Does the urban wage premium differ by pre-employment status?

Silke Hamann, Annekatrin Niebuhr and Jan Cornelius Peters

Appendix

Additional information on data: new employment relationships, censored wages, and instrumental variables

The units of observation in our analysis are new employment relationships. We focus on new employment spells with a length of at least seven days that refer to full-time employment subject to social security contributions outside the public sector and the temporary work sector. Apprenticeships are not considered, nor are new employment relationships that start simultaneously with another employment relationship or with an active labour market programme, as we cannot ensure that this employment is not publicly subsidised. Moreover, we exclude new employment relationships with wages below two times the limit for marginal employment as well as recalls, i.e., cases in which a worker starts to work in an establishment in which she worked at least once during the previous 28 days. If a worker is already employed at the starting date of the new employment relationship in another establishment, we consider the new employment relationship only if the previous employment spell ends within 7 days.

We use the wages of new employment relationships as the dependent variable in the first-stage regression. The first employment spell in the IEB of a new employment relationship ends, at the latest, by December 31st of the year in which the new employment relationship starts. Daily wages are calculated by dividing the reported total earning from this spell by the length of the spell. Information on actual working days or contract hours is not available. Firms report earnings only up to the upper limit for social security contributions such that the wage information in the IEB is right censored. Therefore, we partly impute the wages. We follow Reichelt (2015) and estimate an interval regression, a generalisation of Tobit regression, to predict wages above the threshold (approximately 6% of the observations). See Reichelt (2015) for a detailed description of how interval regression is

applied to impute right-censored wages. The results of our regression analysis do not change when we use the reported wages as dependent variable instead of the imputed wages in the first-stage regression.

We use historical population density and soil characteristics as instrumental variables for current employment density. Historical regional population density is measured in 1871, 1880, 1890, 1900, 1910, 1925, and 1933 and is provided by Rothenbacher (2002). The soil data come from the European Soil Database. The available raster data has been aggregated at the regional-level using the same characteristics as Combes et al. (2010).

Standard errors and two-stage regression model

Combes and Gobillon (2015) and Combes et al. (2008) note that the computation of standard errors poses a problem in the common one-stage estimation approach because the corresponding covariance matrix has a complex structure. This is due to the unobserved regional effects and the mobility of workers across labour markets. For a migrant the one-stage wage equation in first differences includes two different unobserved local shocks which refer to the region of origin and the region of destination. Furthermore, the locations of these shocks vary across migrants because their regions of origin differ and they do not move to the same region of destination. Therefore, it is not possible to sort workers in such a way that results in a simple covariance matrix structure and allows to cluster standard errors at each date by region. The authors propose the two-stage approach to solve this problem. Combes et al. (2008) provide a detailed discussion of this issue.

Additional information on instruments and their validity

The OLS estimate of the elasticity of wages with respect to employment density might be biased because of missing local characteristics or local shocks that influence both the population location and productivity/wages. To tackle this problem via 2SLS, we need instruments that affect local productivity/wages only via their impact on the spatial distribution of population. The instruments should influence the regional labour supply, but not directly local productivity or wages.

More precisely, the instrument variables M_{rt} are valid if they are relevant $[COV(M_{rt}, D_{rt})] \neq 0$ and exogenous, i.e., uncorrelated with the error term $[COV(M_{rt}, e_{rt})] = 0$. In the following we provide a more detailed discussion of these issues with respect to the instruments used in our analysis, i.e., historic pooulation density and soil characteristics. The IV strategy that we apply is well established in the literature on agglomeration effects. Several influential papers use such instruments to estimate unbiased effects of density on different outcome variables (e.g. Ciccone & Hall, 1996; Combes et al., 2008; Guevara-Rosero et al., 2018).

The historic population density is taken from Rothenbacher (2002). We mainly use population figures for the year 1900 (see Table A1). The data is available for administrative units of that time, inter alia, kingdoms and their provinces. We gauge the historic population density of the 141 labour market regions as defined by Kosfeld and Werner (2012) by assigning each labour market region to one of the administrative units for which historic population density is available using historic maps. If a labour market region comprises parts of more than one historic administrative unit, we compute an average of the respective historic polulation densities. Since most historic administrative units were larger than the present labour market regions, some spatial variation in present labour market density is lost in the IV estimation. However, there is still a significant positive correlation between present employment density and the generated historic population density variable (Table A11).

Following Combes et al. (2010), we use the following soil characterisitcs from the European Soil Database as further instruments: mineralogy of subsoil (3 categories), dominant parental material (1st and 2nd aggregate, 7 and 10 categories, respectively), water capacity of sub- and topsoil (4 and 5 categories, respectively), depth of rock (4 categories), erodibility (4 categories), carbon content (3 categories), hydrological class (6 categories), and local terrain ruggedness defined according to Combes et al. (2010) as the difference between the mean of maximum altitudes across all pixels in a region and the mean of minimum altitudes. The soil data is available as raster data with cells of 1 km per 1 km. The information has been aggregated first at NUTS 3 level and in a second step at the level

of the labour market regions. Despite 'ruggedness' the soil characteristics are discrete. Therefore, we used the value that appears most often in each area (ibd.).[1]

Relevance

Using historical population figures to instrument contemporaneous population density has become a standard approach since the seminal contribution by Ciccone and Hall (1996). In the literature on agglomeration effects, historical population data is usually considered to provide highly relevant instruments because the spatial distribution of population and economic activity is highly persistent due to the locations´ housing stock and production sites (Combes et al., 2010). The stability of spatial population patterns in Germany is confirmed by the significant correlation between historical population density and current employment density in Table A11. The partial R^2 of the excluded instruments in the first stage regression for labour market density with historic population density and its spatial lag as the only excluded instruments is 0.22 (Table A6, column (1)). In the corresponding first stage regression for the spatial lag of employment density the partial R^2 is 0.58.

Soil characteristics are supposed to be significant because they determine the fertility of soils and are, therefore, important determinants of early agricultural production. This in turn explains the role of soil chraracteristics as fundamental drivers of population settlements (Combes and Gobillon, 2015). We do not consider pairwaise correlations with current employment density because each soil characteristic is described by several discrete variables.

More importantly, relevance requires a partial correlation of the instrument with the endogenous regressor, namely, the coefficient of the instrument variable should be significant in the first stage regression. Table A12 summarizes the first stage results. The first stage results indicate that the instruments are relevant. The historical population density and the corresponding spatial lag are significant in all first-stage regressions. The soil characteristics are categorial variables and are

[1] We gratefully thank Malte Reichelt for providing us with the information from the European Regional Soil Database at the level of the labour market regions.

included via dummy variables. Although we do not detect an important effect for every single feature characteristic, i.e., every dummy variable, the soil characteristics turn out to be valuable predictors of present employment density. If only soil characteristics are used as instrument variables, the partial R^2 of these instruments in the first stage regression is 0.39 and 0.24, respectively (Table A6, column (10)). However, the soil instruments are weaker when compared with the historical population density, in line with evidence in Combes et al. (2010) who note that most soil characteristics vary rather smoothly across space. But altogether the first stage results indicate that the instrument variables are relevant.

Combes and Gobillon (2015) note that in practice, historical population figures turned out to be extremely relevant instruments. Soil characteristics are also found to be relevant but they have less power to explain current employment density.

Exogeneity

Exogeneity means that the instruments are orthogonal to the error term. This requires that our instruments are not correlated with missing local variables and not determined by productivity or wages. Combes et al. (2010) argues that simultaneity is unlikely if long lags of population density are used as instruments. A simultaneity problem caused by local shocks that influence both the population location and productivity/wages will only persist if these shocks are expected more than 100 years before their appearance and they have determined population location more than 100 years before the incidence. The authors argue that this is extremely unlikely.

