Regional Science and Urban Economics 95 (2022) 103785

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

The spatial decay of human capital externalities - A functional regression approach with precise geo-referenced data

Johann Eppelsheimer [a], Elke J. Jahn [b], Christoph Rust [c,d,*]

[a] Institute for Employment Research (IAB), Nuremberg, Germany
[b] Institute for Employment Research (IAB), Nuremberg and University of Bayreuth, Germany
[c] WU Vienna University of Economics and Business, Austria
[d] University of Regensburg, Germany

ARTICLE INFO

JEL classification:
C13
D62
J24
J31
R10
R12

Keywords:
Human capital externalities
Functional regression
Geodata
Wages

ABSTRACT

This paper analyzes human capital externalities from high-skilled workers by applying functional regression to precise geocoded register data. Functional regression enables us to describe the concentration of high-skilled workers around workplaces as continuous curves and to efficiently estimate a spillover function determined by distance. Furthermore, our rich panel data allow us to address the sorting of workers and disentangle human capital externalities from supply effects by using an extensive set of time-varying fixed effects. Our estimates reveal that human capital externalities attenuate with increasing distance and disappear after 25 km. Externalities from the immediate neighborhood of an establishment are twice as large as externalities from surroundings 10 km away.

1. Introduction

Workers interact with each other within and across firms. They share their knowledge, discuss ideas and adopt procedures and technologies. All of these interactions potentially increase the productivity of workers through *'human capital externalities'* (Davis and Dingel, 2019; Acemoglu, 1996; Lucas, 1988; Marshall, 1890). Although a large amount of empirical literature supports the existence of geographically bounded human capital externalities (Cornelissen et al., 2017; Ciccone and Peri, 2006; Moretti, 2004; Rauch, 1993), little is known about the exact spatial extent of human capital externalities. For several reasons, human capital externalities are likely to decline with increasing distance. For instance, distance increases the costs of planned social interactions, such as meetings. Similarly, distance reduces the likelihood of unintended encounters that lead to the exchange of knowledge. Moreover, because distance generally increases the number of intermediaries between individuals in a social network and an increasing number of intermediaries impedes information flows, distance depresses indirect information flows. Consequently, individuals are likely to benefit more from proximate than distant neighbors.

The aim of this paper is to shed light on the question of how human capital externalities emerging from interactions among workers close to the workplace attenuate with increasing distance. To this end, we draw from a large and novel micro panel dataset that features the exact coordinates of nearly all German establishments and rich information on individual workers covering one and a half decades. To describe the distributions of high-skilled workers we compute spatial functions that relate the share of high-skilled workers to the distance to each workplace. Furthermore, we introduce a new estimation procedure to the urban economic literature that is capable of evaluating such detailed geodata. This method allows us to estimate the spatial attenuation of human capital externalities with high precision. In line with previous studies, we assume that wages reflect the productivity of workers and aim to measure human capital externalities based on external wage effects from the local concentration of high-skilled workers. External wage effects may arise from knowledge exchange (Marshall, 1890;

* Corresponding author. WU Vienna University of Economics and Business, Austria.
E-mail address: christoph.rust@wu.ac.at (C. Rust).

https://doi.org/10.1016/j.regsciurbeco.2022.103785
Received 18 June 2020; Received in revised form 9 March 2022; Accepted 10 March 2022
Available online 15 March 2022

Lucas, 1988) or the diffusion of new technologies (Nelson and Phelps, 1966; Acemoglu, 1998).

Previous empirical research provides initial evidence for spatially decreasing human capital externalities. Using cross-sectional data from the US, Rosenthal and Strange (2008) construct concentric rings around workers that measure the concentration of human capital within 5 miles and between 5 and 25 miles. To explore how human capital externalities attenuate with increasing distance, they regress individual wages on the concentration of human capital within these rings. They find that human capital externalities in the inner ring are notably larger than externalities in the outer ring. A closely related study by Fu (2007) adopts the strategy of Rosenthal and Strange (2008) to analyze cross-sectional data from the Boston metropolitan area. Using more precise geocoded data, Fu (2007) measures the concentration of human capital within finer rings (i.e., 0–1.5, 1.5–3, 3–6 and 6–9 miles). Fu (2007) finds evidence that human capital externalities may vanish after only three miles. Recent findings from the Netherlands in a study using panel data and concentric distance rings of 0–10, 10–40, and 40–80 km suggest that human capital externalities extend to 10 km (Verstraten, 2018). Although these studies provide evidence for the spatial attenuation of human capital externalities, the exact decay of the effects remains unclear because the literature is constrained either by relatively imprecise geo-information or by specific data from a single area. Furthermore, most empirical evidence is restricted to cross-sectional data, which complicates causal inference. Additionally, the described studies overlook the fact that human capital externalities from high-skilled workers are entangled with supply effects (Katz and Murphy, 1992; Card and Lemieux, 2001; Borjas, 2003; Moretti, 2004; Ciccone and Peri, 2006).

To fully exploit the information contained in the exact geocodes of workplaces, we adopt a methodologically fresh approach and measure the magnitude of human capital externalities (or spillovers) with respect to distance in a continuous manner. Recent developments in functional data analysis (FDA) provide particularly suitable frameworks. FDA is a branch of statistics that extends classical statistical methods to random variables with a functional nature, such as curves or surfaces over a continuous domain. Typical examples of such data are temperature curves, growth curves and the continuous evolution of stock prices over time. The continuity of curves indicates that adjacent values are related. In many applications, exploiting this information makes FDA more efficient than classical multivariate methods for discretized data.

While statisticians employ FDA in a wide range of applications (see Ullah and Finch, 2013 for a systematic overview), FDA is applied quite rarely in economics (examples include Ramsay and Ramsey, 2002; Wang et al., 2008 and Caldeira and Torrent, 2017).[1] Therefore, this paper illustrates the potential of FDA for use in economic research with high-dimensional variables. Our approach relies on a functional linear regression model in which a scalar outcome variable (log wage) is regressed on observations of a functional random variable (the share of high-skilled workers as a function of the distance to a worker's workplace). Thus, we augment the classical scalar-on-function regression model to incorporate additional scalar-valued explanatory variables and use an estimation procedure suggested by Crambes et al. (2009), which is based on smoothing splines and makes it possible to model the function-valued spillover parameter very flexibly. The estimated spatial spillover function relates wages to the share of high-skilled workers as a function of distance, which is evaluated at 500-m intervals up to 50 km.

The previous literature that estimates the spatial attenuation of economic effects follows a semi-parametric approach (e.g., Rosenthal and Strange, 2008; Fu, 2007; Verstraten, 2018; Gibbons et al., 2021; Faggio et al., 2019; Faggio, 2019).[2] In the semi-parametric approach, econometricians estimate linear models in which the main explanatory variable is measured in several geographically concentric rings or circles around observations. Although the semi-parametric approach is generally well suited to measure the spatial attenuation of economic effects and is a straightforward application of the linear ordinary least squares (OLS) model, it is less precise compared with our FDA approach. The reason is that multicollinearity issues usually do not allow the estimation of effects from a large and fine-graded series of measurement points. Therefore, to circumvent multicollinearity issues, researchers are forced to construct relatively broad rings or circles that measure the spatial distribution of the explanatory variable. Our FDA approach solves this issue by regularizing the parameter estimates, which enables us to exploit geographically fine-graded data and to estimate the spatial attenuation of economic effects in detail.

To identify external wage effects based on the concentration of high-skilled workers around workplaces, we address a series of confounding factors. We account for unobserved individual heterogeneity as well as locational advantages with worker and local-labor-market-region fixed effects. Two additional challenges in identifying regional human capital externalities are confounding supply effects and the sorting of high-skilled workers into high-wage regions. We aim to address both problems with an extensive set of time-varying fixed effects.

Empirically, it is well established that high- and low-skilled workers are mutually imperfect substitutes (Autor et al., 2008; Ciccone and Peri, 2005; Card and Lemieux, 2001; Krusell et al., 2000). Under this condition, standard production models with high- and low-skilled workers as inputs and productivity-enhancing externalities from the local concentration of high-skilled workers illustrate two channels through which the local concentration of human capital affects wages. First, human capital externalities and, second, labor supply effects that stem from the imperfect substitution of high- and low-skilled workers and thus from changes in their relative scarcity. Although human capital externalities increase the productivity and wages of all workers, supply effects increase only the wages of low-skilled workers but depress the wages of high-skilled workers. Consequently, low-skilled workers unambiguously benefit from an increasing concentration of high-skilled workers. The net effect on high-skilled workers depends on the relative sizes of human capital externalities and supply effects (for more details, see, e.g., Moretti, 2004, Ciccone and Peri, 2006, and Heuermann, 2011). However, in both cases, human capital externalities are entangled with supply effects.

The aim of this paper is to estimate human capital externalities. Consequently, we need to disentangle spillover from supply effects. To do so, we follow Eppelsheimer and Möller (2019) and exploit the different spatial natures of the two effects. While it is plausible that supply effects are spatially equally distributed within local labor markets (i.e., supply effects originating in one part of the city uniformly affecting wages throughout the city), the intensity of spillover effects truly depends on distance (i.e., spillovers affect close neighbors more than distant neighbors). Thus, in the data, we aim to purge spillover effects from supply effects by eliminating the variation common within regional labor markets. To do so, we include time-varying labor-market-area fixed effects in the econometric specification (i.e., a specific intercept for every labor market area in each year). Because supply effects may have different impacts on high- and low-skilled workers, we further interact these labor-market-area-year fixed effects with a skill dummy.

[1] Readers with a general interest in FDA are referred to the textbooks of Ramsay and Silverman (2005); Ferraty and Vieu (2006); Horváth and Kokoszka (2012) and Hsing and Eubank (2015).

[2] Some examples of studies that investigate the spatial patterns of agglomeration effects are Arzaghi and Henderson (2008), who study networking effects within the advertising agency industry in Manhattan; Ahlfeldt et al. (2015), who examine productivity externalities in Berlin; Andersson et al. (2019), who evaluate productivity effects from industry specialization and diversity in Swedish cities; and Faggio (2019) and Faggio et al. (2019), who assess the local labor market impacts of relocations of public sector jobs in the UK and Germany.

Following Cornelissen et al. (2017), who addressed worker sorting in a related context at the firm level (Abowd et al., 1999; Card et al., 2013), we address the sorting of high-skilled workers into high-wage regions (Acemoglu and Angrist, 2000) by including a comprehensive set of fixed effects. In particular, the above-introduced labor-market-area-year fixed effects nullify unobserved regional heterogeneity that might attract high-skilled workers, such as (changes in) average wages, general labor market conditions, and amenities. Importantly, labor-market-area-year fixed effects also cover temporal labor market shocks that might pull or push skilled workers into or out of regions, which is a concern raised by Moretti (2004).

Overall, we estimate wage effects from changes in the concentration of high-skilled workers around workplaces. We exploit the variation in the local composition of high-skilled workers over time due to workers who switch employers or the creation or destruction of surrounding establishments. Due to our extensive set of controls and fixed effects, our estimates are unrelated to observable labor market characteristics of individuals, time-constant unobservable individual heterogeneity, and spatially constant temporal trends within local labor markets.

We find significant spillover effects from the local concentration of high-skilled workers around establishments. Moreover, our estimates reveal that spillover effects decay with increasing distance. Human capital externalities from direct neighbors (i.e., high-skilled workers who are located within a 0.5-km radius) are roughly twice as large as spillovers from high-skilled workers that are located 10 km away. After 25 km, spillover effects vanish completely. Overall, a one-standard-deviation increase in the local share of high-skilled workers leads to wage gains of roughly 4%. The magnitude of this effect is comparable to *classical* estimates at the aggregate level (NUTS-3), for which we provide results in the online appendix. Reassuringly, semi-parametric estimates based on the same data lead to similar results. In general, our findings are in line with the urban economic literature and support the existence of human capital externalities.

