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

01|2023 EN AKM effects for German labour market data 1985-2021

Benjamin Lochner, Stefan Seth, Stefanie Wolter

Bundesagentur für Arbeit

AKM effects for German labour market data

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

Stefan Seth (Institute for Employment Research)

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_8521_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 Bellmann et al. 2020.

Abstract

This FDZ Methodenreport describes the estimation procedure and the processing of the person and establishment fixed wage effects estimated 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 Bellmann et al. 2020.

Keywords

German Administrative Data, Person and Establishment Fixed Effects, AKM

1 Introduction

Card et al. (2013) (henceforce CHK) study the role of establishment-specific wage premia in generating recent increases in West German wage inequality. For four subintervals covering the period from 1985 to 2009, they estimate models with additive fixed effects for workers and establishments following the estimation strategy introduced in Abowd et al. (1999) (hereafter AKM).

The analyses of CHK are based on administrative data for Germany, specifically the IAB Employment History File (BEH). This data set represents the universe of workers subject to social security contributions in Germany and contains information on employment and wages on a daily basis. The BEH is maintained by the Institute for Employment Research (IAB) and is one of the major sources of the data products available at the Research Data Center (FDZ) of the German Federal Employment Agency.

On the one hand, the estimated person fixed effects have been interpreted as a combination of individual skills and other time-invariant factors which are rewarded equally across employers. On the other hand, the estimated establishment fixed effect represent the proportional pay premia (or discounts) that are paid by specific establishments to all of their employees.

This report describes the update and extension of the person and establishment fixed effects (henceforth AKM effects). For the original version recall Card et al. 2015.

The update coversfive periods: 1985-1992, 1993 – 1999, 2000 – 2006, 2007 – 2013 and 2014 – 2021. In contrast to CHK, we include both males and females.The first period includes only West Germany, while the other periods contain the person and establishment fixed effects for both East and West Germany. The FDZ has prepared several files containing the AKM effects, which may be linked to the major data products of the FDZ. To find out, wheter there are AKM effects available for a specific data product, please read the documentation on the specific data product.

This FDZ Methodenreport describes these files and is structured as follows: Section 2 summarizes the changes as compared to the prior version AKM effects 1985 - 2017. Section 3 provides a more detailed overview of the BEH and the sample used by CHK. The estimation strategy is briefly discussed in Section 4. The linkage of the AKM effects to the SIAB data is described in section 5. Section 6 provides information on how to access these files. Finally, Section 7 contains a description of variables.

2 Summary of changes

This update of the AKM effects is closely linked to the previous AKM effects (Bellmann et al. 2020). The following improvements and changes have been implemented:

- **Adaption of periods**. The five periods now cover 1985-1992, 1993 – 1999, 2000 – 2006, 2007 – 2013 and 2014 – 2021. The non-overlapping episodes allow to have more clearly structured observation periods.

- **Exclusion of appprentices.** In the previous version, apprentices were included in the sample, but only if their reported age was 20 or above. This excluded a great many apprentices' employment biographies in whole or in part. Also, apprentices are no fulltime employees. We therefore exclude apprenticies by only keeping obsservations with person group 101 in the sample.
- **Improved wage imputation procedure.** First, we now add bonuses and other one-time payments to the annual income, before imputing right-censored wages. Second, we now account for noise in wages around the censoring limit, i.e., we impute wages which are two percent under the censoring limit (in the previous version: 0.5 percent). The new procedure is more accurate, less prone to misreporting and is closer aligned with the official imputation procedure of the BHP.

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

Before estimating the worker and establishment fixed effects, we impute wages above the social security contribution threshold. The imputation strategy follows CHK (see their online appendix for details). We estimate tobit regressions by year, sex, education and age group. As controls we include: age, mean log wage in other years, fraction of censored wages in other years, number of full time employees at the current establishment and its square, dummy for 11 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, dummy for individuals observed only one year between 1985 and 2021, dummy for employees of one-worker establishments, and a dummy for East Germany.

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

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

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 $\psi_{\boldsymbol{J}(i,t)}$, 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. 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 individuals and establishments that receive an effect in our final sample in each observation period. Within the largest connected set, roughly sixty percent of all employees are stayers (see Bonhomme et al., 2023 for a discussion).

Table 4-1 reports the number of individuals and establishments that we estimated an effect for per interval. Additional we report the share of stayers among the individuals.

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

	1985-1992	1993-1999	2000-2006	2007-2013	2014-2021
Number of individuals	28,562,991	33,104,885	30,078,974	29,070,564	31,821,688
Share of stayer	0.648	0.609	0.618	0.642	0.616
Number of establishments	1,907,297	2,565,930	2,462,246	2,117,452	2,048,134

Table 4-2 presents distributional moments for the person and establishment fixed effects in the data for each interval. To avoid weighting with the number of observations, we use each person and establishment only once.

