The spatial extent of human capital spillovers in a transition country: Evidence from Brazil*

Edilberto T. Almeida† Raul M. Silveira Neto† Roberta M. Rocha†

Abstract

In this paper we investigate the spatial extent of agglomeration economies on the wage earnings within the Brazilian cities. For this, we use a geocoded employer-employee panel data 2006-2014, exogenously determined grid and different geographical distance bands. We deal with the spatial sorting and endogeneity in the wage-agglomeration economies relationship with controls for observable and unobservable individual and establishment characteristics and instrumental variables based on the exogenous expansion of higher education in Brazil in period 1991-2004. Our main results indicate that distance seems to be still more important in developing countries context. Adding 1,000 college-educated workers up to 1 km from the individual's workplace would increase the wages of workers on average by 6.78 percent. On the other hand, if the same number of workers are added to the 1-5 km or 5-10 km range, the wages of workers would increase, on average, by 1.95 and 1.27 percent, respectively. This evidence is robust to different specifications (e.g., including worker-plant or worker-city matches) and shows that the speed of decay of agglomeration economies is higher in Brazil than observed in developed countries.

Keywords: attenuation, human capital spillovers, wages, Brazil

Resumo

Neste artigo analisamos a extensão espacial das economias de aglomeração sobre o salário nas cidades brasileiras. Para tanto, foram usados dados em painel 2006-2014 trabalhador-firma georreferenciados, grids exogenamente determinados e diferentes faixas de distância geográfica. Lidamos com o sorting espacial e com a endogeneidade na relação salário-aglomeração usando controles para características observáveis e não observáveis dos trabalhadores e das firmas e variáveis instrumentais baseadas na expansão exógena da educação superior no Brasil no período 1991-2004. Os principais resultados indicam que a distância parece ser ainda mais importante no contexto dos países em desenvolvimento. Adicionar 1.000 trabalhadores com formação universitária até 1 km do local de trabalho individual aumentaria os salários dos trabalhadores, em média, 6,78%. Por outro lado, se o mesmo número de trabalhadores fosse adicionado ao intervalo de 1-5 km ou 5-10 km, os salários dos trabalhadores aumentariam, em média, em 1,95% e 1,27%, respectivamente. Estas evidências são robustas a diferentes especificações (por exemplo, incluindo o matching trabalhador-firma ou trabalhador-cidade) e mostra que a velocidade do decaimento das economias de aglomeração é maior no Brasil do que a observada nos países desenvolvidos.

Palavras-chave: atenuação, spillovers de capital humano, salários, Brasil

JEL classification: L60, R0, J24

Área 1 - Teoria, métodos e modelos de economia regional

*The authors acknowledge and are grateful for financial support by the Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.

†Federal University of Pernambuco - UFPE. E-mail: edilbertotiago@hotmail.com.

1 Introduction

It is well established in the literature that agglomeration economies are important to understand the distribution of economic activities in geographic space (see, e.g., Duranton and Puga, 2004; Rosenthal and Strange, 2004). The central question in the agglomeration literature is idea that spatial concentration enhances productivity. The action of these forces helps to determine the optimal size of cities (Chauvin et al., 2017). Most of the existing evidence on the subject is based on assumptions that the effects of agglomeration economies are homogeneous in space, implicitly assumed to be "club goods" (see Rosenthal and Strange, 2003), which operate homogeneously within cities and intermediate regions or even at more aggregated levels (see Combes and Gobillon, 2015, for a recent survey). However, some externalities, especially those associated with the interaction between workers (face-to-face), such as knowledge spillovers, may not occur homogeneously in the city as a whole and are stronger at short distances (see, e.g., Rosenthal and Strange, 2008; Fu, 2007; Andersson et al., 2009).

While understanding spatial decay of agglomeration economies in a micro-geographic context is politically relevant and essential to understand their nature,[1] there is little empirical evidence about this. Furthermore, this evidence is almost exclusively for developed countries. Using individual wages, Fu (2007) evaluated the spatial scope of different types of local externalities in the Boston Metropolitan Area and found that the effects of human capital attenuate sharply with distance. The contribution of Rosenthal and Strange (2008) revealed that the effects of agglomeration economies are localized and attenuate rapidly in the US. In particular, that the external returns to education are greater at short distances (around 8 km). Di Addario and Patacchini (2008) provided similar evidence for Italy from urbanization externalities that occur mainly up to 4 km and are attenuated up to 12 km. More recently, Håkansson and Isacsson (2019) indicated that the effects of urbanization are localized (attenuated up to 25 km) and asymmetric across percentile ranks in the wage distribution in Sweden.

In developing countries, aggregated geographic data provide evidence that the effects of agglomeration economies are greater compared with estimates for developed countries (Duranton, 2016; Barufi et al., 2016; Chauvin et al., 2017; Combes et al., 2013; 2020), despite the numerous costs and disadvantages associated with city size. The structure of cities is different in developing countries, where problems associated with the provision of public services such as transport, and consequently commuting time, in general are greater (Glaeser and Henderson, 2017; Thisse, 2018), which can substantially affect the geographic spread of agglomeration externalities. Little is known about the spatial scope of agglomeration economies in these environments. To be more precise, using micro-geographic data, only Li et al. (2020a) provides some evidence for China. The authors evaluated the effects of localization externalities after controlling for urbanization in different distance bands on the share of new firms. The main findings suggested that attenuation is very different among industries, but in general is faster in China than in the US (compared to Rosenthal and Strange (2003)'s results), indicating that agglomeration economies can be spatially attenuated more quickly because of the local urban structure.

This paper seeks to reduce part of this gap in the literature by analyzing the spatial extent of external returns to education in Brazil. For this, we employ finely geocoded employer-employee panel data in the period 2006-2014 from Brazil to assess if the effects of human capital spillovers on workers' productivity (proxied by their hourly wage) change with distance. More specifically, we evaluate the spatial extent of the externalities generated by the concentration of college-educated workers in different distance bands in relation to the individual's workplace.

Beyond these general differences in the economic environment between developed and developing countries, other characteristics of the Brazilian economy make the investigation of this phenomenon in the country particularly interesting. For example, relative to China, one immediate difference is the free interregional mobility of workers, which can strongly affect the spatial scope of agglomeration economies, in particular human capital spillovers, which are more spatially localized.[2] The spatial concentration of high-tech manufacturing is positively correlated with the share of college-educated workers (Almeida et al.,

[1] See, e.g., Rosenthal and Strange (2020) for a discussion of the different geographic scales of operation of micro-foundations of the agglomeration economies.

[2] An example of this type of restriction is China's *hukou* system, which controls population movement by restricting workers' social rights mostly to their birthplace (Combes and Gobillon, 2015).

2020), which may suggest that spillovers occur over short distances. Brazil is historically among the countries with the highest levels of income inequality in the world (Fishlow, 1972; de Mendonça and de Barros, 1995; Narita et al., 2003), but at the beginning of the 21st century, there was a reduction of per capita household income inequality associated with the reduction of educational differentials (Barros et al., 2007a; 2007b; Oliveira and Silveira Neto, 2013; 2016). By analyzing the period 2006-2014 we investigate if there was any change in spatial scope of external return to education consistent with this inequality reduction.

In addition to making a contribution to the scarce empirical literature, unlike previous studies, such as Rosenthal and Strange (2008), we implement a set of tools taking advantage of characteristics of our rich database and exploit exogenous education policy shocks at the national level to identify the causal effect of spillovers in different distance bands. We use panel data, which allows us to control besides the observed characteristics, the unobserved individual and plant heterogeneities, industry-year specific trends, and region fixed effects in wage regressions. Since our data are point data, our geographic units of analysis are squares measuring 1 km^2 and are defined exogenously from Brazil's geographical boundaries. This allows us to explore very small geographic contexts, for example, smaller than neighborhoods, in line with the literature suggesting that human capital spillovers are very local (Fu, 2007; Rosenthal and Strange, 2008). Our choice of units also allows us to minimize potential problems associated with the a priori definition of official administrative areas, such as heterogeneous sizes (Briant et al., 2010). To deal with the potential endogeneity of the human capital variables, we combine exogenous shocks in Brazilian educational policies with the lagged demographic structure of each distance band in a shift-share instrumental variable approach. We also explore how our main results vary when our estimates are conditional to a diverse set of additional controls, such as worker-plant match fixed effects (or job-spell fixed effects), worker-city match fixed effects, mass of low-schooling workers, and the transportation infrastructure surrounding the individual's workplace.

The main results indicate that both the externalities generated by the concentration of general workers and college-educated workers are highly localized and much stronger in the first distance band (0-1 km). The positive effects of external return to education are generally attenuated up to 10 km. For example, in our main specification, adding 1,000 college-educated workers up to 1 km would increase the wages of workers on average by 6.78 percent. On the other hand, if the same number of workers are added to the 1-5 km or 5-10 km range, the wages of workers would increase, on average, by 1.95 and 1.27 percent, respectively. This evidence is robust to different specifications and shows that the speed of decay of agglomeration economies is higher in Brazil than observed in developed countries.

The remainder of the paper is structured as follows. In the next section we present the theoretical framework that provides the basis for our empirical approach. Section 3 describes the empirical strategy. Sections 4 and 5 present the results and final comments.

2 Theoretical framework

In this section we present a theoretical framework to address the spatial scope of external returns to education. We define external returns to education as the effect of an increase in the number of educated workers in a specific location and in neighboring locations on total wages minus the effect due to private returns to education (Acemoglu and Angrist, 2000; Moretti, 2004a). We adapt the theoretical structure proposed by Moretti (2004a) to identify the externalities generated by concentration of educated workers, assuming there are two types of workers, with low and high education, who are imperfect substitutes. But different from that author, we expand the model so that we can capture the spatial scope of human capital externalities. As mentioned by Rosenthal and Strange (2020), the human capital spillovers as envisioned by Marshall (1890) are likely to be highly local. The empirical evidence confirms this hypothesis (e.g., Fu, 2007; Rosenthal and Strange, 2008), and as highlighted by Charlot and Duranton (2004; 2006), although the improvement in information technology allows effective communication with distant partners, there is no evidence that this type of communication replaces face-to-face meetings, but instead is complementary.

