Do Human Capital Externalities Affect Firms' Total Factor Productivity? Evidence from Brazilian Cities

Edilberto Almeida* Raul Silveira Neto†

Abstract

In this paper, we estimate the effect of local human capital externalities on firms' Total Factor Productivity (TFP) in Brazilian cities. To obtain credible results, we consider a unique set of confidential information at the plant level and a shift-share instrumental variable approach that uses the policy-driven external expansion of the human capital of cities. Our set of evidence indicates a significant positive effect of local human capital on firms' TFP, which favors the idea that interactions with skilled workers outside plants benefit firms' economic performance. This effect is underestimated when considering only the OLS estimate and survives different robustness checks, including the presence of other sources of agglomeration gains and different city sizes.
Keywords: Human capital externalities, total factor productivity, shift-share instrumental variable.

Resumo

Neste artigo, estimamos o efeito das externalidades do capital humano local sobre a Produtividade Total dos Factores (TFP) das empresas nas cidades brasileiras. Para obter resultados confiáveis, combinamos um conjunto único de informações confidenciais ao nível das firmas e uma abordagem de variável instrumental shift-share que utiliza a expansão exógena do capital humano das cidades orientada por políticas públicas. O nosso conjunto de evidências indica um efeito positivo significativo do capital humano local na TFP das empresas, o que favorece a ideia de que as interações com trabalhadores qualificados fora das fábricas beneficiam o desempenho econômico das empresas. Este efeito é subestimado quando se considera apenas a estimativa OLS e sobrevive a diferentes testes de robustez, incluindo a presença de outras fontes de ganhos de aglomeração e diferentes dimensões das cidades.
Palavras-chave: Externalidades de capital humano, produtividade total dos fatores, variável instrumental shift-share.

JEL classification: J30, L60, O40.

Área 1 - Teoria, métodos e modelos de economia regional

*Department of Economics, Federal University of Pernambuco and PIMES - UFPE. E-mail: edilberto.almeida@ufpe.br. We thank Carlos Lessa, Juarez Filho and entire CDDI/IBGE team for the availability of access to the database, Claucia Ferreira and Instituto de Pesquisa Econômica Aplicada (IPEA) for making data available. The authors acknowledge and are grateful for financial support by the Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco - FACEPE, (Grant/Award Number AMD-0117-6.03/19).

†Department of Economics, Federal University of Pernambuco and PIMES - UFPE, and CNPq Research Productivity Scholarship. E-mail: raul.silveirant@ufpe.br

1 Introduction

The externalities of human capital in production refer to productivity gains from interaction with more qualified workers. Marshall (1890) was the pioneer in pointing out the importance of these overflows of knowledge and learning in explaining the spatial concentration of productive activity. More recently, Lucas (1988) and Glaeser (1999) argued more formally in defense of these sources of gains to explain disparities in income between countries and productivity between cities, respectively. Mainly focused on the experience of developed countries, the available evidence regarding the importance of these human capital externalities in explaining the productivity of firms and workers justifies the attention given by these authors (see Combes and Gobillon, 2015 for a review).

For example, the evidence presented by Moretti (2004b) for the United States indicates that the productivity of firms located in cities with a high share of skilled workers is greater than the productivity of similar firms located in cities with low share ones. More recently, using aggregate data from the United States, Guo et al. (2018) also found that the level of human capital has a positive effects on productivity. Considering a broad set of countries and different levels of aggregation, Gennaioli et al. (2012) presented evidence on the importance of human capital to explain productivity at the regional and firm levels. The results suggest the importance of education of workers, education of entrepreneurs, and regional externalities for explaining regional development.

Despite the significant effects of general urban agglomeration on wages normally found in developing countries (Duranton, 2016; Barufi et al., 2016; Chauvin et al., 2017; Combes et al., 2020; Silva and Azzoni, 2022), there are very few studies that specifically explore the effects of human capital externalities on agents' productivity. As for the Brazilian case, Chauvin et al. (2017) and Almeida et al. (2021) explored the effect of local human capital on wages and found strong positive effects, but did not consider directly its impact on firms' productivity. The few works that have studied the influence of local human capital on firms' productivity considered the Chinese experience. Using aggregated data, Fleisher et al. (2010) provided evidence of the positive effects of human capital on the growth of the TFP. Working with firm-level data, Liu (2013) also found evidence that firms are more productive by benefiting from the concentration of skilled workers in the city in which it is installed.[1]

In the current investigation, we expand this empirical emergent literature by estimating the effects of local human capital on Brazilian firms' total factor productivity. Considering the period 2007-2017 and the Brazilian most important cities (cities with 100 thousand or more inhabitants), our investigation combines a unique confidential disaggregated data at the level of the firm's production function from Brazilian Institute of Geography and Statistics (IBGE), microdata from RAIS (*Relação Anual de Informações Sociais*, from the Minsitry of Labor and Emplloyment), and a shift-share instrument for the local human capital in a two-step approach. In the first step, our final database allows us to follow the firm's technology in time as well as to use several variables of controls for observed firms' characteristics to estimate their TFP. Next, we use a shift-share regression design considering the local human as the main explanatory variable (possibly endogenous) for firms' TFP, and a shift-share instrument based on the exogenous expansion of

[1]Parallel to the approach of the influences of the human capital concentration on TFP and wages, there are researches with the objective of identifying the urban wage premium. In this context, the effects of economic density on the local wage are studied (see, e.g., Glaeser and Mare (2001); Wheaton and Lewis (2002); Combes et al. (2008, 2010, 2011); Heuermann et al. (2010); Groot et al. (2014) and Duranton (2016)).

schooling in Brazilian cities generated by general government policies.