Endogeneity of the instruments might, however, also arise due to some missing regional characteristics that are determinants of the past population location *and* contemporaneous productivity/wages. In our second stage regressions, we therefore control for a number of local characteristics such as a coastal loaction, climate features and amenities. Following Combes et al. (2010), we assume that contemporaneous determinants of local wages are not associated with the factors behind historical agglomeration patterns given that we control for these factors in the second

stage. There are good reasons to believe that this assumption is fulfilled. The German economy has changed a lot between 1900 and 2005 as indicated by significant changes of the sectoral structure. The share of the agricultural sector in production and employment has declined, the weight of manufacturing increased initially in the period under consideration and declined in recent decades while the service sector gains in importance (see Braun, 1990). Moreover, technological progress radically changed production techniques applied in the German economy (see, e.g., Spoerer & Streb, 2013) and, as argued by Combes et al. (2010), this also changes location requirements of production sites. Additionally, transport costs declined noticeably since 1900 in Germany and Europe and there has been a significant variation in trade barriers in the period under consideration. Furthermore, non-economic factors that influence the choice of residence have likely changed as well since the standard of living increased (Rappaport, 2007). And last but not least, the country experienced two world wars during the twentieth century; see Combes et al. (2010) for a more detailed discussion of these arguments for the case of France.

As regards the exogeneity of soil characteristics, Combes et al. (2010) argue that they have been mainly determined natural forces and not primarily been influenced by human activity. The authors discuss the pros and cons of this argument. Soil characteristics are supposed to be a main determinant of population patterns and economic activity in the past. But soil quality is no longer expected to be important in a country in which agriculture and extractive industries represent only a small share of the economy. Eventually, Combes et al. (2010) note that due to the number of available soil characteristics different sets of instruments can be used and 2SLS results compared to check the robustness (see Table A6 for corresponding results). Moreover, we combine soil characteristics with historic population density in our identification strategy. The availability of several instruments enables us to apply overidentification tests to assess the issue more formally.

We estimate the second stage based on region fixed effects and region-time fixed effects. With respect to 2SLS this implies that we exploit both cross sectional variation and longitudinal variation of the instruments. Making use of the longitudinal variation rests on the assumption that the

historic variation of population in the regions correlates with the contemporaneous change of population. We admit that this assumption is strong and not very likely to be fulfilled. However, we use the approach with region-time effects only as a robustness check and primarily rely on the model based on region fixed effects that does not require this assumption.

Qualification and pre-employment status

The qualification level of the workers and the pre-employment status are correlated.[2] For instance, low-skilled workers are overrepresented in the transitions from long-term unemployment and underrepresented among the job-to-job transition (see Table A8). This correlation might of course affect the results on differences across pre-employment status. If the pre-employment status is just a proxy for ability of the workers, the detected effect heterogeneity might be due to ability differences rather than disparities in pre-employment status. However, the correlation between the two variables turns out to be fairly moderate. The number of job-to-job transitions exceeds the number of new employment relationships of long-term unemployed by more than factor two for the low-skilled workers. Moreover, high-skilled workers are also overrepresented among transitions from long-term unemployment. It is therefore unlikely that the differences in effects across pre-employment status are driven exclusively by the qualification level of the workers. This is confirmed by a two-stage regression where we differentiate by formal qualification (low-, medium- and high-skilled) instead of pre-employment status. We do not detect significant differences in the effect of density across qualification levels in a two-stage model. All skill groups seem to benefit from dense labour markets (see Table A9).

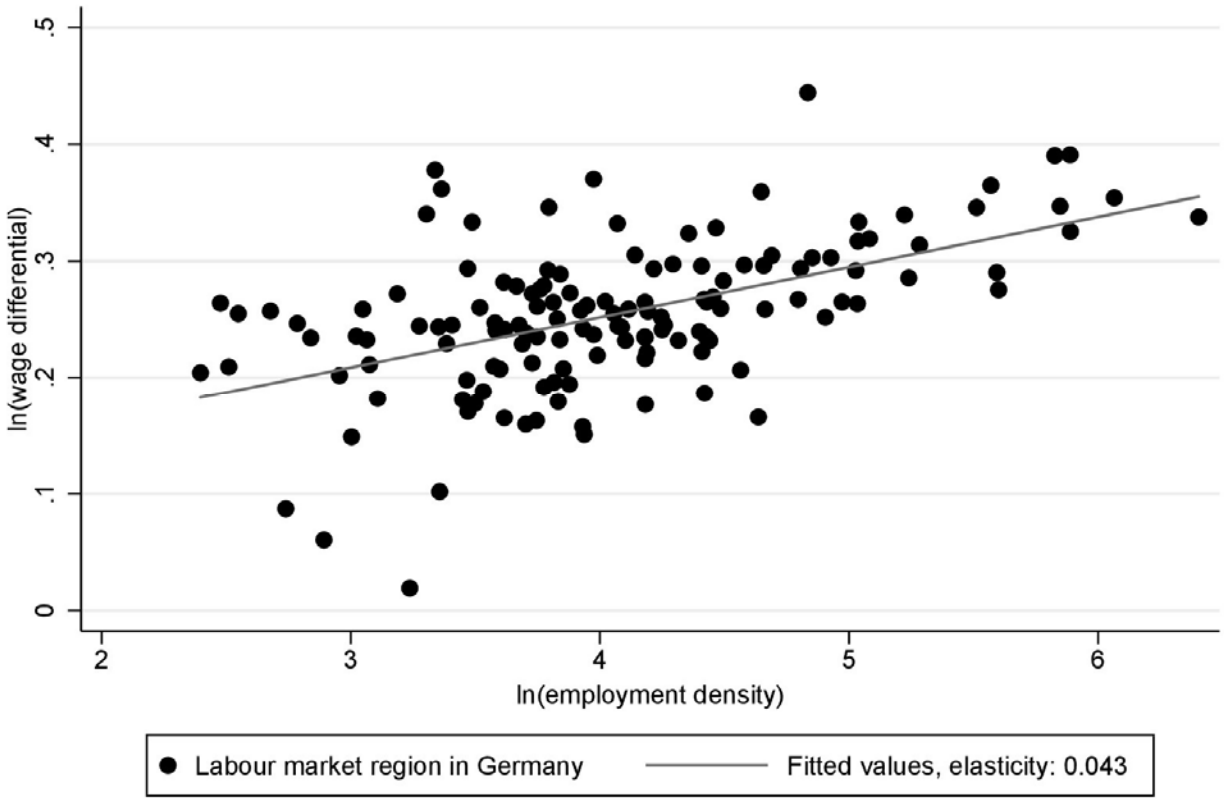
This is also in line with the results of complementary regression analyses. In a one-stage regression model we interact the qualification level of the workers, their pre-employment status and employment density. Thus, we allow for heterogeneous effects with respect to the pre-employment

[2] Formal qualification turns out to be a fairly good proxy for ability as indicated by a comparison of individual fixed effects and formal education (see Figure A4).

status within specific skill groups. The estimates indicate that even for given skill level the effect of density on entry wages significantly differs across pre-employment status (see Table A10). F-tests on equality of the effect of employment density show that there are important differences between job-to-job transitions and the other types of transitions in all skill groups. The differences are also economically significant: The elasticity for the long-term unemployed is only half the size of the effect that we estimate for job-to-job transitions. In contrast, there is no robust evidence on differences between transitions from short- and long-term unemployment.

Figures and Tables

Figure A1: Correlation between employment density and the wage gap between the former long-term unemployed and workers with job-to-job transition



Notes: Gap in average wages after transitions to employment during the period from 2005 to 2011. Workers which were non-employed for at least 365 days are defined as long-term unemployed. Workers with job-to-job transitions were out of job for at most 28 days.