In the model presented by Moretti (2004a), the human capital spillovers are treated as club goods that act homogeneously within cities. Our structure generalizes this hypothesis and allows these effects to be attenuated with geographic distance and therefore allows them to be different within the same city. The production function is given by:

$$\mathbf{Y}_{jz} = (\theta_{1jz}\mathbf{N}_{1jz})^{\alpha_1}(\theta_{2jz}\mathbf{N}_{2jz})^{\alpha_2}\mathbf{K}_{jz}^{1-\alpha_1-\alpha_2} \tag{1}$$

where $\mathbf{Y}_{jz}$ is the output of firm j located at z; $\mathbf{N}_{1jz}$ is the number of workers with high education in firm j located at z; $\mathbf{N}_{2jz}$ is the number of low-schooling workers; $\mathbf{K}_{jz}$ is capital; and θ's are productivity shifters.

We allow for human capital spillovers by letting workers' productivity depend on the number of educated workers in neighboring locales in a continuous space,[3] as well as on their own human capital:

$$\log(\theta_{\ell jz}) = \phi_{\ell jz} + \gamma \sum_{\bar{z}} f(d(z,\bar{z}))\overline{\mathbf{N}}_{1\bar{z}} \qquad \ell = 1,2 \tag{2}$$

where $\phi_{\ell jz}$ is a group-specific effect that captures the direct effect of own human capital on productivity in a specific firm and place ($\phi_{1jz} > \phi_{2jz}$); $\overline{\mathbf{N}}_{1\bar{z}}$ is the number of workers with high education in all firms $k \neq j$ located at $\bar{z}$, so the term $\sum_{\bar{z}} f(d(z,\bar{z}))\overline{\mathbf{N}}_{1\bar{z}}$ captures the effects of Marshallian externalities resulting from the concentration of educated workers in other firms in the same and neighboring localities, weighted by a spatial decay function $f(d(z,\bar{z}))$ with $f(0) = 1$, $f'(d(z,\bar{z})) < 0$; and $d(z,\bar{z})$ is the distance between z and $\bar{z}$. As in Moretti (2004a), if there are positive spillovers, $\gamma > 0$.

Define R as the set of all locations. Now we can define the total of educated workers in R by $\mathbf{N}_{1\text{R}}$ and the total of low-schooling workers by $\mathbf{N}_{2\text{R}}$. Thus, the total number of highly educated workers can be decomposed into $\mathbf{N}_{1jz} + \sum_{\bar{z}} \overline{\mathbf{N}}_{1\bar{z}}$, where the second term is the number of highly educated workers in firms $k \neq j$ in the same (if $z = \bar{z}$) and neighbouring localities (if $z \neq \bar{z}$). If wages are equal to the marginal product of each type of worker and the spillover is external to individual firms in z but internal to the R (take $\overline{\mathbf{N}}_{1\bar{z}}$ as given), the logarithm of wages for highly and low-schooling workers is given by:

$$\begin{aligned}\log(w_{1jz}) = {} & \log(\alpha_1) + \alpha_1\phi_{1jz} + \alpha_2\phi_{2jz} + (\alpha_1+\alpha_2)\gamma\sum_{\bar{z}} f(d(z,\bar{z}))\overline{\mathbf{N}}_{1\bar{z}} \\ & + (\alpha_1 - 1)\log(\mathbf{N}_{1jz}) + \alpha_2\log(\mathbf{N}_{2jz}) + (1-\alpha_1-\alpha_2)\log(\mathbf{K}_{jz})\end{aligned} \tag{3}$$

$$\begin{aligned}\log(w_{2jz}) = {} & \log(\alpha_2) + \alpha_1\phi_{1jz} + \alpha_2\phi_{2jz} + (\alpha_1+\alpha_2)\gamma\sum_{\bar{z}} f(d(z,\bar{z}))\overline{\mathbf{N}}_{1\bar{z}} \\ & + \alpha_1\log(\mathbf{N}_{1jz}) + (\alpha_2 - 1)\log(\mathbf{N}_{2jz}) + (1-\alpha_1-\alpha_2)\log(\mathbf{K}_{jz})\end{aligned} \tag{4}$$

Therefore, the effect of $\mathbf{N}_{1\text{R}}$ on workers' productivity is the sum of three effects: (i) that generated by the number of highly educated workers within firm j (neoclassical effect); (ii) that generated by the number of highly educated workers in other firms k in the same location z; and (iii) that generated by the number of highly educated workers in other firms k in neighboring locations $\bar{z}$. Formally:

$$\frac{\partial \log(w_{\ell jz})}{\partial \mathbf{N}_{1\text{R}}} = \frac{\partial \log(w_{\ell jz})}{\partial \mathbf{N}_{1jz}} + \left.\frac{\partial \log(w_{\ell jz})}{\partial \overline{\mathbf{N}}_{1\bar{z}}}\right|_{\bar{z}=z} + \sum_{\bar{z}\neq z}\frac{\partial \log(w_{\ell jz})}{\partial \overline{\mathbf{N}}_{1\bar{z}}} \qquad \ell = 1,2 \tag{5}$$

Which results in:

$$\frac{\partial \log(w_{1jz})}{\partial \mathbf{N}_{1\text{R}}} = \frac{\alpha_1 - 1}{\mathbf{N}_{1jz}} + (\alpha_1+\alpha_2)\gamma + (\alpha_1+\alpha_2)\gamma\sum_{\bar{z}\neq z} f(d(z,\bar{z})) \quad \forall\, z, \bar{z} \tag{6}$$

$$\frac{\partial \log(w_{2jz})}{\partial \mathbf{N}_{1\text{R}}} = \frac{\alpha_1}{\mathbf{N}_{1jz}} + (\alpha_1+\alpha_2)\gamma + (\alpha_1+\alpha_2)\gamma\sum_{\bar{z}\neq z} f(d(z,\bar{z})) \quad \forall\, z, \bar{z} \tag{7}$$

In words, both types of workers are affected by spillovers generated by proximity to highly educated workers. But note that as in Moretti (2004a), low-schooling workers, w_{2jz}, benefit for two reasons. First, an increase in the number of workers with high education raises low-schooling workers' productivity because of imperfect substitution ($\alpha_1/\mathbf{N}_{1kz} > 0$). Second, the spillover raises their productivity $((\alpha_1+\alpha_2)\gamma + (\alpha_1 +$

[3] To approximate our theoretical framework of our empirical model in the next section, we consider a summation in the second term of equation 2, but the results remain valid when we consider infinitesimal variations (integral).

$\alpha_2)\gamma \sum_{\bar{z} \neq z} f(d(z, \bar{z})) > 0)$. For highly educated workers, w_{1jz}, the impact of an increase in the number of highly educated workers depends on two opposite effects. The first is the conventional supply effect, which makes the economy move along a demand curve $((\alpha_1 - 1)/\mathbf{N}_{1jz} < 0)$, and the second is the spillover effect, which raises productivity.

Unlike Moretti (2004a), in adapting the model proposed in this section, the spillover effects can be attenuated with the distance between z and $\bar{z}$. This can be observed by considering the behavior of $f(d(z, \bar{z}))$ and obtaining the derivative of the return to education with respect to distance between two localities ($z \neq \bar{z}$).

$$\frac{\partial^2 \log(w_{\ell jz})}{\partial \mathbf{N}_{1\mathrm{R}} \partial d(z, \bar{z})} = (\alpha_1 + \alpha_2)\gamma \frac{\partial f(d(z, \bar{z}))}{\partial d(z, \bar{z})} < 0 \qquad \forall z \neq \bar{z}, \ell = 1, 2 \tag{8}$$

It is interesting to note that when z index microgeographic areas of a continuous space, the adaptation of the theoretical structure satisfies to the objective for which it was proposed, because it allows spillover effects to be captured in very small areas, for example, smaller than neighborhoods. In the next section we describe our empirical strategy based on these conclusions about the attenuation of human capital spillovers.

3 Data and empirical strategy

3.1 Data and variables

Our main source of data is the Annual Report of Social Information (*Relação Anual de Informações Sociais*, or RAIS) available each two years in the period 2006-2014, i.e., a total of 5 years, which encompasses all formal workers in Brazil and is available from the Ministry of Labor. This dataset allows us to monitor workers and plants across years and provides detailed information at worker-level such as wages, gender, education, age, tenure, hiring data, number of hours worked, kind of contract, occupation, and identifier of the plant in which the worker is employed (National Register of Legal Entities - CNPJ number). At plant-level, detailed information is available about address, plant size (classification based on the number of workers employed), National Classification of Economic Activities (CNAE), which is compatible with the International Standard Industrial Classification of all Economic Activities (ISIC) revision 4, and opening and closing dates (if applicable). We use the address information available in RAIS data and the Google Maps base to obtain the geographic coordinates of each plant to construct a unique set of point data. We also use other complementary data sources, such as the Census and National Household Survey (*Pesquisa Nacional por Amostra de Domicílios*, or PNAD),[4] all provided by the Brazilian Institute of Geography and Statistics (IBGE) to calculate our instrumental variables (described in subsections 3.3).

Our sample consists of plants engaged in manufacturing. We recognize that human capital externalities can occur in other sectors, but there are reasons for restricting our sample. The most immediate one is that we do not have geocoded data for the other sectors of the economy. Moreover, in the case of the Brazilian economy, industry is the sector with the least informality,[5] which contributes to the representativeness of our study since we use data from the formal labor market. Another interesting characteristic is that, as confirmed in previous studies, manufacturing can benefit more from agglomeration economies (see, e.g., Barufi et al., 2016) and human capital spillovers, which are greater when the sectors are economically close, since they presumably interact more in manufacturing (Moretti, 2004c).

The primary focus of the paper is the spatial scope of human capital spillovers observed from the individual's workplace. In order to achieve this focus, we first define the geographical context of our study. Our geocoded data allow us to freely define spatial units of measurement, so we exogenously divide the Brazilian territory into a uniform set of grid cells (1 km x 1 km) and associate each plant (point data)

[4]This survey provides annually general characteristics of the population, education, labor, income and housing, with household as the unit of survey. It is conducted in nine metropolitan regions (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba and Porto Alegre) chosen to be representative of the country at large.

[5]According to the PNAD data for the year 2014 provided by IBGE, the share of people employed in informal jobs in general industry (including extractive, transformation, electricity and gas, and water and sewage) is 23.9%, while the same measure for the construction sector is 57.9%, agriculture 73%, commerce 36.9%, and other services (excluding domestic) 33.4%.

with its respective cell.[6] By using such small microgeographic units, we do not face the common problems of this type of approach when previously defined areas are used, and we minimize endogeneity problems associated with sorting. For example, Rosenthal and Strange (2008) used the Place of Work Public Use Micro Areas (PWPUMAs) for the US and found that the presence of large PWPUMAs can generate measurement errors in the agglomeration variables, biasing the estimates of the influence of agglomeration towards zero. So we don not need to impose any sample restrictions on the territory because our spatial measurement is homogeneous. In addition, official Brazilian geographic divisions for areas smaller than municipalities, such as census sectors[7] (equivalent to census tracts), are defined based on local factors such as the number of households. The choice of these areas as units of analysis in our approach may generate biased estimates due to simultaneity.

Initially, we had around 8.5 million cells, but not all cells, of course, have plants, and this varies every year as plants are created and/or existing plants are closed/moved. To eliminate cells that are irrelevant for our purpose, we selected only those that have at least one plant and are within metropolitan areas existing in 2006 that encompass all five macro-regions of the country. There are economic and technical factors that justify this restriction. Economically, the externalities generated by the agglomeration economies occur mainly in urban areas, from the concentration of workers and firms. Furthermore, in the Brazilian context the population and economic activity are mainly concentrated in urban areas. Based on data from the 2010 Census, 84.4% of the population lived in urban areas, occupying 1.07% of national territory. Technically, by choosing only the metropolitan areas we are working with spatially smaller municipalities compared to the less urbanized municipalities. This provides more variation in our instrument for the number of college-educated workers, as explained in the next sections.

We define the centroids of these exogenous cells and from them, to capture the agglomeration economies at different distances, we follow Rosenthal and Strange (2003; 2008) by specifying five concentric ring variables, each of which measures the number of workers present at a given distance from the individual's workplace: between 0-1, 1-5, 5-10, 10-20, and 20-40 km. The motivation for choosing the size of the concentric rings is related to the spatial extension of Brazilian cities. The smallest ring can be considered to cover effects at a geographic level smaller than neighborhood. The next two distance ranges, 1-5 and 5-10 km, cover the distances of most common commuting distances within core cities. The two distance bands further away from the centroid, 10-20 and 20-40 km, cover commuting from neighboring cities to the core city and interactions at the level of the metropolitan region as a whole, respectively (see Figure 1).

In this context, another important difference from the strategy employed by Rosenthal and Strange (2008) is their assumption that employment in each PWPUMA is uniformly distributed throughout the given PWPUMA. Then, for each concentric ring, the authors had an approximation weighted by the areas of the PWPUMA forming the concentric ring of the true number of workers. Our microgeographic database provides the location of each plant, so we can get more precise measures of the number of workers in each concentric ring, which consequently minimizes potential endogeneity problems associated with measurement errors. Still in this respect, we also have information for different years, which allows us to observe the trajectory of a plant and its workers over the years, therefore controlling for any observed and unobserved heterogeneity that is fixed in time and minimizing potential endogeneity problems associated with omitted variables. We will return to this issue in the next subsection.

To calculate the individual wages in all models, we impose a few more restrictions on our sample. Hourly wage rates are calculated by dividing monthly wage[8] earnings by the usual number of hours worked per week and the number of weeks worked in the month by male workers between the ages of 18 and 56 who work than 20 hours or more per week. This sample restriction implies that the remaining workers exhibit a lower unobserved variation in possible endogenous decisions to work full time and increases the possibility that any remaining variations are absorbed by the control variables (Rosenthal and Strange,

[6]A similar strategy was used by Larsson (2014) and Andersson et al. (2014; 2019) for Swedish cities and by Li et al. (2020a) for China.

[7]Census data provide official spatial divisions called census sectors. By definition of IBGE, census sectors are the territorial units established for registration control purposes, formed by a continuous area, located in a single urban or rural setting, with the size and number of households that allow the survey by a census taker.

[8]Nominal wages in December deflated by the National Wide Consumer Price Index (*Índice Nacional de Preços ao Consumidor Amplo*, or IPCA) (2017=100).

2008; Håkansson and Isacsson, 2019). To calculate the agglomeration variables in each concentric ring, we used both restricted and unrestricted databases.

3.2 Empirical model specification

The objective of the article is to estimate the spatial scope of human capital spillovers. For this purpose, based on the theoretical structure presented above, which suggests a positive relationship between individual workers' wages and the concentration of college-educated workers, with a decreasing effect with geographic distance, we propose an empirical specification to assess the spatial extent of these effects. Formally our worker-level specification is given by:

$$\mathsf{w}_{izt} = \mathbf{X}_{it}\lambda + \mathbf{H}_{jt}\gamma + \sum_{r} \beta_r \mathsf{S}_{rt} + \alpha_i + \mu_c + \psi_{pt} + \epsilon_{izt} \tag{9}$$

where w_{izt} is the natural log of the real hourly wage of worker i in cell z (in plant j, industry p, metropolitan region c), and year t; $\mathbf{X}_{it}$ is a matrix of worker-level control variables such as age, age squared, tenure, and education; $\mathbf{H}_{jt}$ is a matrix of plant-level control variables, such as size; S_{rt} is our explanatory variable of interest and represents the number of workers with college degree or higher in each concentric ring r;[9] α_i, μ_c and ψ_{pt} are worker, metropolitan region, and industry-year fixed effects, respectively; and ϵ_{izt} is the error term.

Our parameters of interest are β_r, $r = 1, \cdots, 5$. The challenge of the exercise is to identify variation in the individual wages that is driven by concentration of college-educated workers around the individual's workplace and hence exogenous to other factors that affect local wages. If the location of the college-educated workers were random, this parameter would capture the effect of the concentration of college-educated workers in each distance band on the wages. There are, however, different mechanisms that make the hypothesis of S_{zrt}'s exogeneity doubtful.

One source of endogeneity common in these approaches, as explained by Rosenthal and Strange (2008), is measurement error. To deal with potential problem, as we discussed earlier, our geocoded data allow us to set each point (plant) in the exact concentric ring to which it belongs. Therefore, S_{zrt}, $r = 1, \cdots, 5$, is measured more precisely and we considerably reduce the part of the measurement error included in the residual of equation 9, ϵ_{izt}.

Another source of endogeneity is associated with omitted variables correlated with the concentration of college-educated workers in the concentric rings that surround the individual's workplace. Since individuals choose where to live and work, it is obvious that spatial sorting of observable and unobservable characteristics may bias our estimates of the human capital spillover effects. We address the potential endogeneity caused by spatial sorting in three ways. First, our micro-geographic units of one square kilometer are generally outside the set of workers' locational choice and thus the choice of the surrounding concentration of college-educated workers. Second, we introduce a set of control variables for characteristics related to the workers and establishment that may be relevant in this context, such as age, age squared, degree of education, tenure, tenure squared, and occupation for workers (all included in $\mathbf{X}_{it}$); and plant size ($\mathbf{H}_{jt}$). Third, we also include the worker fixed effects that control for unobserved individual characteristics, such as "ambition" or "ability".

In addition to these controls, we also include metropolitan region fixed effects, μ_c, which absorb time-invariant metropolitan region characteristics and conditions, such as geographical location, industrial structure, weather and amenities; and to control for industry common time trends, like sector-specific growth path at 2-digit level, we include industry-by-year fixed effects, ψ_{pt}.

Even using exogenously defined grid and a series of controls in equation 9, β_r can be biased by the influence of unobservable confounding trends. Any unobserved time-varying factors that affect simultaneously both wages and concentration of college-educated workers can make our estimates of human capital spillovers biased, e.g., transitory productivity shocks that attract highly educated workers and raise wages, $\text{cov}(\epsilon_{izt}, \mathsf{S}_{zrt}) \neq 0$. Thus, we cannot guarantee that our human capital variables calculated for each ring are exogenous.

[9] Note that we can generalize this structure and incorporate the urbanization (economic mass) effects measured by the total sum of employment in each ring as in Rosenthal and Strange (2008).

We address these potential concerns by proposing instruments for our potentially endogenous human capital variables. Our identification strategy consists of using shift-share instrumental variables (SSIV). This approach, which uses weighted averages of a common set of shocks, with weights reflecting heterogeneous shock exposure, is increasingly common in many contexts (e.g., Bartik, 1991; Blanchard et al., 1992; Autor et al., 2013) and have had their properties formally discussed in the recent literature (see Borusyak et al., 2018; Adao et al., 2019; Goldsmith-Pinkham et al., 2020). First, we measure the population with college degree or higher based on differences in the demographic structure in each concentric ring in 1991. Next, we predict the number of people with that educational level in each concentric ring by the national change in the share of the population with college degree or higher as a result of changes in the federal government's educational policy in the period 1991-2004, weighted by population with college degree or higher in 1991 by age groups. We present a more detailed discussion of our shift-share regression designs in the next section.

3.3 Educational policy changes and identification

Our analysis exploits large shifts in national education policy between 1991 and 2004 as an exogenous source of variation in the number of college-educated people across concentric rings within Brazilian metropolitan regions to identify the effect of the concentration of college-educated workers in the concentric rings that surround the individual's workplace on the individual wages (our proxy to labor productivity). As the 1991 Census data and the 2004 PNAD data show, in this period there was 39.6% growth in the share of the population with college degree or higher. The changes in the national higher education policy are also clearly seen when we evaluate the growth in the number of higher education institutions and the number of students enrolled in undergraduate programs. For example, the number of higher education institutions and the number of students enrolled in undergraduate programs in the 1995-2003 period grew by 108% and 120%, respectively.[10]

In addition to the large national variation in the number of people with college degree or higher caused by the change in higher education policy ("shift"), the spatial distribution of these people by age group is also heterogeneous across concentric rings. Part of these cross-ring contemporary differences was certainly due to varying economic environment, industrial structures, and labor demand influencing wage and employment growth. To reduce the chances that age structure is itself endogenous, we use lagged weights. That is, as a generalization of the variable used by Moretti (2004a), we use the demographic structure in each concentric ring in 1991 to define our "exposure shares". To the extent that the relative number of people of different cohorts varies across concentric rings, this will lead to differential trends in the number of college-educated people across concentric rings. That is, we construct an educational-policy-driven instrument that retains only the portion of growth in the number college-educated people at concentric ring level attributable to national policy fluctuations.

As Moretti (2004a) pointed out for the US, each new generation has a larger share of more educated individuals – and therefore of the number of educated people when the country's population is growing or remains constant – so young people today are more educated, on average, than young people from previous decades. Particularly in the case of Brazil, as we have mentioned, this trend has been exogenously shifted upward due to educational policies. The identification comes from differences in the relative magnitude of the cohorts that entered and left the labor force between 1991 and 2004. For example, consider two identical concentric rings, except in the age structure. If in one of the rings the number of young adults is higher, then the number of college-educated people is expected to be higher in this ring. This increase in the number of workers depends, thus, on the specific demographic structure in each ring. This is Moretti's argument, since the demographic structure varies between cities (in our case rings), each has its own tendency to increase the share (in our case the number of college-educated workers).

To make our exogenous spatial division compatible with official census data, we used a similar strategy to Rosenthal and Strange (2008) and Verstraten et al. (2018), which is based on the area of municipalities contained in each ring to create geographic weights to be associated with the college-educated population in each cohort.[11] The measure is calculated based on the data from the 1991 Census at the municipal level

[10] Data from Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), a federal research institution linked to the Ministry of Education, available at `http://inep.gov.br`.

[11] For example, if a concentric ring includes all municipality 1 and 20 percent of the area of municipality 2, then population

about the number of workers with college degree or higher in each ring. To instrument the population with college in the ring of 0-1 km, we use data from the distance range 0-5 km; for 1-5 km, we use 20-40 km; for 5-10 km, we use 40-80 km; and for 10-20 km, we use 80-120 km. We highlight two main points that motivated this choice. The first is the concern that the exposure weights are themselves endogenous even though they are lagged in time. To reduce the chances of this occurring, we use both time and space lagged weights. The second is that increasing the width of rings farther from the centroid minimizes multicollinearity problems (Verstraten et al., 2018).

Formally, the instrument for number of workers with college degree or higher is given by:

$$\text{College}_{rt} = \sum_{m} \sum_{c} \omega_{rc} \mathsf{P}_{cm1991} \ln\left(\frac{\mathsf{P}_{mt}}{\mathsf{P}_{m1991}}\right), \text{ with } \omega_{rc} = \frac{\mathsf{A}_{r\cap c}}{\mathsf{A}_c} \tag{10}$$

where $\mathsf{A}_{r\cap c}$ is intersection area between concentric ring r and municipality c; A_c is total area of the municipality; P_{cm1991} is population[12] with college degree or higher in municipality c and cohort m (we defined three age groups: young 16-25, middle-aged 26-50, and old 51-70) in reference year; and P_{mt} is national population with college degree or higher in cohort m and year $t = (2000, \cdots, 2004)$.

3.4 Summary statistics

In this subsection we present some descriptive statistics that show how our data vary across distance bands. Table 1 summarizes descriptive statistics of concentric ring employment variables. Column 2 shows the average number of workers, college-educated workers and number of plants, and column 3 shows the deviations in each concentric ring in 2006. Similarly, columns 4-7 show the same measurements for 2014 and for 2006-2014. It is interesting to note that the average number of college-educated workers increased, on average within five rings by 58.57% while general workers increased only 1.68%. As mentioned, this expansion can be associated with the exogenous increase of the share of population with college degrees in the period 1991-2004. Note also the high standard deviation, indicating the heterogeneity of the spatial distribution of employment within metropolitan areas. The bottom table presents similar measures for the number of plants in each ring, the number of cells in each year (2006 and 2014) and the average number of cells in the period 2006-2014. One possible problem associated with defining ring size is lack of variation of data within each ring, particularly in the smallest rings (0-1 km). Our data show that there is significant variation within the smaller rings, as can be observed both for general workers, such as college-educated workers, and for plants.

As an illustration of the heterogeneity, Figure 1 presents the spatial distribution of college-educated workers in the four largest metropolitan regions of the country (São Paulo, Rio de Janeiro, Belo Horizonte and Porto Alegre). Note that except the Rio de Janeiro Metropolitan Region (RJMR) (Figure 1 (b)), the areas with the highest concentration of college-educated workers in manufacturing are outside the core city. For example, in the São Paulo and Belo Horizonte Metropolitan Regions (SPMR and BHMR), the areas with a high concentration of college-educated workers are in the 10-20 km range (Figures 1 (a) and (c)), while in the Porto Alegre Metropolitan Region the range is 20-40 km range (Figure 1 (d)). The neighbors of the core municipality seem attractive to the large industries as they are still close to the central business district (CBD), so the industrial establishments can still benefit from the positive externalities while avoiding most congestion effects. This location pattern is in conformity with the pattern commonly found in literature on city structure (see, e.g., Anas et al., 1998; Anderson and Bogart, 2001; Coffey and Shearmur, 2002; Billings and Johnson, 2012).

in the ring is set equal to the population in cohort m in municipality 1 plus 20 percent of the population in cohort m in municipality 2. For Brazilian data, the merger between the rings and the old census data is only feasible at the municipal level.

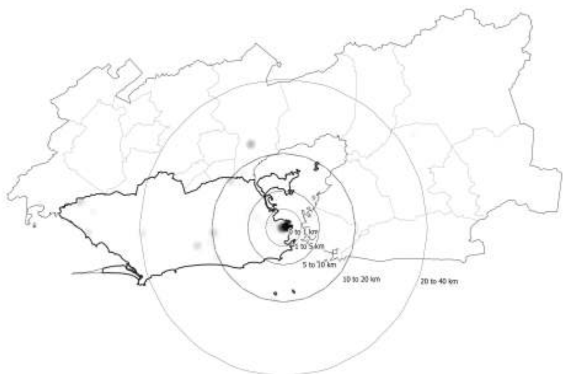
[12] As in Moretti (2004a), the weights are estimated using data from the entire population, since the age structure of the labor force may be endogenous.

Table 1. Descriptive statistics of concentric ring employment variables

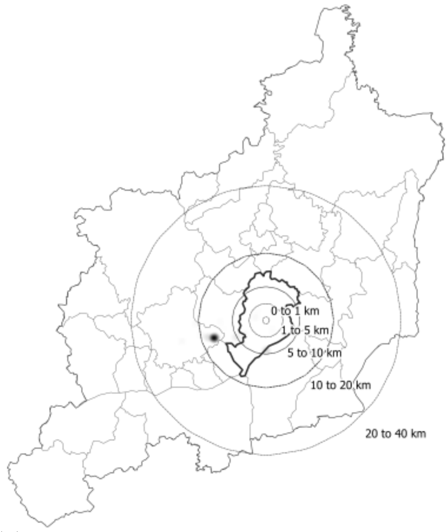
	2006		2014		All sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
# of workers						
Within 0 to 1 km	568.33	1,241.33	573.00	1,238.43	595.26	1,332.80
Within 1 to 5 km	10,077.43	15,398.52	10,109.23	14,596.50	10,540.13	15,800.01
Within 5 to 10 km	24,964.54	38,857.74	25,077.06	36,588.83	26,157.23	39,677.05
Within 10 to 20 km	69,756.78	103,448.40	70,687.17	98,222.49	73,288.39	105,977.80
Within 20 to 40 km	130,705.50	178,786.20	137,873.50	179,009.90	140,154.60	188,226.70
College-or-more, 0 to 1 km	44.78	181.39	70.75	323.24	58.82	282.45
College-or-more, 1 to 5 km	846.75	2,018.22	1,329.71	2,965.81	1,107.00	2,608.39
College-or-more, 5 to 10 km	2,227.10	4,918.54	3,497.00	7,027.02	2,919.03	6,256.79
College-or-more, 10 to 20 km	6,630.83	12,332.47	10,321.55	17,513.41	8,651.46	15,616.67
College-or-more, 20 to 40 km	12,118.67	20,452.63	19,990.25	30,525.30	16,324.91	26,688.37
# of plants						
Within 0 to 1 km	22.77	47.57	24.21	44.68	23.76	47.76
Within 1 to 5 km	417.16	674.58	440.14	645.45	433.58	673.00
Within 5 to 10 km	1,021.34	1,624.74	1,074.58	1,574.11	1,060.68	1,624.94
Within 10 to 20 km	2,741.58	4,037.25	2,897.68	3,975.69	2,845.82	4,058.87
Within 20 to 40 km	4,702.92	6,571.15	5,144.90	6,750.07	4,972.65	6,750.83
# of cells	14,030		16,094		15,099[a]	

Notes: The number of cells in Brazilian metropolitan areas represents the cells with at least one plant. In 2006, for example, our analysis encompasses 14,030 km^2 of urban areas. [a]: average number of cells. Source: Author' computations using information from RAIS.

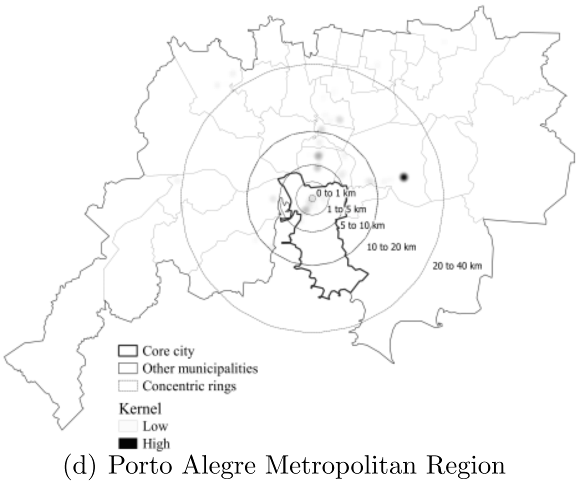
(a) São Paulo Metropolitan Region



(b) Rio de Janeiro Metropolitan Region



(c) Belo Horizonte Metropolitan Region



(d) Porto Alegre Metropolitan Region

Figure 1. Distribution of workers with college-or-more within select metropolitan regions

Notes: Kernel density is estimated using workplace data for workers with college-or-more in 2014. We selected a cell within the core city from which we defined five concentric rings around its centroid.

4 Results

In this section we present and discuss the results. The remaining subsections present our main results about the attenuation of human capital spillovers.

4.1 Spatial scope of human capital spillovers

As pointed out earlier, in the previous subsection we explored a mix of factors that may be correlated with worker productivity. In this subsection we focus on the attenuation of external returns to education. For this, we consider only the number of workers with college degree or more in each concentric ring. The social return to education (private return plus external return) can be associated with different factors, for example, crime reduction, more informed political decisions when voting and enhanced productivity of other workers (see, e.g., Lochner and Moretti, 2004; Milligan et al., 2004; Moretti, 2004b). We focus exclusively on the third example, after discounting the private returns. Proximity to educated workers can enhance productivity of other workers (Moretti, 2004a). Proximity determines the intensity of the effect we are evaluating. So, we now consider only the four closest rings of the individual's workplace, i.e., 0-1, 1-5, 5-10 and 10-20 km. There are two main reasons. First, the human capital externalities occur mainly at short distances (Fu, 2007; Rosenthal and Strange, 2008, 2020; Li et al., 2020b). Second, as presented in Figure 1, part of the 20-40 km ring is outside the metropolitan regions.

Here we provide only the results obtained from the restricted sample (Table 2). Column 1 reports the pooled OLS estimates. In column 1, we address the spatial classification into observables by controlling for the individual employee characteristics presented above; we also control for observed plant-level heterogeneity, to be precise, controlling for plant size; add time-varying industry-specific effects at the 2-digit level (controlling for industry-specific productivity shocks that vary across years and can affect worker earnings); and we add the metropolitan region fixed effect to control for heterogeneity across geographic areas. The coefficients of the first three rings are positive and strongly significant while the coefficient associated with the 10-20 km ring is negative and strongly significant. This exercise indicates a pattern of decay of the externalities generated by the number of college-educated workers, stronger in the range of 0-1 km and smaller in the other rings.

To understand how the pattern of spatial attenuation changes as we include the heterogeneities highlighted above, we also estimate simpler models without controlling for plant, industry-year, and metropolitan region heterogeneities. There are differences between these simple initial specifications, particularly with regard to the magnitude of the coefficients.[13] But a common feature in all is that the positive association between the number of high-skilled workers and wages are greater at short distances (0-1 km).[14] Note also that the coefficient associated with the 10-20 km ring becomes negative when we include the region fixed effects (column 1 in Table 2). One possible interpretation is that two effects can act simultaneously within each ring, a positive one associated with external return to education and a negative one associated with competition between different locations. Nevertheless, when we control for unobserved heterogeneity fixed in time at the metropolitan region level (comparing rings within the same metropolitan region), the positive effect is strong enough to offset the negative only at short distances (up to 10 km).

Since human capital externalities occur mainly through interaction among workers, the spatial decay can be affected (e.g., be stronger) when the frequency of contacts reduces rapidly with distance (Rosenthal and Strange, 2020). So, the heterogeneity of structure and provision of public services in urban areas can help to understand which occurs when the distance from and individual's workplace increases. Consequently, the local public services, including the provision of public transport infrastructure, can affect the interaction of workers further apart. Notice, however, that these factors may help to understand why geographical proximity is important, but do not provide an interpretation for the negative effects. One interpretation of a negative effect, which is also associated with the structure of cities, is the competition between different locations (Håkansson and Isacsson, 2019). The expansion of the number of high-skilled

[13] Due to space limitations, we have not included this results. This information can be made available upon request.

[14] All coefficients associated with the 0-1 km ring are positive and strongly significant. The intensity of decay between ranges, however, is very different. For example, in models OLS 1 to 3 in Table ??, the 0-1 km coefficient for proximity to college-educated workers is, on average, 4 times larger than the corresponding 1-5 km effect and 27.7 times larger than the corresponding 5-10 km effect.

workers at greater distances from an individual's current establishment can be associated with a reduction in the number of workers around the current establishment.

Table 2. Spatial scope of human capital spillovers

Dependent variable: individual hourly wage (in log)			
# of workers with college-or-more	OLS (1)	FE (2)	FE + IV (3)
0 to 1 km	2.49e-05***	6.68e-06***	6.78e-05***
	(4.31e-07)	(3.28e-07)	(6.25e-06)
1 to 5 km	9.04e-06***	-2.57e-07	1.95e-05***
	(1.62e-07)	(2.24e-07)	(2.44e-06)
5 to 10 km	2.21e-06***	-6.90e-07***	1.27e-05***
	(9.43e-08)	(1.44e-07)	(1.55e-06)
10 to 20 km	-3.09e-07***	-5.75e-07***	-1.06e-05***
	(5.20e-08)	(9.08e-08)	(6.62e-07)
Worker-level controls	Yes	Yes	Yes
Plant-level controls	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes
Metropolitan region FE	Yes	Yes	Yes
Kleibergen-Paap rk F[a]			1,466.29
Kleibergen-Paap rk LM[b]			5,346.92
1 st stage F-stat. 0 to 1 km			2,363.64
1 st stage F-stat. 1 to 5 km			6,652.19
1 st stage F-stat. 5 to 10 km			20,791.46
1 st stage F-stat. 10 to 20 km			60,470.27
F-stat.	21,639.14	–	–
R squared	0.6347	0.3827	–

Notes: This table presents the estimates obtained from equation 9 when we consider the number of college-educated workers in each ring. All models are estimated with 2,387,434 observations. Worker-level controls include all the individual characteristics detailed above. Plant-level controls are dummies for plant size. Industry × year effects are dummies for each 2-digit × year combination. Metropolitan region FE are metropolitan region fixed effects. The 1 st stage F-statistic is the F test of excluded instruments. [a]: H_0 - weakly identified model. [b]: H_0 - under-identified model. Standard errors adjusted for clustering are in parentheses. Significance levels: *** $p < 0.01$. Source: Prepared by the author based on estimates.

In columns 2 and 3 we present the results after including the individual-specific fixed effect (FE) and use the shift-share instrument for the number of college-educated workers in the four distance bands presented in equation (10) (FE + IV). The estimates with individual-specific fixed effect (here any individual permanent characteristics are controlled) and all other controls, but without instrumental variables, indicate the importance of geographic proximity of high-skilled workers only 1 km from the individual's workplace. For high-skilled workers farther away from an individual's current establishment, there may be a kind of competition, as in the result for the last concentric ring in column 1. Although all these controls for observable and non-observable variables are important for us to obtain cleaner effects, as mentioned, we will use the nationwide exogenous growth of college-educated workers in a shift-share design to instrument the number of workers with college degree or more in each concentric ring. As can be seen at the bottom of column 3, using the instrumental diagnostic tests mentioned above, the hypotheses of weak instruments and under-identification are strongly rejected. Additionally, the first stage estimates, even after including all control variables and fixed effects presented above, the shift-share IV has strongly significant explanatory power.

The qualitative results regarding the spatial extent of human capital externalities in column 3 resemble those in column 1, but the coefficients for the three rings closest to the current establishment are, on average, 3.5 times larger, indicating the existence of a negative bias in the estimation when we assume exogeneity of our human capital variables. The results of the FE + IV model show that if 1,000 college-educated workers are added at distance up to 1 km (approximately equivalent to the 10/90 spread), wages of workers would increase, on average, by 6.78 percent. On the other hand, if the same number of workers were added to the 1-5 km or 5-10 km range, the wages of workers would increase, on average, by 1.95 and 1.27 percent, respectively.

In column 2, a clear exception is the coefficient of the third ring. The coefficient obtained by the fixed effects estimator without instrumental variables is negative, which may be associated with some kind of simultaneous bias farther from the individual's workplace. These details may indicate it is important to consider endogeneity of proximity of college-educated workers in the wage earnings-human capital spatial distribution relationship. So, summing up all the results of Table 2, the main impression is that in Brazilian cities the external returns to education are positive up to 10 km from the individual's workplace. For high-skilled workers farther away from an individual's current establishment, there may be a kind of competition, as indicated by the results in columns 1-3. These results remain strongly significant even after controlling for the private return to education and other observable characteristics at worker level and plant level, along with productivity shocks specific to industry $\times$ year, metropolitan region and worker fixed effects, and using to instrumental variables.

In general, our main results are in line with previous evidence in the literature. But unlike most of previous studies, as we have discussed so far, we do not assume that the external return to education is homogeneous within a city. Instead, it attenuates rapidly with distance. At more aggregate levels, for example, both for workers (relating local human capital stocks to wages) and for firms[15] (using TFP), geographic proximity of highly skilled workers increases productivity. Moretti (2004a), for example, found that the elasticity of wages to the share of college graduates at the Metropolitan Statistical Area (MSA) level in US was around 1.2, with small variations with different specifications. More recently, Chauvin et al. (2017) indicated that the elasticity was 3.0 to 4.7 for Brazil, 5.2 to 7.2 for China, and 1.9 to 3.2 for India. For all three developing countries, the coefficients are higher than in the US. Similar empirical evidence can be found in other studies on the magnitude of human capital spillovers using different strategies to deal with the endogeneity of aggregate human capital (see, e.g., Moretti (2004b), and more recently Carlino and Kerr (2015) for a detailed review).

The contribution of this paper, however, also provides new insights into these effects within the Brazilian cities, i.e., when our lens narrows to the neighborhood level. In the latter aspect, the evidence is almost exclusively for developed countries. For example, Fu (2007) found that knowledge spillovers were very localized, occurring mainly and strongly at short distances in the Boston Metropolitan Area, within round 2.4 km (in models with 5 rings), which the author called "Smart Café Cities". Rosenthal and Strange (2008) found similar results for the whole US. More specifically, when college-educated workers were less than 8 km away from an individual's workplace, they generated a greater external return than those college-educated workers who were more than 8 km away. With a different empirical strategy but still addressing the same theme, Andersson et al. (2009) evaluated the effects of spatial decentralization higher education in Sweden and found substantial and highly localized spillovers (about 5 km from the new university) in productivity gains.

In addition to providing evidence about the attenuation of human capital spillovers in a context very different from that observed in developed countries, and differently from other previous studies, such as Rosenthal and Strange (2008), our panel data combined with the exogenous shock driven by the Brazilian government's education policy shift allows us to obtain the causal effects more clearly. Furthermore, our results conform very well to the Brazilian economic environment. They are in agreement with the high level of geographic concentration of manufacturing, in particular with the correlation concentration $\times$ share of college-educated workers; with the higher inter-regional mobility of workers, which favors the formation of more specialized and dynamic industrial clusters; and with the high level of educational disparities observed in the country.

4.2 Evidence for different education groups

In the previous subsection, we assumed that the estimates of human capital spillovers are the same for different employees, i.e., when all skill groups are pooled together. The results obtained are, therefore, an average effect across education groups. There are different reasons why this simplification may not be valid. As predicted by a conventional demand and supply models, the effects for less educated workers tend to be greater. Nevertheless, in spillover models, both types of workers can gain from the local increase in the number of college-educated workers. In particular, the less educated workers benefit even in the absence of any spillovers, while on the other hand, the effect on the wages of college-educated workers depends on the existence of the spillovers (see Moretti, 2004a). Thus, to explore these possible differences, we estimated equation 9 by separating the sample into two education groups: less than college degree and college degree or higher.

Table 3 reports the results. In Panel A we follow the same structure previously presented in Table 2, but our outcome variable is only the wages of less educated workers (less than college degree). Similarly, in Panel B we consider college-educated workers. Four important patterns emerge from this table. First, in

[15]See, e.g., Moretti (2004c) and Liu (2013), who related local human capital stocks to TFP in the US and China, respectively.

both cases the proximity to college-educated workers increases an individual's wage. This is indicated by the positive and highly significant coefficients in the first distance band in all columns and for second and third distance bands in most specifications. Second, in both cases it is important to consider the spatial extent of the spillover effect. This effect is more pronounced up to 1 km from an individual's current establishment. Third, comparing the results in the first ring of Panel A with those of Panel B, we can observe a pattern according to a model that includes both conventional demand and supply factors and spillovers. In particular, a greater effect (2.3 times) exists for less educated workers in the first distance band. An exception is column 2, with fixed effects estimation without instrumental variables, where the estimated coefficient is higher for college-educated workers. As mentioned regarding the results in Table 2, in the first distance band the qualitative results in columns 2 and 3 are not different, but in the other concentric rings there are clear differences, highlighting the importance of using instrumental variables. Fourth, according to the results for the second ring, college-educated workers can benefit from greater closeness to other college-educated workers (columns 1 and 3).

Table 3. Spatial scope of heterogeneity of human capital externalities by education groups

	Dependent variable: individual hourly wage (in log)					
# of workers with college-or-more	OLS (1)	FE (2)	FE + IV (3)	OLS (1)	FE (2)	FE + IV (3)
	Panel A: Less than college degree			**Panel B: College degree or more**		
0 to 1 km	5.08e-05***	4.59e-06***	1.10e-04***	3.09e-05***	6.88e-06***	4.58e-05***
	(5.91e-07)	(4.34e-07)	(1.30e-05)	(1.12e-06)	(8.62e-07)	(7.19e-06)
1 to 5 km	4.64e-06***	-3.19e-07	2.63e-05***	1.44e-05***	-1.10e-06	4.10e-05***
	(1.77e-07)	(2.56e-07)	(2.97e-06)	(4.74e-07)	(6.86e-07)	(1.03e-05)
5 to 10 km	2.11e-06***	-4.21e-07**	1.03e-05***	2.81e-06***	-1.62e-06***	-4.25e-06
	(1.00e-07)	(1.65e-07)	(2.44e-06)	(3.06e-07)	(4.17e-07)	(5.88e-06)
10 to 20 km	-1.58e-07***	-7.46e-07***	-1.09e-05***	-1.31e-06***	-2.21e-07	-6.99e-06***
	(5.52e-08)	(1.03e-07)	(7.95e-07)	(1.84e-07)	(2.80e-07)	(1.49e-06)
Kleibergen-Paap rk F[a]			17.39			41.56
Kleibergen-Paap rk LM[b]			69.21			162.73
1 st stage F-stat. 0 to 1 km			1,353.64			640.42
1 st stage F-stat. 1 to 5 km			5,606.02			1,477.87
1 st stage F-stat. 5 to 10 km			16,936.74			2,423.80
1 st stage F-stat. 10 to 20 km			34,172.36			4,161.27
R squared	0.5241	0.3664	–	0.4378	0.3484	–
Observations	2,003,730	2,003,730	1,994,174	187,511	187,511	187,511
Worker-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Plant-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year effect	Yes	Yes	Yes	Yes	Yes	Yes
Metropolitan region FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The industry × year effect is computed at the 2-digit level. All the controls shown at the bottom of this table are included in both panels A and B. The 1 st stage F-statistic is the F test of excluded instruments. [a]: H_0 - weakly identified model. [b]: H_0 - under-identified model. Robust standard errors in parentheses. Significance level: ** $p < 0.05$, *** $p < 0.01$. Source: Prepared by the author based on estimates.

Particularly, there are significant differences in the attenuation pattern between the first and second distance bands between panels A and B when using OLS estimators (column 1). For example, while in Panel A the coefficient of the first ring is 10.9 times larger than the coefficient of the second ring, in Panel B the coefficient of the first ring is 2 times larger than the coefficient of the second ring. This difference, however, is smaller when using FE + IV estimators, as can be seen in column 3 of both panels. When moving farther away from the individual's workplace, the decay of the effect from the second to the third ring is greater for the college-educated group in columns 1. Moreover, when we use instrumental variables, there is no effect in the third ring for the college-educated group. That is, for the less educated group, the human capital externalities are positive and strongly significant up to 10 km. On the other hand, for the college-educated group, human capital externalities occur up to 5 km. A general consideration on these observations is that the attenuation between the first and second ring is larger for the less educated group, and for college-educated workers, in the 1-5 km range external returns are stronger than those observed for less educated workers.

To be more specific, from the results in column 3 for the less educated group (Panel A), the addition of 1,000 college-educated workers in the 0 to 1 km ring implies that wages of a less-educated workers would increase, on average, by 10 percent. For the college-educated sample, the corresponding effect is 4.6 percent. When we look at the second distance band, adding the same number of college-educated workers, wages of less educated group would increase, on average, by 2.6 percent while the wages of the highly skilled group would increase by 4 percent. In the third ring, as mentioned, there is significant effect only for the less educated group, of 1 percent. A possible interpretation of these results is that, as established in the supply and demand models, highly skilled workers are not perfect substitutes for less educated workers, so although they are geographically close, they are not close competitors. On the other hand, they can still generate externalities from the sharing of ideas, which can generate new products and productive processes or improve existing ones. For educated workers, however, geographic proximity (up to 1 km) to other equally skilled workers may generate greater competition since they are perfect substitutes, but when moving further away from the individual's workplace (from 1 km), the competition effect is less.

While these comparisons are interesting, the main findings of this subsection are that the attenuation patterns found in the previous subsection remain largely valid regardless of the subgroup in our sample. This indicates that external returns to education decay with geographic distance within the same city regardless of worker type.

4.3 Robustness checks

In this subsection we report estimates for alternative specifications to check the robustness of our main results. These results were previously reported in column 3 of Table 2 and refer to the pattern of attenuation of human capital spillovers. That is, when we used instrumental variables in addition to a broad set of controls to deal with the potential endogeneity of our human capital variables. So, all results presented below are obtained from FE + IV estimators. Basically, our robustness check consists of exploring what happens to our results about the attenuation patterns when we change and/or add other controls, as well as when we restrict the sample by metropolitan regions and by industries.

One of the most immediate ways to test the robustness of our main results is, for example, to control for industry-specific trends at the 3-digit rather than 2-digit level. The higher the industrial aggregation (e.g., 2-digit), the greater the grouping of different sectors will be, so that specialized industry-specific trends may not be captured. For example, in the official classification of economic activities in Brazil (CNAE), some 2-digit sectors include 3-digit industries with different technological intensity (see Cavalcante, 2014) and therefore have workers with different skills and education. So, to check whether our results are sensitive to these effects, we include 3-digit industry-year specific effects. The results are provided in column 1 of Table 4. Note that although there are small variations in the magnitude of the coefficients, all remain strongly significant and have the same pattern highlighted above. This indicates that our results are not influenced when we control for 3-digit industry-specific productivity shocks that vary over time.