Notice that, beyond the scarcity of evidence for developing countries, the Brazilian context has some characteristics that make the study particularly appealing. Firstly, as compared with other large developing countries, such as China, Brazil has a very high urbanization rate, around 85% (Chauvin et al., 2017), which is probably associated with the absence of restrictions on worker mobility. Since agglomeration gains may be limited in the presence of restrictions on spatial mobility (Au and Henderson, 2006a; 2006b), this characteristic can play an important role in the geographical distribution of economic activity, for example, favoring more market-oriented allocations and affecting the magnitude of agglomeration gains, including human capital spillovers. Actually, recent evidence for Brazil presented by Almeida et al. (2022) indicated strong temporal persistence and higher localization patterns of manufacturing activities than those documented for other developing countries such as China and Russia as well as for developed countries. In addition, Almeida et al. (2022) also found that the local share of college-educated workers is positively associated with the degree of spatial concentration of industries, a finding that suggests that the human capital spillovers may be behind the firms' productivity gains associated with the agglomeration.

Our set of results indicates a significant positive effect of local human capital on firms' TFP. This effect is underestimated when considering only the OLS estimate and robust when also considering other sources of agglomeration gains (associated with the general population density and measures of productive diversity and specialization), different time periods, and exclusion of larger cities (São Paulo and Rio de Janeiro, cities with more than 1 million inhabitants, and capitals of States).

The remainder of the paper is organized as follows. In the next section we present a general spatial equilibrium model that supports our empirical strategy. In sections 3 and 4, we present our econometric strategy and describe our data set, respectively. Finally, the last two sections discuss the results and final considerations, respectively.

2 Conceptual Framework

Similar to Moretti (2004b), we motivate the investigation by presenting a simply analytical framework of based on Roback (1988) spatial equilibrium model. The more simple structure consists of two cities (A and B), two types of workers (skilled and unskilled), two types of goods (one nationally traded commodity denoted by x and locally traded land denoted by z), competition in cities, and fully spatially mobile workers and firms.

Workers maximize utility by choosing the amount of land and of compounded good subject to budget constraint. Given that the composite good is marketed nationally, its price is the same in both cities. Thus, the cost of living depends only on the change in the price of land, p_z, which is the same for all workers in the same city. Solving the problem of workers, the indirect utility function associated for skilled and unskilled workers is given by $V_H(w_H, p_z)$ and $V_L(w_L, p_z)$, respectively.

Consider the production function given by $y = f(A, H, L, K)$ where A is the TFP, H are hours worked by skilled workers, L are hours worked by unskilled workers and K is the physical capital. To introduce the possibility of human capital spillovers, assume that TFP is a function of the aggregate level of human capital in the city given by $A = f(S)$. Where S is the aggregate level of human capital.

To allow the number of skilled workers to be different between cities, Moretti (2004b) assumes that demand for skilled workers is greater in city B, for example. In this case, cities are identical in terms of amenities but differ in technology (see Moretti, 2004a).[2] As the workers are perfectly mobile, the more skilled migrate to city B seeking higher wage. I turn, the average education in city B grows. Even if there are no spillovers, the wages of the two categories of workers increase, the wages of skilled workers increase due to higher productivity, while unskilled workers increase due to complementarity, since they are not perfect substitutes (Katz and Murphy, 1992). In the absence of spillovers effects, $\frac{\partial f(S)}{\partial S} = 0$.

Otherwise, in the presence of spillovers effect, $\frac{\partial f(s)}{\partial S} > 0$. As Moretti (2004b) points out, theoretically, different mechanisms can generates human capital externalities, these simple structure is compatible with most of these mechanisms. For example, more skilled workers concentrated in a city can increase probability and quality of worker-firm matching (Wheeler, 2006; 2008; Freedman, 2008; Greenstone et al., 2010; Abel and Deitz, 2015). A firm located in a city that concentrates more skilled workers can benefit from sharing by better adjusting to indiosincractic shocks (Krugman, 1991; Duranton and Puga, 2004; Overman and Puga, 2010). In addition, the concentration of skilled workers can generate knowledge spillovers through knowledge sharing and formal and informal partnerships generating positive learning externalities for the firm (Charlot and Duranton, 2004; Storper and Venables, 2004; Bacolod et al., 2009; Lychagin et al., 2016; Thisse, 2018).

The equilibrium is obtained when the utilities of the workers are equal in two cities and the firms in different cities have unitary costs. In the presence of human capital externalities, in cities with a larger share of skilled workers, firms can produce the same level of output with less inputs (labor and physical capital). If there are productivity gains in city B, for example, why firms do not migrate until there is no more productivity difference between cities? Moretti (2004b) argues that costs (with wages and rents) are higher in city B, making firms indifferent in migrating. Thus, in this structure proposed by the author, the balance with spillovers is feasible.

3 Empirical strategy and identification

3.1 Model Specification

Similar to Moretti (2004b) and Liu (2013), we consider the effects of human capital spillovers on firm productivity through the following production function:

$$\mathsf{Y}_{pjct} = \mathsf{A}_{pjct}\mathsf{H}_{ct}^{\lambda}(\mathsf{h}_{pjct}\mathsf{L}_{pjct})^{\alpha}\mathsf{K}_{pjct}^{\beta} \tag{1}$$

where Y_{pjct} is output of firm p, engaged in industry j, in city c, and year t; A_{pjct} represents the other factors that can determine productivity, such as local or industry features; H_{ct} is city human capital endowment; h_{pjct} is the number of workers with college degree or higher; L_{pjct} is the number of low-schooling workers; and K_{pjct} it's the physical capital.

We assume that the total factor productivity is given by:

$$\mathsf{TFP}_{pjct} = \frac{\mathsf{Y}_{pjct}}{(\mathsf{h}_{pjct}\mathsf{L}_{pjct})^{\alpha}\mathsf{K}_{pjct}^{\beta}} = \mathsf{A}_{pjct}\mathsf{H}_{ct}^{\lambda} \tag{2}$$

By decomposing the term A_{jct} on characteristics of industry, cities and time, we assume

[2] Another way of assuming a distribution of skilled workers differently between cities is to assume that these are endowed with distinct amenities, so that skilled workers will migrate to the city as better amenities because they value this type of externalities more. For more details, see Moretti (2004a).

the following linearized version of equation (2):

$$\ln(\mathsf{TFP}_{pjct}) = \lambda \ln(\mathsf{H}_{ct}) + \varepsilon_{jt} + \varepsilon_{ct} + \varepsilon_t + \varepsilon_p + u_{pjct} \quad (3)$$

where ε_{jt} represents specific time-varying characteristics of industrial sector, ε_{ct} are time-varying characteristics of cities where firms are located that can influence productivity, ε_t is the time effect, ε_p is the plant-level fixed effect, and u_{pjct} is the error term.

Our parameter of interest is λ, the parameter that captures the effect of local human capital on the firm's TFP. Two empirical challenges arise in obtaining such an estimate. First, we need to obtain reliable valor of TPF and, then, a way to consistently estimate the parameter of interest. Regarding the first challenge, notice that OLS estimations of the TFP through the residues of the production function are likely to be biased, due to the simultaneity between the factors of production and the productivity: more productive firms may allocate their production inputs differently than less productive firms (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). Instead of using firms' investment to control for unobserved productivity shocks as in Olley and Pakes (1996), using plant-level confidential information from IBGE and microdata from RAIS, we apply the TFP estimator of Levinsohn and Petrin (2003), which considers intermediate inputs as a proxy for these productivity shocks.

Even with credible values for firms' TFP, obtaning a consistent estimate of λ is a quite challenging task, since if H_{ct} is correlated with any unobserved features, $\mathsf{Cov}(\mathsf{H}_{ct}, \varepsilon' s) \neq 0$, traditional OLS estimate of λ will be inconsistent. Note that such a situation is far from unlikely, since, for example, more productive companies may choose cities with more human capital or unobserved factors that affect the company's productivity may be associated with local human capital. We face this challenge by taking advantage of the richness of our database (described in the next section), which allows the use of longitudinal information about firms, and adopting a two-step strategy similar to that employed by Heuermann (2011) and Groot et al. (2014). Such a strategy involves obtaining a general local (city) indicator of the firms' TFP (after eliminating the influences of the firms' characteristics) and subsequently investigating the role of local human capital in this indicator using a shift-share instrumental variable approach.

Specifically, in the first step, we estimate the equation 3 without the term H_{ct} but including firm's characteristics, such as dummies of size, sector×year, average worker characteristics, and dummies for city×year:

$$\ln(\mathsf{TFP}_{pjct}) = \varepsilon_{jt} + \varepsilon_t + \mathsf{Age}_{pt} + \sum_{size} \beta_{size} \mathsf{D}^{size}_{pc} + \sum_c \sum_t \epsilon_{ct} \mathsf{D}_{ct} + u_{pjct} \quad (4)$$

where ε_{jt} are dummies for each sector and year, ε_t are time dummies, Age_{pt} is the average age of the firm's employees, D^{size}_{pc} are dummies for size (measured by number of employees), and D_{ct} are dummies for each city and year. In the latter case, note that each D_{ct} generates a city-specific productivity index given by ϵ_{ct}.

Note that these estimated city-specific productivity indices represent the effects of different local factors that affect the performance of firms in cities after discounting the effects of firm characteristics that may be associated with both their performance and the characteristics of cities. In this sense, for example, the effects of possible sorting of cities by firms based on observable or unobservable characteristics are taken into account.

In the second step, we use estimated $\hat{\epsilon}_{ct}$ in the first step as the depend variable. Here, we assume that a city human capital depends essentially on its the share of college-educated workers (denoted by S_{ct}). Similarly to Liu (2013), we also assume that $\mathsf{H}_{ct} = \exp(\mathsf{S}_{ct})$. Our second step, thus, consider a regression of $\hat{\epsilon}_{ct}$ on S_{ct} and additional city-level controls. Specifically, we consider the following specification:

$$\hat{\epsilon}_{ct} = \alpha_1 + \lambda \mathsf{S}_{ct} + \mathsf{X}_{ct}\theta + \mu_{ct} \tag{5}$$

where X_{ct} is a vector of local characteristics that can influence the local productivity index.

Obtaining a credible estimate of λ in the equation 5 remains challenging as other local factors can jointly affect human capital and local productivity indicators. We deal with it by considering a set of controls commonly related to local productivity, X_{ct}, and using an instrumental variable for S_{ct}. In the set of control variables, based on agglomeration literature and previous empirical works, we include cities' density of people, measures of market structure, productive specialization, diversification, and geographic and climate variables.

There is now a certain consensus that, due to different agglomeration gains, denser cities tend to be more productive and attract more qualified workers (Henderson, 2003; Combes et al., 2012; Behrens et al., 2014; Accetturo et al., 2018; Gaubert, 2018). There is, therefore, a double motivation for including this variable as a control: Its inclusion captures the influence of other sources of agglomeration (for example, associated with better labor market matching or greater sharing of inputs and services) that could also act associated with local human capital and represent a control for factors that attract the most qualified workers.[3]

The demographic density, however, is a very generic measure of the source of agglomeration gains. Therefore, we also consider more specific measures of the degree of productive diversification and competition of locations. In the first case, in line with Jacobs (1969)'s arguments that associate local productivity with productive diversification, we follow Combes et al. (2011) and use the diversification indicator given by:

$$\mathsf{Dive}_{ct} = \frac{\mathsf{E}_{ct}^2}{\sum_j \mathsf{E}_{jct}^2} \tag{6}$$

where E_{ct} is the employment of city c in the period t and E_{jct} is the employment of city c in the period t in activity j.

As highlighted by Glaeser et al. (1992), the local degree of competition between firms can affect their economic efficiency and even affect the degree of attractiveness of cities. We, thus, follow the suggeston of this authors and include a competition index given by:

$$\mathsf{Comp}_{ct} = \frac{\mathsf{F}_{ct}}{\mathsf{E}_{ct}} \frac{\mathsf{E}_t}{\mathsf{F}_t} \tag{7}$$

where F_{ct} is the number of firms in the city c in year t; and F_t is the number of firms in the country.

The second group of control variables included in X_{ct} refers to geographic and climatic variables. Such variables can affect the efficiency of firms, directly, by affecting local production conditions and, indirectly, by presenting amenities that affect the sorting of more qualified workers. More specifically, we include in this vector of controls the distance from municipality center to the coast (in km), the average water precipitation between 1931-1990 (100 millimeters per year), the average sunshine during the day between 1931-

[3] As local density can react to productivity, we recognize that this variable can also be endogenous, which makes it a "bad" control. As we show in the next section, however, our results are unaffected by its exclusion.

1990 (100 hours per year), and the average altitude.

Finally, to address sources of endogeneity possibly still present in the estimation of λ in the equation 5, we used an shift-share instrumental variable for S_{ct}.

3.2 The shift-share instrument

Our shift-share instrument (SSIV) for S_{ct} exploits large shifts in the Brazilian national education policy between 1980 and 2010 combined with the past demographic structure of higher educated individuals (the share component). The heterogeneous shock exposure is based on the different demographic structures of cities.

The educational reforms implemented by the Brazilian Federal Government during the period 1980-2010 led to one of the fastest rises in educational attainment on record in history (Bruns et al., 2011; Lindert, 2021). In a political environment of redemocratization of the country, the reforms implemented encompassed both basic (Bourguignon et al., 2003; Cardoso et al., 2004; Bruns et al., 2011) and higher education (Corbucci, 2002; Corbucci et al., 2016). To higher education, two policy changes were particularly important. Firstly, the policy shifts favoring the expansion of higher education began in the 1990s, focusing on private higher education (Ferreyra et al., 2017; Rocha et al., 2017). During this period, the processes of authorization and recognition of courses and institutions were streamlined, leading to a 132% growth in the number of enrollments in undergraduate courses from 1997 to 2003 (Corbucci et al., 2016).[4] Second, the policy of expanding and decentralizing federal universities in the first decade of the 21st century. Between 2000 and 2010, the number of federal universities grew by almost 50% and the largest growth (125%) occurred in federal universities located in the interior of the country (Niquito et al., 2018). Combined, these policies promoted a historically unprecedented increase in the share of the population with a college degree. According to data form the 1980 and 2010 Census, between these two years, there was 187% growth in the share of the population with college degree or higher. The effects of the national higher education reforms are also clear in the number of students enrolled in undergraduate programs which grew 296%. As evidenced by *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) data, the effects of the national education reforms are also apparent in the 296% growth in the number of students enrolled in undergraduate programs during the same period.

As Moretti (2004a) points out to the US, each new generation have a larger share of more educated individuals as young people today are more educated, on average, than young people from previous decades. Consider two identical cities except in the age structure. If in one of the cities the share of young is higher, then the proportion of graduates is expected to be higher in this city. As young people enter the labor market, the share of workers with college degree or higher also shows a growth trend following the trend of increased education. Specifically in the case of Brazil, this trend has been exogenously shifted upward (a supply shock) owing to educational policies. This increase in the number of college-educated workers depends, thus on the specific past demographic structure in each city and of the shift component.

Our proposed instrumental variable uses the exogenous policy oriented schooling expansion in Brazil and the past demographic structure of cities to captures what would have happened to the share of college-educated workers in a given city if the trend of na-

[4] In addition to these policies aimed at the private sector and crucial for encouraging investment in human capital, the Federal Government also expanded subsidized student loans (Rocha et al., 2020; Saccaro and França, 2021).

tional growth had been observed in it. Formally, we propose using the following measure as an instrument for the share of college workers in the cities:

$$\mathsf{IV}_c = \sum_m \omega_{m,c} \times (\mathsf{P}_{2010,m} - \mathsf{P}_{1980,m}) \tag{8}$$

where m indexes the age groups, we defined three age groups: young 16-25, middle-aged 26-50, and elderly 51-70; c indexes the municipalities; ω_{mc} is the share of group m in city c in 1980; and $\mathsf{P}_{2010,m} - \mathsf{P}_{1980,m}$ is the national change in college share for group m between 1980 and 2010.

Notice that the proposed instrument does not change in time. Therefore, we use IV_c to instrument the mean of $\bar{\mathsf{S}}_{ct}$ in time by 2SLS model. The validity of identification in a shift-share instrument approach has been discussed in recent literature and relies on assumptions about the exposure shares and/or shocks. As demonstrated by Goldsmith-Pinkham et al. (2020), strict exogeneity of the shares is a sufficient condition for consistency of the SSIV estimator. Alternatively, Borusyak et al. (2022) show that the exogenous variation of shocks is also sufficient to ensure the validity of the SSIV design. In our context, despite using exposure shares that are 30 years out of date, a causal interpretation of λ be supported by the assumption that absent the large shifts in the Brazilian national education policy more- and less exposed municipalities would have experienced similar trend in the local productivity index. Thus the identifying condition is given by:

$$\text{Cov}(\mathsf{IV}_c, \mu_c) \equiv \sum_m \omega_m \Delta \mathsf{P}_m \text{E}\left[\frac{\omega_{mc}}{\omega_m}\mu_c\right] \xrightarrow{p} 0 \tag{9}$$

where μ_c is the error term in equation 5.

In our approach, a significant identification threat arises from the potential influence of unobserved shocks at the municipality level on the national share of individuals with a college degree or higher. This influence could introduce bias into our results. To be more precise, our identification would be threatened by municipality shocks that both affect local productivity indices and also influence the national share of college-educated people. As outlined in the following sections, we deal with this concern by specifically excluding municipalities with larger populations, which have the potential to exert a disproportionate influence on national shares.