Table A1: Variables – definitions and sources

Variable	Definition	Source
Gross daily wage	Daily wages are calculated by dividing the reported total earning from employment spell by the length of the spell.	Integrated Employment Biographies (IEB)
Educational level of worker	A categorial variable that combines information on highest school leaving certificate, completed vocational training, and university degree. For some employment spells, this information is missing. If so, we use the information from previous employment spells following Fitzenberger, Osikominu, & Völter (2005).	IEB
Gender		IEB
Nationality		IEB
Experience	The difference between the considered date of transition to employment and the date of the first employment spell in the IEB. This variable is left censored because the IEB data do not capture employment spells before January 1, 1975.	IEB
Recent work experience	Years of employment measured on a daily basis for the five years before the considered transition to employment. Marginal employment is not included, nor are employment spells that are combined with active labour market policies. We distinguish total, occupation-specific, and region-specific work experience as well as work experience acquired in agglomerations. Occupation-specific experience is defined with respect to 21 occupational segments (see Matthes, Burkert, & Biersack 2008). Region-specific experience refers to previous employment in the regional labour market in which the new employer is located, and experience acquired in agglomerations is classified based on a typology of the Federal Institute for Research on Building, Urban Affairs and Spatial Development, which distinguished agglomerations, less dense urbanised regions, and rural regions. The classification is based on the population share living in cities, the existence of large cities within the region, and the population density.	IEB
Number of employers	Logarithm of the number of unique establishment identifiers. If there was no previous employer, this variable is set to zero and the dummy variable "First employer" to one which is zero otherwise.	IEB
Pre-employment status	Dummy variables referring to the 28 days before the considered transition to employment • unemployment benefits (Arbeitslosengeld I) • unemployment assistance (Arbeitslosengeld II/Arbeitslosenhilfe) • unemployed and registered as a job seeker • not unemployed but registered as a job seeker • participating in active labour market policy programmes.	IEB
Occupational status	Categorical variable that distinguishes white-collar and blue-collar workers based on the type of pension insurance institution (vom Berge, Burghardt, & Trenkle, 2013). Blue-collar workers are also classified by activity: unskilled workers, skilled workers, and master craftsman/foreman. In December 2011, a new occupational classification was introduced. Therefore, for some observations, the occupational status is unknown.	IEB
Firm characteristics	Logarithmic number of employees, employment growth (dummy variable), share of workers with a university degree, share of workers with no completed vocational training/no university degree. The information refers to the last reference date (June 30) before the considered transition.	Establishment History Panel (BHP)
Industry share	Logarithm of the employment share of the industry (2-digit level: 88 industries) of total regional employment.*	Employment statistics of the Federal Employment Agency (FEA)

Table A1. Continued.

Variable	Definition	Source
Industrial diversity	Logarithm of the inverse Herfindahl index based on the employment shares of industries of total regional employment. The own industry is excluded when the inverse Herfindahl index is calculated.*	FEA
Number of establishments of the local industry	Logarithmic number of establishments with at least one employee subject to social security on June 30 at t-1. Only firms in the same industry and same regional labour market are considered.*	FEA
Human capital of the local industry	Share of workers with a university degree of total employment and share of workers without completed vocational training/university degree in the same industry and regional labour market.*	FEA
Skill-specific unemployment rate of the regional labour market	Logarithmic share of persons registered as unemployed of the number of persons who are registered as unemployed or employed in the region. We distinguish three groups: persons with a university degree, persons with completed vocational training, and persons without completed vocational training/university degree. Information refers to June 30 at t-1	(Un-) Employment statistics of the FEA
Industry fixed effects	Fixed effects for 88 distinct industries (2-digit level according to the industry classification from 2008). In 2008, there was a change in the industry classification. If an establishment is observed before and after 2008, we assign the employment spells from 2005-2007 to the industry that the firm reports in 2008 (or later). If an establishment identifier shows up only for 2005-2007, we use a correlation matrix between the old and new industry classification as described by Eberle, Jacobebbinghaus, Ludsteck and Witter (2011).	IEB
Occupation fixed effects	Fixed effects for 21 distinct occupational segments.	IEB
Employment density	Working population in 1,000 per square kilometre.	Regional Database Germany (RDG) of the Federal Statistical Office
Weather indicators	Information covering the period 1999-2009 collected at 71 weather stations. For each regional labour market we use data from the weather station which is nearest to the geographical centre of the region. We use the average temperature, average number of hours of sunshine, and average precipitation.	Deutscher Wetterdienst
Restaurant workers	Share of restaurant workers defined according to the 1988 classification of occupations (codes 912 - waiters, 411 - cooks) of the total regional population.	FEA and RDG
Share of recreation area	The share of urban green space, parks, allotment gardens, sport fields, and campsites of the total area.	TRDG
Coast	A dummy variable that indicates whether the region is located on the coast.	
Historical population density	Historical population density is available for 111 historic regions. We use this information to approximate the historic population density in 1900 for our 141 regional labour market regions. If one labour market region includes (parts of) several historic regions, we calculate the weighted average of the density of the different historic regions. In column (8) of Table 1 we use data for 1871, 1880, 1890, 1900, 1910, 1925, and 1933 and generate a panel data set with seven waves that is used to instrument for the employment density over 2005-2011.	Rothenbacher (2002)
Soil data	We use the following indicators: topsoil and subsoil mineralogy, dominant parent material (high and low aggregate), topsoil and subsoil water capacity, depth to rock, soil differentiation, erodibility, carbon content, hydrogeological class, and ruggedness. The European Soil Database provides raster data. All indicators (except ruggedness) are categorical variables. Based on the raster data, we choose the modal value to aggregate the information at the regional labour market level.	European Soil Database

* The information refers to June 30th in t-1.