In addition, our results show that the majority of the effect is bounded to the close vicinity of workplaces to high-skilled workers. Individuals working further away still gain from human capital externalities but to a weaker extent. Workers in very remote regions, however, do not benefit from human capital externalities at all. Furthermore, we find that human capital spillovers differ by subgroups. Spillover effects are stronger in metropolitan and urban regions compared to rural regions. Further, human capital externalities are more pronounced for medium- and high-skilled workers than for low-skilled workers. Workers in the manufacturing sector benefit slightly more than workers in the service sector. Finally, human capital externalities are higher for younger than older workers, and females benefit more than males.

The remainder of the paper is organized as follows. The next section explains the estimator and our identification strategy. Section 3 summarizes the data. Section 4 presents our main findings and discusses the identified effects on various subgroups. Section 5 illustrates the statistical properties of the estimator in a simulation study, and section 6 compares our estimation procedure to the semi-parametric approach. Section 7 provides robustness checks. Section 8 presents the conclusion of the paper.

2. Estimation strategy

This paper seeks to measure the spatial attenuation and reach of human capital externalities. Thus, we aim to measure external productivity effects from the local concentration of human capital. As productivity itself cannot be observed directly, we use individual wages as a proxy. To model the concentration of human capital, we compute the share of high-skilled workers around each workplace in our data as a continuous function that depends on distance. Consequently, each workplace provides a unique functional description of the surrounding concentration of human capital. To fully exploit the available information, we model the treatment effect accordingly as a continuous function that depends on distance.

In the following section, we present our estimator, describe how we adjust it to meet the requirements of our application, and define the functional representation of the concentration of high-skilled workers around a workplace. Finally, we specify our identification strategy that addresses endogenous sorting of workers and confounding supply effects. For notational simplicity, we formulate the estimation framework only for the cross-section. The empirical model, of course, considers the panel structure of our data.

2.1. Estimator

The spatial allocation of human capital varies considerably across and within administrative boundaries. For a given location, for example, worker i's workplace, the concentration of high-skilled workers in the immediate neighborhood of worker i's workplace may therefore differ from the concentration in the greater adjacent area. In principle, it is possible to measure the concentration of high-skilled workers at any distance to worker i's workplace. Thus, one can naturally regard the concentration of high-skilled workers with respect to the distance to worker i's workplace as a curve. We use these function-valued observations as an explanatory variable to assess how the concentration of human capital influences productivity in space.

The functional linear regression model with a scalar response variable is a suitable framework to measure such a relationship. With Y_i being the scalar dependent variable, which is the log wage in our empirical analysis, the model is defined as

$$Y_i = \int_0^1 \beta(z) X_i(z)\,\mathrm{d}z + \varepsilon_i,\ i = 1, \dots, n \tag{1}$$

where $X_i \in L^2([a, b])$, $i = 1, \dots, n$, are n identically distributed random functions defined on a common domain, which we set to $[0, 1]$ without loss of generality. In our application, $X_i(z)$ is the share of high-skilled workers among all workers located z units away from worker i's workplace. In principle, it is possible to observe the whole curve X_i, that is, $X_i(t)$ for any $t \in [0, 1]$, but in practice, one must work with a finite number of observation points. These points are denoted by $z_1, \dots, z_p$, and we only consider the case in which these points are equidistant. The function-valued coefficient parameter $\beta \in L^2([0, 1])$ is the quantity of interest and describes the influence of X_i on Y_i, which can be different for different z values. This coefficient function measures the magnitude of the productivity spillover induced by human capital located z units away from worker i's workplace. The error term ε_i is assumed to be independent and identically distributed (iid), have a mean of zero, and be independent of the curves X_i (later, we will consider heteroscedastic and autocorrelated errors).

Model (1) has received considerable attention in the FDA literature (see Morris, 2015, for an overview). Classically, the estimation of β is based on the Karhunen-Loève decomposition of the empirical covariance operator of the observed curves X_i. A drawback of the classical approach is that the expansion of the functional principal component (FPC) estimator heavily depends on the random curves' correlation structure. In this paper, we therefore build upon the smoothing spline estimator proposed by Crambes et al. (2009). This approach has the advantage that the basis functions are independent of the curves X_i, which results in a more 'natural' way of modeling β (as a smooth function). From an asymptotic perspective, both estimators have minimax-optimal convergence rates (Hall and Horowitz, 2007; Crambes et al., 2009).

To estimate β, the approach of Crambes et al. (2009) minimizes the penalized sum of squared residuals

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{p}\sum_{j=1}^{p}\beta(z_j)X_i(z_j)\right)^2 + \rho\left(\frac{1}{p}\sum_{j=1}^{p}\pi_\beta^2(z_j) + \int_0^1 (\beta^{(m)}(z))^2\,\mathrm{d}z\right), \tag{2}$$

which comprises two terms – the first quantifies how well the model can fit the data, and the second measures the curvature of β (via its m-th derivative). The term $\frac{1}{p}\sum_{j=1}^{p}\pi_\beta^2(z_j)$, where $\pi_\beta(z)$ is the best approximation of $\beta(z)$ by a polynomial of degree $m - 1$, is not common in traditional smoothing spline regressions. However, this term is necessary to ensure a unique solution without imposing further assumptions on the random function X_i.

Traditional smoothing splines penalize second derivatives; thus, setting $m = 2$ is also the most natural choice in our situation, which results in an expansion of cubic natural splines with knots at $z_1, \ldots, z_p$. The penalty parameter $\rho \geq 0$ controls the flexibility of the estimated parameter function $\hat{\beta}$. With $\rho = 0$, as one extreme, equation (2) coincides with the least-squares criterion and, with $\rho \to \infty$ as the other extreme, $\hat{\beta}$ is constrained to be a function of which the m-th derivative is zero. Since $m = 2$ in our case, the coefficient function will become a straight line if $\rho \to \infty$.

The estimation framework proposed by Crambes et al. (2009) does not cover the case of additional (scalar) covariates. Therefore, to account for the influence of further explanatory variables, we must expand model (1) with a k-vector of scalar-valued explanatory variables Z_i and a corresponding parameter vector γ:

$$Y_i = \int_0^1 \beta(z)X_i(z)\,\mathrm{d}z + Z_i'\gamma + \varepsilon_i. \tag{3}$$

Accordingly, we augment the smoothing spline estimator of Crambes et al. (2009) to incorporate scalar-valued explanatory variables. We refer the reader to the online appendix B.1 supporting this paper for the technical details of the estimation method and statistical inference and to appendix B.2 for the choice of the smoothing parameter ρ.

2.2. Calculation of curves

A key feature of our analysis is the representation of the spatial density of high-skilled workers around workplaces as curves. To measure the concentration of high-skilled workers, we follow the recent literature on regional human capital externalities (for example Moretti, 2004) and use the share of high-skilled workers relative to all workers. To calculate curves from geocoded data, we compute the values of the functions $X_i(z)$ for each worker i on an equidistant grid $z_1, \ldots, z_p$:

$$X_i(z_j) = \frac{n^{hs}_{i,[z_j-h,z_j)}}{n_{i,[z_j-h,z_j)}}. \tag{4}$$

Here, $n^{hs}_{i,[z_j-h,z_j)}$ refers to the number of high-skilled individuals for which the spherical distance between their working location and the workplace of worker i is at least as large as $z_j - h$ and smaller than z_j. Similarly, $n_{i,[z_j-h,z_j)}$ is the total number of workers (high skilled and non-high skilled) within the same distance window. In other words, the value of the curve X_i at distance z_j indicates the share of high-skilled workers among all workers within the distance window $\left[z_j - h, z_j\right)$, where h is a fixed bandwidth. To ensure that an establishment's own skill structure does not affect measurements of its surroundings, we compute $X_i(z_1)$ without its own number of workers. Thus, we only measure regional human capital externalities without establishment-internal spillovers. To balance the analytical precision and computational costs, we choose a bandwidth of $h = 500$ m and compute $X_i(z_j)$ on the grid $z_j = 500\,m, 1000\,m, \ldots, 50,000\,m$.

2.3. Identification

Following the explanation of the estimator, we now address confounding supply effects and endogenous sorting of individuals. The empirical literature has established that high- and low-skilled labor are imperfect substitutes (e.g., Autor et al., 2008; Ciccone and Peri, 2005; Card and Lemieux, 2001; Krusell et al., 2000). As Acemoglu and Angrist (1999), Moretti (2004) and Ciccone and Peri (2006) illustrate, in addition to potential externalities, changes in the supply of high-skilled labor constitute a market mechanism that affects wages. Due to these supply effects, an increase in the share of high-skilled workers in the labor market decreases the wages of high-skilled workers and increases the wages of low-skilled workers. Consequently, changes in the local concentration of high-skilled workers might simultaneously influence wages through supply effects and human capital externalities.

To disentangle human capital externalities from supply effects, we follow Eppelsheimer and Möller (2019) and exploit the different spatial natures of the two effects. On the one hand, the intensity of human capital externalities should be highly localized and decay with increasing distance. Therefore, we expect larger spillovers from establishments in close vicinity than from those farther away. On the other hand, supply effects arguably uniformly affect larger areas. Consequently, purging the variation common within larger areas from the data potentially eliminates labor supply effects while not affecting highly localized externalities.

In our estimation framework, we aim to achieve such a disentanglement of human capital externalities and supply effects by including local-labor-market-area fixed effects. To identify local labor markets, we follow the definition from the Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR) that groups 402 counties into 258 local labor market areas based on commuter links (Kosfeld and Werner, 2012). These local labor markets can be interpreted as self-contained labor markets. Nevertheless, in case supply effects are not perfectly uniform within each area, some correlation between human capital externalities and supply effects may still remain in the data.

As supply effects vary over time and effects might be different for different skill groups, we expand equation (3) to include time-varying labor-market-area fixed effects for each skill group π_{rst} (i.e., an intercept for each labor market area and skill group in every year). Our full estimation equation is

$$Y_{it} = \int_0^1 \beta(z)X_{it}(z)\,\mathrm{d}z + Z_{it}'\gamma + \delta_i + \tau_t + \omega_o + \pi_{rst} + u_{it}. \tag{5}$$

Here, Y_{it} is the individual log wage of worker i in year t, and $X_{it}(z)$ is the share of high-skilled workers, which is described as a continuous curve around the workplace of individual i that depends on distance z. Note that all workers of a given establishment in year t share the same locational characteristics; specifically, they all have the same curve $X_{it}(z)$. $\beta(z)$ is the associated spillover function that we seek to retrieve from the data. The model controls for time-varying observable individual characteristics Z_{it} and a series of fixed effects. δ_i are individual fixed effects that control for unobserved heterogeneity of workers, τ_t is a year fixed effect and ω_o is an occupation fixed effect. Finally, we also include dummy variables for local labor markets in each time period and for every skill level (π_{rst}) which are crucial to our identification strategy, as discussed above.

A further challenge is that high-wage areas might attract high-skilled workers. Such a trend would reverse the direction of causality in our estimates (Moretti, 2004). Our identification strategy aims to overcome this issue by removing all time-constant and time-varying variations at the local labor market level by including local-labor-market-area fixed effects (π_{rst}). These fixed effects erase push and pull factors that might attract or distract high-skilled workers. Thus, reversed causality is quite unlikely.[3]

In summary, equation (5) allows us to estimate human capital externalities unrelated to the supply effects that are spatially constant within local labor market areas and the endogenous sorting of individuals. The remaining variation in $X_{it}(z)$ in equation (5) stems from temporal intraregional changes in the concentration of high-skilled workers around workplaces.

Note that our approach aims to measure human capital externalities at the workplace, not the place of residence. However, in cases in which workers reside close to their workplaces, we cannot discriminate whether externalities come from the workplace or the residential neighborhood. In this case, we assign the effect to the workplace. However, in Germany, most individuals do not live very close to their places of work (see Dauth and Haller, 2018). Thus, such a bias is likely to be small.

3. Data and descriptive statistics

3.1. Data

In the empirical analysis, we combine administrative data on almost all German establishments and rich data from a representative sample of workers over a period of 15 years. Our panel data include exact geo-coordinates of establishments and therefore allow us to describe the distribution of high-skilled workers as spatial functions around individual workplaces. We evaluate the share of high-skilled workers at 500-m intervals up to a distance of 50 km.

Our main meso-level data sources are the *Establishment History Panel* (BHP 7516) and *IEB GEO* from the Institute for Employment Research (IAB).[4] The *Establishment History Panel* comprises all German establishments with at least one employee on June 30 of each year. The dataset provides establishment-level information on the total number of employees and the number of employees with tertiary education, among other metrics. To measure the distribution of high-skilled workers, we classify employees holding a degree from a university or a university of applied sciences as high-skilled.[5]

We expand the dataset with exact geo-coordinates from IEB GEO. IEB GEO is a novel data source that includes the addresses of establishments in the *Establishment History Panel* between 2000 and 2014 as geo-coordinates. In Germany, firms are obliged to register at least one of their establishments per municipality and industry. In general, the registration of one establishment per municipality provides a detailed description of the geographic landscape of workplaces. In some cases, however, firms might actually have multiple establishments within the same industry in a single municipality, which they do not report. In these cases, we cannot confirm that individuals work where they are registered. We therefore exclude the following chain-store industries from our data: construction, financial intermediation, public service, retail trade, temporary agency work, and transportation. With the remaining set of establishments, we compute the density of high-skilled workers as spatial functions around establishments as described in section 2.2. Nevertheless, if the skill-structure of satellite offices differs from that of the headquarters, there still might be some measurement error. For instance, if lower-skilled workers near satellite offices are wrongly assigned to the headquarters, the share of high-skilled workers would be too small at the headquarters and too large near satellite offices. Consequently, this would lead us to overestimate human capital externalities near the headquarters and underestimate externalities near satellite offices for workers who work close to the headquarters. Since the reverse is true for workers employed farther away from headquarters, it is not clear whether these two estimation errors offset each other in the aggregate.

In the econometric analysis of human capital externalities, we merge the constructed spatial functions of high-skilled workers with micro-level data from the *Sample of Integrated Labour Market Biographies* (SIAB 7514).[6] The Sample of Integrated Labour Market Biographies is a 2% random sample of social security records. The dataset contains, among other data, information on wages, age, work experience and education with daily precision. To combine the individual-level data with the establishment-level data, we transform the individual-level dataset into a yearly panel with June 30 as the reference date and link workers and establishments with their unique establishment identifiers.

Because employers face legal sanctions for misreporting, information on wages in German social security data is generally highly reliable. However, one limitation is that roughly 10% of earnings are right-censored at the social security maximum. Therefore, we impute top-coded wages following Dustmann et al. (2009) and Card et al. (2013) (see section B.3 in the online appendix supporting this paper for details). Further, we improve the information on education following Fitzenberger et al. (2005) and restrict the sample to full-time workers aged between 18 and 64. As we are only interested in the effects on individuals in regular employment, we exclude apprentices, interns, marginally employed workers, and trainees. The final dataset consists of 3,498,536 observations from 539,179 individuals between 2000 and 2014.

To assign workplaces to local labor markets, we use the *de facto* standard definition of local labor market areas in Germany from the Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR). The goal in designating these local labor market areas is to design regions with strong internal commuter links but clear detachment from other areas. The construction is based on Kosfeld and Werner (2012), who use factor analysis of commuter flows to identify local labor market areas in Germany. The BBSR groups the 402 counties in Germany into 258 local labor market areas with an average radius of 21 km. The sizes of these local labor market areas correspond well to the findings of Manning and Petrongolo (2017), implying that 80% of the effects of local labor demand shocks are measurable within 20 km. As a rule of thumb, the authors further suggest that treatment areas for labor demand shocks should be 2.5 times the median commute. In our case, the rule of thumb suggests 24 km, which is close to the actual size of our local labor markets (Dauth and Haller, 2018, own calculations).

Local labor markets consist of multiple counties (*NUTS-3)*. For robustness checks, we complete our dataset with county-level indicators of the population density, unemployment rate, number of hotel beds (as a proxy for amenities and infrastructure), and net rents per square meter (as a proxy for housing cost) from the BBSR.

[3] One might be tempted to think that reversed causality also threatens identification within local labor markets. However, it does not seem plausible that high-skilled workers systematically sort into specific high-wage workplace areas. Instead, high-skilled workers plausibly sort into high-wage firms. Differently stated, it seems unlikely that workers choose firms based on the characteristics of all firms in a neighborhood, but instead, workers choose employers that match their preferences and skills. At the treatment level, such a sorting process would not materialize into the wages of neighboring firms and would thus not reverse the direction of causality. Note that since we are measuring externalities at the workplace, the sorting of individuals into residential neighborhoods containing similar individuals does not affect our estimates (see, e.g., Ananat et al., 2018). Furthermore, individual fixed effects would nullify such a bias.

[4] For a detailed description of the Establishment History Panel, see Schmucker et al. (2016).

[5] There are two types of universities in the German tertiary education system: traditional universities and universities of applied sciences (*Fachhochschulen*). Compared to traditional universities, universities of applied sciences focus more on practical topics. Universities of applied science usually also have a stronger focus on engineering and technology. Both kinds of universities award bachelor's and master's degrees.

[6] For a detailed description of the Sample of Integrated Labour Market Biographies, see Antoni et al. (2016).

3.2. Descriptive statistics

Fig. 1 provides an overview of the distribution of high-skilled workers in German establishments. For data-protection reasons, the map shows the share of high-skilled workers at the workplace in 1 × 1-km grid cells. Note that the data used in the econometric analysis are more precise and offer exact coordinates.

The map illustrates the considerable diversity in the distribution of high-skilled workers in Germany. For instance, among the largest cities, there are massive concentrations of high-skilled workers in Munich, Hamburg and Berlin. By contrast, Nuremberg and Bremen exhibit significantly lower shares of high-skilled workers. Moreover, apart from metropolitan areas, several other hot spots of skilled labor exist. For example, in Erlangen (15 km north of Nuremberg), Darmstadt (25 km south of Frankfurt) and Jena (70 km south east of Leipzig), over 30% of full-time workers hold a degree from a university or university of applied sciences. Moreover, the distribution of high-skilled firm neighborhoods also varies considerably within administrative regions. The upper-right panel of Fig. 1 shows a substantial cluster of high-skilled workers in the city center of Berlin. Additionally, there are several smaller clusters along the main traffic connections. The bottom-right panel focuses on the Rhein-Ruhr area. While high-skilled workplaces are evenly distributed in Essen and Dortmund, they appear to be very concentrated in the city centers of Düsseldorf, Cologne and Bonn. Numerous small hot spots also exist between the cities.

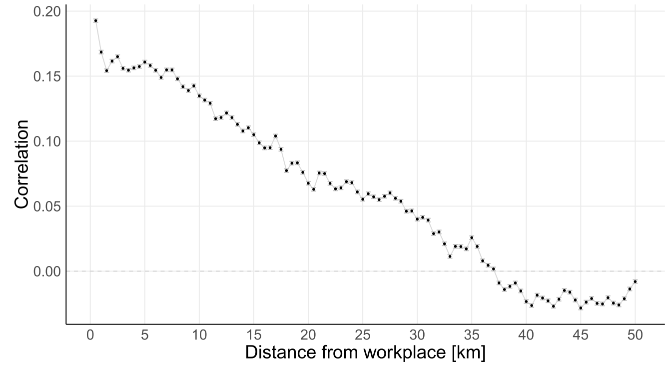

Fig. 2. Correlation of individual wages with the share of high-skilled workers around workplaces.
Notes: The figure illustrates the correlation between log wages and the shares of high-skilled workers around establishments within distance windows $[z_j - 500m, z_j]$, $z_j = 500m, 1000m, \ldots, 50000m$. The graph suggests that the correlation between individual earnings and the intensity of human capital attenuates with increasing distance. Note that the magnitude of the correlation coefficients cannot be interpreted directly.

To capture the heterogeneous distribution of high-skilled workers, we compute a spatial function that relates the share of high-skilled workers to the distance from each workplace in our data (see section B.4 in the online appendix for examples of such curves).

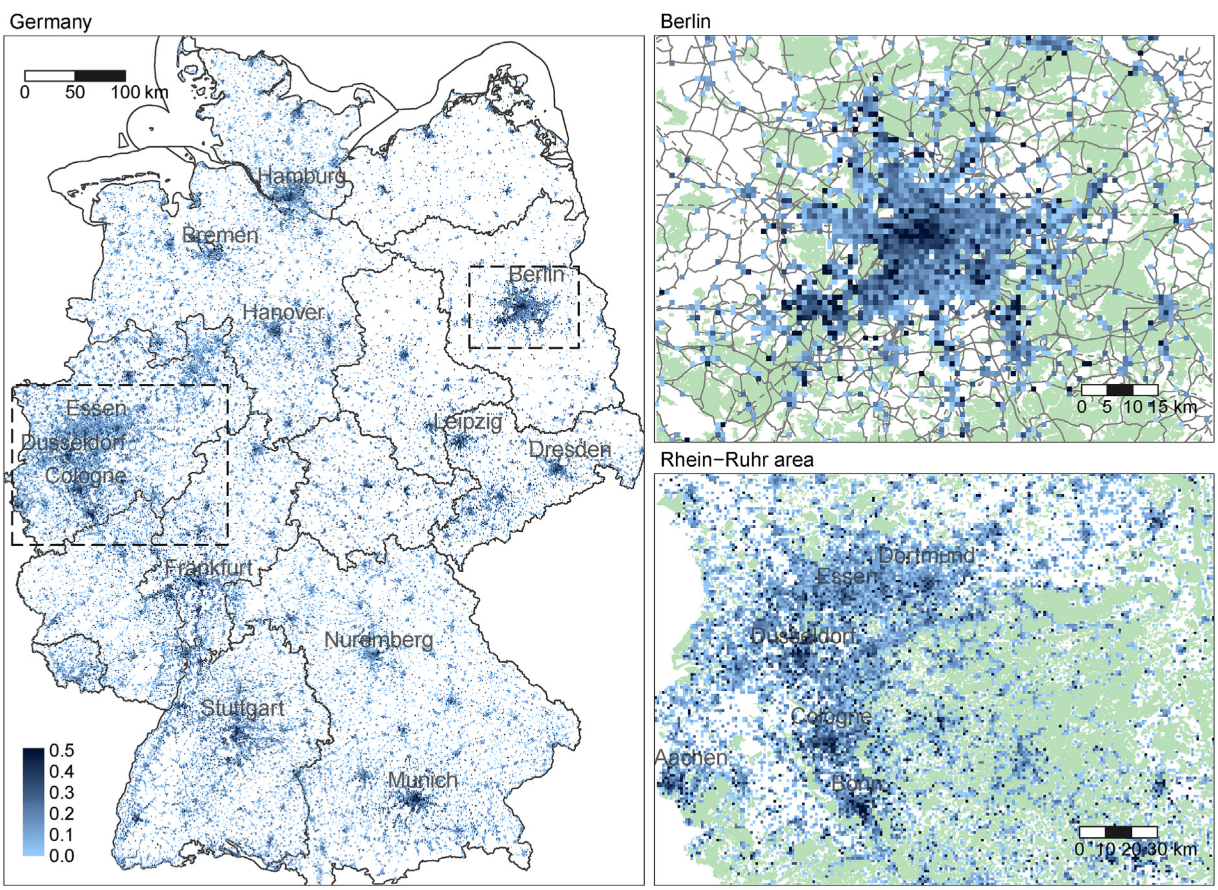

Fig. 1. Distribution of high-skilled workers in Germany.
Notes: The figure depicts the shares of high-skilled workers at the workplace in 1 × 1-kilometer grid cells in Germany (left panel), Berlin (upper-right panel), and the Rhein-Ruhr area (bottom-right panel) in 2014. For data-protection reasons, the maps depict aggregated data in grid cells. For the same reason, we have removed cells containing fewer than four establishments from the graphs. Note that the data used for our statistical analysis are more precise and provide the exact coordinates. Light-blue cells indicate low shares of high-skilled workers, and dark cells signal high shares (see the scale at the bottom left). For the sake of clarity, values are capped at 50%. In the left panel, black lines depict the boundaries of federal states. In the right panels, green areas depict forests, and in the upper-right panel, gray lines and dashed gray lines illustrate streets and railways, respectively.

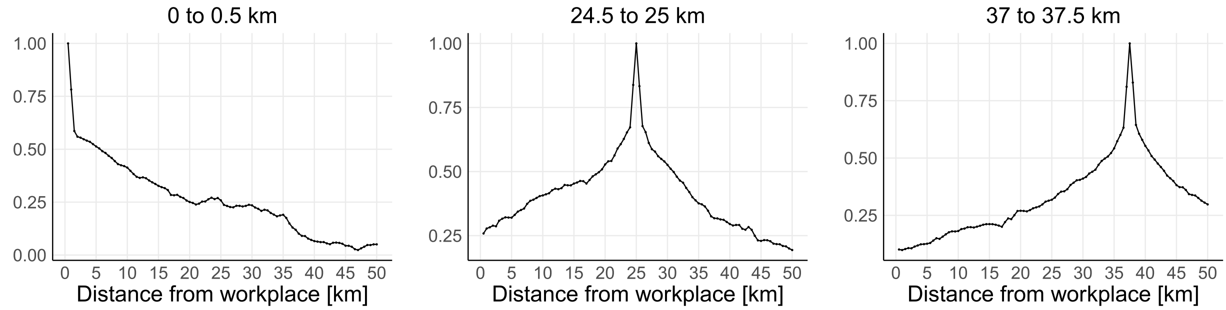

Fig. 3. Spatial autocorrelations at selected measurement points.
Notes: The graphs show the spatial autocorrelations of the spatial functions of high-skilled workers at different measurement points. For instance, the panel in the middle shows the correlation of the share of high-skilled workers 24.5 to 25 kilometers away from workplaces with the share of high-skilled workers at the other 99 measurement points. The focal points in the remaining two panels are 0 to 0.5 and 37 to 37.5 kilometers, respectively. As is typical with functional data, values close to the focal points have high correlations. The correlation declines with increasing distance from the focal point. Note that the three selected focal points well illustrate the general pattern of the underlying three-dimensional correlation function.

To obtain a first impression of the relationship between individual earnings and the spatial concentration of human capital, Fig. 2 shows the correlation between log wages and the share of high-skilled workers around establishments within distance windows $[z_j - 500\ m, z_j), z_j = 500\ m, 1000\ m, \ldots, 50,000\ m$. While the magnitude of the *ordinary* correlation has no direct interpretation, the declining trend signals that the relationship between income and the spatial concentration of high-skilled labor decays with increasing distance.

One reason that the magnitude of the correlation coefficients has no direct interpretation is that the functions expressing the shares of high-skilled workers are spatially autocorrelated. Fig. 3 illustrates this issue. The graph depicts the correlation between the share of high-skilled workers in three selected distance windows with the remaining 99 measurement points. For instance, the first panel presents the correlation of the share of high-skilled workers between measurement point t_1 and the random curve's value at $t_2, \ldots, t_{100}$. As the figure shows, adjacent values have very high correlations compared to more distant measurement points. In principle, it is possible to use all grid values of the functional predictor as regressors to measure the partial effect. However, as shown in the next section, the strong correlations between adjacent measurement points lead to a multicollinearity problem. As a consequence, the effects can no longer be measured.

For additional summary statistics on individual wages and other covariates in our dataset, we refer the reader to appendix A.

4. Results

Our main results show that spillover effects from the local concentration of high-skilled workers around establishments significantly increase individual wages. The spillover effects decay with increasing distance, and the point estimates suggest that after 10 km, the effects are reduced by half. Beyond 25 km, the effects are no longer distinguishable from zero. In the following, we present the estimation results and discuss our findings. Additionally, we present spillover effects differentiated by various subgroups.

4.1. Main findings

We illustrate estimates of the spatial intensity of human capital externalities from high-skilled workers in Figs. 4 and 5. Fig. 4 depicts an unrestricted estimate of equation (5) (i.e., setting $\rho = 0$ when solving (2)), which coincides with standard OLS regression.[7] Fig. 5 presents

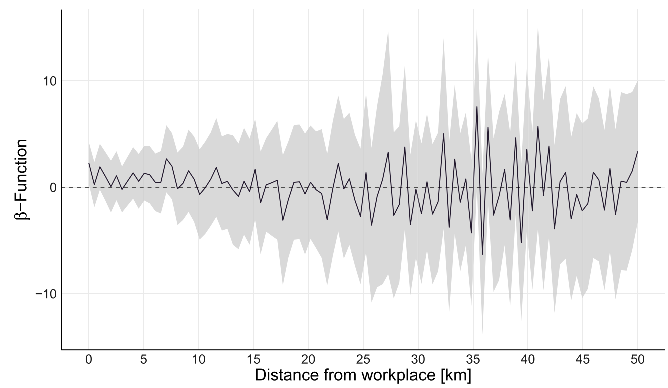

Fig. 4. Unrestricted estimates of spatial human capital externalities from high-skilled workers.
Notes: The figure presents an unrestricted estimation of spatial human capital externalities from high-skilled workers affecting individual log wages (equation (5)). We measure the concentration of high-skilled workers as the share of high-skilled workers within distance z. The black line illustrates the estimated spillover function ($\beta(z)$), and the gray area indicates the 99% confidence band. The unrestricted estimator coincides with the standard OLS estimator. Due to multicollinearity and overfitting, the estimator cannot retrieve valid estimates of $\beta(z)$ from the data. The underlying model controls for worker fixed effects, skill-specific yearly labor-market-area fixed effects, occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). $N = 3,498,536$.

penalized estimates of equation (5) (i.e., $\rho > 0$). Both estimates control for supply effects and endogenous sorting of individuals with an extensive set of fixed effects. In addition to standard controls from the labor literature, our models include skill-specific yearly local labor market fixed effects. In the graphs, black lines display the estimated spillover functions. The gray area indicates the associated 99% confidence band.

As Fig. 4 shows, the unpenalized estimate of equation (5) identifies no significant link between the spatial concentration of high-skilled workers and individual earnings. The point estimates are unstable, and the confidence bands include the null over the whole domain. There are two reasons for the unstable behavior of the curve. First, as described in the previous section, the measurement points of the share of high-skilled workers are highly correlated. Because the unrestricted estimator is (up to a scale) identical to the standard OLS estimator, high correlations among a large set of regressors pose multicollinearity problems. Consequently, the estimates exhibit high variance. Second, an unrestricted estimator allows the computation of unnecessarily complex functions and is therefore potentially prone to overfitting the data due to modeling noise.

[7] Note that OLS estimates of equation (5) would be scaled by the number of discretization points in $X_{it}(z)$. By contrast, our main estimates provide an approximation via a Riemann sum and are thus scaled such that the number of discretization points does not affect the scaling of the (discretized) β.

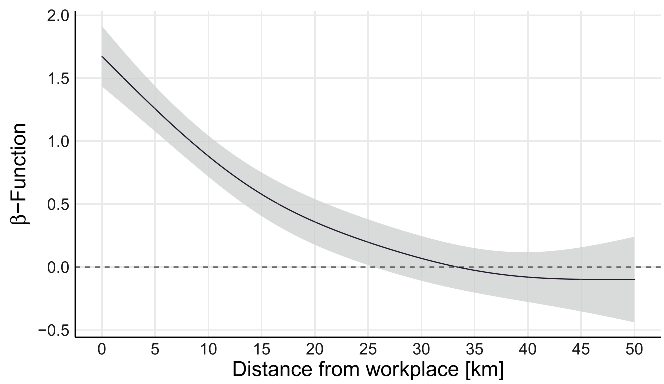

Fig. 5. Spatial human capital externalities from high-skilled workers.
Notes: The figure shows spatial human capital externalities from high-skilled workers affecting individual log wages. We measure the concentration of high-skilled workers as the share of high-skilled workers z units away from individual workplaces. To compute the spatial spillover function ($\beta(z)$) we estimate equation (5) with the smoothing spline estimator. The black line illustrates the estimated spillover function ($\beta(z)$), and the gray area indicates the 99% confidence band. The graph shows significant spillover effects that decay with increasing distance. The effect of a p-percentage-point increase in the share of high-skilled workers within distance z_0 and z_1 (in a 0 to 1 range) is p times the area below the estimated spillover function from z_0 to z_1. For instance, a 20-percentage-point increase in the concentration of high-skilled workers within 5 kilometers ($z_0 = 0, z_1 = \frac{5}{50}$) leads to a wage gain of 2.95%. The underlying model controls for worker fixed effects, skill-specific yearly labor-market-area fixed effects, occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). $N = 3,498,536$.

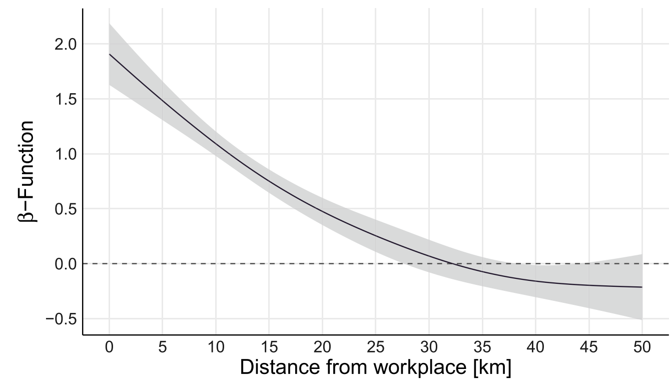

Fig. 6. Spurious estimates of spatial human capital externalities from high-skilled workers.
Notes: The figure presents estimates of the spatial human capital externalities from high-skilled workers affecting individual log wages without nullifying supply effects that stem from the imperfect substitution of high- and low-skilled workers. Specifically, the graph depicts estimates of the spatial spillover function ($\beta(z)$) from equation (5) without skill-specific yearly labor-market-area fixed effects (π_{rst}). The black line illustrates the estimated spillover function ($\beta(z)$), and the light-gray area indicates the 99% confidence band. The underlying model controls for worker effects, skill-specific yearly labor-market-area fixed effects, occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). $N = 3,498,536$

By contrast, the penalized estimates in Fig. 5 reveal a clear influence of the spatial concentration of high-skilled workers around establishments on individual wages. The spillover effects decay with increasing distance and vanish after approximately 25 km. The magnitude of the effects from direct neighbors of establishments is roughly twice as large as effects from high-skilled workers located 10 km away.[8] In the graph, the effect of a p-percentage-point increase in the share of high-skilled workers within distance z_j and $z_{j'}$ (in a 0 to 1 range) is p times the area below the estimated spillover function from z_j to $z_{j'}$. For instance, a 20-percentage-point increase in the concentration of high-skilled workers within 5 km leads to wage gains of 2.95% ($\approx 20 \times \{1.25 \times \frac{5}{50} + \frac{1}{2}\left[(1.7 - 1.25) \times \frac{5}{50}\right]\}$). A spatially evenly distributed ten-percentage-point (one standard deviation) increase in the share of high-skilled workers over the whole domain increases individual wages by 4.25% ($\approx 10 \times \frac{1}{2}\left(1.7 \times \frac{25}{50}\right)$).

Reassuringly, *classical* estimates at an aggregate level, wherein we use OLS to model the wage effect of the share of high-skilled workers within counties and identical covariates as in equation (5), suggest effects of the same magnitude (see section B.6 of the online appendix supporting this article). Our results are also similar to the findings of Rosenthal and Strange (2008) for the US, in which the authors regress wages on the number of workers with a college degree or higher education within a 5-mile distance and within 5- to 25-mile distances. They report that spillovers from high-skilled workers within a 5-mile distance are up to 3.5 times larger than spillovers from high-skilled workers 5 to 25 miles away. Averaging our estimates within the same distance windows yields a ratio of 2.5. Although we follow a different estimation approach with different data, our findings seem consistent with those of Rosenthal and Strange (2008).

If we compare our results to findings from studies that analyze human capital externalities on administrative levels, our estimates are at the lower end of the range. The reason is likely our demanding battery of fixed effects, which lead to rather conservative estimates. Our results imply that an evenly distributed one-percentage-point increase in the share of high-skilled workers increases the wages of other workers by 0.4%. Based on data from US metropolitan regions, Moretti (2004) found that a one-percentage-point increase in the share of college graduates increases wages by between 0.4% and 1.2%.[9] Using Russian survey data, Muravyev (2008) found a 1.5%-response of wages. Based on similar data to those used in our analysis, Heuermann (2011) estimated a wage reaction of 1.8% for highly qualified workers and 0.6% for other workers in Germany.

Let us now briefly discuss the importance of removing supply effects when estimating human capital externalities. Fig. 6 reports estimates of our model (equation (5)) without skill-specific yearly labor-market-area fixed effects (π_{rst}) and thus includes supply effects that stem from the imperfect substitution of high- and low-skilled labor (see Moretti, 2004; Ciccone and Peri, 2006). Compared to our main findings, the estimated relationship between individual wages and the concentration of high-skilled workers appears stronger in these estimates. Specifically, a global upward shift of the estimated $\beta(z)$ by roughly a factor of 1.2 is observed. Although π_{rst} also nullifies other confounders (e.g., temporal effects from the sorting of high-skilled workers), the uniform upward shift of $\beta(z)$ corresponds well to Ciccone and Peri (2006). The authors also find that the bias from supply effects in Mincerian estimates of human capital externalities are significant.

4.2. Heterogeneities of the results

Our rich dataset allows us to analyze human capital externalities by subgroups. To this end, we interact the curves of high-skilled workers around establishments with corresponding group identifiers and estimate extended versions of model (5). In particular, we differentiate spillover effects by region type, skill, sector, age, and gender.

[8] A table containing results of the full curve can be found in the online appendix B.5.

[9] Note that the following selection only includes studies with comparable measurements of human capital. Clearly, there are other important contributions using, for instance, average years of schooling to measure human capital.

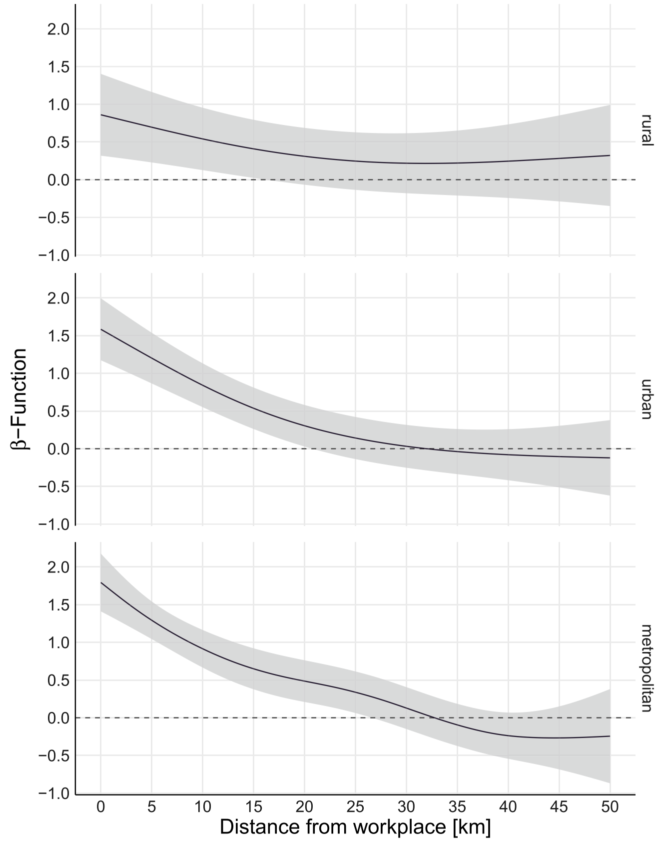

Fig. 7. Spatial human capital externalities by region type.
Notes: The figure shows spatial human capital externalities from high-skilled workers affecting individual log wages for workers working in different types of regions. The black lines illustrate the estimated spillover function ($\beta(z)$) for each skill group, and the gray area indicates the 99% confidence band. The underlying model controls for worker fixed effects, skill-specific yearly labor-market-area fixed effects, occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). The subgroups have sizes $N = 1,947,436$ (metropolitan), $N = 1,163,916$ (urban), and $N = 387,184$ (rural).

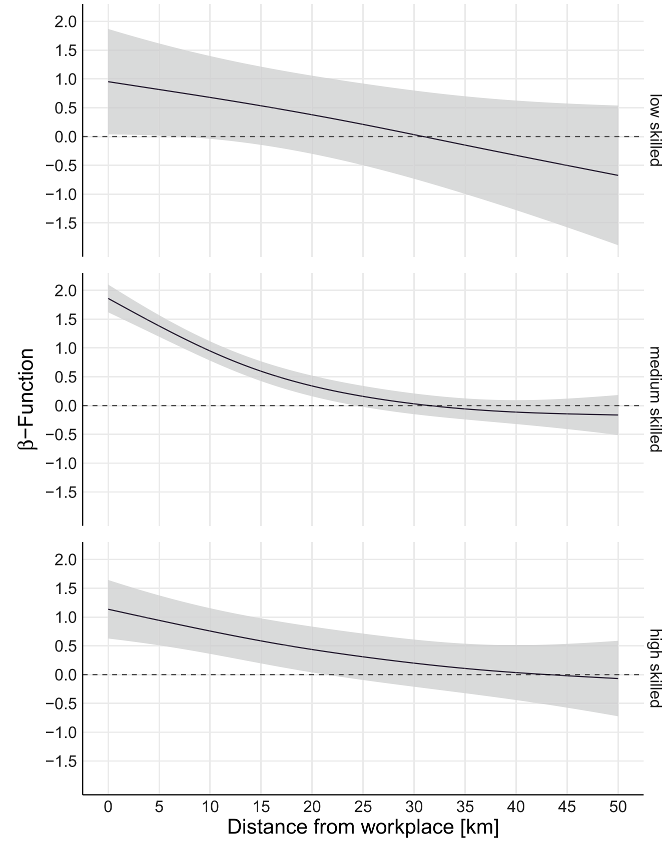

Fig. 8. Spatial human capital externalities by skill groups.
Notes: The figure shows spatial human capital externalities from high-skilled workers affecting the individual log wages of different skill groups. The black lines illustrate the estimated spillover function ($\beta(z)$) for each skill group, and the gray area indicates the 99% confidence band. The underlying model controls for worker fixed effects, skill-specific yearly labor-market-area fixed effects, occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). The subgroups have sizes $N = 281,993$ (low-skilled), $N = 2,552,942$ (medium-skilled), and $N = 663,661$ (high-skilled).

4.2.1. Region type

Plausibly, an urban infrastructure facilitates knowledge exchange. Therefore, we expect stronger externalities from high-skilled workers in urban than in rural areas. To examine whether the degree of urbanization affects human capital externalities, we incorporate the interactions of indicators, provided by the BBSR, for metropolitan, urban, and rural counties with the concentration of high-skilled workers into our main specification.[10]

Fig. 7 displays the results. In line with our expectations, human capital externalities are the strongest in metropolises, followed by urban counties. Spillover effects in rural counties are comparably small. The same pattern holds for the spatial decay of human capital externalities. In metropolitan counties, human capital externalities reach furthest. Their decay is strongest in rural counties. Specifically, our estimates suggest that human capital externalities are almost twice as large in metropolises as in rural counties. These findings suggest that interactions between workers that lead to human capital externalities might be more frequent and less costly in metropolitan and urban than those in rural areas.

4.2.2. Skill

Fig. 8 presents estimates by skill groups. The panel at the top depicts human capital externalities for workers without vocational training (low-skilled), the panel in the middle presents effects on workers who have completed vocational training (medium-skilled), and the bottom panel displays spillover effects for workers with a degree from a university or university of applied science (high-skilled). Generally, the estimates are in line with our main findings and show that human capital externalities attenuate with increasing distance.

Moreover, the results indicate that the size of human capital externalities differs by skill group. Effects on low-skilled workers are statistically insignificant over most of the domain. Effects on medium-skilled workers are statistically highly significant and considerably larger than effects on low-skilled workers. Additionally, effects on high-skilled workers are statistically and economically significant, albeit smaller than the effects on medium-skilled workers. The gap between medium- and high-skilled workers is largest near establishments.

Our finding that medium-skilled workers benefit the most from human capital externalities might imply that knowledge exchange from the better-educated to the less-educated – in a teacher-student fashion – explains a considerable proportion of human capital externalities. This type of knowledge transmission may be particularly pronounced in Germany, where the level of education between medium- and high-skilled workers does not greatly differ due to the well-regarded dual education system.

[10] In total, we have 402 counties which can be devided into 136 metropolitan, 17 urban, and 179 rural counties. Information on the region type is aggregated on the labor market level. Consequently, the county type corresponds to the labor market region type.

In contrast, the permeability to low-skilled workers, which comprises 8% of our observations, seems to be less pronounced. Presumably, there is considerably less exchange between high-skilled and low-skilled workers than between high-skilled and medium-skilled workers.[11]

The results of previous studies regarding the effect of human capital externalities on workers' wages by skill groups are ambiguous. For the US, Moretti (2004) finds that a one-percentage-point increase in the regional concentration of college graduates increases the wages of high school drop-outs, high school graduates and college graduates by 1.9%, 1.6% and 0.4%, respectively. Thus, our results are in line with Moretti's (2004) findings regarding the relative size of effects on medium- and high-skilled workers but deviate from his findings concerning low-skilled workers. Apart from institutional differences between the US and Germany, this deviation might be explained by supply effects. As Moretti (2004) illustrates, his estimates include not only human capital externalities but also labor supply effects. While supply effects increase the wages of non-highly skilled workers due to their relative scarcity, they decrease the wages of high-skilled workers. Consequently, Moretti's (2004) estimates of human capital externalities might be too large for low-skilled workers and too small for high-skilled workers. In contrast, we try to disentangle human capital externalities form supply effects by removing constant variation within local labor markets from the data.

For Germany, Heuermann (2011) reports that external wage effects from an increase in the regional share of high-skilled workers are stronger for high-skilled workers than for non-high-skilled workers (medium-skilled and low-skilled workers combined). If we follow Heuermann's (2011) definition, we find stronger effects for non-high-skilled than for high-skilled workers. A reason for this difference might be the period under consideration. Heuermann (2011) investigates external wage effects with data from 1995 to 2002. During this observation period, the German labor market can be characterized as rather rigid and dominated by collective wage agreements. In contrast, during our observation period (2000–2014), the German labor market experienced fundamental structural reforms that increased the labor market flexibility considerably. Additionally, the coverage of collective wage bargaining agreements decreased rapidly in all sectors. Both changes affected low-skilled workers the most.

4.2.3. Sector

To investigate whether spillover effects differ by sector, we expand our main specification with the interaction between the share of high-skilled workers around establishments and the indicators of individuals in manufacturing and the service industry. The results (see figure B4 in the online appendix) show significant spillover effects from the local concentration of high-skilled workers on both sectors. As in our baseline specification, human capital externalities decrease with increasing distance and disappear after roughly 25 km. Additionally, we find slightly larger effects on workers in manufacturing than on workers in the service sector. For instance, a ten-percentage-point increase in the share of high-skilled workers over the whole domain increases individual wages of workers in manufacturing by 4.4%, while workers in the service industry only gain a 3.7% increase.

The stronger impact of human capital externalities in the manufacturing industry confirms the results of Heuermann (2011), who also found larger spillover effects in manufacturing industries. One possible explanation of this pattern might be the importance of physical capital in the manufacturing sector. Acemoglu (1998) argues that high-skilled workers facilitate the adoption of new technologies. For manufacturing firms, this often facilitates investments in physical capital, that potentially boosts the productivity of a broad set of workers.

4.2.4. Age

We also investigate differences in human capital spillover by age groups (see figure B5 in the online appendix). Thus, we interact the predictor curves with a dummy variable that takes the value of one if the worker is younger than 40 years old. Our estimates show that for young workers, spillover effects are slightly larger near the workplace and are more widespread than those for older workers. Plausibly, young workers interact more often. Additionally, they might use modes of communication that reach further in space. Moreover, standard human capital theory implies that, due to lower work experience, learning on the job and informal learning might be more pronounced for younger workers.

