Boston College Department of Economics.

Drechsler, Jörg; Ludsteck, Johannes (2025): Imputation Strategies for Rightcensored Wages in Longitudinal Datasets, arXiv:2502.12967v1.

Drechsler, Jörg; Ludsteck, Johannes; Moczall, Andreas (2023): Imputation der rechtszensierten Tagesentgelte für die BeH. FDZ-Methodenreport 05/2023 (de), Nürnberg, DOI:10.5164/IAB.FDZM.2305.de.v1

Guimaraes, Paulo (2016): GROUP2HDFE: Stata module to compute number of restrictions in a linear regression model with two high-dimensional fixed effects, Statistical Software ComponentsS458180, Boston College Department of Economics.

Lochner, Ben; Seth, Stefan; Wolter, Stefanie (2023): AKM effects for German labour market data 1985-2021. (FDZ-Methodenreport, 01/2023 (en)), Nürnberg, 13 S.

Schmucker, Alexandra; Vom Berge, Philipp (2025): Sample of integrated labour market biographies (SIAB) 1975-2023. (FDZ-Datenreport, 02/2025 (en)), Nürnberg, 77 S.

Vom Berge, Philipp; Schmucker, Alexandra (2025): Creating cross-sectional data and biographical variables with the Sample of Integrated Labour Market Biographies 1975-2023 * Programming examples for Stata. (FDZ-Methodenreport, 01/2025 (en)), Nürnberg, 13 S.

Imprint

FDZ-Methodenreport 05|2025 EN

Date of publication

30 May 2025

Publisher

Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nuremberg

Download

https://doku.iab.de/fdz/reporte/2025/MR_03-25_EN.pdf

Documentation version

DOI: 10.5164/IAB.FDZM.2503.en.v1

All publications in the series "FDZ-Methodenreport" can be downloaded from

https://fdz.iab.de/en/research/publications/fdz-methodenreport-series/

Website

https://fdz.iab.de

Corresponding author

Stefanie Wolter
Phone: +49 911 179-1615
Email: Stefanie.Wolter@iab.demailto: