Worker Sorting, Industry Sorting, and Agglomeration Effects*

Wolfgang Dauth

Institute for Employment Research Nuremberg (IAB) &

University of Bamberg

Michael Moritz

Institute for Employment Research Nuremberg (IAB)

Anja Rossen

Institute for Employment Research Nuremberg (IAB)

Very preliminary and incomplete version.

Please do not cite or circulate

This Version: February 25, 2026

* Order of Coauthorship will be randomized: https://www.aeaweb.org/journals/policies/random-author-order

Wolfgang Dauth: Institute for Employment Research, Regensburger Str. 104, 90478 Nuremberg. Michael Moritz: Institute for Employment Research, Regensburger Str. 104, 90478 Nuremberg. Phone: +49-911-179-2133. E-mail: michael.moritz@iab.de (corresponding author). Anja Rossen: Institute for Employment Research Lower Saxony-Bremen, Röpkestraße 3, 30173 Hannover.

Worker Sorting, Industry Sorting, and Agglomeration Effects

Very preliminary and incomplete version.

Please do not cite or circulate

Abstract

Significant spatial wage disparities can be observed in virtually all countries with free market economies. In particular, larger cities offer higher wages compared to more rural areas. There are at least two major explanations for this observation: (1) people with characteristics that are related to higher wages prefer to live in larger cities and (2) the same worker becomes more productive if she or he is located in a larger rather than a smaller city. In this paper, we shed light on the relative importance of those explanations and demonstrate that, after controlling for worker and firm sorting, there is still a significant agglomeration effect, which we call the true urban wage premium, that makes wages increase with the size of population. Following Card, Rothstein, and Yi (2024, 2025), we find a so-called hierarchy effect, i.e. the mobility of workers between cities of different sizes is usually not between representative firms. Furthermore, we demonstrate how the wage decomposition by Abowd, Kramarz, and Margolis (1999) can be adapted to identify the true urban wage premium. We split the AKM plant effects into an industry-specific component and a region-specific component. Those premia reveal the various sources of wage disparities and are informative to (local) policymakers. We come to the conclusion that on the one hand the decomposition is nearly fully absorbed by the regional component. On the other hand, bigger cities do not offer a more favorable industry mix.

JEL classification: J31, R12, R23, R38

Keywords: urban wage premium, economies of agglomeration, wages

1. Introduction

The urban wage premium is a well-known and understood fact. Significant spatial wage disparities can be observed in virtually all countries with free market economies. In particular, larger cities offer higher wages compared to more rural areas. In Germany, for example, the unconditional elasticity of nominal wages with respect to city size is 0.04 (Ciccone, 2002), which means that doubling city size implies a wage difference by four percent.

By using an extract of the Integrated Employment Biographies for the years from 2011 to 2019, Figure 1 is mapping the average nominal wages in 141 German local labor markets (LLM) showing the highest wages in West Germany's big cities Munich, Stuttgart, Frankfurt am Main and Düsseldorf. Figure 2 corroborates the positive correlation between the population size and average wages in the local labor markets. The wage level is lower in most of the East German regions, but the slopes of the regression lines are nearly parallel, i.e. a similar relationship is observable for both West and East Germany.