Table A2. Summary statistics, first-stage variables

	All transitions				Only transitions that are considered on the first stage with individual FE			
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
Individual characteristics								
ln(gross daily wage)	4.135	0.499	3.267	7.573	4.124	0.487	3.267	7.573
Education								
Secondary/intermediate school leaving certificate								
without completed vocational training	0.095	0.293	0.000	1.000	0.086	0.281	0.000	1.000
with completed vocational training	0.660	0.474	0.000	1.000	0.682	0.466	0.000	1.000
Upper secondary school leaving certificate								
without completed vocational training	0.016	0.125	0.000	1.000	0.013	0.112	0.000	1.000
with completed vocational training	0.084	0.278	0.000	1.000	0.082	0.275	0.000	1.000
Degree of university of applied sciences	0.045	0.207	0.000	1.000	0.044	0.205	0.000	1.000
College/university degree	0.100	0.312	0.000	1.000	0.093	0.302	0.000	1.000
Female worker	0.331	0.471	0.000	1.000	0.295	0.456	0.000	1.000
Foreign worker	0.075	0.264	0.000	1.000	0.074	0.262	0.000	1.000
Lifetime work experience (in years)	15.009	9.270	0.003	36.975	14.947	8.933	0.000	36.969
Work experience previous 5 years (in years)	3.399	1.541	0.000	4.999	3.447	1.429	0.000	4.999
Length of employment spell in the year of transition (in months)	6.107	3.634	0.033	12.000	5.849	3.561	0.033	12.000
Occupation specific work experience (in years)	2.348	1.962	0.000	4.999	2.353	1.884	0.000	4.999
Work experience in region (in years)	2.218	1.924	0.000	4.999	2.166	1.847	0.000	4.999
Work experience in agglomerations (in years)	1.819	1.984	0.000	4.999	1.807	1.939	0.000	4.999
ln(Number of previous employers)*	1.279	0.813	0.000	5.100	1.405	0.803	0.000	5.100
First employer	0.031	0.173	0.000	1.000	0.028	0.165	0.000	1.000
Unemployment benefit (ALG I)	0.249	0.432	0.000	1.000	0.293	0.455	0.000	1.000
Unemployment assistance (ALG II, ALHI)	0.069	0.275	0.000	1.000	0.076	0.268	0.000	1.000
No unemployment benefit/assistance	0.672	0.470	0.000	1.000	0.631	0.483	0.000	1.000
Unemployed and registered as a job seeker	0.322	0.467	0.000	1.000	0.364	0.481	0.000	1.000
Not unemployed but registered as a job seeker	0.099	0.297	0.000	1.000	0.102	0.302	0.000	1.000
Not registered as a job seeker	0.579	0.494	0.000	1.000	0.534	0.499	0.000	1.000
Participation in measures of active labour market policy	0.055	0.229	0.000	1.000	0.057	0.231	0.000	1.000
Occupational status								
Unskilled worker	0.240	0.427	0.000	1.000	0.253	0.435	0.000	1.000
Skilled worker	0.232	0.422	0.000	1.000	0.259	0.438	0.000	1.000
Master craftsman, foreman	0.009	0.096	0.000	1.000	0.009	0.097	0.000	1.000
Employee	0.437	0.496	0.000	1.000	0.407	0.491	0.000	1.000
unknown (only 2011)	0.082	0.273	0.000	1.000	0.072	0.258	0.000	1.000
Establishment characteristics								
ln(Number of workers)	3.942	1.940	0.000	10.875	3.802	1.879	0.000	10.875
Share of high-skilled workers	0.116	0.205	0.000	1.000	0.107	0.199	0.000	1.000
Share of low-skilled workers	0.155	0.215	0.000	1.000	0.158	0.220	0.000	1.000
Increasing employment (Y/N)	0.415	0.493	0.000	1.000	0.417	0.493	0.000	1.000
Regional characteristics								
ln(Employment share of local industry)	-3.532	1.055	-12.732	-0.855	-3.539	1.047	-12.732	-0.855
ln(Number of establishments of local industry)	6.342	1.667	0.000	9.646	6.374	1.642	0.000	9.646
Industrial diversity	3.019	0.263	1.444	3.551	3.013	0.264	1.444	3.551
Share high-skilled workers of local industry	0.106	0.109	0.000	1.000	0.100	0.106	0.000	1.000
Share low-skilled workers of local industry	0.187	0.090	0.000	1.000	0.190	0.092	0.000	1.000
ln(Local unemployment rate among high-skilled labour)*	1.921	0.429	0.294	2.838	1.917	0.428	0.294	2.838
ln(Local unemployment rate among skilled labour)	2.340	0.474	0.981	3.484	2.344	0.478	0.981	3.484
ln(Local unemployment rate among low-skilled labour)	3.374	0.370	2.245	4.293	3.366	0.370	2.245	4.293
Transitions	1,005,316				646,477			

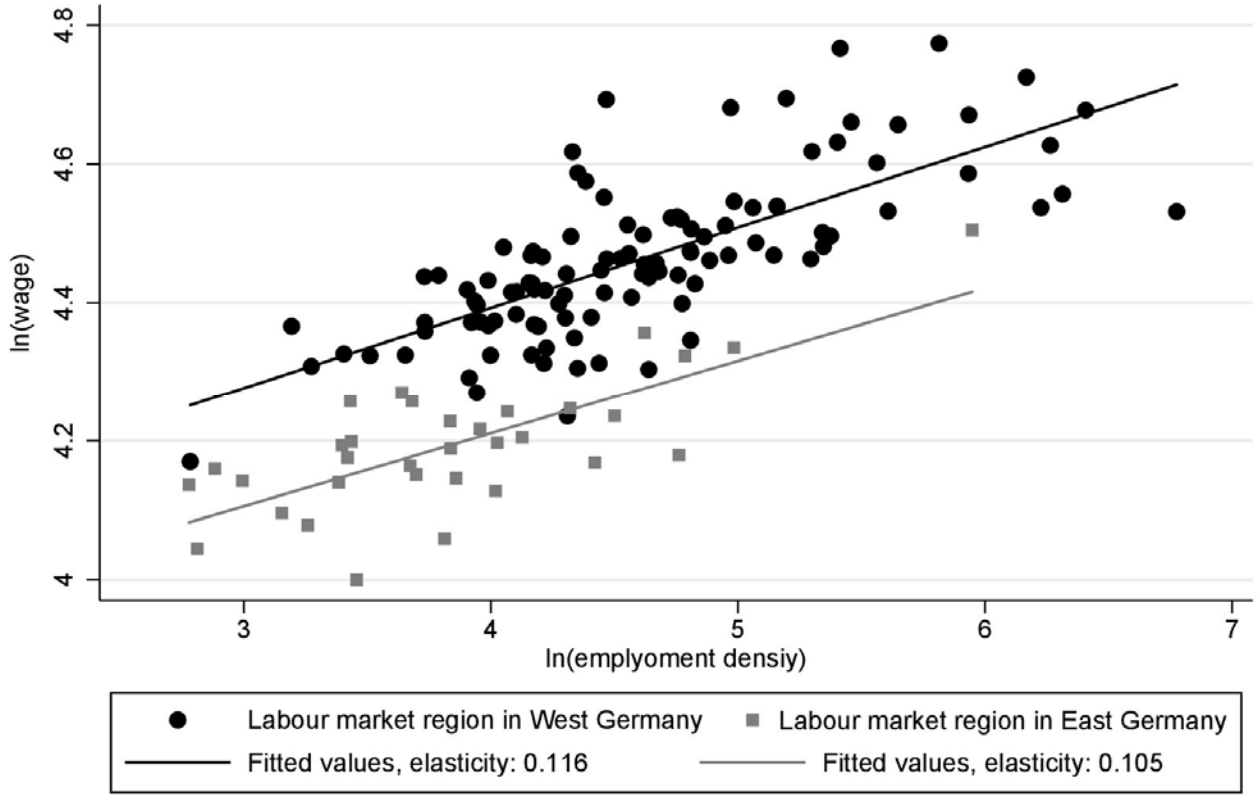
* The statistics on the local unemployment rate among high-skilled labour base only on observations of workers with a university degree. The same applies to the local unemployment rates of the other skill groups.

Table A3: Summary statistics, second stage variables

	Mean*	SD	Min.	Max.
ln(density)	-2.494	0.787	-4.152	-0.130
ln(density), spatial lag	-2.257	0.590	-3.878	-0.730
East Germany	0.234	0.424	0.000	1.000
Average annual precipitation amount 1999-2009	828.043	309.266	466.250	1855.150
Average annual hours of sunshine 1999-2009	1677.156	111.832	1357.610	1916.750
Average temperature 1999-2009	9.196	1.809	2.950	11.360
Coast (Yes/No)	0.085	0.280	0.000	1.000
Restaurant workers per 1,000 inhabitants	69.488	23.193	0.000	137.965
Share of recreation area	1.371	1.177	0.199	6.675
ln(historical population density 1900)	4.687	0.580	3.738	7.690
ln(historical population density 1900) , spatial lag	4.862	0.580	3.912	7.024
Observations	141			

* Except for climate indicators and historic population density the data refers to the year 2005.

Figure A2: Correlation between employment density and wages in new employment relationships



Notes: Average wages based on transitions to employment during the period from 2005 to 2011. Gross daily wages are measured in 2011 prices. Regional labour markets along the former inner-German border are considered West German regions based on their economic centres.