Table 4-2 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

5 AKM effects and linkage

The AKM effects can be directly merged to the respective data set using the provided 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 7521)

Next, we provide exemplary STATA code for mergingthe AKM effects to the SIAB 7521. 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 7521 data (see Schmucker et al. 2023) by using the *persnr_siab* variable in order to link the person effect, respectively the *betnr_siab* variable to link the establishment effects. Tthe 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 Eberle and Schmucker (2019).

```
use SIAB_7521_v1.dta, clear

merge m:1 persnr_siab using SIAB_7521_v1_akm_pers.dta
```

```
use SIAB_7521_v1.dta, clear

merge m:1 betnr_siab using SIAB_7521_v1_akm_estab.dta
```

Table 5-1 and Table 5-2 provide an overview on the number of linked AKM effects in the SIAB 7521 data. We show the number of individuals and establishments that have any match in the respective interval. The second part of Table 5-1 reports the number and share of individuals that merge with the effect if we restrict the SIAB to the same criteria that we use for the AKM estimation (fulltime employed, age 20-60, no apprenticies). Note that for the unrestricted link the number is lower as SIAB covers also the period before 1985. After 2011 the part-time share increases, which decreases the share of unrestricted matches.

Table 5-1 Number of linked AKM - Establishment effects in the SIAB 7521 data

	No restrictions		Restricted to: only BeH-spells, fulltime, age 20-60	
Interval	Total individuals	Percentage	Total	Percentage
Any match in the SIAB data	1,299,877	65.7	566,317	97.9
1985 – 1992	752,578	86.5	459,844	97.6
1993 – 1999	830,198	82.7	487,752	99.0
2000 – 2006	870,578	84.6	374,876	99.3
2007 – 2013	882,787	85.6	357,469	99.4
2014 – 2021	891,714	77.6	400,120	99.2

Table 5-2 Number of linked AKM establishment effects

Interval	Total establishments	Percentage
Any match	2,105,594	81.9
1985 – 1992	545,768	94.1
1993 – 1999	683,552	95.0
2000 – 2006	782,924	90.7
2007 – 2013	770,575	87.2
2014 – 2021	833,941	83.5

6 Data access

The AKM effects will be provided to researchers upon request via e-mail by FDZ. 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 execution, 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
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 Effekt 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 – 1999 (peff_1993_1999)

Category	Description
Variable label	Person effect 1993-1999
Variable name	peff_1993-1999
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 1999.

7.2.3 Person Effect 2000 – 2006 (peff_2000_2006)

Category	Description
Variable label	Person effect 2000 – 2006
Variable name	peff_2000_2006
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 2000 to 2006.

7.2.4 Person Effect 2007 – 2013 (peff_2007_2013)

Category	Description
Variable label	Person effect 2007-2013
Variable name	peff_2007_2013

Category	Description
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 2007 to 2013.

7.2.5 Person Effect 2014 – 2021 (peff_2014_2021)

Category	Description
Variable label	Person effect 2014-2021
Variable name	peff_2014-2021
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 2014 to 2021.

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 – 1999 (feff_1993_1999)

Category	Description
Variable label	Establishment Effect 1993-1999
Variable name	feff_1993-1999
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 1993 to 1999.

7.3.3 Establishment Effect 2000 – 2006 (feff_2000_2006)

Category	Description
Variable label	Establishment Effect 2000-2006
Variable name	feff_2000_2006
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2000 to 2006.

7.3.4 Establishment Effect 2007 – 2013 (feff_2007_2013)

Category	Description
Variable label	Establishment Effect 2007-2013
Variable name	feff_2007-2013
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2007 to 2013.

7.3.5 Establishment Effect 2014 – 2021 (feff_2014_2021)

Category	Description
Variable label	Establishment Effect 2014-2021
Variable name	feff_2014-2021
Category	Establishment variable
Data type	Numerical
Hierarchy	None
Detailed description	This variable contains the estimated establishment effects for the period 2014 to 2021.

Literature

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

Bellmann, Lisa; Lochner, Ben; Seth, Stefan; Wolter, Stefanie (2020): AKM effects for German labour market data. (FDZ-Methodenreport, 01/2020 (en)), Nürnberg, 24 S.

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.

Eberle, Johanna; Schmucker, Alexandra (2019): Creating cross-sectional data and biographical variables with the Sample of Integrated Labour Market Biographies 1975-2017 * Programming examples for Stata. (FDZ-Methodenreport, 04/2019 (en)), Nürnberg, 16 S.

Schmucker, Alexandra; Seth, Stefan; Vom Berge, Philipp (2023): Sample of integrated labour market biographies (SIAB) 1975-2021. (FDZ-Datenreport, 02/2023 (en)), Nürnberg, 82 S.

Imprint

FDZ-Methodenreport 01|2023 EN

Date of publication
21 March 2023

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

Rights of use
This publication is published under the following Creative Commons licence:
Attribution – ShareAlike 4.0 International (CC BY-SA 4.0)
https://creativecommons.org/licenses/by-sa/4.0/deed.de

Download
https://doku.iab.de/fdz/reporte/2023/MR_01-23_EN.pdf

Documentation version
DOI: 10.5164/IAB.FDZM.2301.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.de