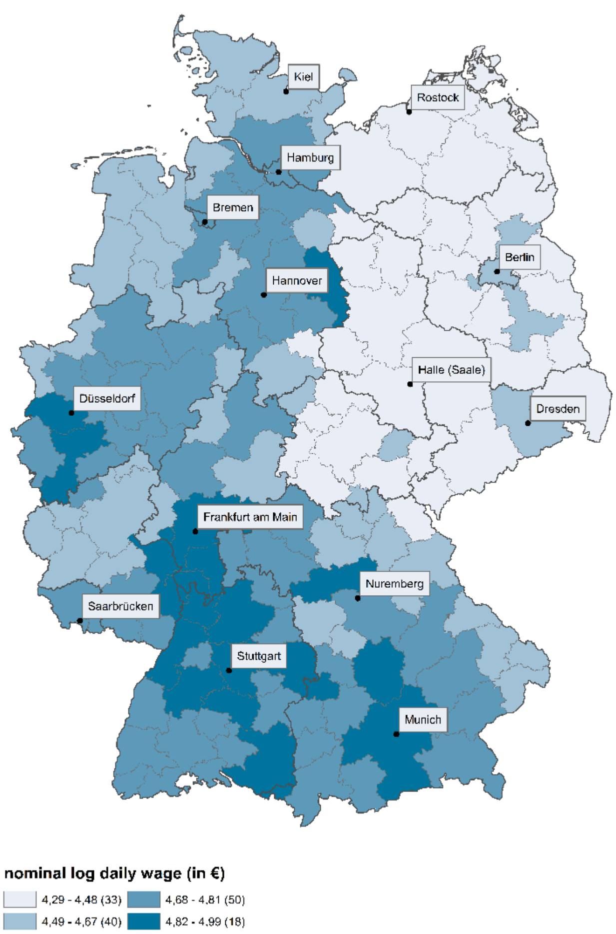

Figure 1: Average nominal wages in German local labor markets (LLM)

Source: IEB; own calculations.

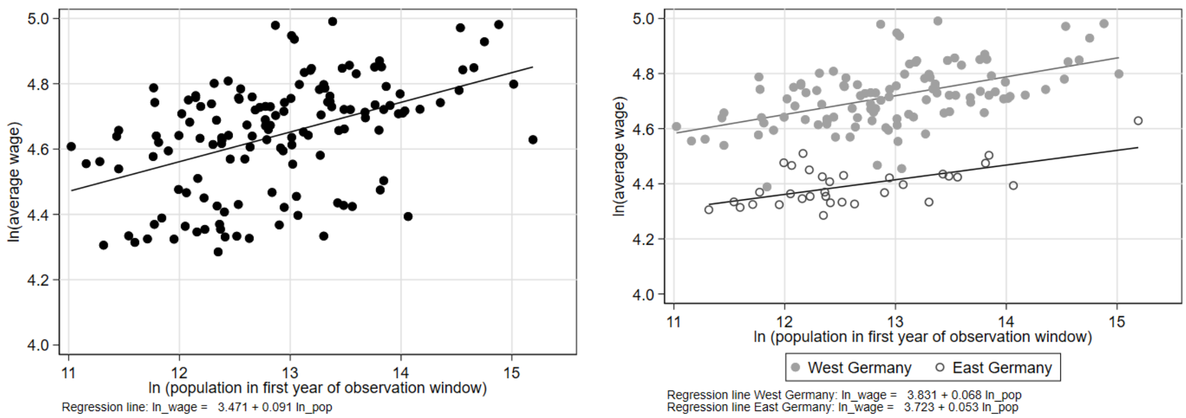

Figure 2: Scatterplot with regression line of ln (average wage) vs. ln (population in first year of observation window) for the whole of Germany (left panel) and separated for East and West Germany (right panel).

Source: IEB; own calculations.

There are at least two major explanations for this observation. The first is that people with characteristics that are related to higher wages prefer to live in larger cities. This has been documented, amongst others, by Combes, Duranton, and Gobillon (2008) and D'Costa and Overman (2014) and explains round one half of the city-size premium. The other explanation is that the same worker becomes more productive if she or he is located in a larger rather than a smaller city. This may be either because regions differ with respect to their composition and more productive firms or industries sort into larger cities (cf. Dauth et al., 2022 or Ellison and Glaeser, 1997, respectively) or because of externalities of agglomeration that raise productivity in bigger cities (e.g. Duranton and Puga, 2004).

In this paper, we use a full sample of all workers subject to social security in Germany for the years 2011-19. We shed light on the relative importance of those explanations and demonstrate that, after controlling for worker and plant sorting, there is still a significant agglomeration effect that makes wages increase with the size of local population.[1] To this end, we adapt a novel decomposition of the sources of spatial wage disparities for the German labor market. Our empirical approach uses the decomposition proposed by Card, Rothstein, and Yi (2024, henceforth CRY). This decomposition stems from the seminal work by Abowd, Kramarz, and Margolis (1999, henceforth AKM) that additively decomposes individual log wages into a worker-specific component and a firm-specific component. A huge number of studies have adopted this approach and demonstrated that this decomposition approximates observed wages very well. For example, Card, Heining, and Kline (2013) apply the AKM model to German data and report an R-squared of up to 0.93, which is only slightly smaller than a model with worker/plant-match-specific effects with an R-squared of 0.95. CRY modify this approach to the

[1] In our empirical analysis, we use the terms "plant", "firm", and "establishment" interchangeably.

spatial setting by decomposing log wages into a worker component and an industry component. CRY establish the so-called “hierarchy effect”, i.e. the systematic asymmetry in the pattern of workers’ mobility. Typically, workers move from relatively higher-paying firms in the origin industry to relatively lower-paying firms in the destination industry, if the latter industry is superiorly ranked in the wage ladder compared to the former industry. In another paper by Card, Rothstein, and Yi (2025), besides the worker component, the authors use a component that is idiosyncratic to region/industry-cells, referred to as local industries. By further decomposing these local industry effects and taking the hierarchy effect into account, they find that around 87 percent are attributable to true region and true industry effects. They conclude that this does not leave much room for an agglomeration effect that is specific to local industries.