Table A4: First stage results with region fixed effects for ln (imputed gross daily wage)

	All transitions				Job-to-Job transitions	After short-term non-employment	After long-term non-employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual Characteristics							
Female worker	-0.210***	-0.203***					
	(0.001)	(0.001)					
Foreign worker	-0.011***	0.006***	-0.001	-0.001	-0.003	0.001	-0.001
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.006)	(0.022)
Education, reference: Secondary/intermediate school leaving certificate with completed vocational training							
Secondary/intermediate school leaving certificate without completed vocational training	-0.029*	-0.037**	-0.008	-0.002	0.042	0.011	-0.120
	(0.014)	(0.014)	(0.018)	(0.018)	(0.030)	(0.031)	(0.148)
Upper secondary school leaving certificate without completed vocational training	0.048**	0.079***	-0.078***	-0.074***	0.007	-0.072+	-0.217
	(0.015)	(0.014)	(0.021)	(0.021)	(0.033)	(0.043)	(0.151)
Upper secondary school leaving certificate with completed vocational training	0.105***	0.108***	0.021***	0.019***	0.018**	0.005	0.033
	(0.002)	(0.002)	(0.004)	(0.004)	(0.006)	(0.011)	(0.028)
Completion of a university of applied sciences	0.314***	0.350***	0.181***	0.166***	0.126***	0.164***	0.240**
	(0.009)	(0.009)	(0.011)	(0.011)	(0.017)	(0.034)	(0.076)
College/university degree	0.462***	0.499***	0.236***	0.219***	0.175***	0.184***	0.268***
	(0.010)	(0.009)	(0.011)	(0.011)	(0.017)	(0.035)	(0.075)
Experience	0.022***	0.013***	0.053***	0.046***	0.072***	0.015***	0.039*
	(0.000)	(0.000)	(0.002)	(0.002)	(0.003)	(0.004)	(0.016)
Experience^2	-0.000***	-0.000***	-0.001***	-0.001***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Length of employment spell in year of transition		0.009***	0.007***	0.007***	0.007***	0.006***	0.007***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Work experience in previous 5 years		0.055***	0.038***	0.031***	0.033***	0.012***	0.002
		(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
Occupation specific work experience (prev. 5 years)		0.018***	0.006***	0.006***	0.004***	0.007***	0.019***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.004)
Work experience in the region (prev. 5 years)		-0.015***	-0.004***	-0.003***	-0.003***	-0.002*	-0.010+
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.005)
Work experience in agglomerations (prev. 5 years)				0.009***	0.007***	0.005**	0.016*
				(0.001)	(0.001)	(0.002)	(0.006)
ln(Number of previous employers)				0.022***	-0.001	0.019***	0.053**
				(0.002)	(0.003)	(0.003)	(0.016)
First employer				-0.015***	-0.164***	-0.011	-0.045
				(0.004)	(0.170)	(0.007)	(0.034)
Public assistance benefits, reference: no benefit							
Unemployment benefit (ALG I)		-0.033***	-0.010***	-0.009***	-0.002	0.003	0.001
		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.012)
Unemployment assistance (ALG II, ALHI)		-0.030***	-0.007***	-0.007***	-0.002	-0.004	-0.015
		(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.011)
Pre-employment status, reference: not registered as job seeker							
Unemployed and registered as a job seeker		-0.069***	-0.030***	-0.030***	-0.041***	0.001	-0.006
		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.010)
Not unemployed, but registered as a job seeker		-0.080***	-0.024***	-0.024***	-0.037***	0.009*	-0.003

Table A4. Continued.

	All transitions				Job-to-Job transitions	After short-term non-employment	After long-term non-employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual Characteristics							
		(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.009)
Participation in measures of active labour market policy		-0.034***	-0.019***	-0.019***	-0.017***	-0.013***	-0.010
		(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.009)
Occupational status, reference: low-skilled worker							
Skilled worker	0.077***	0.043***	0.019***	0.019***	0.012***	0.019***	0.029**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.009)
Master craftsman, foreman	0.298***	0.237***	0.061***	0.061***	0.039***	0.062***	0.097*
	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)	(0.010)	(0.038)
Employee	0.227***	0.175***	0.024***	0.024***	0.017***	0.022***	0.030*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.013)
unknown (only 2011)	0.244***	0.163***	0.063***	0.063***	0.050***	0.049***	0.055*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.022)
Establishment characteristics							
ln(Number of workers in establishment)	0.042***	0.037***	0.016***	0.016***	0.011***	0.021***	0.026***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)
Share of high skilled in establishment	0.229***	0.214***	0.061***	0.058***	0.055***	0.026***	0.049+
	(0.006)	(0.006)	(0.003)	(0.003)	(0.004)	(0.008)	(0.026)
Share of low skilled in establishment	-0.076***	-0.067***	-0.030***	-0.029***	-0.022***	-0.033***	-0.052***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.015)
Increasing employment in establishment	-0.030***	-0.014***	-0.004***	-0.004***	-0.002	-0.003**	-0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)
Regional characteristics							
ln(Employment share of local industry)				0.007***	0.007***	0.009***	-0.002
				(0.001)	(0.002)	(0.002)	(0.009)
ln(Industrial diversity)				0.010	0.017+	-0.022+	0.096+
				(0.006)	(0.009)	(0.013)	(0.050)
ln(Number of establishments in local industry)				-0.007***	-0.010***	-0.006**	-0.001
				(0.001)	(0.001)	(0.002)	(0.008)
Share high-skilled workers within in local industry				0.083***	0.092***	0.025	0.105
				(0.009)	(0.013)	(0.024)	(0.072)
Share low-skilled workers in local industry				-0.006	-0.011	0.002	0.002
				(0.008)	(0.013)	(0.017)	(0.067)
ln(Local unemployment rate among high-skilled labour)	-0.059***	-0.049***	-0.075***	-0.073***	-0.057***	-0.049**	-0.070+
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.016)	(0.041)
ln(Local unemployment rate among skilled labour)	-0.023***	-0.003	-0.020***	-0.023***	-0.021**	-0.004	-0.028
	(0.005)	(0.005)	(0.004)	(0.005)	(0.008)	(0.008)	(0.035)
ln(Local unemployment rate among low-skilled labour)	-0.015*	0.002	-0.011+	-0.013*	-0.023*	-0.004	0.019
	(0.006)	(0.006)	(0.007)	(0.007)	(0.011)	(0.011)	(0.052)
Constant	3.738***	3.596***	3.357***	3.469***	3.289***	3.799***	3.054***
	(0.015)	(0.014)	(0.030)	(0.035)	(0.058)	(0.072)	(0.280)
Observations	1005316	1005316	646477	646477	262782	173532	12607
Adjusted R^2	0.547	0.602	0.135	0.136	0.128	0.091	0.172

	All transitions				Job-to-Job transitions	After short-term non-employment	After long-term non-employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual fixed effects	No	No	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses. (1)-(2) standard errors clustered at firm level. (3)-(7) Huber/White/sandwich estimator. All models include time fixed effects, region fixed effects, industry fixed effects as well as occupation fixed effects

Table A5: Second stage results for region fixed effects, limited information maximum likelihood (LIML)

	All transitions		Job-to-Job transition	Transition after short-term non-employment	Transition after long-term non-employment
	(1)	(2)	(3)	(4)	(5)
ln(density)	0.013	0.013*	0.010	0.024	-0.049
	(0.024)	(0.006)	(0.011)	(0.062)	(0.102)
ln(density), spatial lag	0.010	0.010+	0.018*	0.006	0.025
	(0.016)	(0.006)	(0.008)	(0.019)	(0.048)
Observations	141	987	141	141	141
R^2	0.832	0.647	0.772	0.803	0.312
Adjusted R^2	0.821	0.642	0.756	0.790	0.265
First stage: Individual characteristics	Yes	Yes	Yes	Yes	Yes
First stage: Biography	Yes	Yes	Yes	Yes	Yes
First stage: Worker fixed effects	Yes	Yes	Yes	Yes	Yes
First stage: Agglomeration variables	Yes	Yes	Yes	Yes	Yes
2SLS: Amenities + East Germany	Yes	Yes	Yes	Yes	Yes
2SLS: F-test for density	17.659	27.998	17.659	17.659	17.659
2SLS: F-test for spatial lag	17.163	22.842	17.163	17.163	17.163
2SLS: Kleibergen-Paap LM rk statistic (p-value)	0.007	0.006	0.007	0.007	0.007
2SLS: Kleibergen-Paap Wald rk F statistic	17.436	25.332	17.436	17.436	17.436
2SLS: Sargan statistic (p-value)	0.318	0.238	0.365	0.224	0.254

Notes: + p<0.1, * p<0.05. Bootstrap standard errors in parentheses (500 replications). F-test: Angrist-Pischke multivariate F-test of excluded instruments. Instruments: historic population density, spatial lag of the historic population density, information on soil characteristics from the European Soil Data base. All first-stage regression models include time-varying worker and job characteristics, worker fixed effects, information on labour market biographies and local industry and regional labour market conditions. All second-stage regression models include controls for first and second nature amenities (see Table A3), a constant, and a dummy variable for East Germany, specification (2) in addition time fixed effects.