4 Data

Plant-Level Production Functions Data. The main data sets at firm level is the Annual Industrial Survey (PIA - *Pesquisa Industrial Anual*). Provided by the Brazilian Institute of Geography and Statistics (IBGE), the PIA presents statistics of the extractive and manufacturing firms and its main objective is to provide the data to characterize the Brazilian industrial structure and to monitor its transformations in time. The PIA data are restricted and can only be accessed through the IBGE restricted room. The data are available from 1996 onwards and comprise all companies with 30 or more people employed in the extractive and manufacturing industries. The other companies have a small expression in the overall economic activity and are treated through sampling. Different from other sources of information about firms in Brazil, such as RAIS, through the PIA, it is possible to access information on firms' production functions, including the set of all factors and inputs.

More specifically, by working with the PIA database at the firm level, besides more traditional characteristics such as employment and wages, we are able to measure the use of intermediate inputs and build a proxy for the physical capital of companies. For example, data are available on purchases of industrial machinery and equipment, depreciation, improvements and losses of industrial machinery and equipment, purchases of electricity, fuel, raw materials, auxiliary materials and stocks. Regarding the structure of revenue and production value, data on total revenue, industrial production value, and added value are available. We use this set of information to estimate the TPF of firms in the manufacturing industry.

Formal Labor Market Data. We also use information from the Annual Report on Social Information (RAIS - *Relação Anual de Informações Sociais*) provided by the Ministry of Labor. The RAIS is an annual report with information on the universe of formal workers. With this data, it is possible to follow formal workers and firms in time. The so-called RAIS-identified micro-data contains detailed information on the employment relationship and characteristics of the worker, such as age, gender, qualification, and remuneration, and on firms' number of employees. When estimating TFP, RAIS data at the worker level is utilized to categorize each firm's employment according to educational level, enabling differentiation of the labor production factor between workers with college degree or higher and low-schooling workers. RAIS data is also used to calculate labor market indicators at the municipal level in the other steps of the empirical strategy.

Census Microdata. We use data from the 1980 and 2010 censuses on the demographic structure of the municipalities. Specifically, we considered education data of the population by age group at the municipal level in our shift-share regression design outlined in the following section.

Other Sources. Additionally, we also use data on climate characteristics of municipalities, such as average precipitation levels and daily sunshine hours, obtained from the National Institute of Geology (INGEO).

Table 1 presents the descriptive statistics of variables at city level used in the second step estimation (city-level variables). Compared to their average values, we highlight the large variances in the local productivity index, the share of those most educated, and the employment density. In Table 5 we also present the values of these variables by some percentiles of their respective distributions across cities.

5 Results

Our main results are obtained after different initial estimations. The first of them refers to the firms' TFP using the estimator proposed by Levinsohn and Petrin (2003). These estimates are presented in Table 6 in the Appendix, which presents the estimated coefficients for capital, qualified labor, and unskilled labor. The estimated coefficients allow obtaining the TFP values of the firms used in our first step regression 3 that generate the local (city-level) productivity indices. The estimates of the coefficients of the variables in this equation 3 are also presented in the appendix (see Table 7).

From these initial estimates, in the following subsection, we present the estimates of the impact of local human capital on firms' TFP (equation 5) using both OLS and 2SLS estimators (our second step). In the second subsection, we present different robustness checks for the results obtained.

Table 1. Descriptive statistics - City-level variables

Variable[a]	Mean	Std. Dev.	Min	Max
$\bar{\hat{\epsilon}}_c$	-0.188	0.240	-1.516	0.547
# college	1851.025	5540.230	0.500	76588.710
College share	0.005	0.006	0.000	0.041
Population	395,940	883,789	102,137	11,800,000
Employment (E_c)	15,973.020	31,789.360	61.000	438,538.400
Employment density	59.711	151.491	0.014	1,630.125
# firms (F_c)	659.343	1,557.447	6.500	23,043.140
Comp_c	1.229	0.713	0.156	4.433
Dive_c	10.751	5.552	1.169	26.412
Distance o the coast (km)	2.394	3.401	0.002	22.854
Average altitude	3.978	3.517	0.01	11.96
Average water precipitation	13.852	3.636	3	27
Average sunshine	19.682	3.658	12	27

Note: This table presents the descriptive statistics for variables of second step estimation. Due to the confidential nature of the data in the first-step estimation, we do not report these informations. [a] All variables represent the mean annual values over the period 2011-2017. # college is the number workers with college-or-more and # firms is the number of firms. Distance o the coast is the distance from municipality center to the coast (in km), average water precipitation is calculated between 1931-1990 (100 millimeters per year), average sunshine during the day is calculated between 1931-1990 (100 hours per year).

5.1 Baseline results

The main results of the paper for the effect of local human capital on firms' TFP are present in the following Table 2.[5] In Panel A, column (1) reports OLS estimates and columns (2)-(3) 2SLS estimates using our shift-share instrument for the local share of college-educated workers. In Panel B, we present first-stage regression results when using this SSIV.

Firstly, from Panel B of Table 2, note that our that first-stage SSIV regressions results indicate, as expected, a positive association between the share of college-educated workers and the instrument. Furthermore, the value of the F statistic of this first stage (exceeds 10 in all specifications) suggests that such an instrument is strong, therefore favoring the implemented strategy.