Table 4. Robustness checks

Dependent variable: individual hourly wage (in log)						
# of workers with college-or-more	FE + IV					
	(1)	(2)	(3)	(4)	(5)	(6)
0 to 1 km	5.59e-05***	3.12e-05***	6.60e-05***	6.78e-05***	4.28e-05***	2.43e-04***
	(6.66e-06)	(6.14e-06)	(6.35e-06)	(6.25e-06)	(5.62e-06)	(3.82e-05)
1 to 5 km	3.10e-05***	2.85e-05***	2.20e-05***	1.95e-05***	2.68e-05***	2.13e-04***
	(2.42e-06)	(3.64e-06)	(2.45e-06)	(2.45e-06)	(4.73e-06)	(2.58e-05)
5 to 10 km	7.76e-06***	1.43e-05***	1.23e-05***	1.26e-05***	6.86e-06***	7.42e-05***
	(1.45e-06)	(1.71e-06)	(1.48e-06)	(1.74e-06)	(1.81e-06)	(7.31e-06)
10 to 20 km	-9.29e-06***	-1.26e-05***	-1.07e-05***	-1.06e-05***	-5.60e-06***	-3.37e-05***
	(5.78e-07)	(6.18e-07)	(6.56e-07)	(6.96e-07)	(9.40e-07)	(3.44e-06)
Worker-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Plant-level controls	Yes	Yes	Yes	Yes	Yes	Yes
2-digit × year FE	No	No	Yes	Yes	Yes	Yes
3-digit × year FE	Yes	Yes	No	No	No	No
Metrop. region FE	Yes	Yes	No	Yes	Yes	Yes
Worker-plant FE	No	Yes	No	No	Yes	Yes
Worker-Metrop. FE	No	No	Yes	No	No	No
City population	No	No	No	Yes	No	No
Low-schooling workers	No	No	No	No	Yes	Yes
Transport × year FE	No	No	No	No	No	Yes
Kleibergen-Paap rk F[a]	1,050.46	635.25	1,449.44	1,385.05	1,192.05	30.37
Kleibergen-Paap rk LM[b]	3,867.09	2,561.33	5,311.26	5,086.24	1,643.85	117.77
Observations	2,387,434	2,223,550	2,378,812	2,387,434	2,223,550	2,191,252

Notes: Due to computational limitations, we do not report the F test of excluded instruments in the first stage. Robust standard errors in parentheses. [a]: H_0 - weakly identified model. [b]: H_0 - under-identified model. Robust standard errors in parentheses. Significance level: *** $p < 0.01$. Source: Prepared by the author based on estimates.

Another interesting issue refers to match-specific productivity. Although evidence has been reported that the worker-plant match can influence estimates in wage equations (see, e.g., Krishna et al. (2014) and Araújo and Paz (2014) for Brazil), studies of the external returns to education generally do not control for this effect. As highlighted by Woodcock (2015), in the absence of distortions, good workers will match with good firms, and if match-specific productivity is important in wage determination, the absence of this effect on wage regression can generate biased estimates of the returns to observable worker and firm characteristics. For example, in equation 9, firms' unobservable time invariant characteristics can influence our estimates. Firms self-selection of larger cities may arise if only the most productive firms survive in large urban centers. This additional concern can be addressed directly when we modify our wage equation to include firms' fixed effects by considering a wage specification similar to Abowd et al. (1999). We check if our results change when we control for worker-plant matched fixed effects (or job-spell fixed effects). The results reported in column 2 show that, in terms of attenuation, there is no changes. All coefficients still remain strongly significant and larger at shorter distances, particularly in the first ring, attenuating up to 10 km. Note, however, that the coefficient of the first ring is less than those of the other models (e.g., the 0-1 km ring coefficient in column 1 is 2 times greater).

As discussed earlier, workers choose the city where their skills are most valued. We deal with this problem using the exogenous expansion of higher education in Brazil. Another way to test the robustness of our instruments is to include worker-metropolitan-region matching effects. That is, everything that is specific to a worker-metropolitan-region pair is absorbed by the fixed effect. In this case, the variation that comes from movers is absorbed and the identification is based on stayers and comes from changes in the number of college-educated workers in each ring over time (Moretti, 2004a). It is expected that the results will not be highly sensitive to the inclusion of worker-metropolitan-region matching effects. Otherwise, if the results are highly sensitive, doubts would be cast on the validity of our instruments. Conditional estimates of worker-metropolitan-region matching are reported in column 3. There is no evidence that unobserved individual ability and return to unobserved ability across cities affect the attenuation results.

We also can evaluate what happens to the results when we include simultaneously other effects of agglomeration in each ring. The human capital variables are usually correlated with density (Combes and Gobillon, 2015). Therefore, other effects besides human capital spillovers can be captured by human capital variables when not controlling for the presence of other types of workers. To test the robustness of our results, we proceed in two ways: (i) in column 4 we include the city population as a control for these mechanisms, assuming that they act homogeneously within the same city; and (ii) in column 5 we include worker-plant matching fixed effects and the number of low-schooling workers in each ring to control for the presence of possible gains from density that vary with distance. Rosenthal and Strange (2008), for example, showed that the number of low-schooling workers has a negative effect on wages. Here we do not report the estimates (both for population and low-schooling workers) because these variables are potentially endogenous and therefore should be analyzed only as a robustness test. We conclude that other effects associated with the city size and concentration of low-schooling workers at different distances from the individual's workplace are not a major sources of bias. Our main results remain largely robust to these effects.

The transportation infrastructure around establishments can affect worker productivity. So, here we also include the same control variables for the transportation infrastructure around the individual's workplace (distance in kilometers from the cell's centroid to the nearest railroad, federal highway, state highway, airport and port). But unlike before, here we interact these time-invariant controls with the time effect to capture trends specific to each cell's transportation infrastructure improvement. In column 6, we include these control variables with worker-plant matching fixed effects and number of low-schooling workers effects. As can be seen, the estimated coefficients remain largely significant, and more importantly, again show that the pattern of attenuation remains robust, as expected.

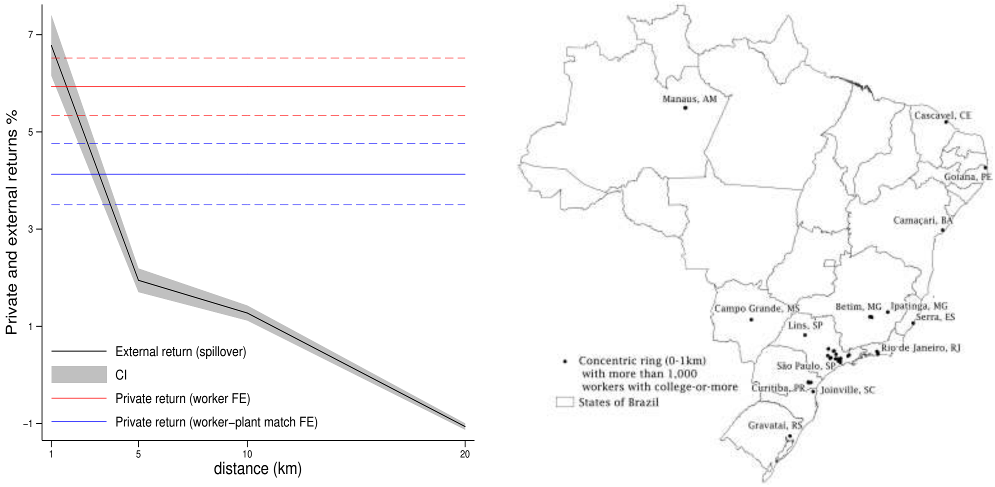
4.4 External versus private returns to education

So far we have presented estimates for the external return to education. In this subsection we compare the external returns to education estimated in the previous subsections with the private returns to education. Following Rosenthal and Strange (2008), we report the estimates for the private returns to education in Table 5, where we omit the agglomeration variables but retain all other controls. This also allows us to compare the relationship between private and external returns to education in Brazil with those obtained by those authors for the US. To do this, in column 1 the estimates are obtained by OLS. Consistent with the literature on the private returns to education, the incremental contribution of a college degree beyond that of a high school diploma (complete high school) on an individual's wage is, on average, 56.26 percent, a larger private return than that obtained by the authors for the US (roughly 30 percent). However, when we control for the worker or worker-plant matched fixed effects (columns 2 and 3), the incremental gain is much smaller, roughly 6 and 4 percent, respectively. This is expected, given that the unobserved heterogeneity of workers and worker-plant matching omitted in the OLS estimates are biasing the private return upward.

Table 5. Private returns to education

	OLS	Worker FE	Worker-plant FE
Dependent variable: individual hourly wage (in log)	(1)	(2)	(3)
Illiterate (reference category)			
Incomplete primary school	0.0608***	-0.0127**	-0.0095
	(0.0047)	(0.0057)	(0.0061)
Incomplete high school	0.1823***	-0.0142**	-0.0114*
	(0.0047)	(0.0057)	(0.0061)
Complete high school	0.3142***	-0.0160***	-0.0105*
	(0.0047)	(0.0058)	(0.0061)
Incomplete college	0.6366***	0.0042	0.0036
	(0.0050)	(0.0061)	(0.0065)
College degree or more	0.8772***	0.0433***	0.0308***
	(0.0050)	(0.0059)	(0.0063)
Worker-level controls	Yes	Yes	Yes
Plant-level controls	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes
Metropolitan region FE	Yes	Yes	Yes
F-stat.	22,173.68	3,453.35	–
R squared	0.6389	0.3825	0.3622

Notes: This table presents the estimates obtained from equation 9 when we omit the agglomeration variables. All models are estimated with 2,387,434 observations. Worker-level controls include all the individual characteristics detailed above. Plant-level controls are dummies for plant size. Industry × year effects are dummies for each 2-digit × year combination. Metropolitan region FE is metropolitan region fixed effects. Standard errors adjusted for clustering are in parentheses. Significance levels: *** $p < 0.01$. Source: Prepared by the author based on estimates.



(a) Comparison

(b) Concentric ring (0 to 1 km)location in 2014

Figure 2. External versus private returns

Two interesting patterns emerge from this evidence. First, as we have shown in the previous sections, adding 1,000 college-educated workers within 1 km would increase an individual's wage by roughly 6.8 to 24 percent, depending on the included controls (considering both the results in column 3 in Table 2 and those in Table 4). These effects are comparable to 12 to 43 percent of incremental private returns associated with obtaining a college degree in the OLS model. The equivalent percentage for the US is 20 to 50. That is, looking only at the OLS results, the external return measured as a share of the private returns to education is lower in Brazil. On the other hand, when we analyze the results of private returns to education conditional on the worker or worker-plant matching fixed effect, in most cases the external returns to education exceed the private returns, but only at short distances.

Consider the example in Figure 2 (a), where we plot the average percent change (with confidence interval - CI) in workers' wages given an increase of 1,000 college-educated workers in different distance bands (results of column 3 in Table 2). To compare external versus private returns, we also plot the average percentage change in wages associated with obtaining a college degree following high school in both worker FE (solid red line with dashed lines representing the CI) and worker-plant matching FE models (solid blue line with dashed lines representing the CI). Note that for distances up to 1 km from the current establishment, human capital spillovers can outperform private returns, provided there is a certain increase (minimum 900) in the number of workers with college degree or higher. These results are particularly interesting in the context of public education policies and the efficiency of investments in education, because they suggest that the external return not only contributes to increasing the social returns to education, but also may be greater than the private return at short distances. We also report where these effects can occur in Figure 2 (b), which shows the location of 121 rings (0 to 1 km) in 2014 with 1,000 or more college-educated workers. Most of the rings are in (or near) the SPMR (73) and the remaining ones are located in industrial clusters in the other regions of the country (e.g., Campo Grande, MS; Camaçari, BA; Manaus, AM; and Joinville, SC). Moreover, these results are also in line with Moretti (2004a), who also found similar results with aggregated geographic data.

