University of Zurich[UZH]

URPP Equality of Opportunity

Urban Wage Premia and Heterogeneous Sorting

Julius Lüttge

Equality of Opportunity Research Series #7
November 2022

University of Zurich[UZH]

URPP Equality of Opportunity

URPP Equality of Opportunity Discussion Paper Series No.7, November 2022

Urban Wage Premia and Heterogeneous Sorting

Julius Lüttge
University of Zurich
julius luettge@econ.uzh.ch

URPP Equality of Opportunity, University of Zurich, Schoenberggasse 1, 8001 Zurich, Switzerland
info@equality.uzh.ch, www.urpp-equality.uzh.ch

Urban Wage Premia and Heterogeneous Sorting

Julius Lüttge*

November 18, 2022

Abstract

Wages are higher in urban regions. These urban wage premia may be driven by sorting of more productive workers into urban regions, by a static productivity advantage, and by higher wage growth. This paper documents a size-earnings elasticity of five per cent in Germany. Sorting of more productive workers into larger regions explains 40 per cent of this elasticity. The remainder is driven in equal parts by a static productivity effect and a dynamic learning effect. The urban wage premium is strongly increasing in educational attainment. This is largely driven by different degrees of sorting between education groups. Similarly, there are large urban wage premium differences between occupation groups, which are entirely driven by differences in sorting.

*University of Zurich, Department of Economics. julius.luettge@uzh.ch.

1 Introduction

Regional wage disparities are a feature of any economy. What drives them is one of the key questions in urban economics. In part, higher wages in cities and other more densely populated regions can be attributed to these regions attracting more productive workers. At the same time, there are reasons why cities also might make workers more productive: they offer more job opportunities, allow more specialization, and provide more opportunities for skill development. To understand the benefits that cities provide, however, we need to understand which types of workers benefit most from these factors, and separate these benefits from skill-biased sorting into cities.

This paper separates the portion of regional wage disparities that is due to worker sorting from worker benefits that are tied to the location in which they work. These benefits can either materialize immediately upon taking up a job in a city (a static wage level effect) or accumulate over time as a worker specializes or acquires particularly valuable skills (a dynamic effect). The size of the wage premium and the role of its components likely differs by worker characteristics. Further analysis assesses these heterogeneities by education and occupational cluster: do workers of higher education levels or in more interactive-intensive occupations benefit more from density? And do the sources of these benefits differ, too?

Higher wages in more densely populated regions have been documented in many contexts (see e.g., Duranton and Puga, 2004; Combes et al., 2008; Glaeser and Resseger, 2010; Glaeser, 2012; Matano and Naticchioni, 2016). In many settings, a sizeable part of the level difference in wages between regions has been found to be explained by the sorting of inherently more productive workers into urban regions (e.g. Combes et al., 2008; D'Costa and Overman, 2014; Carlsen et al., 2016). The benefit that is not explained by worker sorting is a true benefit that accrues to workers. The most exhaustive study of both the static and the dynamic wage benefits that urban workers experience has been put forward by De la Roca and Puga (2017). Further work also directly studies the wage growth of workers in urban areas (Glaeser, 1999; Glaeser and Maré, 2001; Yankow, 2006; Baum-Snow and Pavan, 2012). Another branch of the literature is concerned with the heterogeneity of wage premia of workers with different characteristics. Grujovic (2020) documents the wage premium to be higher for workers who work in jobs with higher abstract task-intensity in Germany. This finding is in line with the results presented by Andersson and Thulin (2013) who find higher premia for non-routine jobs in Sweden. Studies on the variation in wage premium between

education groups have previously found smaller differences than those studying differences between occupations (Carlsen et al., 2016; Korpi and Clark, 2019).

This paper first documents the size of urban wage premia in Germany, separating between sorting, static, and dynamic effects. The method follows De la Roca and Puga (2017). To estimate each of the factors (sorting, static, dynamic) in the urban wage premium, I use German administrative data that follow individuals over time and across locations, so that it is possible to account for experience previously acquired in other locations. The first step of the analysis is to estimate urban wage premia using static OLS. While ignoring sorting and dynamics, this shows the overall size of the urban wage premium. The next step controls for individual characteristics. This filters out the contribution of sorting on ability on the wage premium but is subject to a bias from dynamic effects. In a third step, the specification allows the wage premium to vary by where a worker's previous experience has been acquired. This estimation can distinguish between the static part of the urban wage premium and the dynamic part, which is acquired over time.

Using Spanish data, De la Roca and Puga (2017) estimate that an urban wage premium of ca. 5 per cent per doubling in city size. They find that sorting of more productive workers into cities is not a driver of the observed urban wage premium. The other two forces, the immediate static productivity gain and a dynamic gain that accrues over time both explain roughly half of the urban wage premium. In replicating the results of De la Roca and Puga (2017), this study finds a similar magnitude of the total urban wage premium for the German case: a doubling in region size is associated with a roughly 5 per cent increase in wages, conditional on observables. However, the role of the forces behind this premium differs from the Spanish case. Sorting of inherently more productive workers explains about 1.5 percentage points of this effect. The remaining 3.5 percentage points can - as in the Spanish case - be attributed roughly equally to a static and a dynamic gain.

The second part of the analysis assesses to what extent sorting, the static productivity gain, and the value of local experience differ by a worker's education group and by occupation cluster. This goes beyond the analysis in De la Roca and Puga (2017) who only consider the average population effect as their data does not include detailed education or occupation information. Grujovic (2020), on the other hand, does consider heterogeneity by task content in occupations but her analysis abstracts from the dynamic element of urban wage premia, and does not explicitly study differential sorting of different types of workers.

An assessment of the heterogeneity of the urban wage premium between different education and occupation groups allows two types of comparisons between

groups. First, it shows considerable heterogeneity in the size of the urban wage premium: the total urban wage premium is increasing in educational attainment. It also differs between occupation groups: the urban wage premium in production occupations is about twice as large as in manual services, and significantly higher again in cognitive/interactive occupations. Second, it allows a comparison of the relative importance of the drivers behind this relationship. This shows that sorting on unobserved ability is strongly increasing in educational attainment. This stronger sorting of workers with more education explains most of the wage premium difference between education groups. The static urban wage premium is also higher for workers with higher education levels, while the dynamic urban wage premium is similar for each of the groups. The picture is even starker between occupation groups: the differences in the total urban wage premium are almost entirely driven by sorting: the static and dynamic benefits of major cities are very similar between occupation groups. Sorting, however, is skewed: inherently more productive workers in cognitive/interactive occupations tend to move to larger cities. This type of sorting is much weaker in production occupations. In manual services, it is - if anything - reversed: workers negatively select into larger cities.

The paper proceeds as follows: Section 2 presents the data, sample restrictions, and summary statistics. Section 3 describes the estimation first of static and then of dynamic urban wage premia. Section 4 presents the results of the static estimation (section 4.1 for the full sample and heterogeneity in section 4.2) and of the dynamic estimation (overall results in section 4.3 and heterogeneity in section 4.4). Section 4.5 discusses what the results imply for the relative role of the drivers of the urban wage premium.

2 Data

Estimating dynamic urban wage premia requires panel data that follow individual workers over time and space. German social security data fulfill this requirement, with the additional benefit that they include relatively detailed worker information, such as education and occupation.

I use the factually anonymous version of the Sample of Integrated Labour Market Biographies (SIAB Regional File 7517, Antoni et al., 2019), provided by the Institute for Employment Research (IAB). This is a 2 per cent random sample drawn from the Integrated Employment Biographies (IEB), which cover all employees employed in the years 1975 to 2017. The data do not include civil servants, military personnel, or the self-employed.

For each worker, the data include the location of the establishment in one of

328 regions. These are based on the 401 German administrative districts ('Landkreise' and 'kreisfreie Städte'), which are aggregated such that each region contains at least 100,000 inhabitants. As a measure of population density, I compute population within 10 km of the average resident for each region, using grid cell data from GEOSTAT (2006). I also use population density and population totals as alternative measures. However, the population within 10 km of the average inhabitant is my preferred measure because it is less affected by where administrative boundaries are drawn.

For the estimation of dynamic premia, I classify cities into size groups, combining all three measures: the largest size group contains all cities that are among the top 10 in all of population total, population density, and population within 10 km of the average resident. The second size group contains all cities that are in the top 20 for each measure (but not in group 1). The third group includes all other regions.[1] As a robustness check, I also use an alternative classification of regions: ranking regions by their population within 10 km of the average resident, I classify the largest regions as major cities, including regions that together contain 25 per cent of the population. Those regions that contain the next 25 per cent of the population are classified as medium cities. The regions that contain the remaining 50 per cent of population are classified as rural. Using this alternative measure, the size classes contains 38, 62, and 167 regions, respectively.

The SIAB data set includes a relatively high level of demographic detail. This sets it apart from other social security datasets, which often lack education or occupation information. The R7517 version of the data includes 120 occupations. I group these into the three clusters 'manual services', 'production', and 'cognitive/interactive'. The occupation clusters are defined based on similarity of task content. This is done by using information on task use from the BIBB Survey on Qualification and Working Conditions of 2006 (Hall and Beermann, 2006). For more details on the clustering method, see Lüttge (2022).

Education is coded into five groups in the raw data, which I combine into three groups for easier interpretability: workers without vocational training, workers with vocational training or Abitur, and university graduates (including graduates of universities of applied sciences).

Wages in the SIAB data are top-coded at the threshold for social security contributions. To address this, I simulate censored wages, following the procedure by Card et al. (2013), which is based on Dustmann et al. (2009). I fit a series

[1]The largest size class includes Cologne, Essen, Frankfurt, and Munich. The second size class includes Düsseldorf, Hamburg, Nuremberg, and Stuttgart. Eastern German regions are not included: see below.

of Tobit models to log daily wages, and then impute an uncensored value for each censored observation using the estimated parameters of these models and a random draw from the associated (left-censored) distribution. Wages are then deflated to 2015 prices using the CPI.

2.1 Sample restrictions

Estimating the value of experience obtained in different locations requires workers' full labour histories. So we need to exclude all workers who are not observed for earlier parts of their careers. First, I exclude workers who were born before 1957, since they were over 18 in the first year of data in 1975 so that their first years of potential experience are not observed. Second, I restrict the sample period to 1995 to 2017. This way, all years include at least some workers with 20 years of experience or more. However, since all regressions include year fixed-effects, this is not strictly necessary, but due to the restriction on date of birth, earlier years include few observations.

Third, since Eastern Germany is included in the data only from 1992, I exclude all workers who likely gained unobserved experience in Eastern Germany before that date. This means, if workers are not observed before 1990 in the West (that is, before the border opened on 9 November 1989), they are only included if their age suggests that they will not have worked before 1992 at all, i.e., they are born in 1974 or after. Experience gained in the East is observed from 1992, and I include it to compute past experience. However, I exclude workers who currently work in the East, as I would only be able to include workers who are originally from the West but for the latest years. With these restrictions, the only workers included with unobserved experience will be those who moved West-East between November 1989 and December 1991 and returned West afterwards. However, this group is unlikely to be significant.

Similarly, experience gained abroad is also not observed. Since this is likely higher among foreign citizens, I exclude them from the sample. However, the citizenship variable is not very reliable (for some workers, it jumps back and forth from German to non-German), so I count all workers as German whose citizenship is listed as German for at least half their spells.

I also exclude sectors where wages are highly regulated, and therefore unlikely to respond to market forces in the same way as others: public administration, education, medicine, and agriculture. For similar reasons, workers on apprenticeship contracts are excluded. Note that while these categories are excluded from the regressions, experience in these categories is still counted as prior experience.

2.2 Summary Statistics

Table 1. Summary sample statistics

Observation share in %	Manual services	Production	Cognitive/ interactive	Total
Manual services	1.8	2.6	0.7	5.2
Production	16.9	35.8	27.7	80.4
Cognitive/interactive	0.5	0.5	13.5	14.5
Total	19.2	38.9	41.9	100

Mean wage in EUR	Manual services	Production	Cognitive/ interactive	Total
Manual services	74	87	93	83
Production	89	106	132	111
Cognitive/interactive	110	118	165	161
Total	88	104	142	117

Notes: The table shows summary statistics of the sample, weighted by employment spell duration. The sample includes 3,823,785 observations of 227,977 workers, covering an average of 10.4 years per person.

Table 1 shows summary statistics of the sample. We have 227,977 workers in the sample who are observed for an average total duration of 10.4 years, which results in a total of 3,823,785 observations. The mid education group is the largest, both overall and in each of the three occupation groups. The highest education group is heavily skewed towards cognitive/interactive occupations. Mean wages are, as one would expect, increasing in education. Between occupation groups, wages show the same pattern for all education groups: wages are lowest in manual services, and highest in cognitive/interactive occupations.

Figure 1 shows raw daily median wages by region density. It is clearly visible that wages are higher in denser regions. They are highest in Munich and Frankfurt as well as in certain cities home to large corporations[2]. The figure also shows Eastern regions in red, which are, however, excluded from further analysis. The wage gap to regions in the West is clearly visible.

[2]such as Erlangen, Wolfsburg, Ingolstadt, and Ludwigshafen (which are home to the headquarters of Siemens, VW, Audi, and BASF, respectively). Much of the high wages in these cities is explained by worker characteristics, in particular the industry they work in.

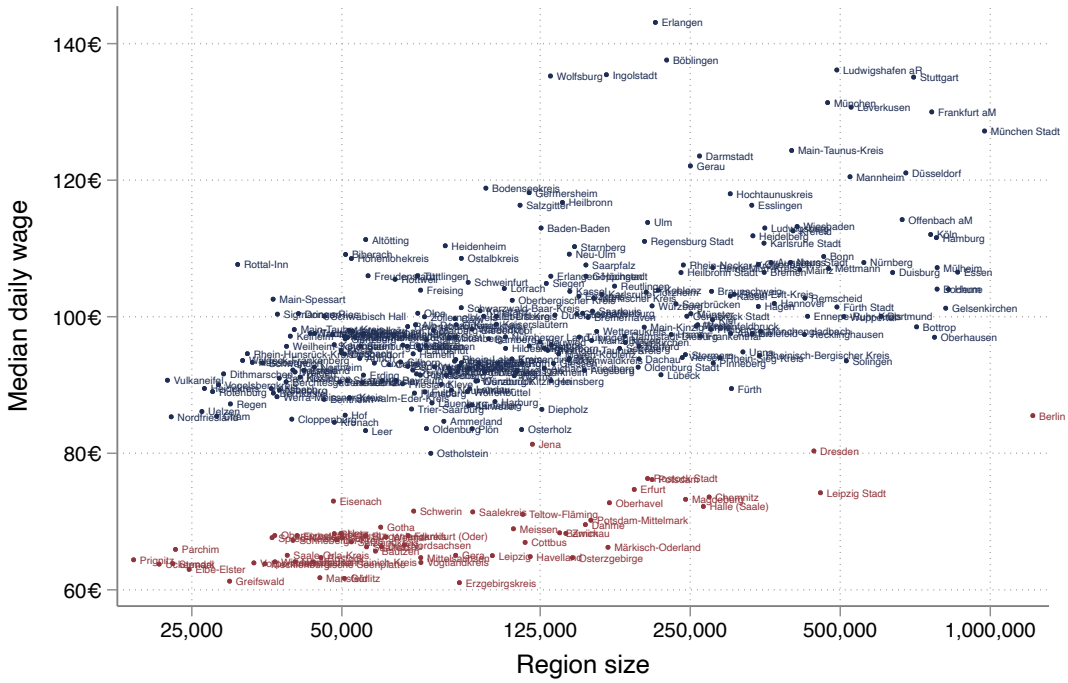

Figure 1. Daily median wages and region size, 1995-2017.

Notes: This figure plots raw median daily wages in each region, pooling years 1995 to 2017 and deflated to 2015 Euros. Region size is defined as the number of residents within 10 km of the average resident and shown on a log scale. The regions in red are in East Germany, which are excluded from the main analysis.

3 Estimation strategy

3.1 Static estimation of urban wage premia

I estimate the static urban wage premium following the two-step procedure in De la Roca and Puga (2017)[3]. In the first step, I regress log wages of individuals i who work in region r at time t ($\ln w_{irt}$) on a region dummy, a year fixed effect, and a vector of characteristics of worker i at time t, x_{it}, which includes education, occupation group, industry, experience, experience^2, tenure, and tenure^2:

$$\ln w_{irt} = \alpha_r + \delta_t + x'_{it}\psi + \epsilon_{irt} \tag{1}$$

$\hat{\alpha}_r$ estimates each city's wage level adjusted for composition and time effects. In the second step, I regress the estimated region fixed effects on the log of the region size measure. The estimated coefficient $\hat{\beta}$ then is the urban wage premium,

[3]This procedure is commonly used when including worker fixed-effects due to the complex variance-covariance structure that would result from including city size as a regressor in equation 1 (for a discussion, see Combes et al., 2008).

and indicates the wage gain that is associated with a doubling in region size:

$$\hat{\alpha_r} = \beta \ \ln(\text{regionsize})_r + u_r \tag{2}$$

Equation 1 does not include an individual fixed effect. As a result, any unobserved personal characteristics that affect wages are subsumed into the error term. If, say, inherently more productive workers tend to be located in cities, then the α_r estimates from equation 1 will be increasing in city size, and the resulting β estimate will be biased upwards. To address this, we can include a worker fixed effect, estimating equation 3 instead:

$$\ln w_{irt} = \alpha_r + \delta_t + \mu_i + x'_{it}\psi + \epsilon_{irt} \tag{3}$$

The worker fixed effect in equation 3, μ_i, captures all time-invariant personal characteristics, observed or unobserved. As a result, the other coefficients are only estimated based on time-varying personal characteristics. For our main parameters of interest, the region fixed-effects, this means that they are estimated only based on workers who moved location at least once during their careers.

By comparing urban wage premia results based on wage regressions 1 and 3, we can therefore get a sense of the role of sorting for urban wage premia. Previous studies have interpreted the different estimates from equations with and without individual fixed effects as evidence of sorting (e.g., Glaeser and Maré (2001), Combes et al. (2008)). Note, though, that while equation 3 addresses sorting, it does not account for dynamic effects, so may still be biased (De la Roca and Puga, 2017). In other words, $\hat{\alpha_r}$ does not separate the immediate wage gain associated with moving to region r from the wage growth effect that workers in region r experience over time. I address this issue in section 3.2.

To assess how the relationship between wages and region size differs for different education and occupation groups, we can specify a third version of the wage regression, replacing the region fixed effect α_r in equation 1 with a separate effect for each subgroup, α_{re} for each education group e or α_{ro} for each occupation group o. This allows us to estimate a corresponding second step for each education and occupation group separately. Instead of regression 1, we then estimate regressions

$$\begin{aligned} \ln w_{irt} &= \alpha_{re} + \delta_t + x'_{it}\psi + \epsilon_{irt} \\ \ln w_{irt} &= \alpha_{ro} + \delta_t + x'_{it}\psi + \epsilon_{irt} \end{aligned} \tag{4}$$

The region fixed effect α_{re} captures how much higher or lower wages are in region r for a worker in education group e, in comparison to workers with the same

personal characteristics in other locations. Personal characteristics in the first regression include occupation, so that we are comparing the city-education effect within the same occupation group. Similarly, the second regression includes the education group in the vector of personal characteristics, so that we are comparing the city-occupation effect within the same education group. If we included both interactions (α_{re} and α_{ro}), this would be cumbersome to interpret: the region-education fixed effect would capture how much higher of lower wages are in region r for a worker in education group e, in comparison to workers with the same personal characteristics in other locations, while holding fixed the deviation of wages in region r in the worker's occupation group o from similar workers' wages.

In the corresponding second step, we regress the region fixed effect on city size for each subgroup separately, i.e.,:

$$\begin{aligned} \hat{\alpha}_{re} &= \beta_1 \ \ln(\text{regionsize})_r + u_{re} \\ \hat{\alpha}_{ro} &= \beta_2 \ \ln(\text{regionsize})_r + u_{ro} \end{aligned} \tag{5}$$

for each education group e and each occupation group o.

3.2 Dynamic estimation of urban wage premia

The estimation method described above captures only the static urban wage premium. If the labour market experience acquired in urban areas is more valuable, then the static estimates are biased as they capture both the immediate urban wage premium and the portion that materializes over time. To distinguish these two components, we can follow the method employed by De la Roca and Puga (2017). We need to estimate an equation that allows the value of experience to vary by where it is acquired and where it is used. In general terms, we can do this by enhancing equation 1 by adding a function of the experience worker i has obtained in each location j up to point t and is using in location $r(it)$ where they now work, $f(exp_{it,j=1}, \ldots, exp_{it,j=J}, r(it))$:

$$\ln w_{irt} = \alpha_r + \mu_t + \mu_i + f(exp_{it,j=1}, \ldots, exp_{it,j=J}, r(it)) + x'_{it}\psi + \epsilon_{irt} \tag{6}$$

Here, α_r captures the static wage premium in region r. μ_t is a time fixed effect. μ_i is an individual-fixed effect. x_{it} is a vector of individual characteristics, such as education, industry, occupation, tenure, and tenure^2.

A natural functional form for $f(exp_{it,j=1}, \ldots, exp_{it,j=J}, r(it))$ is

$$\begin{aligned} &f(exp_{it,j=1}, \ldots, exp_{it,j=J}, r(it)) = \\ &\sum_{j=1}^{J}\sum_{k=1}^{K}(\gamma_{jk} exp_{itj}\mathbb{1}(r_{it}=k) + \delta_{jk} exp_{itj} exp_{it}\mathbb{1}(r_{it}=k)) \end{aligned} \tag{7}$$

In this form, γ_{jk} captures the value of experience acquired in region j if it is used in region k for each region-region combination j-k. δ_{jk} captures the decay of the value of local experience. This is similar to a squared experience term exp^2_{itj}. However, exp^2_{itj} would not allow for continued decay in the value of local experience from location j after moving to new location k. The expression $exp_{itj}exp_{it}$, on the other hand, lets the value of experience in region j decay continuously, albeit at possibly different speeds depending on where it is used due to the $\mathbb{1}(r(it) = k)$ term.

Rather than using 267^2 combinations of regions, I group regions into three size classes. As described in section 2, the largest size class includes Cologne, Essen, Frankfurt, and Munich. The second size class includes Düsseldorf, Hamburg, Nuremberg, and Stuttgart. The third class includes all other regions. I define region 3 as the base region and restrict the value of experience to be the same whether it used in city group 1 and city group 2, i.e., combine city groups $k = 1$ and $k = 2$ (but only in terms of where experience is used - not in terms of where it is acquired). The regression then is:

$$\begin{aligned} \ln w_{irt} =& \alpha_r + \mu_t + \mu_i \\ &+ \gamma_{33}\text{exp}_{it} + \phi_{32}\text{exp}_{it}\mathbb{1}[r \in \{1,2\}] \\ &+ \phi_{13}\text{exp}_{it,j=1} + \phi_{12}\text{exp}_{it,j=1}\mathbb{1}\,[r \in \{1,2\}] \\ &+ \phi_{23}\text{exp}_{it,j=2} + \phi_{22}\text{exp}_{it,j=2}\mathbb{1}[r \in \{1,2\}] \\ &+ \delta_{33}\text{exp}^2_{it} + \eta_{32}\text{exp}^2_{it}\mathbb{1}[r \in \{1,2\}] \\ &+ \eta_{13}\text{exp}_{it,j=1}\text{exp}_{it} + \eta_{12}\text{exp}_{it,j=1}\text{exp}_{it}\mathbb{1}[r \in \{1,2\}] \\ &+ \eta_{23}\text{exp}_{it,j=2}\text{exp}_{it} + \eta_{22}\text{exp}_{it,j=2}\text{exp}_{it}\mathbb{1}[r \in \{1,2\}] \\ &+ x'_{it}\psi + \epsilon_{irt} \end{aligned} \tag{8}$$

Here, γ_{33} captures the value of experience acquired in a region of size 3 (any region that is not among the largest 8 cities) and used in a region of size 3.[4] $\hat{\gamma}_{33}$ is expected to be positive because it is simply the value of local experience. ϕ_{13} captures the additional value of experience acquired in one of the largest 4 cities.

[4]For this illustration, we abstract from the squared terms in experience, which are captured by δ_{33} and the η_{jk} terms.

However, we again distinguish by where the experience is used, so city experience used in a smaller size 3 region is $\gamma_{13} = \gamma_{33} + \phi_{13}$. Major city (size 1) experience used in a size 1 city is $\gamma_{11} = \gamma_{33} + \phi_{13} + \phi_{12}$.

If experience acquired in one of the largest 4 cities is more valuable than non-urban experience, no matter where it is used, then $\hat{\phi}_{13}$ will be positive. This would also indicate that the additional value of city experience is portable. If it is perfectly portable, city experience used in the city will have the same value, and $\hat{\phi}_{12}$ will be zero. If, however, major city experience is even more valuable when used in a major city, $\hat{\phi}_{12}$ will be positive. If major city experience is more valuable than size 3 experience but only if used in a major city, i.e., if it is not portable at all, then $\hat{\phi}_{13}$ will be zero and $\hat{\phi}_{12}$ will be positive.

The parameters ϕ_{23} and ϕ_{22} estimate the corresponding value of experience acquired regions of size 2 (the second largest 4 cities). Above, we abstracted from possible non-linearities in the value of experience of each type. Potential decay in the experience terms is accounted for by the squared terms in experience, which are captured by δ_{33} and the η_{jk} terms.

Assessing heterogeneity by education and occupation in the dynamic case is similar to the static case. We can add an education group dummy and an occupation group dummy to each of the γ_{jk}, ϕ_{jk}, δ_{jk} and η_{jk} terms.[5]

4 Results

4.1 Static Urban Wage Premia Estimation Results

Table 2 shows the results from the static urban wage premia estimation of regressions 1 and 2. Column 1 shows results of a regression without worker fixed effects. As expected, wages increase in education and task level, and are concave in experience and tenure. Column 2 shows the results from regression 2, showing that a doubling in city size is associated with a 0.0505 log point (i.e., 5.2 per cent) increase in wages.

Figure 2 shows what is behind column 2 of Table 2, plotting the region fixed effects against region population density. The urban wage premium estimate of 0.0505 log points is in line with those found in other settings (0.046 for Spain in De la Roca and Puga (2017), 0.051 for France in Combes et al. (2010), and 0.041 for the U.S. in Glaeser and Resseger (2010)).

Controlling for observables, and in particular for industry and occupation fixed effects, removes the outlier status of the industrial cities from Figure 1. Note also

[5] Alternatively, one could add a dummy for each education-occupation cell. However, since some cells are very small (see Table 1), this is not feasible to estimate.

Table 2. Estimation of the static region size wage premium

	(1) Log real wage	(2) Region FE	(3) Log real wage	(4) Region FE
Experience	0.0366***		0.0481***	
	(0.0003)		(0.0005)	
Experience^2	-0.0006***		-0.0007***	
	(0.0000)		(0.0000)	
Firm tenure	0.0154***		0.0055***	
	(0.0002)		(0.0002)	
Firm tenure^2	-0.0004***		-0.0001***	
	(0.0000)		(0.0000)	
Education: High	0.5835***		0.3367***	
	(0.0041)		(0.0058)	
Education: Mid	0.1443***		0.1028***	
	(0.0027)		(0.0039)	
Occ: Cogn./interactive	0.3542***		0.0919***	
	(0.0023)		(0.0021)	
Occ: Production	0.0545***		0.0224***	
	(0.0018)		(0.0018)	
Log region size		0.0505***		0.0235***
		(0.0047)		(0.0025)
City FEs	Yes		Yes	
Year FEs	Yes		Yes	
Industry FEs	Yes		Yes	
Worker FEs	No		Yes	
R^2	0.452	0.349	0.775	0.258
Observations	3,758,979	267	3,745,495	267

Notes: The table shows the urban wage premia estimation with and without worker fixed effects. Column 1 shows estimates from wage regression 1, where log wages are regressed on personal characteristics and region dummies. Column 2 shows estimates from regression 2, where the region dummy estimates from regression 1 are regressed on the region's size measure. For column 3, the regression includes individual fixed effects. Column 4 shows the results of regression 2, using column 3's region coefficients. All specifications include a constant term. Columns 1 and 3 include a year-indicator, sector dummies, and occupation dummies. Coefficients are reported with standard errors in parentheses, which are clustered by worker in columns (1) and (3). ***, **, and * indicate statistical significance at the 1, 5, and 10% levels. Worker values of experience and tenure are computed to daily accuracy and are expressed in years.

that region size accounts for a third of the earnings differences between locations over and above individual observables (R^2 of 0.349 in column 2).

Since regression 1 does not control for individual unobserved characteristics,

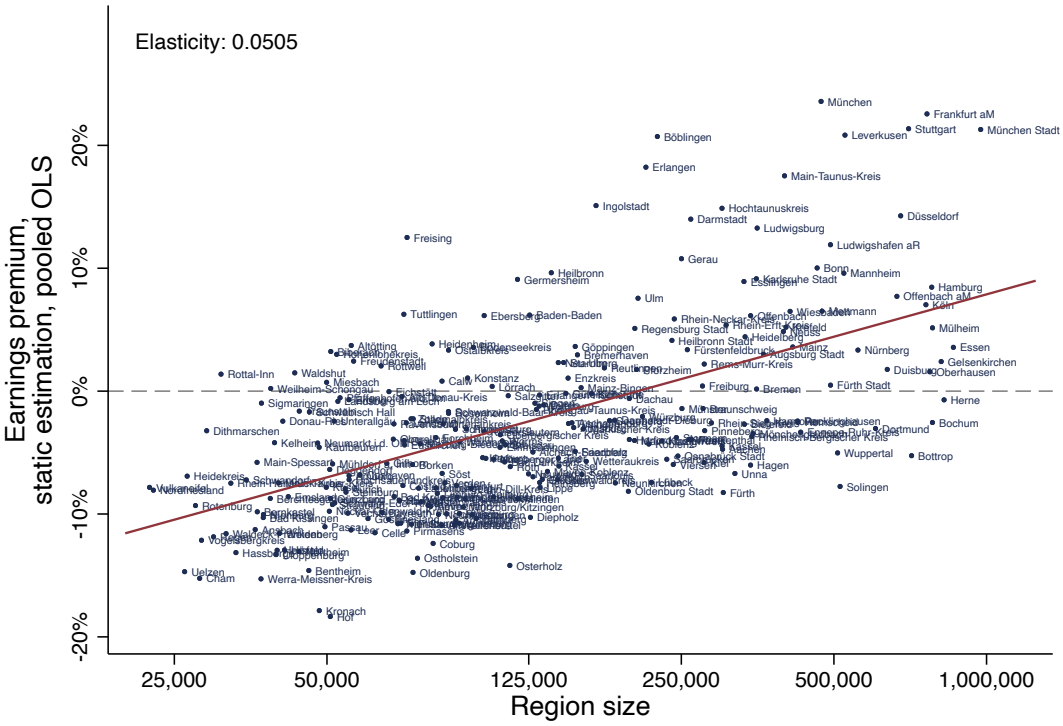

Figure 2. Estimation of the static regional wage premium

Notes: This figure plots the region fixed effects of column 1 in Table 1 against region size. The trend line is the regional wage premium, which is also shown in column 2 of Table 1. Region size is defined as the number of residents within 10 km of the average resident and is shown on a log scale.

part of the estimated wage differentials are likely due to worker selection. In order to address worker sorting across cities on unobservables, column 3 reports estimates of a version of regression 1 that includes worker fixed effects. In this regression, the education and occupation coefficients are only estimated for workers who have changed education and occupation, respectively, at least once during their careers. Similarly, the region fixed-effects are only estimated for workers who are observed in more than one location during their career. The associated region-size wage elasticity is reported in column 4, which is roughly half that reported in column 2. This may be interpreted as revealing that roughly half of the static urban wage premium is explained by more productive workers sorting into larger cities. Note, though, that neither specification controls for dynamic bias. This will be addressed in section 4.3. The drop in the elasticity when including worker fixed effects found here is in line with those found in other studies: Mion and Naticchioni (2009) report a drop of 66% for Italy, Combes et al. (2010) report a drop of 35% for France, De la Roca and Puga (2017) report a drop of 47% for Spain, Grujovic (2020) finds a drop of 51% for Germany.

4.2 Heterogeneity in static urban wage premia

The estimates of Table 2 and Figure 2 are based on workers of all education and occupation groups. In the following, I investigate how the static urban wage premium differs between education and occupation groups. This is done by estimating variations of equation 4 and the corresponding equation 5.

The results of the first step is shown in Table 3. In column 1, the region fixed effect is interacted with an education dummy. In column 2, the region fixed effect is interacted with an occupation dummy. Columns 3 and 4 show the corresponding results where the regressions include a worker fixed effect.

The results in Table 3 are all but identical to those of Table 2. The resulting region-education and region-occupation fixed effects from columns 1 and 2 are shown in Figures 3 and 4, respectively. In both figures, column 1 shows the same values as Figure 2, while the other columns show the region fixed effects for each subgroup.

The estimates of the urban wage premium regression 5 display significant heterogeneities. These are also shown in Figure 5. Figure 3 and panel 1 of Figure 5 show that the urban wage premium is increasing in education. For a worker with no vocational training, a doubling in city size is associated with two per cent higher wages. For a worker with vocational training, the urban wage premium is more than twice as large, at about five per cent. For university graduates, the wage premium is even steeper at roughly eight per cent.

Figure 4 and panel 2 of Figure 5, showing the step 2 results corresponding to column 2 of Table 3, show differential urban wage premia by occupation group. Here, we see higher urban wage premia in production occupations, and highest premia in cognitive/interactive occupations.

Panels 3 and 4 of Figure 5 are equivalent to panels 1 and 2, but now the underlying wage equations include worker fixed-effects (Table 3, columns 3 and 4). As expected, controlling for worker sorting attenuates any urban wage premia estimates. Interestingly, only the wage premia between education groups in panel 3 remain meaningfully different between each other. However, the difference between groups loses statistical significance. Similarly to the regressions for all groups reported in Table 4, worker sorting explains around half of the static urban wage premium for each education group.

For occupation groups, this is different. In panel 4, all the urban wage premium estimates for all occupation groups are reduced to the same level of around .02. Consequently, worker sorting explains the entirety of the differential urban wage premium between occupation categories.

Grujovic (2020) also estimates heterogeneity in static urban wage premia. Her

Table 3. Wage regressions for estimation of region size wage premium by education and occupation

	(1) Log real wage	(2) Log real wage	(3) Log real wage	(4) Log real wage
Experience	0.0366***	0.0365***	0.0480***	0.0488***
	(0.0003)	(0.0003)	(0.0005)	(0.0006)
Experience^2	-0.0006***	-0.0006***	-0.0007***	-0.0008***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Firm tenure	0.0153***	0.0154***	0.0054***	0.0028***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Firm tenure^2	-0.0004***	-0.0004***	-0.0001***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Occ: Cognitive/interactive	0.3555***		0.0918***	
	(0.0023)		(0.0021)	
Occ: Production	0.0536***		0.0222***	
	(0.0018)		(0.0018)	
Educ: mid		0.5760***		0.3119***
		(0.0041)		(0.0066)
Educ: high		0.1441***		0.1021***
		(0.0027)		(0.0047)
Region x Educ FEs	Yes		Yes	
Region x Occ FEs		Yes		Yes
Year FEs	Yes	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes	Yes
Worker FEs	No	No	Yes	Yes
R^2	0.456	0.456	0.775	0.785
N	3,758,979	3,758,979	3,745,495	2,823,446

Notes: The table shows wage regression results of variations on equation 4. All specifications include a constant term, as well as year- and industry-dummies. Column 1 includes region-by-education fixed effects. Column 2 includes region-by-occupation fixed effects. Columns 3 and 4 are equivalent to columns 1 and 2 but also include worker fixed effects. The corresponding second step coefficients (the region wage premia estimates) are shown in Figure 5. The sample in column 4 is restricted to workers who have not changed their occupation in the past two years. Coefficients are reported with standard errors in parentheses, which are clustered by worker. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels. Worker values of experience and tenure are computed to daily accuracy and are expressed in years.

results are not directly comparable but overall in line with those presented here: she also finds the urban wage premium to be increasing in education, and this difference not to be fully explained by sorting. She also considers heterogeneity by occupation category, and finds that the urban wage premium is higher in

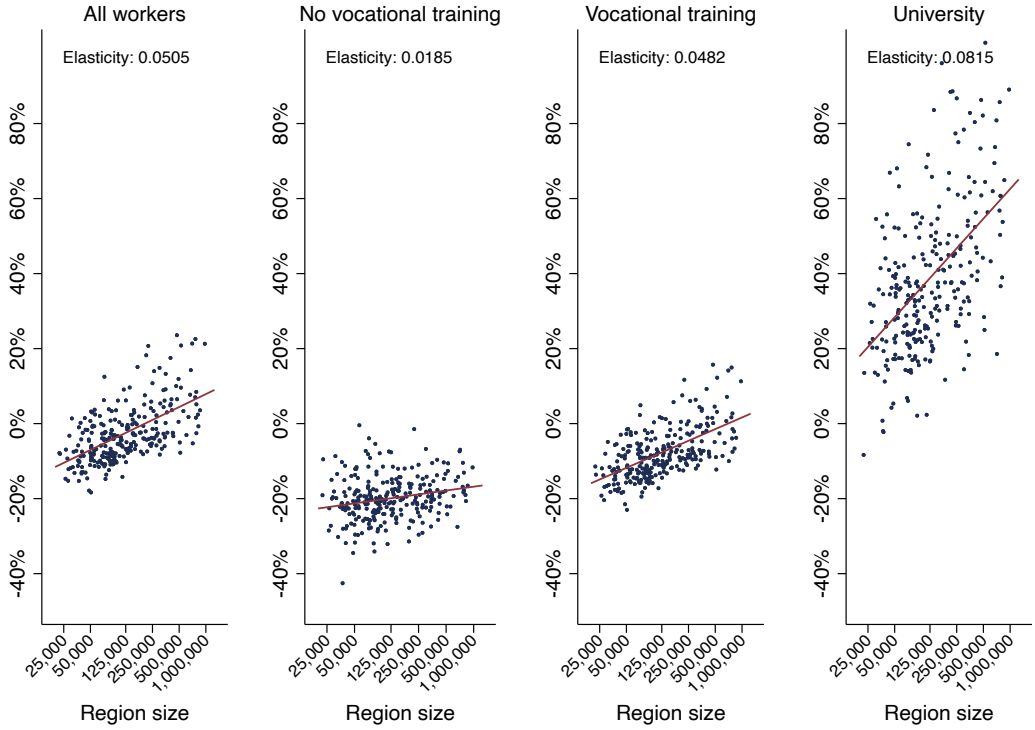

Figure 3. Region fixed effects of static urban wage premium estimation: by education group

Notes: This figure plots the region fixed effects against region size. The left panel reproduces the values shown in Figure 2. The other panels show the region fixed effects for each education group. These are estimated in regression 4, whose other coefficients are shown in column 1 of Table 3. The region fixed effects corresponding to column 3 of Table 3 are shown in Appendix Figure A2.1. The trend lines are the regional wage premium. Region size is defined as the number of residents within 10 km of the average resident and is shown on a log scale.

occupations with higher abstract task intensity. The occupation groups used in this study (manual services, production, cognitive/interactive) are not defined by abstract task intensity but are grouped into categories such that their component occupations are similar to each other in task content. This method is a priori agnostic about the types of tasks within occupation groups. Ex-post, however, the main tasks in the cognitive/interactive occupation group are abstract tasks[6]. Therefore, the result that the urban wage premium is highest in this group is also consistent with the results in Grujovic (2020) who finds that the static urban wage premium is higher in occupations with higher abstract task intensity.

[6]The main tasks are 'working with computers', 'consulting, advising', 'gathering information, investigating', and 'organizing, planning, working out operations'.

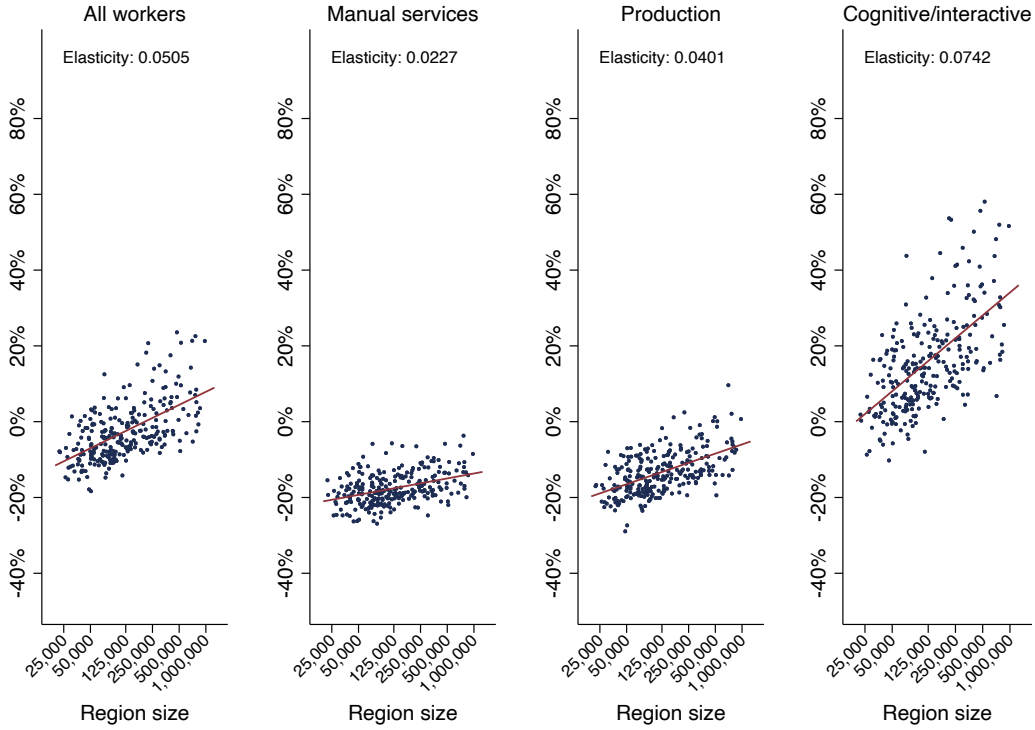

Figure 4. Region fixed effects of static urban wage premium estimation: by occupation group

Notes: This figure plots the region fixed effects against region size. The left panel reproduces the values shown in Figure 2. The other panels show the region fixed effects for each occupation group. These result from regression 4, whose other coefficients are shown in column 2 of Table 3. The region fixed effects corresponding to column 4 of Table 3 are shown in Appendix Figure A2.2. The trend lines are the regional wage premium. Region size is defined as the number of residents within 10 km of the average resident and is shown on a log scale.

4.3 Dynamic Urban Wage Premium

The above results ignore any dynamic component of the urban wage premium. In the following, we let the value of local experience vary by where it is acquired and where it is used. This allows simultaneous estimation of both the immediate urban wage premium and the dynamic urban wage premium.

The results of the dynamic wage premium specification of equation 8 are shown in column 1 of Table 4. The regression includes the same individual characteristics as in Table 2. The coefficients indicate that the first year of experience in one the 4 largest cities (size class 1) is associated with a 2.4 percentage point higher wage increase ($exp(\gamma_{33}+\phi_{13}+\phi_{12}+\delta_{33}+\eta_{13}+\eta_{12})-1 = exp(0.0463+0.0194+0.0034+(-0.0007-0.0002-0.0001))-1 = 7.0\%$) than the first year of experience in a smaller region (size class 3, $exp(\gamma_{33}+\delta_{33})-1 = exp(0.0463+(-0.0007))-1 = 4.6\%$).

The estimates also tell us whether it matters where experience is used in addition to where it is acquired. The interaction terms of where experience is acquired

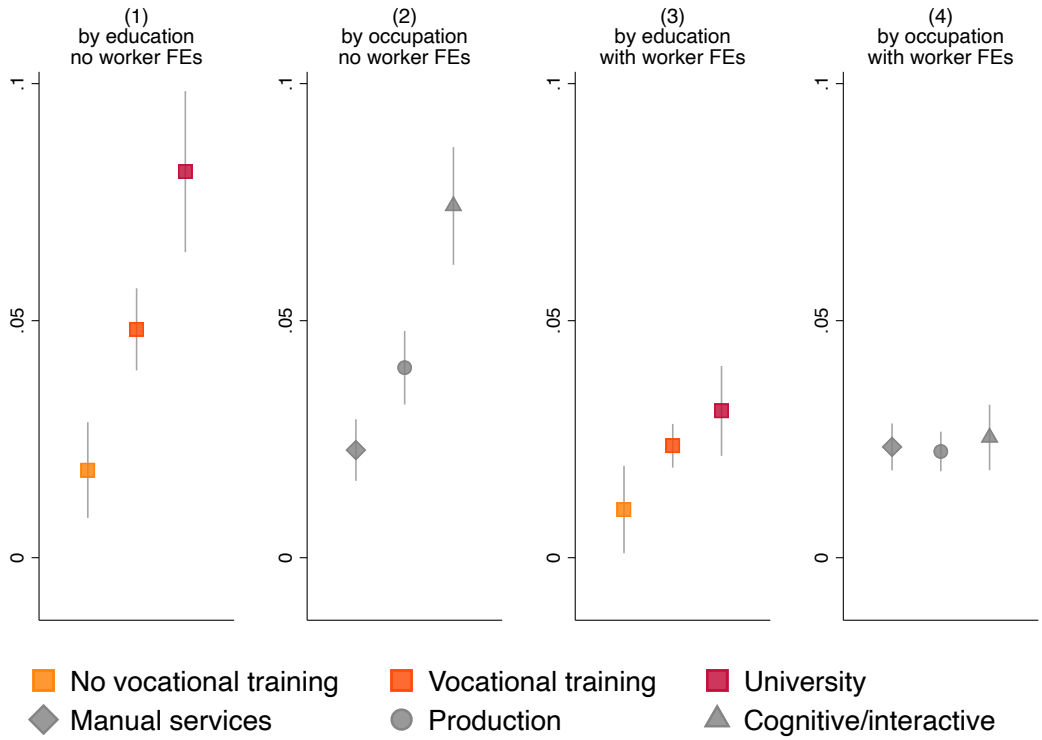

Figure 5. Estimated static urban wage premia by education and occupation groups
Notes: This figure shows the static urban wage premium by education and occupation group. These are the β coefficients of regression 2, where the region fixed effects are from regression 4 (whose results are reported in Table 3). For each education and occupation group, the region fixed effects are regressed separately on region size. The vertical lines show 95 per cent confidence intervals. The columns correspond to those in Table 3: in column 1, the underlying regression includes only region-by-education fixed effects. In column 2, it includes only region-by-occupation fixed effects. Columns 3 and 4 are equivalent to columns 1 and 2 but the underlying first step also includes worker fixed effects.

and where it is used (e.g., "Experience urban size 1 $\times$ now in size 1 or 2" for experience acquired in size class 1 and now working in size class 1 or 2) indicate that large city experience is more valuable when used in a large city. Similarly, experience acquired in a smaller region is also more valuable when used in a larger city. Part of this effect may be explained by a move-premium, which would arise, for instance, if workers only moved if the resulting wage gain compensated them for the move. To control for this potential bias, the regression includes a squared polynomial of the count of a worker's location moves.

Overall, the evidence shows that experience acquired in large cities is portable. The same is true for moves from smaller to larger locations: previous experience is more valuable in the new job location.

The results in column 1 of Table 4 show that where workers acquire experience matters more than where they use it. So while moving from a smaller region to a large city brings additional benefits, the main effect is that any additional

experience gained large cities is more valuable and remains valuable even if the worker moves again later.

Column 2 of Table 4 shows the coefficient of regression 2, analogously to column 2 of Table 2. This can be interpreted as the initial static premium after moving from the median sized city to a city twice as large. It abstracts from any future earnings gains that only materialize over time. Column 3 displays the medium-term premium, which incorporates the static and dynamic premia by adding the value of a worker's experience evaluated at the average experience in a location in our sample (16.2 years). Comparing column 3 of Table 4 to column 2 of Table 4, this exercise shows that in the medium term about half of the urban wage premium is static ($.0191/.0387 = 49\%$), while the other half is dynamic.

De la Roca and Puga (2017) report estimates for Spain by the same method, and find the immediate urban wage premium to be very similar at 0.022. However, the dynamic component is larger in the Spanish case where the urban wage premium rises to 0.051 in the medium term.

Admittedly, the estimates in Table 4 are cumbersome to interpret because it requires taking the squared experience terms into account. Figure 6 shows the results visually. The solid blue line shows the earnings profile of a person who works in the largest city, Munich, in relation to a person who works in a median size city, Konstanz, for the first ten years of his career. In this example, there is no significant initial, static, Munich wage premium. However, the gap to an equivalent worker in Konstanz grows over time, so that after ten years, the Munich worker's wage is predicted to be ca. 24% higher.

The dashed blue line shows what happens if the worker moves from Munich to Konstanz after 5 years. If urban experience were not portable, the worker's premium would drop to zero immediately. Instead, her experience acquired in Munich is portable and earns her a higher wage than an equivalent Konstanz worker who never worked in Munich. Over time, this big city-experience premium depreciates but only slowly, so that at the end of the period, she still earns ca. 10% more than an equivalent Konstanz worker who never worked in Munich.

The dashed red line shows the wage profile of a worker who spent the first five years of her career in Konstanz, and then moves to Munich. When she moves, her wage jumps by 5 per cent. This represents the static gain from moving to Munich and the higher value of experience in Munich. So smaller-region experience is also more highly valued in the larger city. However, her wage does not jump to the level of a worker who spent all of her career in Munich. This is because big-city experience is more highly valued in large cities than small-region experience is. While the 95%-confidence intervals of the dashed red line (the Konstanz-Munich

Table 4. Estimation of static and dynamic region size wage premia

	(1) Log real wage	(2) Initial premium (Region indicator coeffs. of col. (1))	(3) Medium-term premium (initial + mean local experience)
Experience	0.0463***		
(γ_{33})	(0.0005)		
Firm tenure	0.0055***		
	(0.0002)		
Location Moves	-0.0041***		
	(0.0002)		
Exp. city size 1	0.0194***		
(ϕ_{13})	(0.0012)		
Exp. city size 2	0.0182***		
(ϕ_{23})	(0.0013)		
Exp. city size 1 × now in size 1 or 2	0.0034***		
(ϕ_{12})	(0.0012)		
Exp. city size 2 × now in size 1 or 2	0.0006		
(ϕ_{22})	(0.0012)		
Exp. city size 3 × now in size 1 or 2	0.0107***		
(ϕ_{32})	(0.0008)		
Experience^2	-0.0007***		
(δ_{33})	(0.0000)		
Exp. city size 1 × exp.	-0.0002***		
(η_{13})	(0.0000)		
Exp. city size 2 × exp.	-0.0003***		
(η_{23})	(0.0000)		
Exp. city size 1 × exp.× now in size 1 or 2	-0.0001***		
(η_{12})	(0.0000)		
Exp. city size 2 × exp.× now in size 1 or 2	-0.0000		
(η_{22})	(0.0000)		
Exp. city size 3 × exp.× now in size 1 or 2	-0.0003***		
(η_{32})	(0.0000)		
High Education	0.3389***		
	(0.0058)		
Mid Education	0.1115***		
	(0.0039)		
Occ: Cognitive/interactive	0.0915***		
	(0.0021)		
Occ: Production	0.0203***		
	(0.0018)		
Log region size		0.0191***	0.0387***
		(0.0025)	(0.0057)
Region FEs	Yes		
Year FEs	Yes		
Industry FEs	Yes		
Worker FEs	Yes		
R^2	0.776	0.187	0.270
Observations	3,745,495	267	267

Notes: All specifications include a constant term. Column 1 includes a year-indicator, industry dummies, and occupation dummies. It also includes squared terms in firm tenure, and location moves, which are not shown to improve readability. Column 1 shows estimates from regression 8, where log wages are regressed on personal characteristics, a person's region-experience history, and a region dummy. Column 2 shows estimates from the corresponding regression 2, where the region dummy estimates are regressed on the region's size measure. Column 3 displays the medium-term premium, which incorporates the static and dynamic premia by adding the value of a worker's experience evaluated at the average experience in a location in our sample. Coefficients are reported with standard errors in parentheses, which are clustered by worker in column (1). ***, **, and * indicate statistical significance at the 1, 5, and 10% levels. Worker values of experience and tenure are computed to daily accuracy and are expressed in years.

mover) and the solid blue line (the Munich stayer) are overlapping, note that much of the imprecision is due to the regions' time-constant urban premia, which depend on the specific region-pair we consider. The dynamic component, i.e., the slopes of the lines are estimated off of grouped regions, and are quite precisely estimated.

Overall, Figure 6 highlights two key facts: first, much of the urban wage premium is not immediate. It accumulates over time by the acquisition of more valuable experience. Second, since urban experience is highly portable, it matters more where experience is acquired than where it is used.

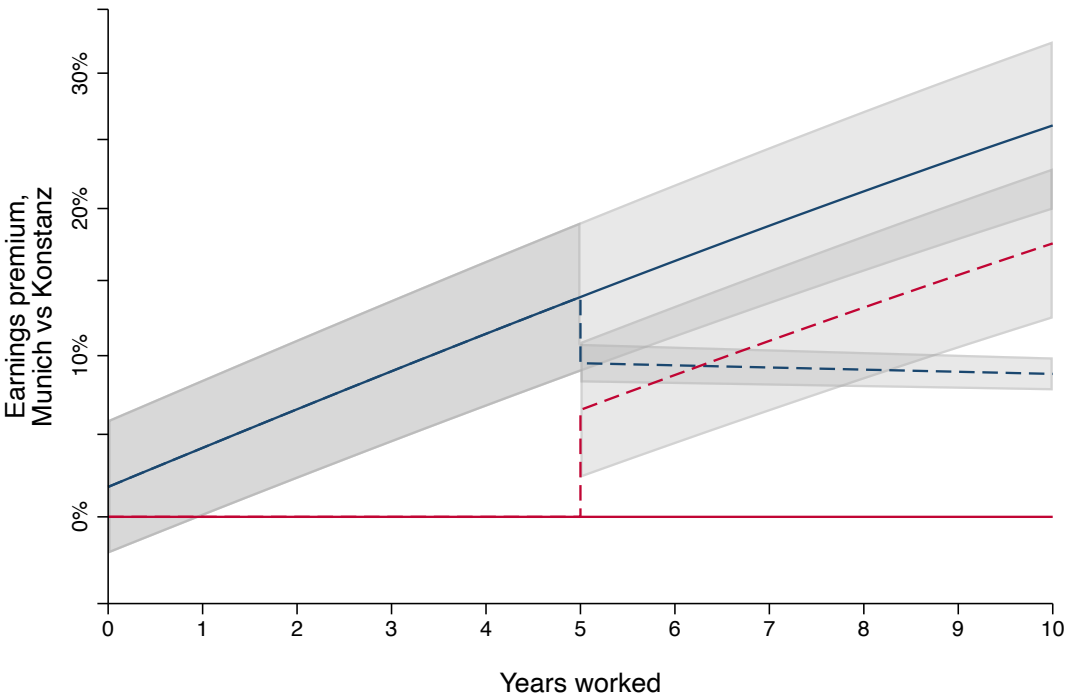

Figure 6. Earnings profiles: major city relative to a smaller region

Notes: The figure shows wage profiles of four hypothetical workers, based on the results shown in column 1 of Table 4. The red line represents a hypothetical worker who starts his career in Konstanz and stays there for ten years. The profiles are shown relative to this Konstanz worker, for whom the profile is therefore flat. The blue solid line shows the profile of a worker who starts hiss career in Munich and works there for ten years. The blue dashed line represents a worker who starts his career in Munich and moves to Konstanz after five years. The red dashed line represents a worker who starts his career in Konstanz and moves to Munich after five years. Munich and Konstanz are chosen as examples because Munich is the largest city, and Konstanz is roughly median-size in the sample. The earnings premium is estimated in log points, so the y-axis labels mark $\ln(1.1) = 0.0953$ for 10%, etc.

4.4 Dynamic premia by education and occupation

When we add an education group dummy or an occupation group dummy to each experience term of equation 8, we can assess the heterogeneity in wage profiles by education and occupation groups.

The wage regression has too many coefficients to be reasonably displayed in a table (three variations of each experience coefficient). Ultimately, we are interested in the resulting short-term and medium-term wage premia. These are shown in Tables 5 and 6. Table 5 shows that the immediate urban wage premium is increasing in education (columns 1 to 3), as is the medium-term premium (columns 4 to 6). The middle education group is by far the largest, so it is not surprising that the immediate and medium term wage premia for this group are almost identical to those of all education groups combined reported in Table 4: a move to a region twice as large is associated with an immediate wage gain of .020 log points. Over time, this wage gain increases in comparison to the counterfactual scenario in which the worker did not move, so that after about 16 years (the mean worker experience in the sample), the wage gain is 0.036 log points.

For a worker without vocational training, the wage premia are lower: the immediate wage gain is estimated at .011 log points, and grows to .021 log points over the medium term. For university graduates, the immediate premium is higher at .027 log points but the medium term premium is only slightly higher than the middle group's at .038 log points.

This pattern of increasing premia in education is in line with the findings of the static premia reported in Figure 2. In fact, the immediate premium from the full dynamic estimation is not much lower than the wage premium from the static estimation. This suggests that the dynamic bias in the static estimation, which ignores the dynamic part of the premium, is small.

Steeper wage profiles for workers with higher levels of education have been widely documented (an early example is Lazear, 1976). The results of dynamic urban wage premia add another dimension to this phenomenon: while workers with higher education levels experience faster wage growth over their careers than workers with lower education levels, this is particularly true in larger cities.

To get a sense of the quantitative importance of the urban component of the education wage premium, consider Figure 7, which shows estimated absolute wage profiles for workers in each education group. In each education group panel, the blue line shows the estimated wage profile of a hypothetical worker who starts his career in one of the largest cities and works there for ten years. The red line shows the same for a worker in a smaller region. The profiles abstract from the wage effects of job mobility or any career changes. For all groups, the wage profiles in

major cities are considerably steeper.

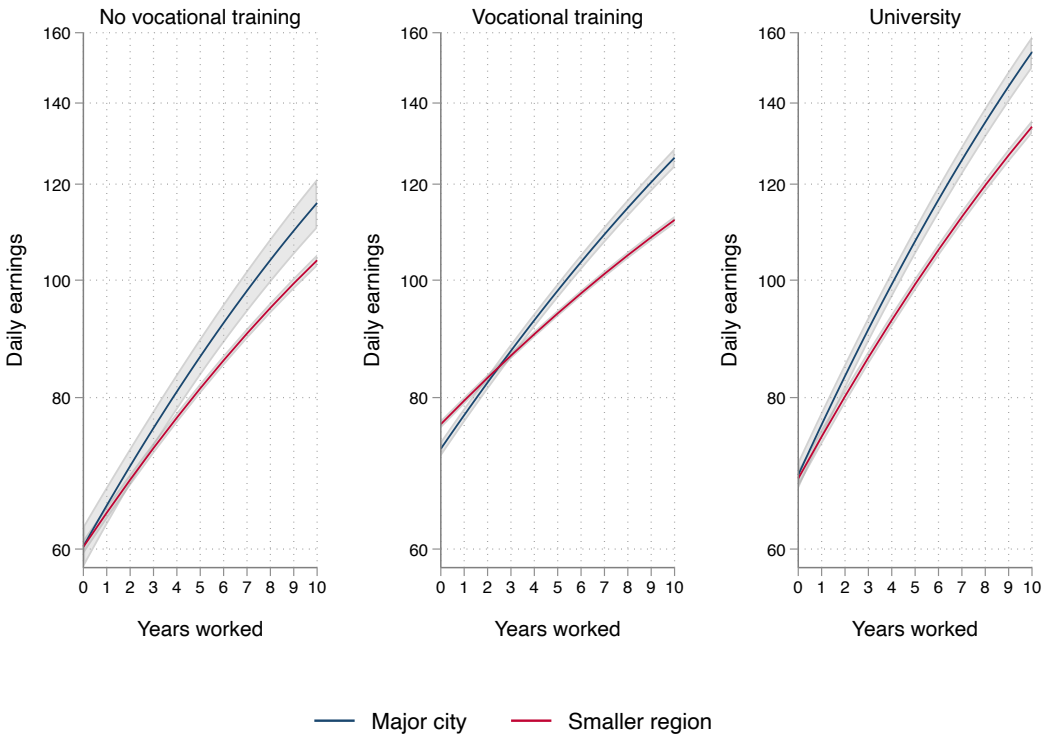

Figure 7. Earnings profiles by education group: major city and smaller region
Notes: The figure shows wage profiles for each education group of two hypothetical workers, based on the results from a version of regression 8 where the experience terms are interacted with an education dummy. The red line represents a worker who starts his career in a smaller region and stays there for ten years. The blue solid line shows the profile of a worker who starts his career in Munich and works there for ten years. The vertical axis shows daily earnings on a log scale.

The heterogeneity by occupation group are shown in Table 6. Unlike between education groups, there is no meaningful difference between the urban wage premium for the different occupation groups in either the immediate premium or the medium-term premium. The immediate urban wage premium is the same for workers of all education groups - between 0.022 and 0.024.

4.5 Drivers of the urban wage premium

The previous sections showed urban wage premia estimates based on different specifications. We can interpret the first specification (regression 1), which ignored dynamic forces and sorting, as yielding the total wage premium. The resulting estimate is biased by sorting and by dynamic forces. For the sample of all workers, this was estimated to be .0505 log points (Table 2).