Considering now the estimates in Panel A of Table 2, we highlight four relevant evidences from the results. First, note that in all specifications our coefficient of interest is positive and statistical significant. This is in line with the idea that the spatial concentration of human capital generates external gains for firms. More specifically, based on our preferred specification (that of column 3 of Table 2), we observe that a ten percentage point increase in the proportion of college-educated workers in cities increases, on average, companies' total factor productivity approximately by 1%. Note that this is a very important effect. For example, assuming a percentage variation (in logs) in the share of

[5]The results of Table 2 were obtained by TFP estimation when the intermediary inputs is the consumption of raw materials. The results do not change if we estimate the main results using TFP estimation in column 2 of Tables 6 and 6 in the Appendix.

more educated corresponding to the difference between the 50th percentile and the 75th percentile of the distribution of this variable among cities (see Table 5), the value of the coefficient implies an increase on average of 8.1% in TFP of firms.[6] Such economic significance is consistent with the available evidence regarding the importance of human capital externalities in developing countries obtained in conventional regressions using salary as the dependent variable (Chauvin et al., 2017; Duranton, 2016).

Table 2. The impact of local human capital on firms' TFP - Results of second step estimation

	OLS	2SLS	
	(1)	(2)	(3)
Panel A: Second Stage - Dep. var. is $\bar{\hat{\epsilon}}_c$			
ln(*College Share*)	0.042**	0.097***	0.096***
	(0.011)	(0.034)	(0.026)
ln(*Dens*)	0.026***		0.010
	(0.004)		(0.013)
Panel B: First Stage - Dep. var. is ln(*College Share*)			
IV_c	–	1,916***	1,041***
	–	(651.3)	(92.42)
1^{st} F-stat.	–	17.46	126.89
Controls	Yes	No	Yes
Observations	264	264	264

Note: Table 2 shows the results for estimation of distinct specifications for equation 5 when the dependent variable is a average over time of the variable $\hat{\epsilon}_{ct}$. Column (1) present the OLS results including all controls variables that encompass the local factors at the municipal level as previously described, such Dive_{ct}, Comp_{ct} and the distance from municipality center to the coast (in km), average water precipitation between 1931-1990 (100 millimeters per year), average sunshine during the day between 1931-1990 (100 hours per year) and average altitude, as well as employment density. Columns (2) and (3) present the 2SLS results without additional controls and including all control variables, respectively. Standard errors corrected for clustering at the macro-region-level are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$.

Second, the values of our estimates for the effect of local human capital on firm productivity using the instrument are approximately double that obtained using the OLS estimator. Chauvin et al. (2017) found a similar pattern using wage regressions and IV estimates based on a demographic instrument for Brazil. Such evidence is consistent with the presence of unobservable characteristics associated with local human capital which negatively affect firm productivity. We do not have a definitive explanation here, but the result seems consistent with the higher presence of more educated people in environments

[6] $8.1\% = (\ln(0.007) - \ln(0.003)) \times 0.096 \times 100$. Using estimates from wage regression, Quintero and Roberts (2023) found that the change in city's average years of schooling from the 25th to the 75th percentile implies an estimated productivity increase of 18%. Using our estimates, a similar change would bring an average increase of about 19% in firms' TFP.

with greater urban service congestion. Third, it is worth highlighting that the value of the coefficient estimated in the specifications using the instrument changes little with the use of controls, which suggests that our SSIV fully does its job of bringing exogeneity to the variable of interest.

Finally, despite the positive value for the estimated coefficient of density variable (as expected), this is not significant in the specification with SSIV. Note that Chauvin et al. (2017), for Brazil, and Quintero and Roberts (2023), for Latin America countries, also found that external effects of human capital on wages are more important than general agglomeration effects of density. This specific result suggests that the concentration of more educated workers is more relevant to firm productivity than other sources of agglomeration gains associated with local greater density of people.

Although they are not directly comparable, as they measure firms' productivity differently, our results are in line with those already obtained for the USA by Moretti (2004b) and for China by Liu (2013). Thus, for Brazil too, the local concentration of more qualified workers can have an important positive effect on firms' productivity. Our results are also in line with those already obtained by Chauvin et al. (2017) and Almeida et al. (2022), for Brazil, and Quintero and Roberts (2023), for LAC countries, using wages as the dependent variable and not firm total factor productivity. By considering companies' TPF instead of wages, our set of evidence, in a more general and direct way, favors the role of local human capital as an important factor in firm performance.

5.2 Robustness checks

We now provide different robustness checks for our main results. Specifically, we verify the reliability of our results through two additional sets of results: using the local productivity measure of each year of the sample and varying the sample of cities according to their size.

Remember that, because our instrument does not vary over time, in the second step of the strategy we use the temporal average of local productivity indicators as the dependent variable (see equation 5). In the first set of new evidence, we repeat the second step and obtain new estimates for the impact of local human capital on firms' TFP using as a dependent variable the value of such a productivity index for each year (therefore, seven different estimates are generated for the coefficient of interest). These new estimates are presented in the following Table 3.

As can be seen from the values presented in this table, in addition to evidence once again favorable to the use of the instrument (Panel B of Table 3), the new estimated values of the coefficient of interest are, in all years, close to those previously obtained (Panel A of Table 3). And, except for 2017, all the new estimates are statically significant. Therefore, our strategy of choosing the average of local productivity indicators instead of that of any year without clear criteria is far from compromising the information provided by the estimate.

The second set of new estimates considers different samples depending on the size of the city. In an influential paper, De La Roca and Puga (2017) showed evidence that workers' learning gains would be more significant in larger cities as the experiences would be more productively valuable. Our second set of estimates, thus, explores this idea by considering possible favorable productivity differentials for firms in the largest cities and its influence on our main results.