4.2.5. Gender

In our last investigation of heterogeneous effects, we explore human capital externalities by gender (figure B6 in the online appendix). In line with our baseline model, human capital externalities are quite pronounced and decay with increasing distance for both men and women. Interestingly, spillover effects are larger for female workers than for male workers but are less far-reaching. According to our data, women are slightly younger and more often medium-skilled, which might explain the higher spillover effects in close proximity to the workplace for this group. A further explanation could be that women more often interact close to the workplace due to family responsibilities (e.g., while picking up children from school). Differences in commuting behavior might explain the faster decay of human capital externalities, as commutes to the workplace are, on average, shorter for women than for men (Dauth and Haller, 2018).

5. Simulation study

From a theoretical perspective, drawing local inference about the slope parameter β in a regression model with a functional predictor is a difficult issue (see the online appendix B.1 for a discussion). Therefore, the following Monte-Carlo simulation exercise evaluates the statistical properties of our estimation framework. The results show that although our estimation framework yields locally biased estimates, it is reliable in the sense that it reproduces the structure of the true curve well. Additionally, we show that the inference procedure controls size when the true functional coefficient is linear under the null hypothesis.

In the simulation study, we consider four scenarios. First, we evaluate the estimator's properties in a case in which the data-generating process (DGP) resembles our particular real-world problem. For this purpose, we select the DGP from the preferred estimate (Fig. 5). Additionally, we incorporate parameter estimates from all covariates and generate artificial observations of the dependent variable based on iid errors drawn from $N(0, \hat{\sigma}_u^2)$. Here, $\hat{\sigma}_u$ denotes the standard error of the residuals of the estimated model. Therefore, the structure of the simulated dataset (e.g., the sample size, number of establishments, and number of workers per establishment) is the same as that of the original sample. The remaining three scenarios assess the statistical properties of the estimator in different extreme situations. Here, we simulate data with a similar structure to that of the real dataset. In particular, we replicate the first two moments of the original data. The second and third scenarios evaluate the accuracy of the inference procedure when the null is the zero function or a linear function. The fourth and most extreme setting analyzes the performance of the estimator when the true parameter is a non-smooth step function. To assess the statistical properties of the estimator, we simulate 1000 replications in each scenario.

Fig. 9 summarizes the results of the four simulations. In each panel,

[11] While 33% of low-skilled workers are foreigners, this is only true for 5% of medium- and high-skilled workers. Thus, language barriers might further impede the transmission of human capital from high-to low-skilled workers. Contrarily, 67% of foreign low-skilled workers, compared to 58% of German low-skilled workers work, work in metropolitan areas. This might increase the strength of human capital externalities for foreign low-skilled workers.

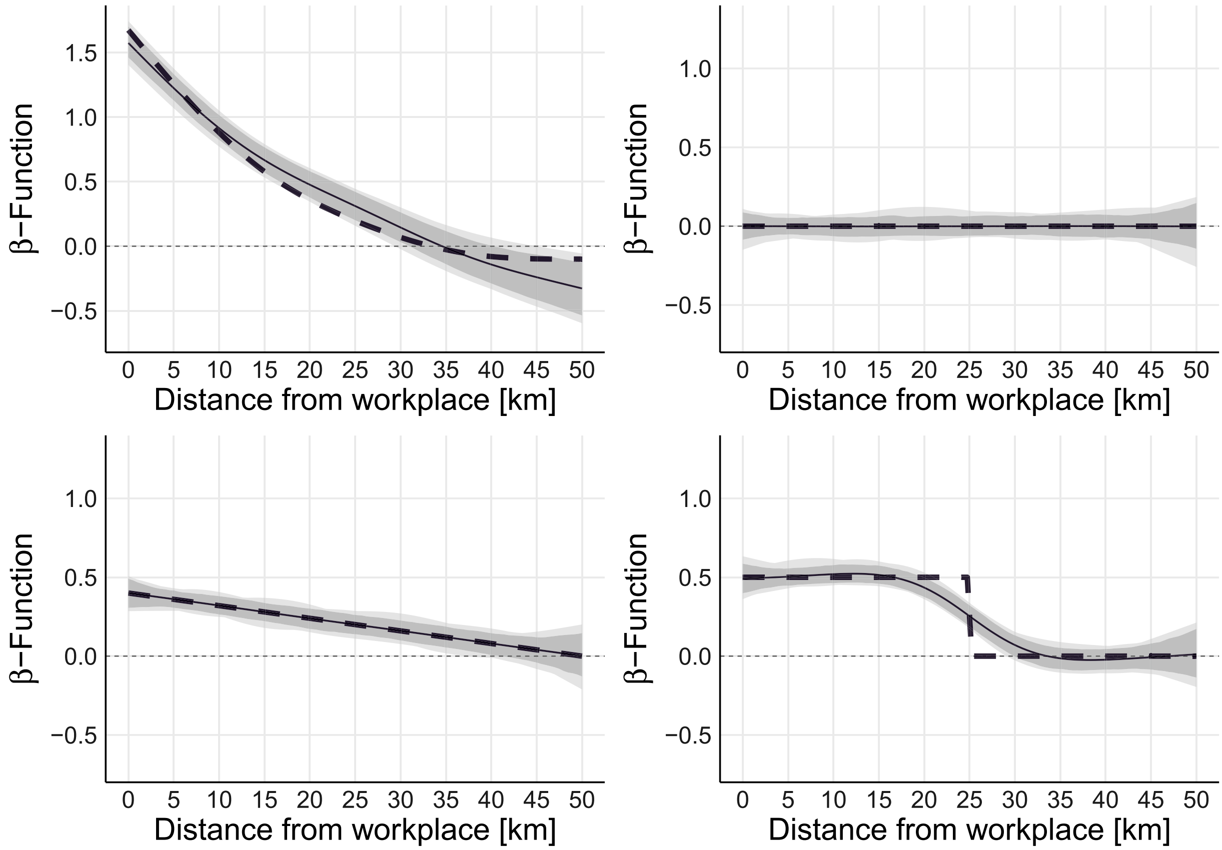

Fig. 9. Performance of the estimator in different simulations.
Notes: The figure shows four Monte-Carlo simulations. The bold dashed line depicts the true parameter function $\beta_0(z)$, the light-gray areas show pointwise minimum and maximum values of all estimates, and the dark-gray areas show the first and 99th percentiles of all estimates of the parameter function. The solid line represents the pointwise mean over all replications. Simulated replications of the estimator are obtained by estimating model (5) based on simulated data. The model setup corresponding to the top-left panel uses the predictors from the real-data application, and observations of the dependent variable are simulated based on estimated coefficients and iid normally distributed errors. All other model setups are based solely on simulated data that mimic the original sample but use different specifications for the functional parameter $\beta(z)$. In the top-right panel, $\beta(z) = 0$; in the bottom-left panel, $\beta(z) = 0.4(1 - z)$; and in the bottom-right panel, $\beta(z) = 0.5 \cdot \mathbb{1}(z < 0.5)$.

Table 1
Performance measurements in different simulations.

	Specification for β_0			
	I	II	III	IV
Integrated squared bias	0.0096	0.0000	0.0000	0.0055
Integrated variance	0.0030	0.0009	0.0009	0.0010
Coverage probability of 99%-CIs	0.7290	0.9920	0.9930	0.0000

Notes: The table contains the integrated variance, integrated squared bias and coverage probability of the confidence bands of the parameter estimates for the functional coefficient from all four model setups considered in the simulation exercise. In the first model setup, the data are generated based on the regressors and functional predictors with corresponding coefficients taken from the original estimate. The other model setups are based solely on simulated data with similar characteristics to those of the original data. In setup II, the functional coefficient of the DGP is zero, and in setup III, it is a linear function. The coefficient in the last setup (column IV) is discontinuous and possesses a discrete jump in the interior of its domain. We compute integrated variance as $1000^{-1} \int \sum_{r=1}^{1000} \left(\hat{\beta}_r(z) - \overline{\beta}(z)\right)^2 \mathrm{d}z$ and integrated squared bias as $\int \left(\overline{\beta}_r(z) - \beta_0(z)\right)^2 \mathrm{d}z$, where $\overline{\beta}(z) = 1000^{-1} \sum_{r=1}^{1000} \hat{\beta}_r(z)$.

the bold dashed line depicts the true parameter function $\beta_0(z)$ of the DGP, the light-gray areas show the pointwise minimum and maximum values of all estimates, and the dark-gray areas show the first and 99th percentiles of all estimates of the parameter function. The solid line represents the pointwise mean over all replications. In general, the estimates follow the true parameter function well, and no replication deviates substantially from the DGP. However, as is typical for penalized (or nonparametric) models, the estimates deviate from the true curve in regions with complex structures (i.e., in regions with strong nonlinearity). In such regions, the estimator possesses a local bias. As one might expect, this behavior is especially pronounced at the jump discontinuity of the step function in the bottom-right panel of Fig. 9. By construction, however, the smoothing splines estimator never produces estimates different from zero in regions where the true curve is zero in a larger neighborhood. Therefore, if the underlying functional shape of the spatial decay of human capital externalities is monotonically decreasing and zero beyond a certain distance, the regularized estimation captures the true curve well. This assumption appears to be reasonable for our application.

Table 1 provides the integrated squared bias, integrated variance, and the coverage probability of the confidence bands for each scenario. The integrated (squared) bias is largest for the setup in which the function-valued parameter is taken from the real-data application because the true parameter is curved over the whole domain (column 1). Similarly, the variance is the largest in this model setup. The two scenarios with linear parameter functions based on the construction of the estimator show favorable properties and exhibit the lowest variance and no bias (columns 2 and 3). In this situation, confidence bands computed with equation (B.6) of the online appendix have proper coverage probability; however, this coverage probability does not longer hold for more complex parameter functions. In the most extreme case (discontinuous β_0), the bias at the jump discontinuity is so large that the confidence bands are unable to cover the true parameter over the whole domain (column 4).

The implications from the simulation study for our main findings are as follows. If the true spatial decay of human capital externalities is not too complex, our estimates and confidence bands are generally reliable. However, because the estimator is locally biased in regions with a more complex β_0, identifying the exact distance at which human capital externalities cease is difficult. A conservative strategy would be to choose a threshold somewhat lower than that indicated by the confidence bands. Regarding our main findings, such a strategy suggests that human capital externalities might already be statistically insignificant after 22–25 km.

6. Semi-parametric OLS estimates with broader rings

The previous literature that measures the spatial attenuation of economic effects in various contexts uses a semi-parametric framework, in which the main explanatory variable is measured in a series of concentric rings or circles. The outcome variable is then regressed on the series of measurements (e.g., Rosenthal and Strange, 2008; Fu, 2007; Verstraten, 2018; Gibbons et al., 2021; Faggio et al., 2019; Faggio, 2019). The beauty of the semi-parametric framework is that it is a straightforward application of the linear OLS model and, in principle, can be applied to any geographical data. The drawback of the semi-parametric framework compared to our FDA approach is that estimates of the spatial attenuation of economic effects are less precise. The reason for this difference is that multicollinearity issues (usually) do not allow the estimation of economic effects from a large or fine-graded series of measurements. To circumvent multicollinearity issues, researchers construct relatively broad rings or circles that measure the spatial distribution of the explanatory variable. We corroborate our main findings by applying the semi-parametric framework to our research question. Specifically, we estimate the effects from the shares of high-skilled workers within 0–1, 1–5, 5–10, 10–25 and 25–50 km from workplaces on log wages using OLS. Albeit less precise, the estimated economic effects are of similar magnitude to our main findings and support our procedure.

Before explaining the corresponding econometric specification, let us briefly discuss the properties of the semi-parametric approach by means of a small simulation exercise. To this end, we generate 1000 replications of the DGP (1) using predictors resembling the first and second moments of our real data application. The functional coefficient β_0 corresponds to the dashed line of Fig. 10. Then, we compute the averages of the simulated curves with respect to larger intervals of the domain.[12] We obtain the spillover parameters by regressing the (simulated) dependent variable on these averages and normalizing the respective coefficient by the ring width. The widths of the rings are consistent with the specifications of the semi-parametric regression.

In Fig. 10 we illustrate the results of the simulation study. The coefficient function of the DGP is depicted by the dashed line, and the vertical solid lines indicate the boundaries of the rings used in our specification. The gray areas illustrate the first and 99th percentiles of all replications, and the horizontal black lines represent the mean over all replications. In general, the results show that the approximation obtained via a Riemann sum performs quite well. However, the outcome heavily depends on the bandwidths of the rings. In addition, the semi-parametric approach does not facilitate learning how the coefficient function behaves inside the intervals from the data.