Table A6: Second stage results for region fixed effects, different sets of instruments (2SLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ln(density)	0.016	0.019*	0.014	0.015	0.017+	0.016	0.015	0.015+	0.015	0.010+	0.014*
	(0.010)	(0.008)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.005)	(0.006)
ln(density), spatial lag	0.008	0.006	0.008	0.009	0.008	0.009	0.009	0.009	0.009	0.013+	0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.005)
First stage for ln(density)											
Angrist-Pischke F statistic	51.501	7.518	13.819	13.803	13.400	18.605	9.571	20.125	31.563	6.427	17.659
Partial R^2 of exluded instruments	0.225	0.292	0.258	0.253	0.237	0.250	0.266	0.284	0.252	0.391	0.538
First stage for spatial lag of ln(density)											
Angrist-Pischke F statistic	159.071	15.034	32.662	40.567	43.493	55.993	30.595	54.310	94.087	1.703	17.163
Partial R^2 of exluded instruments	0.575	0.615	0.589	0.582	0.582	0.579	0.602	0.584	0.582	0.240	0.702
Sargan statistic (p-value)	n.a.	0.328	0.011	0.630	0.039	0.595	0.105	0.521	0.674	0.256	0.316
Instruments											
Historic population density + spatial lag	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Soil characteristic (see Table A6)	None	Dominant parent material	Top- & sub-soil available water capacity	Depth of rock	Soil erodibility	Topsoil organic carbon content	Hydro-geological class	Subsoil mineralogy	Ruggedness + subsoil mineralogy	All	All

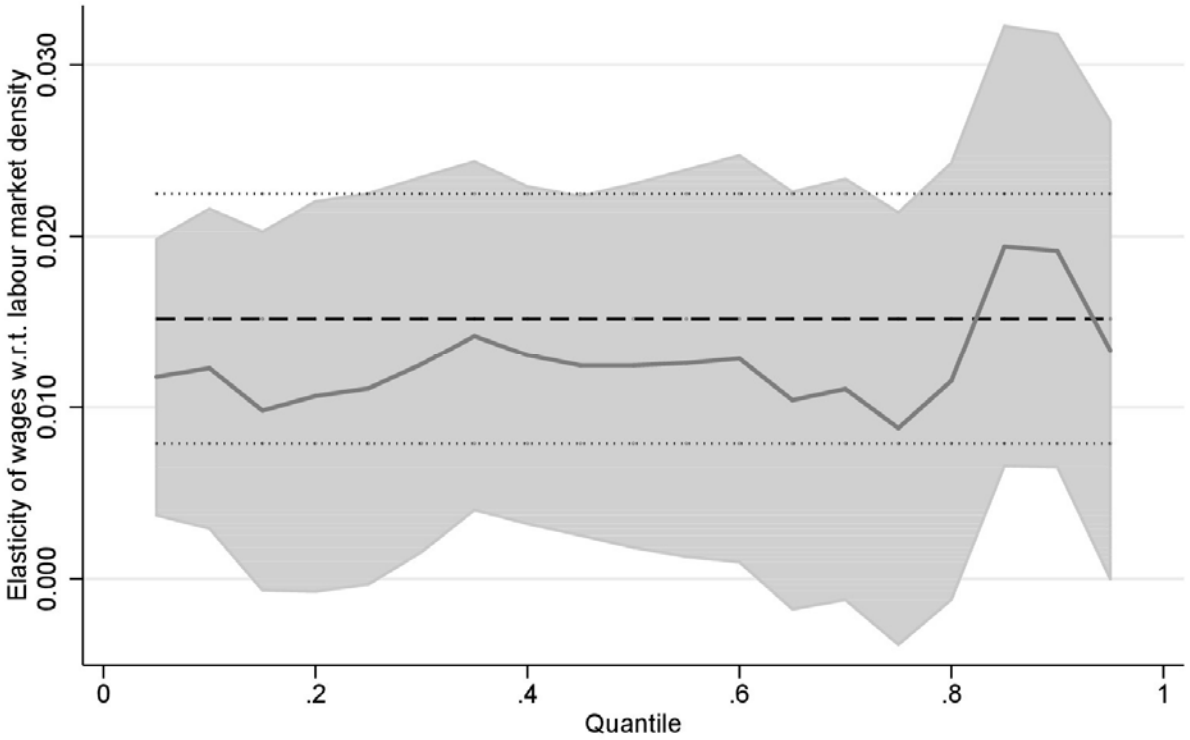
Notes: + p<0.1, * p<0.05. 141 observations for each regression. All regressions include a constant, a dummy variable for East Germany, and amenity variables (see Table A3).

Table A7: One-stage regression with interaction effects: density effect by pre-employment status

Effect of employment density on ln(gross daily wage)	(1)	(2)	(3)
Overall effect (reference: long-term non-employed)	0.012***	0.005*	0.006+
	(0.002)	(0.002)	(0.003)
Additional effect for job-to-job	0.016***	0.016***	0.015***
	(0.002)	(0.002)	(0.003)
Additional effect for short-term non-employed	0.004*	0.005**	0.005+
	(0.002)	(0.002)	(0.003)
Observations	646,477	646,477	646,477
Adjusted R^2	0.136	0.138	0.138
Individual characteristics	yes	yes	yes
Biography	yes	yes	yes
Worker fixed effects	yes	yes	yes
Agglomeration variables	yes	yes	yes
Amenities	no	yes	yes

Notes: + p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses clustered by worker. Model (2) includes amenity indicators (see Table A3) without interactions while in model (3) interaction effects between the amenities and the three pre-employment groups are included in order to allow for group-specific effects of amenities.

Figure A3: Impact of labour market density on wages – quantile regression results



Notes: The solid line represents the coefficients of a bootstrapped quantile regression with increments of 0.05 and 500 replications. The shaded area indicate the 95% confidence interval. The dashed lines refer to the corresponding OLS results given in column (5) in Table 1.

Table A8: Correlation of educational degree and pre-employment status

Frequency; *Expected frequency*	Job-to-Job	Short-term non-employment	Long-term non-employment	Total
Low-skilled	38,453 51,579.2	41,563 33,542.8	18,511 13,405.0	98,527
Medium-skilled	393,529 394,228.3	266,520 256,373.5	93,009 102,456.4	753,058
High-skilled	94,304 80,478.6	34,170 52,336.7	25,257 20,915.7	153,731
Total	526,286	342,253	136,777	1,005,316

Pearson chi2(4) = 3.2e+04 Pr = 0.000

Notes: The education degree refers to the highest educational level obtained by a person in the observation period. Workers with university degree or degree in applied sciences are considered high-skilled, workers with completed vocational training medium-skilled, and all other workers low-skilled.