To receive an unbiased estimate of the static urban wage premium, we con-

Table 5. Dynamic region size wage premium, by education groups

	Initial premium			Medium-term premium		
	(1) Low Education	(2) Mid Education	(3) High Education	(4) Low Education	(5) Mid Education	(6) High Education
Log region size	0.011**	0.020***	0.027***	0.021***	0.036***	0.038***
	(0.004)	(0.002)	(0.005)	(0.006)	(0.005)	(0.006)
R^2	0.019	0.223	0.123	0.062	0.333	0.188
Observations	267	267	267	267	267	267

Notes: The table shows short-term and medium-term urban wage premia for each education group. This is based on regression 2 using the region-education fixed effect coefficients from a variant of regression 6 where the region FE and local experience is interacted with an education dummy. Column 1 shows results for the immediate urban wage premium for the low education group, column 2 for the mid education group, and column 3 for the high education group. Columns 4 to 6 show the corresponding medium term premia for low, mid, and high education, respectively. As in Table 4, medium-term is defined as the average experience in a location in our sample (about 16 years). Coefficients are reported with robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels.

Table 6. Dynamic region size wage premium, by occupation groups

	Initial premium			Medium-term premium		
	(1) Manual services	(2) Production	(3) Cognitive/ interactive	(4) Manual services	(5) Production	(6) Cognitive/ interactive
Log region size	0.023***	0.024***	0.022***	0.035***	0.034***	0.034***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)
R^2	0.240	0.317	0.171	0.319	0.357	0.273
N	267	267	267	267	267	267

Notes: The table shows short-term and medium-term urban wage premia for each occupation group. This is based on regression 2 using the region-occupation fixed effect coefficients from a variant of regression 6 where the region FE and local experience is interacted with an occupation dummy. Column 1 shows results for the immediate urban wage premium for workers in manual services, column 2 for the workers in production occupations, and column 3 for workers in cognitive/interactive occupations. Columns 4 to 6 show the corresponding medium term premia for low, mid, and high education, respectively. As in Table 4, medium-term is defined as the average experience in a location in our sample (about 16 years). Coefficients are reported with robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels.

trolled for worker sorting by including individual fixed effects and controlled for dynamic forces by letting the value of local experience vary over time (equation 6). For the full sample, this gives us an estimate of .0191 log points (column 2 of Table 4). This is the immediate wage gain a worker is estimated to receive when moving to a region twice as large.

Since urban experience is valued more than non-urban experience, the urban

wage premium grows over time. This leads to the urban wage premium growing to an estimated .0387 log points after the mean local sample experience of 16.2 years.

The difference between the immediately realized urban wage premium of .0191 and the dynamically realized urban wage premium of 0.0387 can be interpreted as the dynamic portion of the total urban wage premium. The remainder to the total urban wage premium of .0505 can then be attributed to worker sorting on unobservables. Figure 8 visualizes this interpretation: the first panel shows each component of the urban wage premium using the full sample estimates: of the total urban wage premium, 23 per cent can be attributed to sorting, 38 per cent is static, and 39 per cent is dynamic.

The other three panels in Figure 8 show the component drivers of the urban wage premium for each education group. This shows that the static component is increasing in education, but the dynamic component is largest in the mid-education group. However, the biggest difference is in the role of sorting: the sorting of more productive workers (within education group) into urban regions is much stronger among university graduates. In fact, sorting explains the entirety of the larger total urban wage premium for this group in comparison to the mid-education group. For workers without vocational training, we see a different picture: the static and dynamic urban wage premia account for all of the total urban wage premium. In fact, the dynamic estimation yields a slightly higher urban wage premium than the static estimation, suggesting slightly negative sorting if any.

Figure 9 shows the breakdown of the drivers of the urban wage premium for each occupation group. Unlike between education groups, the static and dynamic components are almost constant across occupation groups. However, there is again a big difference in the role of sorting. Within cognitive/interactive occupations, larger regions attract those workers with higher inherent productivity, so that sorting accounts for more than half of the overall urban wage premium for this group. For production workers, on the other hand, there is little sorting on unobservables. For workers in manual services, however, the estimates suggest negative sorting: the workers in larger cities tend to be those with lower inherent productivity. Note that unlike education level, which rarely changes over the course of the career, workers may change between occupation groups more. However, the occupation groups are defined such that they differ maximally in task content, so moves between occupation groups are rare. Appendix Table A2.1 confirms that moves across occupations are rare. It shows a transition matrix of moves between occupation clusters, conditional on workers moving between jobs.

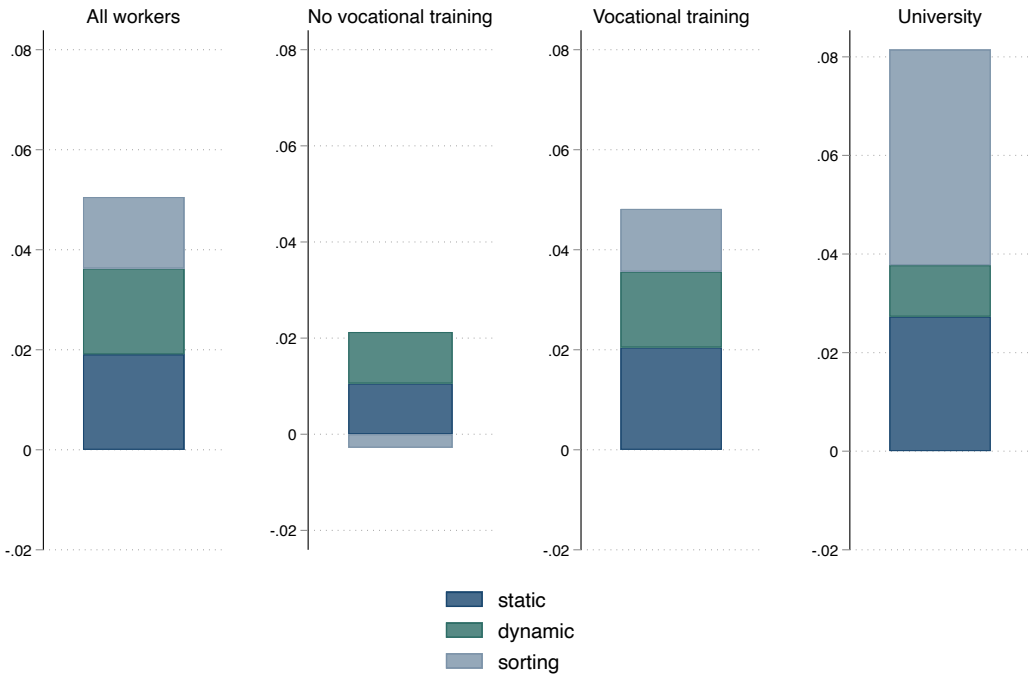

Figure 8. Drivers of the urban wage premium: by education group

Notes: The urban wage premium is decomposed using coefficients from different regression specifications. The static urban wage premium is estimated controlling for potential bias from sorting and dynamic effects. Its estimates are reported in Table 4 for all workers and in columns 1 to 3 of Table 5 by education group. The dynamic urban wage premium is estimated taking the dynamic effects into account. Its estimates are reported in Table 4 for all workers and in columns 4 to 6 of Table 5 by education group. The overall urban wage premium is estimated without controlling for sorting and dynamic effects and is reported in Table 2 and in Figure 5 by education group.

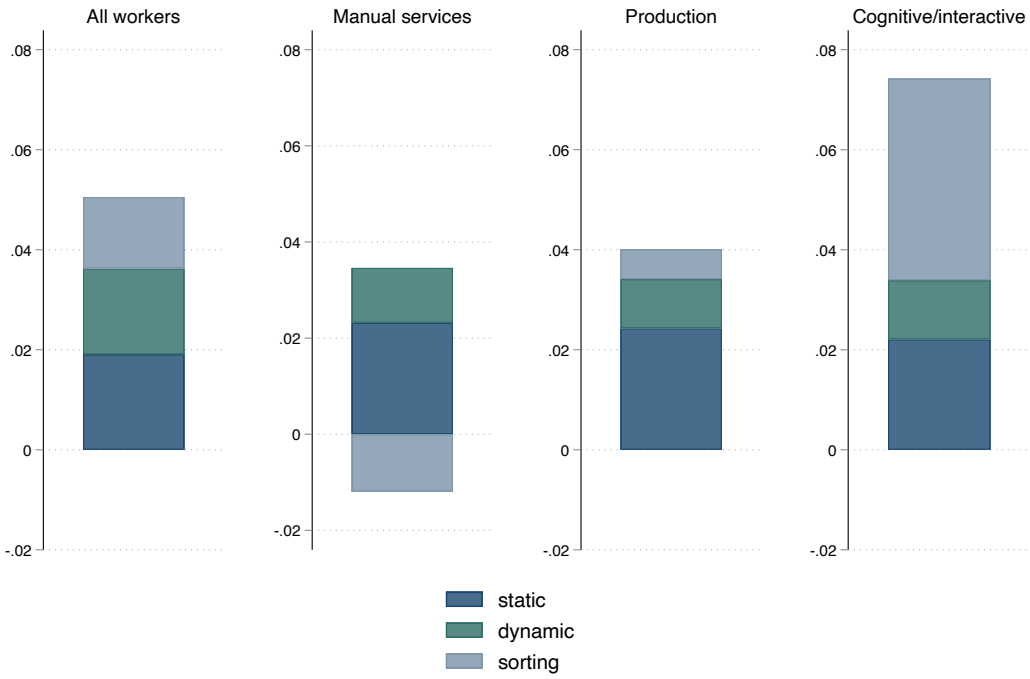

Figure 9. Drivers of the urban wage premium: by occupation group

Notes: The urban wage premium is decomposed using coefficients from different regression specifications. The static urban wage premium is estimated controlling for potential bias from sorting and dynamic effects. Its estimates are reported in Table 4 for all workers and in columns 1 to 3 of Table 6 by occupation group. The dynamic urban wage premium is estimated taking the dynamic effects into account. Its estimates are reported in Table 4 for all workers and in columns 4 to 6 of Table 6 by occupation group. The overall urban wage premium is estimated without controlling for sorting and dynamic effects and is reported in Table 2 and in Figure 5 by occupation group.

5 Conclusion

This paper estimates urban wage premia for Germany, and separates the forces that drive it: sorting, static premia, and dynamic premia. Replicating the results presented by De la Roca and Puga (2017) for Spain, this paper shows that both static and dynamic forces significantly contribute to the observed total urban wage premium. Unlike in the Spanish case, however, sorting also plays a major role. In the German case, a doubling in region size is associated with a 5 per cent increase in wages and each of the components contributes to about a third to the total. This exercise highlights the importance of sorting on unobservables and considering both wage level and wage growth effects when assessing the benefits of urban labour markets.