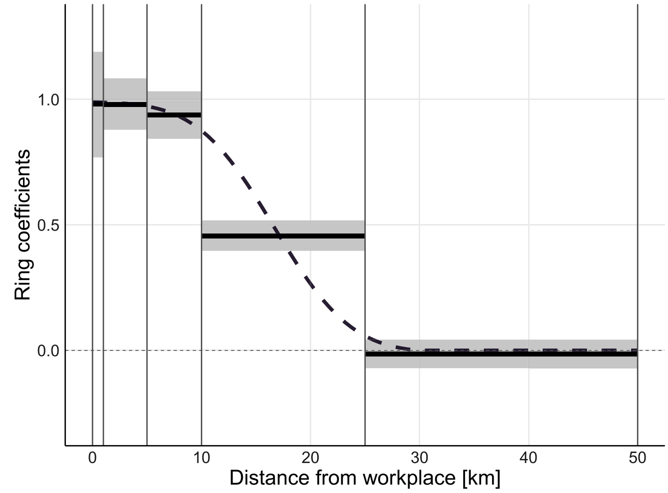

Fig. 10. Simulation results of semi-parametric OLS estimates.
Notes: The figure shows a Monte-Carlo simulation of the semi-parametric OLS estimation. The bold dashed line depicts the true parameter function $\beta_0(z)$. The vertical solid lines depict the boundaries of the rings, and the horizontal black lines illustrate the mean over all replications of the approximation of the functional coefficient via a Riemann sum. The gray areas reflect the range between the 1st and 99th percentiles of all estimated coefficients of the Riemann sum. The Riemann sum coefficients are obtained by dividing the raw regression coefficients of the aggregated rings by the ring width. Simulated replications were obtained by estimating equation (6) based on data generated by DGP (1) with the same predictors used in the Monte-Carlo exercise described in section 5.

Let us now compare our estimates to those obtained by the semi-parametric approach. To this end, we estimate the following model:

$$\begin{aligned} Y_{it} &= \alpha_1 x_{1\mathrm{km},it} + \alpha_2 x_{5\mathrm{km},it} + \alpha_3 x_{10\mathrm{km},it} + \alpha_4 x_{25\mathrm{km},it} + \alpha_5 x_{50\mathrm{km},it} \\ &\quad + Z_{it}'\gamma + \delta_i + \tau_t + \omega_o + \pi_{rst} + u_{it}. \end{aligned} \tag{6}$$

Here, Y_{it} is the individual log wage of worker i in year t. $x_{1\mathrm{km}}$ is the share of high-skilled workers within a 0–1 km distance from i's workplace, $x_{5\mathrm{km}}$ is the share of high-skilled workers within a 1–5 km distance from i's workplace, $x_{10\mathrm{km}}$ is the share of high-skilled workers within a 5–10 km distance from i's workplace, etc. Accordingly, $\alpha_1, \ldots, \alpha_5$ are the spillover coefficients that we seek to estimate. In line with our main model, we control for time-varying observable characteristics of individuals, establishments and regions (Z_{it}) and a series of fixed effects. δ_i are individual fixed effects, τ_t is a year fixed effect and ω_o is an occupation fixed effect. Finally, we also include dummy variables for labor market areas in each time period (π_{rst}).

Table 2 summarizes the results. Columns 1 and 2 of Table 2 show the strengths of human capital externalities from five different distances (i.e., 0–1 km, 1–5 km, 5–10 km, 10–25 km and 25–50 km). The estimates in the first column are obtained without controlling for supply effects (π_{rst}), while the estimates in the second column are controlled for supply effects. The estimates of human capital externalities are statistically significant through the ring covering a distance of 10–25 km.

Due to the use of different bandwidths, we cannot directly compare the magnitudes of the raw estimates. As an illustration, consider that the parameter estimate in the first ring measures wage effects from a one-percentage-point increase in the share of high-skilled workers within 1 km around individuals. The parameter estimate in the second ring expresses the effects of a one-percentage-point increase within one to 5 km. Both estimates implicitly assume that the one-percentage-point increase in the share of high-skilled workers is uniformly distributed within each band area (i.e., the share of high-skilled workers increases by one-percentage-point within each kilometer distance). Thus, by construction, the second ring captures a treatment five times

[12] By aggregating the curves in such a manner, the resulting *rings* no longer reflect the shares of high-skilled workers within each ring but indicate a weighted average in which, assuming a uniformly populated area, more central observations receive larger weights compared to more distant observations in each ring. In our real data application, we are, of course, able to compute the shares of high-skilled workers within the distance windows.

Table 2
Semi-parametric OLS estimates with broader rings.

	raw		per km	
	(1)	**(2)**	**(3)**	**(4)**
Share of high-shilled workers in ...				
0–1 km	0.071***	0.058***	0.071***	0.058***
	(0.008)	(0.005)		
1–5 km	0.100***	0.086***	0.025***	0.022***
	(0.011)	(0.010)		
5–10 km	0.146***	0.103***	0.029***	0.021***
	(0.012)	(0.012)		
10–25 km	0.154***	0.070***	0.010***	0.005***
	(0.014)	(0.020)		
25–50 km	−0.035	−0.016	−0.001	−0.001
	(0.021)	(0.026)		
Worker fixed effects	Yes	Yes	Yes	Yes
LLM × year × skill fixed effects	No	Yes	No	Yes

Notes: The table summarizes the estimates of the human capital externalities from high-skilled workers in broad concentric rings on individual log wages. The estimates replicate our main model in a less precise manner and provide a comparison of the magnitude of the effects. The first two columns show raw coefficient estimates. Columns three and four show estimated effects within 1-km bands. The underlying models also control for occupation and time fixed effects, and worker characteristics (age, work experience, tenure, and the respective second-order polynomials). Cluster-robust standard errors are shown in parentheses. *** indicates significance at the 0.1%-level. $N = 3,498,536$.

stronger than that of the first ring. To make the parameter estimates comparable across rings, we therefore divide the raw estimates by their underlying bandwidths in columns 3 and 4. The results show the effect of a one-percentage-point increase in the share of high-skilled workers within 1 km within the corresponding bandwidth.

In line with our main findings, columns 3 and 4 show that human capital externalities decay with increasing distance. Also similar to our main findings, human capital externalities lose their economic significance at distances greater than 25 km. Additionally, the magnitudes of the estimated effects are similar to those from our main model. For instance, according to our main model, a 20-percentage-point increase in the share of high-skilled workers within 5 km leads to wage gains of 2.95%. According to our semi-parametric estimates with broader rings, the same increase in the share of high-skilled workers raises wages by 2.88%. The difference between the two estimates is minor. In summary, the semi-parametric estimates buttress our main findings.

7. Robustness checks

We apply several robustness checks to corroborate our findings. In this section, we briefly summarize the results of these exercises.

7.1. Establishment fixed-effects

As a first robustness check, we expand our estimation equation (5) with establishment fixed effects. Consequently, the estimates become very conservative and should be regarded as the lower bounds of human capital externalities. With establishment fixed effects, we aim to strengthen the validity of our main findings against a series of concerns.

The first concern is that the endogenous sorting of firms might threaten our estimates of human capital externalities. Although initial evidence against the spatial sorting of firms (Combes et al., 2012) exists, one might argue that if highly productive firms systematically sort into high-skilled firm neighborhoods, we might mistake firm-specific wage premia for externalities. Establishment fixed effects nullify such a bias because they purge the data from establishment-specific variations.

Another concern is that neighborhood characteristics correlated with firm productivity and the concentration of high-skilled workers might bias our estimates. Examples of such neighborhood characteristics are the proximity to infrastructure and market access. In a similar vein, the accessibility of establishments might also introduce bias. Specifically, if the receipt of compensating wage premia for commuting and the accessibility of establishment locations by workers and the concentration of high-skilled workers are correlated, our estimates might be biased. Because neighborhood characteristics and the accessibility of establishments are tied to their locations, establishment fixed effects, which are also tied to the locations, aid in removing such biases.

Figure B7 of the online appendix shows the estimates of human capital externalities with establishment fixed effects. Generally, the coefficient function has a similar structure to that of the coefficient function from our main specification. However, the estimated strength and reach of human capital externalities are smaller and disappear already after 17 km. A conservative interpretation of this robustness exercise would be to expect the *real* spillover function to be somewhere in between the two curves. Importantly, the robustness exercise indicates that our findings are not driven by the sorting of firms and neighborhood characteristics.

7.2. Supply effects and housing costs

In our main specification, we assume that supply effects are constant within local labor markets. Accordingly, we aim to disentangle human capital externalities from supply effects by removing all variation from the data that is constant within local labor markets. However, if supply effects are partly localized, we cannot guarantee that removing all variation common within local labor markets fully disentangles human capital externalities from these effects. To address this issue, we expand our main specification with regional control variables at the county level, which is the smallest administrative unit for which data are available. Specifically, we add controls for the log population density, the unemployment rate, and the number of hotel beds (as a proxy for amenities and infrastructure). Adding these controls at the county-level should at least partially address the described concerns. Reassuringly, our main results are not affected by including county-level controls (see figure B9 in the online appendix).

A related concern might be that housing costs affect our estimates. Throughout the paper we implicitly assume that local-labor-market-area fixed effects cover housing costs to a large extent. To investigate whether differences in housing costs might still affect our estimates, we now further expand our specification by including the log asking rent per square meter. Asking rents are available at the county level and are kindly provided by the BBSR (Bundesinstitut für Bau-, Stadt- und Raumforschung, 2020). As asking rents are only available from 2004 onward, we limit the timeframe of the robustness exercise from 2004 to 2014. Estimates within shorter timeframes are generally smaller than those within the full timeframe. However, estimates for the shorter observation period with and without including asking rents are almost identical. Consequently, we believe that our findings are not influenced by housing prices.[13]

7.3. Non-border regions

Because we have no data on workers outside of Germany, measurements of the distribution of high-skilled workers in border regions are partly truncated. For instance, establishments in the city center of Passau are only 2 km from the Austrian border. Therefore, past 2 km' distance, we observe the concentration of high-skilled workers only in southwest to northeast directions. Consequently, information on the distribution of high-skilled workers comes solely from these data points. Ignoring the partial truncation, we implicitly assume that the distribution on the Austrian side of the border is the same as on the German side of the border and that there are no costs from frictions in information flows across the border. To assess whether these assumptions influence our estimates, we remove border regions from our dataset and re-estimate our main model with establishments that are at least 50 km from the German border. Reassuringly, results without the border region do not differ notably from our main specification. The corresponding figures are available upon request.

7.4. Global labor market shocks

Another concern may be that global labor market shocks influence our findings through local industry or occupation clusters. For instance, if wages and the demand for skilled labor temporarily increase within a sector containing establishments that tend to cluster locally, our estimates would capture a spurious relation between wages and the local concentration of high-skilled workers. To rebut these concerns, we augment our model with year-specific industry and occupation fixed effects. Reassuringly, including industry and occupation trends does not affect our results. See section B.8.3 in the online appendix for more details.

Generally, including occupation fixed effects in wage regressions might be a concern since occupation is attached to the job and thus might be endogenous. In a further robustness check, we therefore exclude occupation fixed effects from our main specification. The according estimates are almost identical to our main results (see figure B8 in the online appendix.)

8. Conclusions

This paper studies the impact of human capital externalities from the spatial concentration of high-skilled workers around establishments on the individual wages of workers. We use, for the first time, precise geocoded register data from an entire economy and a novel estimation method from the field of FDA to compute the spatial decay of human capital externalities. We find significant spillover effects from the concentration of high-skilled workers around establishments that attenuate with increasing distance. The effects of human capital externalities from the direct neighborhood of establishments are roughly twice as large as those from high-skilled workers located 10 km away. After 25 km, the effects vanish. Overall, a spatially evenly distributed one-standard-deviation increase in the local share of high-skilled workers leads to wage gains of 4.25%.