Table A9: Second-stage results for region fixed effects by skill level

	High-skilled			Medium-skilled			Low-skilled		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln(density)	0.016^{+}	0.030***	0.025*	0.015**	0.021***	0.022**	0.017	0.024*	0.038**
	(0.009)	(0.009)	(0.012)	(0.005)	(0.005)	(0.007)	(0.012)	(0.011)	(0.012)
ln(density), spatial lag	0.006	0.004	0.016	0.007	0.006	0.004	-0.002	-0.002	-0.010
	(0.009)	(0.009)	(0.011)	(0.005)	(0.005)	(0.006)	(0.010)	(0.010)	(0.011)
East Germany	-0.065***	-0.061***	-0.059***	-0.123***	-0.120***	-0.121***	-0.079***	-0.073***	-0.069***
	(0.011)	(0.011)	(0.012)	(0.008)	(0.008)	(0.008)	(0.016)	(0.016)	(0.016)
Constant	-0.107	-0.072	-0.042	-0.174**	-0.153*	-0.162*	-0.220*	-0.187^{+}	-0.149
	(0.075)	(0.074)	(0.077)	(0.060)	(0.060)	(0.069)	(0.102)	(0.102)	(0.119)
Observations	141	141	141	141	141	141	141	141	141
R^2	0.450	0.521	0.516	0.864	0.871	0.871	0.500	0.504	0.494
Adjusted R^2	0.412	0.489	0.482	0.854	0.862	0.862	0.466	0.470	0.460
First stage: Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage: Biography	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage: Worker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage: Agglomeration variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Second stage: Amenities	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2SLS: F-test for density			17.659			17.659			17.659
2SLS: F-test for spatial lag			17.163			17.163			17.163
2SLS: Kleibergen-Paap LM rk statistic (p-value)			0.007			0.007			0.007
2SLS: Kleibergen-Paap Wald rk F statistic			17.436			17.436			17.436
2SLS: Sargan statistic (p-value)			0.229			0.228			0.423

Notes: + p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Bootstrap standard errors in parentheses (500 replications). F-test: Angrist-Pischke multivariate F-test of excluded instruments. Instruments: historic population density, spatial lag of the historic population density, information on soil characteristics from the European Soil Data base. All first-stage regression models include time-varying worker and job characteristics, worker fixed effects, information on labour market biographies and local industry and regional labour market conditions (see Table A2). All second-stage regression models include controls for first and second nature amenities (see Table A3). The skill level refers to the highest educational degree reported in the observation period. Workers with university degree or degree in applied sciences are considered high-skilled, workers with completed vocational training medium-skilled, and all other workers low-skilled.

Table A10: One-stage regression with interaction effects: density effect by pre-employment status and skill level

Effect of employment density on ln(gross daily wage)	(1)	(2)	(3)
Job-to-job # low-skilled	0.019***	0.008+	0.016*
	(0.004)	(0.004)	(0.007)
Job-to-job # medium-skilled	0.029***	0.020***	0.022***
	(0.001)	(0.002)	(0.002)
Job-to-job # high-skilled	0.025***	0.024***	0.012*
	(0.003)	(0.003)	(0.005)
Short-term non-employment # low-skilled	0.005	-0.002	-0.001
	(0.004)	(0.005)	(0.008)
Short-term non-employment # medium-skilled	0.016***	0.010***	0.015***
	(0.001)	(0.002)	(0.002)
Short-term non-employment # high-skilled	0.026***	0.014**	-0.004
	(0.003)	(0.004)	(0.007)
Long-term non-employment # low-skilled	0.000	-0.013*	-0.014
	(0.004)	(0.006)	(0.010)
Long-term non-employment # medium-skilled	0.011***	0.007**	0.013***
	(0.002)	(0.002)	(0.003)
Long-term non-employment # high-skilled	0.029***	0.012*	-0.009
	(0.004)	(0.005)	(0.009)
Observations	646,477	646,477	646,477
Adjusted R^2	0.137	0.138	0.139
Individual characteristics	yes	yes	yes
Biography	yes	yes	yes
Worker fixed effects	yes	yes	yes
Agglomeration variables	yes	yes	yes
Amenities	no	yes	yes
F-tests on equality of effect of employment density (p-values)			
Job-to-job # low-skilled = Short-term non-employment # low-skilled	0.0000	0.0138	0.0192
Short-term non-employment # low-skilled = Long-term non-employment # low-skilled	0.0994	0.0412	0.1653
Job-to-job # low-skilled = Long-term non-employment # low-skilled	0.0000	0.0002	0.0029
Job-to-job # medium-skilled = Short-term non-employment # medium-skilled	0.0000	0.0000	0.0000
Short-term non-employment # medium-skilled = Long-term non-employment # medium-skilled	0.0034	0.1159	0.5559
Job-to-job # medium-skilled = Long-term non-employment # medium-skilled	0.0000	0.0000	0.0017
Job-to-job # high-skilled = Short-term non-employment # high-skilled	0.7394	0.0170	0.0202
Short-term non-employment # high-skilled = Long-term non-employment # high-skilled	0.3530	0.6711	0.6248
Job-to-job # high-skilled = Long-term non-employment # high-skilled	0.2159	0.0184	0.0237

Notes: + p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses clustered by worker. Model (2) includes amenity indicators (see Table A3) without interactions while in model (3) interaction effects between the amenities and the nine skill-pre-employment groups are included in order to allow for group-specific effects of amenities. The skill level refers to the highest educational degree reported in the observation period. Workers with university degree or degree in applied sciences are considered high-skilled, workers with completed vocational training medium-skilled, and all other workers low-skilled.

Table A11: Correlation between employment density 2005 and population density 1900

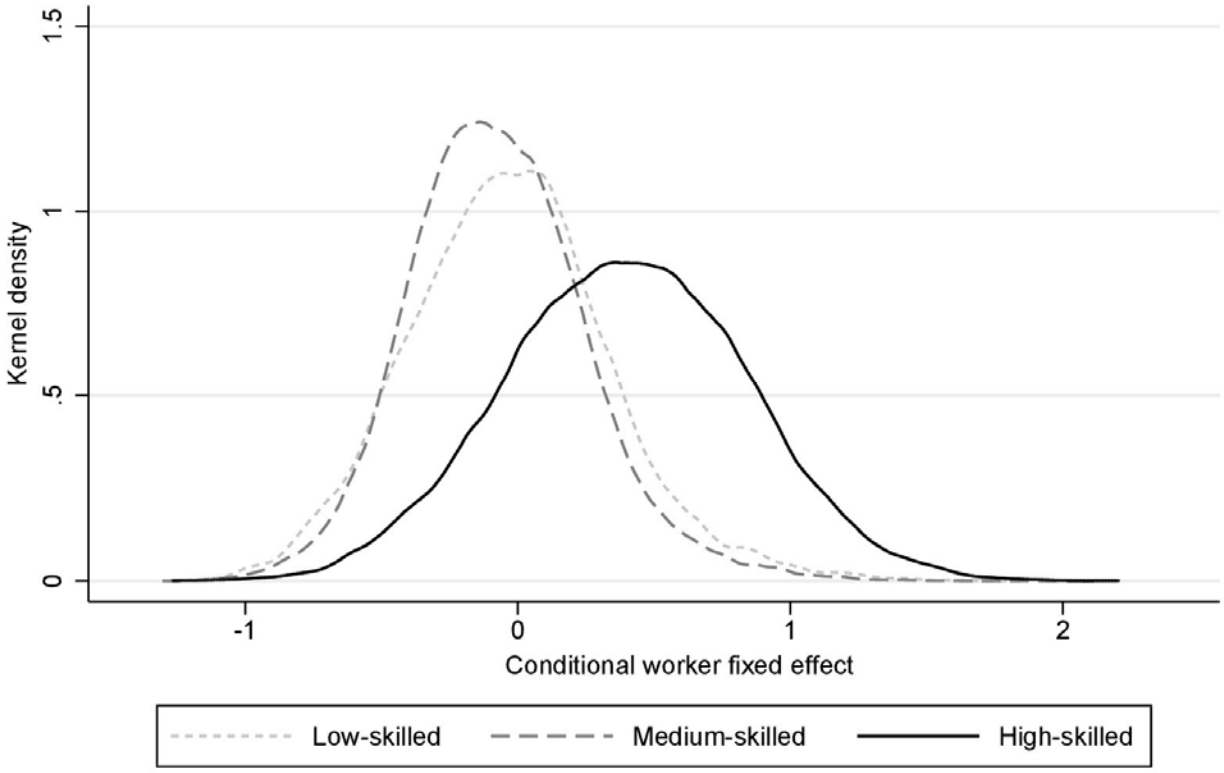
	Coefficient of correlation
ln(density)	0.614
ln(density), spatial lag	0.557
Labour market regions	141