However, we argue that this conclusion based on comparing the R-squared of different models provides only an incomplete picture. We add two additional insights: First, we refer to CRY and demonstrate that mobility between German cities of different sizes is usually not between representative firms, the above-mentioned “hierarchy effect”. This implies that we may underestimate the actual effect of mobility between small and large cities. In a further step, we estimate the unconditional urban wage premium getting the result of 6.5 % in the unweighted case and 7.9 % in the case of weighting by employment size. Moreover, all estimated fixed effects (by person, region, firm and hierarchy) are positively correlated with the population size of the local labor market region. Second, we adapt the approach of Card, Rothstein, and Yi (2024, 2025) by decomposing the AKM plant effects into an industry-mix effect and a region-specific effect. This allows us to analyze if bigger cities offer a more favorable industry mix and to identify the true urban wage premium. We show that the local plant-specific wage is more or less fully absorbed by the regional component, which in turn increases with the size of the local population by both a statistically and economically significant extent. This provides strong evidence for the presence of so-called Marshall-Arrow-Romer-type agglomeration effects.

In the following sections we explain the data and the empirical strategy (section 2), illustrate and discuss the results (section 3) and conclude with section 4.

2. Empirical Analysis

We use data from the Integrated Employment Biographies (IEB, version 16.00.00-202012) of the Institute for Employment Research (IAB). With regard to employment spells, it covers the full job history of the universe of employees, excluding civil servants and self-employed persons. We focus on male full-time employees, aged 18 to 60, who are subject to social security contributions in Germany. In our extract for the years from 2011 to 2019, the IEB includes 110,389,868 worker-year observations and a total of 18,063,799 individuals and 1,910,676 plants. Worker wages are defined as the log of average gross labor earnings per day. Means and standard deviations for all our variables are reported in Table 1. Our main geographical unit of analysis are 141 local labor markets (LLM) in Germany, following Kosfeld and Werner (2012), who base their classification (*Arbeitsmarktregionen*) on commuter flows (rather than just administrative borders). Therefore, the distinct regions are similar to US commuting zones used in CRY and the closest approximation to local labor markets. Industries are defined as 2-digit ISIC Rev.4 industry codes.

Columns (1) to (2) show statistics for employees in all 141 LLM, whereas columns (3) to (4) restrict to the largest 10 LLM. These larger LLM record a higher daily wage on average, a higher share of university graduates as well as a higher share of foreigners. Columns (5) to (7) show the results of univariate regressions of the listed worker characteristics on the log workforce size in the LLM. Log workforce size is highly correlated with the share of university graduates, and, vice versa, negatively with the share of workers with vocational training.

Table 1: Summary statistics of wages, population and worker characteristics

	ALL LLM's		Largest 10 LLM's		Univariate Regression on log workforce size		
	Mean	Std. Dev	Mean	Std. Dev	Coeff.	Std. Err	R-sq.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
a. Mean Worker Characteristics							
log daily wage	4.507	0.472	4.593	0.526	0.035	0.000	0.032
no vocational training	0.046	0.210	0.059	0.236	0.007	0.000	0.056
vocational training	0.830	0.375	0.770	0.421	-0.039	0.000	0.346
university or university of applied science	0.123	0.329	0.171	0.376	0.032	0.000	0.402
Age: 20-29 years	0.000	0.006	0.000	0.007	0.000	0.000	0.001
Age: 30-39 years	0.106	0.308	0.105	0.307	0.007	0.000	0.005
Age: 40-49 years	0.485	0.500	0.494	0.500	0.008	0.000	0.030
Age: 50 years and more	0.409	0.492	0.401	0.490	-0.014	0.000	0.017
German citizenship	0.925	0.264	0.891	0.311	-0.015	0.000	0.067
foreign citizenship	0.075	0.263	0.108	0.310	0.015	0.000	0.064
no information in citizenship	0.000	0.020	0.001	0.027	0.000	0.000	0.101
Population (LLM141 level)	1,194	1,032	2,657	769	1,030	1,112	0,745
Population (district level)	457	690	1,068	1,077	426	1,137	0,361
b. Decomposition of mean wages from regression model							
LLM wage residual	0.000	0.157	0.000	0.125	0.041	0,000	0,050
LLM predicted wage	0.000	0.152	0.000	0.121	0.025	0,000	0,020
c. Add 2-digit Industry Effects to wage model	4.507	0.000	4.593	0.000	0.000	0,000	
LLM wage residual	0.000	0.282	0.000	0.309	0.040	0,000	0,015
LLM predicted wage	0.000	0.249	0.000	0.257	0.027	0,000	0,009
d. Add 3-digit Industry Effects to wage model	4.507	0.000	4.593	0.000	0.000	0,000	
LLM wage residual	0.000	0.292	0.000	0.322	0.040	0,000	0,014
LLM predicted wage	0.000	0.258	0.000	0.269	0.027	0,000	0,008
Mean (3-digit) Ind. Composition Effect	4.507	0.000	4.593	0.000	0.000	0,000	
Share of Total Sample	0.000		0.000		-0.004	0.000	0.000

Notes: The table refers to the AKM estimation sample, consisting of 18,063,799 persons, 110,389,868 person-year observations, and 1,910,676 plants.
Regression models in column (5)-(7) are fit to LLM-means of row variable using log(LLM workforce size) as explanatory variable.
Source: IEB; own calculations.

Empirical Strategy

In order to apply the approach of the "hierarchy effect" by Card, Rothstein, and Yi (2024), we initially adopt the wage decomposition proposed by Card, Rothstein, and Yi (2025) in the following equation:

$$y_{it} = \alpha_{1i} + \psi_{p(it)} + X'_{it}\beta + \varepsilon_{it}, \tag{1}$$

where y_{it} denotes the log average daily wage, α_{1i} is an individual-specific fixed effect, $\psi_{p(it)}$ signifies a fixed effect for the plant p at which worker i is employed at date t, X_{it} represents a vector covering education, age and foreigner profiles and ε_{it} stands for the error term.

In contrast to Card et al. (2024), who calculate the difference between the firm fixed effect and the average premium in the related industry, we apply a different approach of the "hierarchy effect". In our definition, the hierarchy effect is the gap between the firm fixed effect and the fixed effect in the associated region. Consequently, we run a second AKM model, this time with the region fixed effect γ_r:

$$y_{it} = \alpha_{2i} + \gamma_{r(it)} + X'_{it}\beta + \varepsilon_{it} \tag{2}$$

Afterwards, we can calculate the hierarchy effect at the plant level h_p by equation (3):

$$h_p = \psi_p - \gamma_{r(p)} \tag{3}$$

In a next step, we perform several separate estimations by individually regressing the averages of log wages and the fixed effects gained from the estimations of equations (1) and (2), each at regional level, on the log of population size in the local labor markets. Additionally, we include a dummy variable for the eastern German regions:

$$\overline{premium\ k_r} = \tau\, pop_r + \varphi\, east_r + \mu_r \tag{4}$$

This means that we estimate the impact of population size (measured in the year 2010) on different *premia k*, e.g. in the case of average regional wages, *k* stands for $\bar{y}_r$, the unconditional regional wage . In the other cases, *k* stands for the mean values of α_{1i}, α_{2i}, $\psi_{p(it)}$, $\gamma_{r(it)}$, and h_p from equations (1), (2) and (3). In each case, we estimate two versions, once with unweighted observations, once with observations weighted by the number of employees.

In the final step of our analysis, we initially perform a regression at worker level using the AKM plant effects $\psi_{p(it)}$ estimated in equation (1). Our goal is to split the AKM plant effects into region and industry fixed effects, in order to identify the true urban wage premium:

$$\widehat{\psi_{p(it)}} = \omega_{1r(it)} + \delta_{1j(it)} + \nu_{1p(it)}, \tag{5}$$

where $\omega_{1r(it)}$ stands for the region FE, i.e. the true urban wage premium and $\delta_{1j(it)}$ denotes the industry FE. In a separate specification, we control for plant size (in the original form and as a squared polynomial):

$$\widehat{\psi_{p(it)}} = \omega_{2r(it)} + \delta_{2j(it)} + \lambda_1 plant\_size + \lambda_2 plant\_size\_squared + \nu_{2p(it)} \tag{6}$$

3. Results

a. Hierarchy effect

First of all, we are interested in the development of the relative wage position of inter-regional plant movers. Table 2 tabulates the relations between the origin and destination region ordered by quartiles (from Q1 to Q4) of the region-specific AKM fixed effects. The former quartile value in a cell signifies the average percentile of the AKM fixed effects in the corresponding origin regions, the latter quartile value denotes the same for the destination regions. By way of example, movers from a Q1 region to a Q3 region are on average ranked in the 47th percentile in the origin region and in the 39th percentile in the destination region. This is quite plausible as average AKM plant fixed effects, i.e. wages, are on average higher in Q3 regions compared to Q1 regions. Hence, upward movers (according to the quartiles) are going to earn more on average even when going down the percentile ladder. When looking at the sorting of origin and destination percentile values by quartile, we observe a reasonable pattern of average AKM fixed effects.

Table 2: Average percentile of AKM FE of origin and destination plant by quartile of origin and destination region

		Destination of plant movers (average AKM plant effect)			
		Q1	*Q2*	*Q3*	*Q4*
Origin of plant movers (average AKM plant effect)	*Q1*	*38.072/43.175*	*43.470/39.323*	*46.245/38.346*	*48.733/37.235*
	Q2	*34.938/46.735*	*38.021/42.229*	*41.228/40.272*	*45.967/40.964*
	Q3	*34.538/48.268*	*36.924/44.671*	*37.802/41.553*	*41.592/41.696*
	Q4	*32.515/49.283*	*37.755/50.028*	*37.719/45.339*	*38.362/42.952*

Note: Average percentile of AKM FE must be region-specific.
Source: IEB; own calculations.

b. Urban wage premium

In a next step, we individually regress the average wage and the fixed effects gained from the estimations of equations (1) and (2), each at regional level, on the log of population size in the local labor markets. Additionally, we include a dummy variable for the eastern German regions. Table 3 shows the results, once for the unweighted version (Panel A), once for version when observations are weighted by the number of employees (Panel B). In all cases, the outcome for the coefficient of population size is significantly positive with slight differences for the person fixed effects from the estimations of equation (1) (avg. person FE I) and equation (2) (avg. person FE II). The coefficients for the average regional wage denote the unconditional urban wage premium, i.e. a doubling of the population size leads to a higher average wage of 6.5 or 7.9 percent. The outcome for the average hierarchy effect indicates the difference between the results for the plant fixed effects and the region fixed effects. In larger cities, the average AKM plant effect is considerably higher than the region FE. This result means that workers on average move to relatively high-paying firms in regions with a lower average wage level compared to the origin region or, vice versa, to relatively low-paying firms in regions with a higher average wage level compared to the origin region. In other words, the position of the new firms in the wage ladder of the destination region is on average lower than in the origin region when moving to a high-wage region and, vice versa, higher when moving to a low-wage region.

Table 3: Table with various specifications at region level

	dependent variable					
	avg. ln wage	*avg. person FE I*	*avg. person FE II*	*region FE*	*avg. AKM plant FE*	*avg. hierarchy effect*
Panel A: unweighted						
ln pop	*.0649****	*.0487****	*.0354****	*.0077****	*.0214****	*.0137****
Panel B: weighted						
ln pop	*.0794****	*.0547****	*.0462****	*.0140****	*.0231****	*.0091****

Source: IEB; own calculations.

c. Industry mix vs. true UWP

In this step, we start by performing a regression at the worker level using the estimated AKM plant fixed effects gained in equation (1). The aim is to decompose the AKM plant FE into an industry specific and a true urban specific component. In version 1, we regress the AKM plant FE on region FE and industry FE (*region_fe_1 and industry_fe_1*). In version 2, we additionally control for plant size as a squared polynomial (*region_fe_2 and industry_fe_2*). The correlation of region FE and industry FE lies between 0.6 and 0.7 (see Table 4). Afterwards, we save the coefficients of those FE and aggregate them at the region level. Again, we regress the fixed effects once on unweighted population (Panel A), once on the population weighted by the number of employees (Panel B).

Table 5 and Table 6 show quite similar results. The decomposition of the AKM plant fixed effect into an industry specific and a true urban specific component is more or less completely absorbed by the latter element. In case of the weighted population, the outcome effect on the industry FE is even significantly negative. This means that the larger the city the more unfavorable is the industry mix for the workers' wages. One reasonable explanation could be that large manufacturing plants providing well-paid jobs are located in less densely populated areas, whereas low-paid service jobs agglomerate in big cities.

Figure 3 and Figure 4 show the region FE and the average industry FE, in each case categorized in five groups. The distribution of the region FE in Figure 3, i.e. the true urban wage premium, reveals a clear pattern with the highest values in the economically strong cities, e.g. Munich, Stuttgart, Frankfurt am Main and Düsseldorf. With the exception of Berlin and some regions bordering on the German capital city, nearly all East German LLM are classified in the lowest group. This is not a really surprising result, as it more or less reflects the regional distribution of wage levels in Germany. The spatial distribution of industry fixed effects in Figure 4 displays that not the big cities offer the highest wages with regard to the industry mix, but the regions close to them. Many of the LLM with the highest industry FE are locations of large manufacturing firms and their suppliers, e.g. in the automobile industry.

Table 4: Correlation coefficients of region FE and industry FE

```
             | region~1 region~2 indust~1 indust~2
-------------+------------------------------------
 region_fe_1 |   1.0000
 region_fe_2 |   0.9945   1.0000
industry_f~1 |   0.6573   0.6526   1.0000
industry_f~2 |   0.6884   0.6844   0.9859   1.0000
```

Table 5: Table with various specifications at region level – version 1

		dependent variable	
	avg. AKM plant FE (from Tab. 3)	*region_fe_1*	*industry_fe_1*
Panel A: unweighted			
ln pop	*.0214****	*.0213****	*.0000*
Panel B: weighted			
ln pop	*.0231****	*.0272****	*-.0041***

Table 6: Table with various specifications at region level – version 2

		dependent variable	
	avg. AKM plant FE (from Tab. 3)	*region_fe_2*	*industry_fe_2*
Panel A: unweighted			
ln pop	*.0214****	*.0190****	*-.0003*
Panel B: weighted			
ln pop	*.0231****	*.0247****	*-.0033***

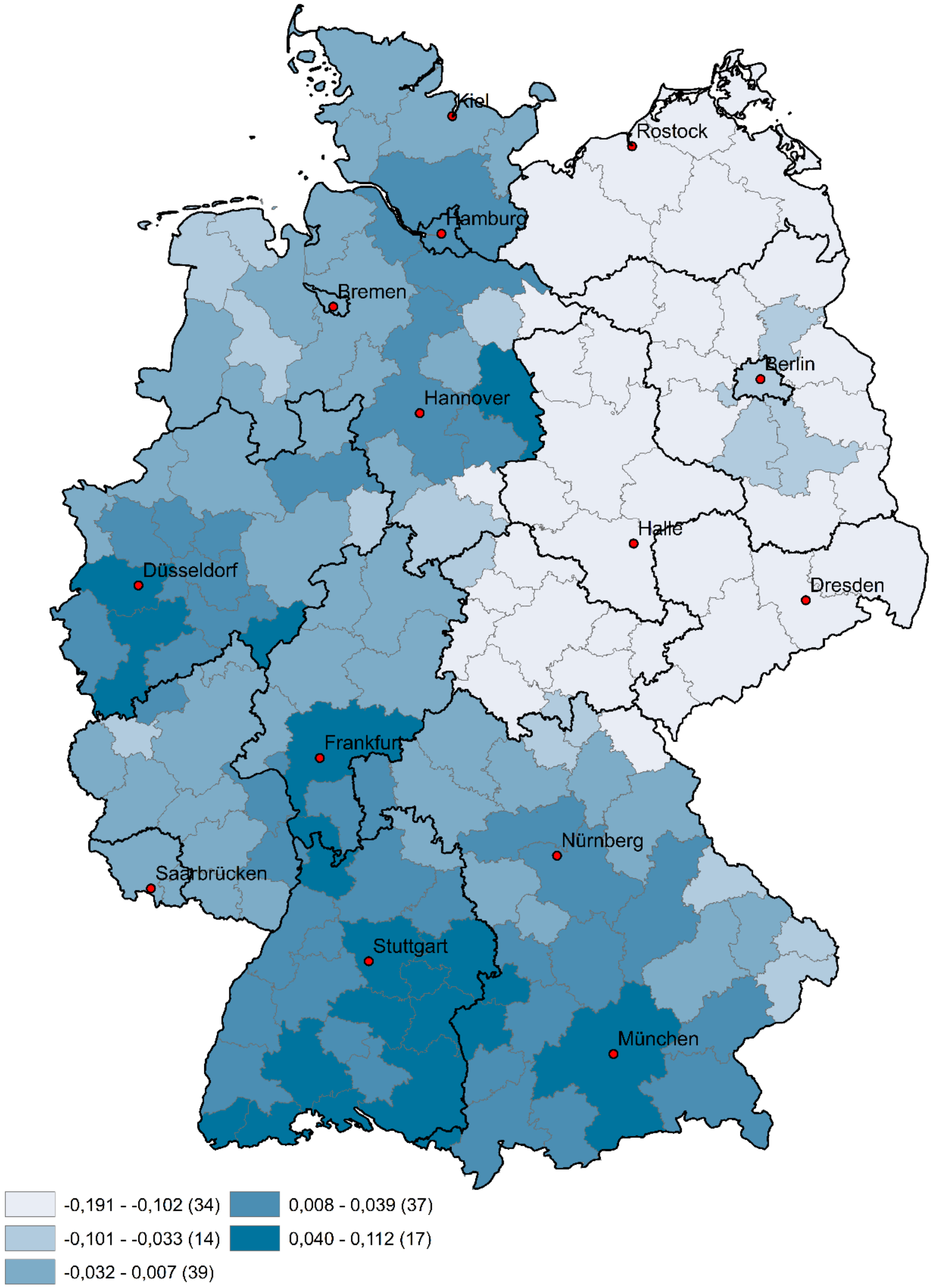

Figure 3: Region fixed effects

Notes: This figure shows the results for the region fixed effects of equation (5) across 141 LLM.
Source: IEB.