Additionally, we find that human capital externalities differ by various subgroups. For instance, spillover effects are stronger in metropolitan and urban regions than in rural areas. Human capital externalities are also more pronounced for medium- and high-skilled workers than for low-skilled workers. Workers in the manufacturing sector benefit slightly more than workers in the service sector. Finally, wage increases from human capital externalities are higher for woman and young workers than for men and older workers.

Two developments in modern social science are primarily responsible for our ability to derive a precise functional relationship between the concentration of high-skilled workers and individual earnings. First, the availability of exact geospatial data enables us to describe the distributions of high-skilled workers around workplaces as functional objects with high resolution. Specifically, we evaluate the concentration of high-skilled workers every 500 m within a radius of 50 km around almost all establishments in Germany. Second, FDA provides tools to fully exploit such detailed data. We employ the estimator of Crambes et al. (2009) to regress a scalar outcome (log wage) on a continuous functional variable (the concentration of high-skilled workers based on distance). Our application illustrates the potential of FDA for use in economic research. FDA is particularly beneficial when the variable of interest can be regarded as a function over some continuum.

Generally, our findings imply that education creates positive externalities in local labor markets. Thus, regions benefit from attracting and training skilled workers. Moreover, to maximize these external effects, establishments should be located close to one another. Although spillover effects are far-reaching, workers and firms benefit most from the skill distribution in their immediate neighborhood of establishments. Because the effects vanish after 25 km, firms in remote regions do not gain from human capital externalities. Overall, our findings support Rosenthal and Strange (2008), who argue that the physical concentration of human capital remains important for economic development. Among other agglomeration effects, human capital externalities help explain differences in productivity between densely populated cities and rural areas.

Author statement

Johann Eppelsheimer: Conceptualization, Methodology, Writing – original draft, Writing-Review and editing, Data curation, Software, Validation. Elke Jahn: Writing-Review and editing, Validation, Data curation. Christoph Rust: Conceptualization, Methodology, Writing – original draft, Writing-Review and editing, Data curation, Software, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Martin Abraham, Annette Bergemann, Gerard van den Berg, Linda Borrs, Frank Cörvers, Wolfgang Dauth, Matthias Dorner, Gilles Duranton, Laurent Gobillon, Peter Haller, Daniel Heuermann, Moritz Kuhn, Christian Merkl, Florian Lehmer, Joachim Möller, Michael Oberfichtner, Veronika Püschel, Uta Schönberg, Tobias Seidel, Heiko Stüber, Rolf Tschernig, Erwin Winkler, Anthony Yezer, four anonymous referees and participants in the Regional Disparity Workshop 2019, ERSA Congress 2018, ERSA Congress 2017, NARSC Conference 2017, Statistical Week 2018, and seminars at the Institute for Employment Research (IAB) Nuremberg and the University of Regensburg for many helpful comments and suggestions. Johann Eppelsheimer acknowledges financial support from the graduate program of the IAB and the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (GradAB).

[13] Estimates for the observation period 2004 to 2014 with and without asking rents are available upon request.

Appendix

A. Summary statistics

The dataset used in our econometric analysis covers 15 years and consists of 3.5 million records of almost 540,000 workers. Table A1 summarizes the dependent variable (log wage) and numerical control variables. In the dataset, the mean daily wage is 111 euros, and the first and second quartiles range from 68 to 129 euros. On average, the individuals are 41 years old and have 15 years of work experience. The median population density is 119 inhabitants per square kilometer (exp(4.78)). Furthermore, 36% of the observations are from females, and 7% are from workers with foreign nationality. The shares of low-, medium- and high-skilled workers are 8%, 73% and 19%, respectively.

Table A.1
Summary statistics

	Mean	Std. Dev.	25th Perc.	Median	75th Perc.
Daily wage	111.37	78.05	68.17	94.64	129.02
Daily log wage	4.55	0.56	4.22	4.55	4.86
Age	41.14	10.65	33.00	41.00	49.00
Work experience (days)	5528.31	3305.44	2860.00	5105.00	7974.00
Tenure (days)	3059.98	2796.97	883.00	2160.00	4398.00
Log population density	3.71	2.38	0.97	4.78	5.66
Log hotel beds	3.16	0.70	2.68	3.14	3.53
Unemployment rate (%)	8.74	4.11	5.60	7.90	11.00
Log asking rent (log EUR/m^2)	1.82	0.25	1.63	1.75	1.99

Notes: The table presents the summary statistics of wages and (numerical) control variables. The underlying dataset contains 3,498,536 observations of 539,179 individuals over a period of 15 years. The regional characteristics are from 402 counties.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.regsciurbeco.2022.103785.

References

Abowd, John, M., Kramarz, F., David, N. Margolis, 1999. High wage workers in high wage firms. Econometrica 67 (2), 251–333.

Acemoglu, D., 1996. A microfoundation for social increasing returns in human capital accumulation. Q. J. Econ. 111 (3), 779–804.

Acemoglu, D., 1998. Why do new technologies complement skills? Directed technical change and wage inequality. Q. J. Econ. 113 (4), 1055–1089.

Acemoglu, D., Angrist, J., 1999, December. How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws. National Bureau of Economic Research. Working Paper 7444.

Acemoglu, D., Angrist, J., 2000. How large are human-capital externalities? Evidence from compulsory schooling laws. In: Bernake, B.S., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2000, vol. 15. MIT Press, pp. 9–74.

Ahlfeldt, G.M., Redding, S.J., Sturm, D.M., Wolf, N., 2015. The economics of density: evidence from the Berlin wall. Econometrica 83 (6), 2127–2189.

Ananat, E., Shihe, F., Ross, S.L., 2018. Race-specific urban wage premia and the black-white wage gap. J. Urban Econ. 108, 141–153.

Andersson, M., Larsson, J.P., Wernberg, J., 2019. The economic microgeography of diversity and specialization externalities — firm-level evidence from Swedish cities. Res. Pol. 48, 1385–1398.

Antoni, M., Ganzer, A., vom Berge, P., 2016. Sample of Integrated Labour Market Biographies (SIAB) 1975-2014. Institute of Employment Research, Nuremberg. FDZ-Datenreport 04/2016.

Arzaghi, M., Henderson, J.V., 2008. Networking off Madison avenue. Rev. Econ. Stud. 75 (4), 1011–1038.

Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in U.S. wage inequality: revising the revisionists. Rev. Econ. Stat. 90 (2), 300–323.

Borjas, G.J., 2003. The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. Q. J. Econ. 118 (4), 1335–1374.

Bundesinstitut für Bau-, Stadt- und Raumforschung, 2020. Wohnungs- und Immobilienmärkte in Deutschland 2020. (Bonn.

Caldeira, J., Torrent, H., 2017. Forecasting the US term structure of interest rates using nonparametric functional data analysis. J. Forecast. 36 (1), 56–73.

Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of West German wage inequality. Q. J. Econ. 128 (3), 967–1015.

Card, D., Lemieux, T., 2001. Can falling supply explain the rising return to college for younger men? A cohort-based analysis. Q. J. Econ. 116 (2), 705–746.

Ciccone, A., Peri, G., 2005. Long-run substitutability between more and less educated workers: evidence from U.S. states, 1950-1990. Rev. Econ. Stat. 87 (4), 652–663.

Ciccone, A., Peri, G., 2006. Identifying human-capital externalities: theory with applications. Rev. Econ. Stud. 73 (2), 381–412.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The productivity advantages of large cities: distinguishing agglomeration from firm selection. Econometrica 80 (6), 2543–2594.

Cornelissen, T., Dustmann, C., Schönberg, U., 2017. Peer effects in the workplace. Am. Econ. Rev. 107 (2), 425–456.

Crambes, C., Kneip, A., Sarda, P., 2009. Smoothing splines estimators for functional linear regression. Ann. Stat. 37 (1), 35–72.

Dauth, W., Haller, P., 2018. Berufliches Pendeln zwischen Wohn- und Arbeitsort. IAB-Kurzbericht, Institute for Employment Research (IAB), Nuremberg.

Davis, D.R., Dingel, J.I., 2019. A spatial knowledge economy. Am. Econ. Rev. 109 (1), 153–170.

Dustmann, C., Ludsteck, J., Schönberg, U., 2009. Revisiting the German wage structure. Q. J. Econ. 124 (2), 843–881.

Eppelsheimer, J., Möller, J., 2019. Human capital spillovers and the churning phenomenon: analysing wage effects from gross in-and outflows of high-skilled workers. Reg. Sci. Urban Econ. 78, 103461.

Faggio, G., 2019. Relocation of public sector workers: evaluating a place-based policy. J. Urban Econ. 111, 53–75.

Faggio, G., Schluter, T., vom Berge, P., 2019. Interaction of Public and Private Employment: Evidence from a German Government Move. Working Papers 19/09. Department of Economics, City University London.

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis - Theory and Practice. Spinger, New York.

Fitzenberger, B., Osikominu, A., Völter, R., 2005. Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. Institute of Employment Research, Nuremberg. FDZ-Methodenreport 03/2005.

Fu, S., 2007. Smart Café Cities: testing human capital externalities in the Boston metropolitan area. J. Urban Econ. 61 (1), 86–111.

Gibbons, S., Overman, H.G., Sarvimäki, M., 2021. The local economic impacts of regeneration projects: evidence from UK's single regeneration budget. J. Urban Econ. 122, 103315.

Hall, P., Horowitz, J.L., 2007. Methodology and convergence rates for functional linear regression. Ann. Stat. 35 (1), 70–91.

Heuermann, D., 2011. Human capital externalities in Western Germany. Spatial Econ. Anal. 6 (2), 139–165.

Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer, New York.

Hsing, T., Eubank, R., 2015. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley & Sons, Chichester.

Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963-1987: supply and demand factors. Q. J. Econ. 107 (1), 35–78.

Kosfeld, R., Werner, A., 2012. Deutsche Arbeitsmarktregionen - Neuabgrenzung nach den Kreisgebietsreformen 2007-2011. Raumforsch. Raumordn. 70 (1), 46–64.

Krusell, P., Ohanian, E., Rios-Rull, Lee, J.-V., Giovanni, L. Violante, 2000. Capital-skill complementarity and inequality: a macroeconomic analysis. Econometrica 68 (5), 1029–1053.

Lucas, R.E., 1988. On the mechanics of economic development. J. Monetary Econ. 22, 3–42.

Manning, A., Petrongolo, B., 2017. How local are labor markets? Evidence from a spatial job search model. Am. Econ. Rev. 107 (10), 2877–2907.
Marshall, A., 1890. Principles of Economics. MacMillan, London.
Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. J. Econom. 121, 175–212.
Morris, J.S., 2015. Functional regression. Annu. Rev. Stat. Appl. 2, 321–359.
Muravyev, A., 2008. Human capital externalities: evidence from the transition economy of Russia. Econ. Transit. 16 (3), 415–443.
Nelson, R.R., Phelps, E.S., 1966. Investment in humans, technological diffusion, and economic growth. Am. Econ. Rev. 56 (1/2), 69–75.
Ramsay, J.O., Ramsey, J.B., 2002. Functional data analysis of the dynamics of the monthly index of nondurable goods production. J. Econom. 107 (1–2), 327–344.
Ramsay, J., Silverman, B., 2005. Functional Data Analysis, 2. ed. Springer, New York.
Rauch, J.E., 1993. Productivity gains from geographic concentration of human capital: evidence from the cities. J. Urban Econ. 34 (3), 380–400.
Rosenthal, S.S., Strange, W.C., 2008. The attenuation of human capital spillovers. J. Urban Econ. 64 (2), 373–389.
Schmucker, A., Seth, S., Ludsteck, J., Eberle, J., Ganzer, A., 2016. The Establishment History Panel 1975-2014. Institute of Employment Research, Nuremberg. FDZ-Methodenreport 03/2016.
Ullah, S., Finch, C.F., 2013, Mar. Applications of functional data analysis: a systematic review. BMC Med. Res. Methodol. 13 (1), 43.
Verstraten, P., 2018, September. The Scope of the External Return to Higher Education. CPB Netherlands Bureau for Economic Policy Analysis. Discussion Paper 381.
Wang, S., Jank, W., Shmueli, G., 2008. Explaining and forecasting online auction prices and their dynamics using functional data analysis. J. Bus. Econ. Stat. 26 (2), 144–160.