Table A12: IV estimation – first stage results

Dependent variable on first stage:	Logarithm of employment density		Logarithm of spatial lag of employment density	
Historic population density 1900				
Logarithm of population density	0.499***	(0.092)	0.111+	(0.057)
Logarithm of spatial lag of population density	0.159+	(0.081)	0.754***	(0.047)
Soil characteristics				
Subsoil mineralogy, reference: Swel. & non swel.2/1 Minerals				
2/1 & 1/1 Minerals	0.556**	(0.208)	0.268	(0.188)
2/1 & 2/1/1 non swelling Minerals	0.802***	(0.160)	0.121	(0.183)
Dominant parent material, reference: metamorphic rocks				
Consolidated elastic sedimentary rocks	-0.186	(0.220)	-0.073	(0.164)
Sedimentary rocks	0.160	(0.412)	-0.015	(0.266)
Igneous rocks	-0.447	(0.392)	-0.458+	(0.271)
Unconsolidated deposits	-0.477*	(0.234)	-0.017	(0.175)
Unconsolidated glacial deposits/glacial drift	-0.496	(0.320)	-0.156	(0.209)
Eolian deposits	-0.417+	(0.247)	0.028	(0.172)
Dominant parent material, 2nd level, reference: fluvial clays, silts and loams				
Psammite or arenite	-0.565*	(0.237)	-0.296	(0.208)
Pelite, lutite or argilite	-0.310	(0.262)	-0.158	(0.225)
Calcareous rocks	-0.348	(0.324)	-0.265	(0.212)
Weakly metamorphic rocks	-1.430***	(0.318)	-0.296	(0.283)
Acid regional metamorphic rocks	-0.626*	(0.298)	-0.183	(0.239)
Residual and redeposited loams from silicate rocks	-0.234	(0.276)	-0.133	(0.168)
Morainic deposits	-0.101	(0.250)	-0.050	(0.147)
Glaciofluvial deposits	0.143	(0.174)	-0.123	(0.127)
Loess	0.038	(0.230)	-0.129	(0.169)
Subsoil available water capacity, reference: medium (100 – 140 mm/m)				
Low (< 100 mm/m)	-0.113	(0.237)	0.023	(0.167)
High (140 – 190 mm/m)	0.118	(0.231)	-0.036	(0.123)
Very high (> 190 mm/m)	0.256	(0.365)	-0.065	(0.199)
Topsoil available water capacity, reference: very high (> 190 mm/m)				
Medium (100 – 140 mm/m)	-0.089	(0.155)	-0.042	(0.125)
High (140 – 190 mm/m)	-0.048	(0.162)	0.084	(0.114)
Depth of rock, reference: shallow (< 40 cm)				
Moderate (40 – 80 cm)	0.468	(0.417)	0.065	(0.220)
Deep (80 – 120 cm)	0.371	(0.446)	0.041	(0.191)
Very deep (> 120 cm)	0.460	(0.450)	-0.017	(0.183)
Soil erodibility class, reference: weak				
Moderate	0.121	(0.193)	0.158	(0.143)
Strong	-0.089	(0.156)	0.124	(0.111)
Very strong	-0.090	(0.171)	0.025	(0.152)
Topsoil organic carbon content, reference: low (1 – 2 %)				
Medium (2 – 6 %)	0.411+	(0.234)	0.071	(0.127)
Very low (< 1 %)	0.326	(0.217)	0.040	(0.120)
Hydrogeological class, reference: 1M, 4W				
1C	0.708*	(0.318)	-0.421	(0.295)
1S	0.816*	(0.345)	-0.033	(0.273)
1L	0.686*	(0.332)	-0.205	(0.283)
1H	0.901**	(0.296)	-0.287	(0.356)
2	0.562	(0.355)	-0.148	(0.297)
Local terrain ruggedness	0.000	(0.000)	0.001*	(0.000)
Labour market regions	141		141	
Angrist-Pischke multivariate F test of excluded instruments	17.659		17.163	
Partial R-squared of excluded instruments	0.538		0.702	

Notes: Both regressions include a constant, dummy for East Germany, and six amenity variables: sunshine, precipitation, temperature, coast, restaurant workers per 1,000 inhabitants, and share of recreation area. Robust standard errors in parentheses. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.

Figure A4: Distribution of worker fixed effects by educational level



Notes: The worker fixed effects refer to specification (2) in Table A7. The skill level denotes the highest educational degree reported in the observation period. Workers with university degree or degree in applied science are considered high-skilled, workers with completed vocational training medium-skilled, and all other workers low-skilled.

References

Andersson, M., Klaesson, J., & Larsson, J. P. (2014). The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science*, *93*(4), 727–747.

vom Berge, P., Burghardt, A., & Trenkle, S. (2013). *Sample of integrated labour market biographies: Regional file 1975-2010 (SIAB-R 7510)*. FDZ-Datenreport 09/2013 (en), Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

Braun, H.-J. (1990): *The German Economy in the Twentieth Century*. Routledge contemporary economic history of Europe series, Routledge, London.

Carlsen, F., Rattso, J., & Stokke, H. (2016). Education, experience, and urban wage premium. *Regional Science and Urban Economics*, 60(C), 39–49.

Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity. American Economic Review, 86(1), 54–70.

Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, *63*(2), 723–742.

Combes, P.-P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In E. L. Glaeser (Ed.), *Agglomeration Economics*, chap. 1, (pp. 15–66). Chicago and London: University of Chicago Press.

Combes, P.-P., & Gobillon, L. (2015*). The empirics of agglomeration economies*. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), Handbook of Regional and Urban Economics, vol. 5, (pp. 247–348). Elsevier.

Eberle, J., Jacobebbinghaus, P., Ludsteck, J., & Witter, J. (2011*). Generation of time-consistent industry codes in the face of classification changes - Simple heuristic based on the Establishment History Panel (BHP)*. FDZ-Methodenreport 05/2011, Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

Fitzenberger, B., Osikominu, A., & Völter, R. (2005). *Imputation rules to improve the education variable in the IAB employment subsample*. FDZ-Methodenreport 03/2005, Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

Guevara-Rosero, G. C., Riou, S. & Autant-Bernard, C. (2018): Agglomeration externalities in Ecuador: Do urbanization and tertiarization matter? *Regional Studies* (online first), DOI: 10.1080/00343404.2018.1470325.

Kosfeld, R., & Werner, A. (2012). Deutsche Arbeitsmarktregionen - Neuabgrenzung nach den Kreisgebietsreformen 2007-2011 [German labour market regions – New delimitation after the territorial reforms of counties 2007-2011]. *Raumforschung und Raumordnung*, *70*, 49–64.

Matthes, B., Burkert, C., & Biersack, W. (2008). *Berufssegmente: Eine empirisch fundierte Neuabgrenzung vergleichbarer beruflicher Einheiten [Occupational segments: A new empirical definition of comparable occupational units]*. IAB-Discussion Paper 35/2008, Institute for Employment Research (IAB) of the Federal Employment Agency (BA), Nuremberg.

Rappaport, J. 2007. Moving to nice weather. *Regional Science and Urban Economics 37* (3): 375–398.

Reichelt, M. (2015*). Using longitudinal wage information in linked data sets - The example of ALWA-ADIAB.* FDZ-Methodenreport 01/2015, Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

Rothenbacher, F. (2002). *The European Population, 1850-1945*. Societies of Europe. Basingstoke: Palgrave Macmillan.

Spoerer, M., Streb, J. (2013): *Neue deutsche Wirtschaftsgeschichte des 20. Jahrhunderts.* Oldenburg Verlag, München.