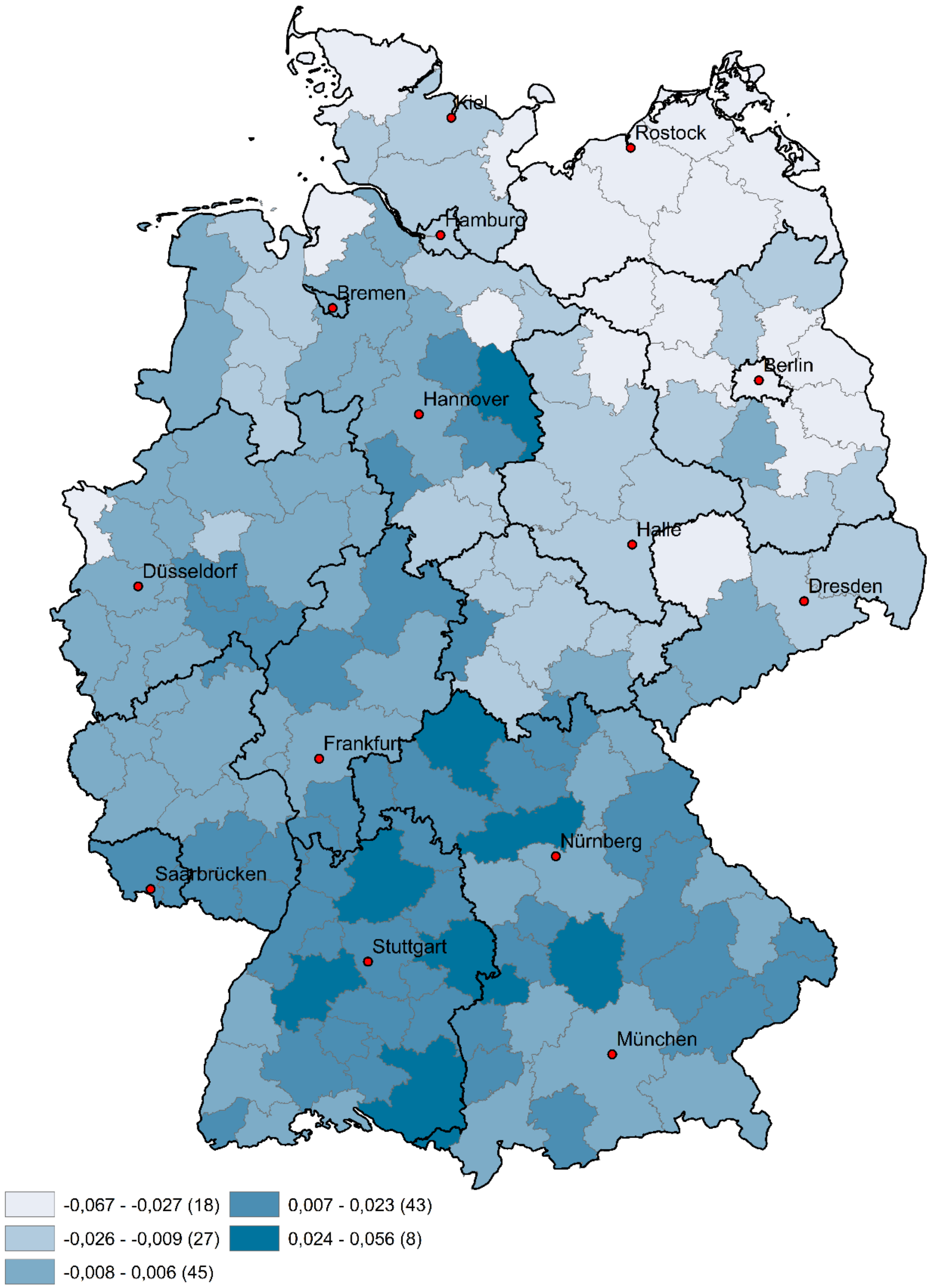

Figure 4: Spatial distribution of industry fixed effects

Notes: This figure shows the results for the spatial distribution of industry fixed effects of equation (5) across 141 LLM.
Source: IEB.

4. Conclusions

In this paper, we tie in with the studies by Card, Rothstein, and Yi (2024, 2025) who analyze the mobility of workers between industries and commuting zones in the United States. Using data from the Integrated Employment Biographies, we find evidence for the existence of a hierarchy effect concerning the mobility of German workers. We demonstrate that mobility from cities between different sizes is usually not between representative firms. Furthermore, based on AKM regressions, we reveal that the decomposition of the plant fixed effects into an industry specific and a true urban specific component is more or less completely absorbed by the latter element. Bigger cities do not offer a more favorable industry mix.

References

Abowd, John M., Francis Kramarz, and David N. Margolis (1999): High Wage Workers and High Wage Firms, *Econometrica*, 67, 251–333.

Card, David, Jörg Heining, and Patrick Kline (2013): Workplace Heterogeneity and the Rise of West German Wage Inequality, *Quarterly Journal of Economics*, 128, 967–1015.

Card, David, Jesse Rothstein, and Moises Yi (2024): Industry wage differentials: A firm-based approach, *Journal of Labor Economics*, 42(S1), S11–S59.

Card, David, Jesse Rothstein, and Moises Yi (2025): Location, Location, Location, *American Economic Journal: Applied Economics*, 17(1): 297–336.

Ciccone, Antonio (2002): Agglomeration Effects in Europe, *European Economic Review* 46(2), 213–227.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008): Spatial Wage Disparities: Sorting Matters!, *Journal of Urban Economics*, 63, 723–742.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum (2022): Matching in Cities, *Journal of the European Economic Association*, 20(4), 1478–1521.

D'Costa, Sabine, and Henry G. Overman (2014): The Urban Wage Growth Premium: Sorting or Learning?, *Regional Science and Urban Economics*, 48, 168–179.

De La Roca, Jorge, and Diego Puga (2017): Learning by Working in Big Cities, *Review of Economic Studies*, 84, 106–142.

Duranton, Gilles, and Diego Puga (2004): Micro-foundations of urban agglomeration economies. In: J. Vernon Henderson and Jacques-François Thisse (Eds.), Handbook of Urban and Regional Economics, vol. 4: 2016-2117. North-Holland, Amsterdam.

Ellison, Glenn, and Edward L. Glaeser (1997): Geographic concentration in US manufacturing industries: a dartboard approach, *Journal of Political Economy*, 105(5), 889–927.

Kosfeld, Reinhold, and Alexander Werner (2012): German Labour Markets – New Delineation after the Reforms of German District Boundaries 2007–2011 (*published in German:* Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007-2011), *Spatial Research and Planning (Raumforschung und Raumordnung)*, 70, 49–64.