5 Concluding remarks

The objective of this paper has been to analyze the spatial extent of human capital spillovers within Brazilian cities. Some studies have explored this topic, but they are almost exclusively for developed countries, as cited throughout the text. Beyond the lack of evidence, the present examination contributes to the discussion on the subject by evaluating this phenomenon in an economic environment very different from that of developed countries. For this purpose, we have exogenously divided the all Brazilian geographic areas into cells of one square kilometer and used a unique and rich microgeographic panel dataset to calculate the number of college-educated workers in four different distance bands from the geographic centroid of each cell. In addition to using panel data and a broad set of controls for observed and unobserved heterogeneities, our identification strategy is based on a shift-share IV for the federal government's education policy shocks in Brazil in the period 1991-2004.

The main set of results provides more detailed evidence when our focus is on the externalities generated by the concentration of college-educated workers. The proposal to isolate this specific effect aims to assess how the external return to education is attenuated with the distance from the current establishment. We used the exogenous expansion of public education in Brazil over the past two decades as an instrument for the number of college-educated workers in each of the rings, besides a broad set of controls including unobservable characteristics of workers, plants, industry, metropolitan region and worker-plant matching to deal with potential endogeneity. The external returns from education are also highly localized and therefore consistent with the idea that interaction between workers (face-to-face) can generate productivity gains from knowledge spillovers. We also found evidence that unskilled workers can obtain higher returns by being spatially close to skilled workers, in line with demand and supply models with spillover (Moretti, 2004a).

The evidence provided here is very consistent with the characteristics of the economic environment in Brazil. We can highlight some of these characteristics with which our results conform very well. The highly localized human capital spillovers we find are consistent with the high geographic concentration of the manufacturing. It is also consistent with the positive correlation of concentration $\times$ share of college-educated workers found by Almeida et al. (2020). The larger effect at very short distances also conforms very well with the absence of restriction on worker mobility, which favors the formation of highly specialized and dynamic local labor markets; and with the low quality of urban infrastructure (e.g., public transportation), which can hinder interaction.

The results presented here are clearly in line with the consensus that urban activities involve increasing returns and hence become more efficient than can be attained in isolation. But it also provides insight on the spatial pattern of the effects within cities in a neighborhood context. The importance is clear of a better understanding of these forces for public policymaking. Urban infrastructure can play an important role, for example, by providing the inputs for collaboration among more distant workers. In developing countries like Brazil, this importance is even greater, since the cities in general have greater structural problems.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9–59. https://doi.org/10.1086/654403.

Adao, R., Kolesár, M., and Morales, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Almeida, E. T., Silveira Neto, R. M., and Rocha, R. M. (2020). Manufacturing location patterns in Brazil. *48° Encontro Nacional de Economia da ANPEC*.

Anas, A., Arnott, R., and Small, K. A. (1998). Urban spatial structure. *Journal of Economic Literature*.

Anderson, N. B. and Bogart, W. T. (2001). The structure of sprawl: Identifying and characterizing employment centers in polycentric metropolitan areas. *American Journal of Economics and Sociology*.

Andersson, M., Klaesson, J., and Larsson, J. P. (2014). How local are spatial density externalities? neighbourhood effects in agglomeration economies. *Regional studies*, 50(6):1082–1095.

Andersson, M., Larsson, J. P., and Wernberg, J. (2019). The economic microgeography of diversity and specialization externalities-firm-level evidence from Swedish cities. *Research Policy*, 48(6):1385–1398.

Andersson, R., Quigley, J. M., and Wilhelmsson, M. (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics*, 66(1):2–15.

Araújo, B. C. and Paz, L. S. (2014). The effects of exporting on wages: An evaluation using the 1999 brazilian exchange rate devaluation. *Journal of Development Economics*, 111:1–16.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–68.

Barros, R. P. d., Franco, S., and Mendonça, R. (2007a). Discriminação e segmentação no mercado de trabalho e desigualdade de renda no brasil. *TD, IPEA*.

Barros, R. P. d., Franco, S., and Mendonça, R. (2007b). A recente queda da desigualdade de renda e o acelerado progresso educacional brasileiro da última década. *TD, IPEA*.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Barufi, A. M. B., Haddad, E. A., and Nijkamp, P. (2016). Industrial scope of agglomeration economies in Brazil. *The Annals of Regional Science*, 56(3):707–755.

Billings, S. B. and Johnson, E. B. (2012). A non-parametric test for industrial specialization. *Journal of Urban Economics*, 71(3):312–331.

Blanchard, O. J., Katz, L. F., Hall, R. E., and Eichengreen, B. (1992). Regional evolutions. *Brookings papers on economic activity*, 1992(1):1–75.

Borusyak, K., Hull, P., and Jaravel, X. (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3):287–302.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and innovation. In *Handbook of Regional and Urban Economics*, volume 5, pages 349–404. Elsevier.

Cavalcante, L. R. (2014). Classificações tecnológicas: uma sistematização.

Charlot, S. and Duranton, G. (2004). Communication externalities in cities. *Journal of Urban Economics*.

Charlot, S. and Duranton, G. (2006). Cities and workplace communication: some quantitative french evidence. *Urban Studies*, 43(8):1365–1394.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98:17–49. https://doi.org/10.1016/j.jue.2016.05.003.

Coffey, W. J. and Shearmur, R. G. (2002). Agglomeration and dispersion of high-order service employment in the Montreal metropolitan region, 1981-96. *Urban studies*, 39(3):359–378.

Combes, P.-P., Démurger, S., and Li, S. (2013). Urbanisation and migration externalities in China.

Combes, P.-P., Démurger, S., Li, S., and Wang, J. (2020). Unequal migration and urbanisation gains in China. *Journal of Development Economics*, 142:102328.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 5, pages 247–348. Elsevier.

de Mendonça, R. S. P. and de Barros, R. P. (1995). A evolução do bem-estar e da desigualdade no Brasil desde 1960. *Revista Brasileira de Economia*, 49(2):329–352.

Di Addario, S. and Patacchini, E. (2008). Wages and the city. evidence from Italy. *Labour Economics*.

Duranton, G. (2016). Agglomeration effects in Colombia. *Journal of Regional Science*, 56(2):210–238.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. Elsevier.

Fishlow, A. (1972). Brazilian size distribution of income. *The American Economic Review*, 62:391–402.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics*, 61(1):86–111. https://doi.org/10.1016/j.jue.2006.06.002.

Glaeser, E. and Henderson, J. V. (2017). Urban economics for the developing world: An introduction. *Journal of Urban Economics*, 98:1–5.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Håkansson, J. and Isacsson, G. (2019). The spatial extent of agglomeration economies across the wage earnings distribution. *Journal of Regional Science*, 59(2):281–301.

Krishna, P., Poole, J. P., and Senses, M. Z. (2014). Wage effects of trade reform with endogenous worker mobility. *Journal of International Economics*, 93(2):239–252.

Larsson, J. P. (2014). The neighborhood or the region? reassessing the density–wage relationship using geocoded data. *The Annals of Regional Science*, 52(2):367–384.

Li, J., Li, L., and Liu, S. (2020a). Attenuation of agglomeration economies: Evidence from the universe of Chinese manufacturing firms. *Working paper*.

Li, J., Liu, S., and Wu, Y. (2020b). Identifying knowledge spillovers from universities: Quasi-experimental evidence from urban China. *Working paper*.

Liu, Z. (2013). Human capital externalities in cities: evidence from chinese manufacturing firms. *Journal of Economic Geography*, 14(3):621–649.

Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1):155–189.

Marshall, A. (1890). *Principles of Economics*. Macmillan London.

Milligan, K., Moretti, E., and Oreopoulos, P. (2004). Does education improve citizenship? evidence from the united states and the united kingdom. *Journal of Public Economics*, 88(9-10):1667–1695.

Moretti, E. (2004a). Estimating the external return to higher education: Evidence from cross-sectional and longitudinal data. *Journal of Econometrics*, 120(1-2):175–212.

Moretti, E. (2004b). Human capital externalities in cities. In *Handbook of Regional and Urban Economics*.

Moretti, E. (2004c). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

Narita, R., Gonzaga, G., and Firpo, S. (2003). Decomposicão da evolucao da desigualdade de renda no brasil em efeitos idade, periodo e coorte/explaining income inequality in brazil: age, period and cohort effects. *Pesquisa e Planejamento Economico*, 33(2):211–252.

Oliveira, R. C. and Silveira Neto, R. d. M. (2013). Escolaridade, políticas sociais e a evolução da desigualdade regional de renda no Brasil entre 2003 e 2011: uma análise a partir das fontes de renda. *Revista Econômica do Nordeste*, 44(3):651–670.

Oliveira, R. C. and Silveira Neto, R. D. M. (2016). Expansão da escolaridade e redução da desigualdade regional de renda no Brasil entre 1995 e 2011: progressos recentes e desafios presentes. *PPE*.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2119–2171. Elsevier.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.

Rosenthal, S. S. and Strange, W. C. (2020). How close is close? the spatial reach of agglomeration economies. *Journal of Economic Perspectives*, 34(3):27–49.

Thisse, J.-F. (2018). Human capital and agglomeration economies in urban development. *The Developing Economies*, 56(2):117–139. https://doi.org/10.1111/deve.12167.

Verstraten, P., Verweij, G., and Zwaneveld, P. J. (2018). Complexities in the spatial scope of agglomeration economies. *Journal of Regional Science*, 59(1):29–55.

Woodcock, S. D. (2015). Match effects. *Research in Economics*, 69(1):100–121.