Note, however, that the limited size of our sample of cities imposes restrictions on the investigation. Therefore, we adopted the strategy of excluding larger cities and checking

Table 3. The impact of local human capital on firms' TFP - Results by year

	2011	2012	2013	2014	2015	2016	2017
Panel A: Second Stage - Dependent variable is $\hat{\epsilon}_{ct}$							
$\ln(College\ Share)$	0.082***	0.117***	0.079*	0.095**	0.084***	0.120***	0.069
	(0.029)	(0.033)	(0.043)	(0.038)	(0.031)	(0.027)	(0.066)
$\ln(Dens)$	0.024	-0.009	0.001	-0.011	0.004	-0.002	-0.000
	(0.015)	(0.014)	(0.014)	(0.018)	(0.013)	(0.024)	(0.028)
Panel B: First Stage - Dependent variable is $\ln(College\ Share)$							
IV_c	1124.61***	1106.06***	1,042.14***	1,077.29***	948.66***	1,067.53***	917.24***
	(138.83)	(111.85)	(120.52)	(113.89)	(102.71)	(69.72)	(50.40)
1st F-stat.	65.62	97.78	74.76	89.47	85.30	234.40	331.15
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	262	264	262	262	263	263	261

Note: Table 3 shows the results for estimation of equation 5 when the dependent variable is a average over time of the variable $\hat{\epsilon}_{ct}$. Columns 1-7 (2011-2017) present the results of estimations equivalent to those in column 3 of Table 2 by different years. Controls include variables that encompass the local factors at the municipal level as previously described, such Dive_{ct}, Comp_{ct} and the distance from municipality center to the coast (in km), average water precipitation between 1931-1990 (100 millimeters per year), average sunshine during the day between 1931-1990 (100 hours per year) and average altitude, as well as employment density. Standard errors corrected for clustering at the macro-region-level are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4. Results excluding different municipalities

	Excluding			
	SP and RJ	pop. > 1 million	States capital	The smallest 10%
	(1)	(2)	(3)	(4)
Panel A: Second Stage - Dependent variable is $\bar{\hat{\epsilon}}_c$				
$\ln(College\ Share)$	0.092***	0.087***	0.080***	0.096***
	(0.029)	(0.019)	(0.019)	(0.031)
$\ln(Dens)$	0.011	0.013	0.014**	0.008
	(0.014)	(0.010)	(0.007)	(0.016)
Panel B: First Stage - Dependent variable is $\ln(College\ Share)$				
IV_c	1,083.71***	1,220.37***	1,354.22***	987.56***
	(195.82)	(69.70)	(215.23)	(150.22)
1st stage F-stat.	30.62	306.53	39.59	43.22
Controls	Yes	Yes	Yes	Yes
Observations	262	249	239	238

Note: Table 4 shows the results for estimation of distinct specifications for equation 5 when the dependent variable is a average over time of the variable $\hat{\epsilon}_{ct}$ and we excluding some municipalities. Controls include variables that encompass the local factors at the municipal level as previously described, such Dive_{ct}, Comp_{ct} and the distance from municipality center to the coast (in km), average water precipitation between 1931-1990 (100 millimeters per year), average sunshine during the day between 1931-1990 (100 hours per year) and average altitude, as well as employment density. Standard errors corrected for clustering at the macro-region-level are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$.

possible changes in the estimates. Columns (1)-(3) in Panel A of Table 4 present the new estimates excluding from the sample of cities the mega-cities of São Paulo and Rio de Janeiro, cities with more than 1 million inhabitants, and state capitals, respectively. As it can immediately notes by comparing the numbers of Tables 4 and 2, once again, the new estimates for the impact of local human capital on firm productivity are quite close to that already shown. Our results, therefore, seem far from being explained by greater gains from learning in the largest cities.

Since it is not possible to rule out the possibility that smaller cities are leading to an underestimation of the influence of local human capital on firm productivity, we also perform a final exercise obtaining additional evidence through a sample excluding the smallest 10% cities. This new estimate is presented in column (4) of Panel A of Table 4. The evidence obtained in this last exercise does not differ from our main estimates either.

6 Concluding Remarks

Focusing mainly on the experience of developed countries, there is an important body of evidence indicating the relevance of human capital externalities in explaining productivity of firms and workers. The evidence for developing countries, however, is rarer and almost non-existent when considering the effects directly on firms' TFP. In this research we contribute to fill both gaps by providing unique estimate of the impact of local human capital on firms' TFP in Brazilian cities.

Our set of evidence indicates that an increase in the share of college-educated workers in Brazilian cities generates important productivity gains for firms. For example, a variation between the 25th and 75th percentiles of the distribution of this share generates gains close to 19% in the firms' TFP. The results are in line with previous evidence obtained for developed countries and with evidence for Brazil obtained through traditional wage regressions. Importantly, we also showed that the effect of local human capital on firms' TFP in Brazil appears more important than other sources of agglomeration gains and it is not explained by larger learning gains in the biggest cities of the country.

It should be noted that the evidence obtained in the work adds a potentially important element in explaining the country's regional disparities in income and productivity. As recently shown by Oliveira and Silveira Neto (2022), differences in share of college-educated individuals across urban centers play an essential in explaining regional income inequality in Brazil. Our results indicate that these differences may also increase this inequality by their external impact on firms' performance.

References

Abel, J. R. and Deitz, R. (2015). Agglomeration and job matching among college graduates. *Regional Science and Urban Economics*, 51:14–24. https://doi.org/10.1016/j.regsciurbeco.2014.12.001.

Accetturo, A., Di Giacinto, V., Micucci, G., and Pagnini, M. (2018). Geography, productivity, and trade: Does selection explain why some locations are more productive than others? *Journal of Regional Science*, 58(5):949–979. https://doi.org/10.1111/jors.12393.

Almeida, E., Silveira Neto, R. M., and Rocha, R. M. (2021). The spatial extent of human capital spillovers in a transition country: Evidence from Brazil. *Available at SSRN 4676751.*

Almeida, E. T. d., Silveira Neto, R. d. M., and Rocha, R. d. M. (2022). Manufacturing location patterns in Brazil. *Papers in Regional Science*, 101(4):839–873. https://doi.org/10.1111/pirs.12672.