The study proceeds to investigate heterogeneity in urban wage premia by education and occupation groups. In the static estimation, it confirms the findings by Grujovic (2020): the total urban wage premium is increasing in educational attainment, and also differs by occupational task content, being highest in cognitive/interactive occupations. However, examining the forces behind this heterogeneity reveals that a lot of this difference between groups is driven by sorting: over 50 per cent of the total observed wage difference for university graduates can be attributed to sorting of inherently more productive workers to larger regions. Sorting for workers with vocational training is lower, and virtually non-existent for those with no vocational training. Even when controlling for the outsize role of sorting, the wage premium is higher for workers with more education, and this is driven both by a static and by a dynamic effect.

Sorting also plays a large role for the heterogeneity in the total urban wage premium between occupation groups. Here, however, it explains the entirety of the variation. After controlling for sorting on unobservables, the size of the wage premium between occupation groups is virtually indistinguishable.

Bibliography

Andersson, Martin and Per Thulin, "Does Spatial Employment Density Spur Inter-Firm Job Switching?," *The Annals of Regional Science*, August 2013, *51* (1), 245–272.

Antoni, Manfred, Philipp vom Berge, and Andreas Ganzer, "Factually Anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB-Regionalfile) – Version 7517 V1," Technical Report, Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) 2019.

Baum-Snow, Nathaniel and Ronni Pavan, "Understanding the City Size Wage Gap," *The Review of Economic Studies*, January 2012, *79* (1), 88–127.

Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 2013, *128* (3), 240–266.

Carlsen, Fredrik, Jørn Rattsø, and Hildegunn E. Stokke, "Education, Experience, and Urban Wage Premium," *Regional Science and Urban Economics*, September 2016, *60*, 39–49.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, March 2008, *63* (2), 723–742.

_ , _ , _ , and Sébastien Roux, "Estimating Agglomeration Economies with History, Geology, and Worker Effects," in "Agglomeration Economics," University of Chicago Press, 2010.

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

De la Roca, Jorge and Diego Puga, "Learning by Working in Big Cities," *The Review of Economic Studies*, January 2017, *84* (1), 106–142.

Duranton, Giles and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies," in J. Vernon Henderson and Jacques-Francois Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Elsevier, 2004, pp. 2063–2117.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg, "Revisiting the German Wage Structure," *The Quarterly Journal of Economics*, 2009, *124* (2), 843–881.

GEOSTAT, Eurostat, "GEOSTAT 2006 Grid Dataset," https://ec.europa.eu/eurostat/web/gisco/ geodata/reference-data/population-distribution-demography/geostat 2006.

Glaeser, Edward L., "Learning in Cities," *Journal of Urban Economics*, 1999, (46), 254–277.

_ , *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, New York: Penguin Books, 2012.

_ **and David C. Maré**, "Cities and Skills," *Journal of Labor Economics*, April 2001, *19* (2), 316–342.

_ **and Matthew G. Resseger**, "The Complementarity between Cities and Skills," *Journal of Regional Science*, February 2010, *50* (1), 221–244.

Grujovic, Anja, "Tasks, Cities and Urban Wage Premia," 2020.

Hall, Anja and Beate Beermann, "BIBB/BAuA-Employment Survey 2006," Technical Report 2006.

Korpi, Martin and William A.V. Clark, "Migration and Occupational Careers: The Static and Dynamic Urban Wage Premium by Education and City Size: Migration and Occupational Careers," *Papers in Regional Science*, February 2019, *98* (1), 555–574.

Lazear, Edward, "Age, Experience and Wage Growth," *The American Economic Review*, 1976, *66* (4), 548–558.

Lüttge, Julius, "Occupational Change and Wage Inequality in Germany," 2022.

Matano, Alessia and Paolo Naticchioni, "What Drives the Urban Wage Premium? Evidence along the Wage Distribution.," *Journal of Regional Science*, March 2016, *56* (2), 191–209.

Mion, Giordano and Paolo Naticchioni, "The Spatial Sorting and Matching of Skills and Firms," *Canadian Journal of Economics/Revue canadienne d'économique*, February 2009, *42* (1), 28–55.

Yankow, Jeffrey J., "Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium," *Journal of Urban Economics*, September 2006, *60* (2), 139–161.

A Appendix

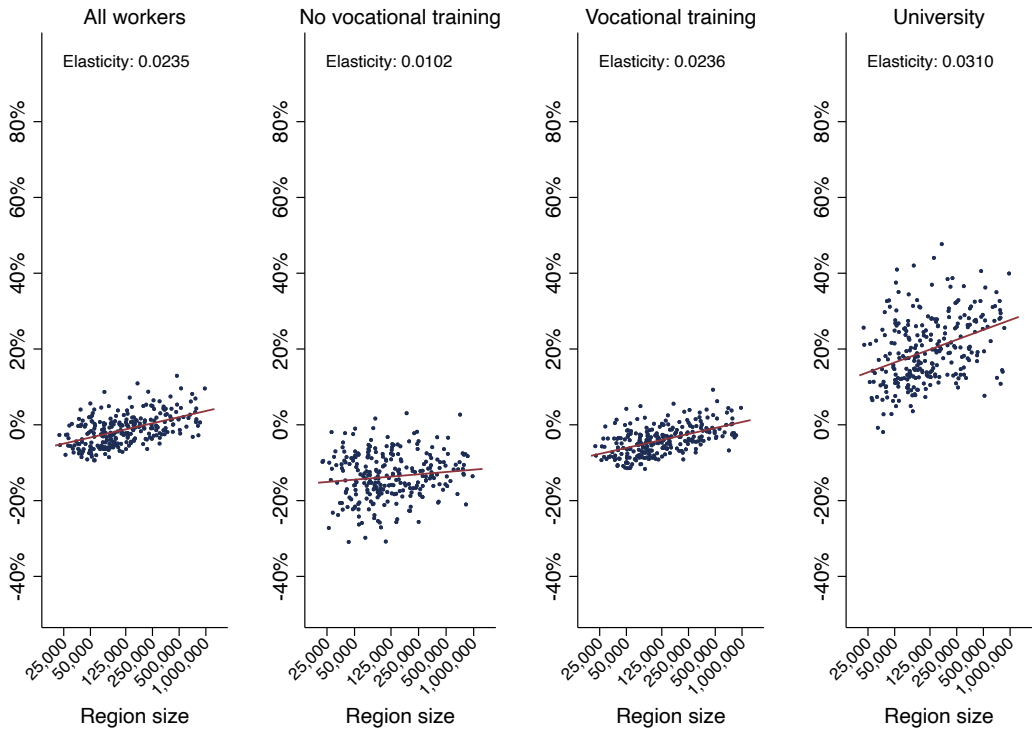

Figure A2.1. Static urban wage premium estimation, with worker FEs: by education group

Notes: This figure plots the region fixed effects against region size, where the first stage regression includes worker fixed effects. The left panel is based on the fixed effects from regression 3. The other panels show the region fixed effects for each education group. These result from regression 4, whose other coefficients are shown in column 3 of Table 3. The region fixed effects corresponding to column 1 of Table 3 are shown in Figure 3. The trend lines are the regional wage premium. Region size is defined as the number of residents within 10 km of the average resident and is shown on a log scale.

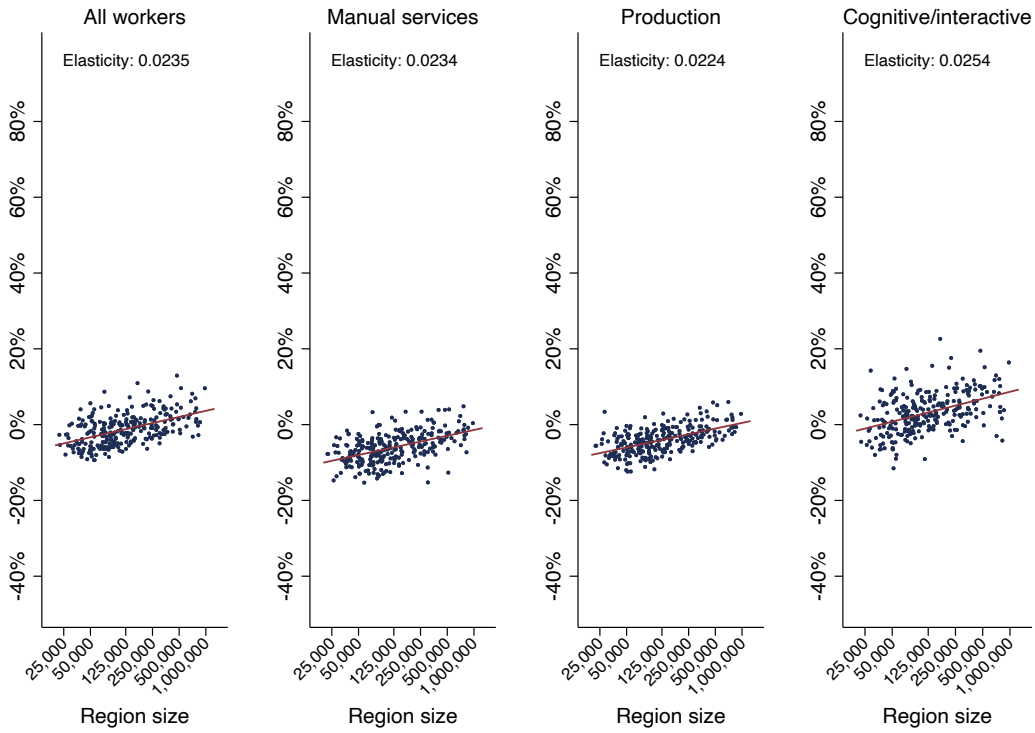

Figure A2.2. Static urban wage premium estimation, with worker FEs: by occupation group

Notes: This figure plots the region fixed effects against region size, where the first stage regression includes worker fixed effects. The left panel is based on the fixed effects from regression 3. The other panels show the region fixed effects for each occupation group. These result from regression 4, whose other coefficients are shown in column 3 of Table 3. The region fixed effects corresponding to column 1 of Table 3 are shown in Figure 4. The trend lines are the regional wage premium. Region size is defined as the number of residents within 10 km of the average resident and is shown on a log scale.

Table A2.1. Occupational transitions

Origin occ cluster	Manual services	Production	Cognitive/ interactive	Total
Manual services	77.4	17.7	12.2	33.8
Production	9.8	74.8	3.9	18.3
Cognitive/interactive	12.8	7.5	83.9	47.8
Total	100.0	100.0	100.0	100.0

Notes: This table shows a transition matrix between occupation clusters. For all workers who move between two jobs, it shows the percentage of each occupation cluster conditional of the occupation cluster in the old job. It is computed using data of all workers in employment on 31 December of each year. Transition to the new job is computed to 31 December of the following year.