Au, C.-C. and Henderson, J. V. (2006a). Are chinese cities too small? *The Review of Economic Studies*, 73(3):549–576. https://doi.org/10.1111/j.1467-937X.2006.00387.x.

Au, C.-C. and Henderson, J. V. (2006b). How migration restrictions limit agglomeration and productivity in china. *Journal of Development Economics*, 80(2):350–388. https://doi.org/10.1016/j.jdeveco.2005.04.002.

Bacolod, M., Blum, B. S., and Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65(2):136–153. https://doi.org/10.1016/j.jue.2008.09.003.

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. https://doi.org/10.1007/s00168-016-0768-3.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553. https://doi.org/10.1086/675534.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213.

Bourguignon, F., Ferreira, F. H., and Leite, P. G. (2003). Conditional cash transfers, schooling, and child labor: micro-simulating Brazil's Bolsa Escola program. *The World Bank Economic Review*, 17(2):229–254. https://doi.org/10.1093/wber/lhg018.

Bruns, B., Evans, D., and Luque, J. (2011). *Achieving world-class education in Brazil: The next agenda*. World Bank Publications.

Cardoso, E., Souza, A. P., et al. (2004). The impact of cash transfers on child labor and school attendance in Brazil. *Vanderbilt University Working Paper No. 04-W07*.

Charlot, S. and Duranton, G. (2004). Communication externalities in cities. *Journal of Urban Economics*, 56(3):581–613. https://doi.org/10.1016/j.jue.2004.08.001.

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.

Combes, P.-P., Démurger, S., Li, S., and Wang, J. (2020). Unequal migration and urbanisation gains in China. *Journal of Development Economics*, 142:102328. https://doi.org/10.1016/j.jdeveco.2019.01.009.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742. https://doi.org/10.1016/j.jue.2007.04.004.

Combes, P.-P., Duranton, G., and Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2):253–266. https://doi.org/10.1093/jeg/lbq038.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594. https://doi.org/10.3982/ECTA8442.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics*, pages 15–66. University of Chicago Press. https://ssrn.com/abstract=1141634.

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. https://doi.org/10.1016/B978-0-444-59517-1.00005-2.

Corbucci, P. R. (2002). Avanços, limites e desafios das políticas do mec para a educação superior na década de 1990: ensino de graduação. *Texto para discussão 869 - IPEA*. https://repositorio.ipea.gov.br/handle/11058/2102.

Corbucci, P. R., Kubota, L. C., and Meira, A. P. B. (2016). Reconfiguração estrutural da educação superior privada no Brasil: nova fase da mercantilização do ensino. *Texto para discussão 2256 - IPEA*. https://repositorio.ipea.gov.br/handle/11058/7336.

De La Roca, J. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

Duranton, G. (2016). Agglomeration effects in Colombia. *Journal of Regional Science*, 56(2):210–238. https://doi.org/10.1111/jors.12239.

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.

Ferreyra, M. M., Avitabile, C., Paz, F. H., et al. (2017). *At a crossroads: higher education in Latin America and the Caribbean*. World Bank Publications.

Fleisher, B., Li, H., and Zhao, M. Q. (2010). Human capital, economic growth, and regional inequality in China. *Journal of Development Economics*, 92(2):215–231. https://doi.org/10.1016/j.jdeveco.2009.01.010.

Freedman, M. L. (2008). Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics*, 64(3):590–600.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–53. https://doi.org/10.1257/aer.20150361.

Gennaioli, N., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2012). Human capital and regional development. *The Quarterly Journal of Economics*, 128(1):105–164. https://doi.org/10.1093/qje/qjs050.

Glaeser, E. L. (1999). Learning in cities. *Journal of Urban Economics*, 46(2):254–277.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6):1126–1152.

Glaeser, E. L. and Mare, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624. DOI:10.1257/aer.20181047.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Groot, S. P., de Groot, H. L., and Smit, M. J. (2014). Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, 54(3):503–523.

Guo, J., Roys, N., and Seshadri, A. (2018). Estimating aggregate human capital externalities. *University of Wisconsin-Madison*.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53(1):1–28. https://doi.org/10.1016/S0094-1190(02)00505-3.

Heuermann, D. (2011). Human capital externalities in western germany. *Spatial Economic Analysis*, 6(2):139–165. https://doi.org/10.1080/17421772.2011.557775.

Heuermann, D., Halfdanarson, B., and Suedekum, J. (2010). Human capital externalities and the urban wage premium: Two literatures and their interrelations. *Urban Studies*, 47(4):749–767. https://doi.org/10.1177/0042098009352363.

Jacobs, J. (1969). *The economy of cities*. New York: Random House.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Krugman, P. R. (1991). *Geography and trade*. MIT press.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Lindert, P. H. (2021). *Making Social Spending Work*. Cambridge University Press.

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

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42. https://doi.org/10.1016/0304-3932(88)90168-7.

Lychagin, S., Pinkse, J., Slade, M. E., and Reenen, J. V. (2016). Spillovers in space: Does geography matter? *The Journal of Industrial Economics*, 64(2):295–335.

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

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

Niquito, T. W., Ribeiro, F. G., and Portugal, M. S. (2018). Impacto da criação das novas universidades federais sobre as economias locais. *Planejamento e Políticas Públicas*, (52):367–394.

Oliveira, R. and Silveira Neto, R. d. M. (2022). Brazil's south–north labour income gap re-examined: evidence across purchasing power adjusted wage distributions. *Regional Studies*, 56(5):818–838.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, pages 1263–1297.

Overman, H. G. and Puga, D. (2010). Labor pooling as a source of agglomeration: An empirical investigation. In *Agglomeration Economics*, pages 133–150. University of Chicago Press.

Quintero, L. E. and Roberts, M. (2023). Cities and productivity: Evidence from 16 latin american and caribbean countries. *Journal of Urban Economics*, 136:103573.

Roback, J. (1988). Wages, rents, and amenities: differences among workers and regions. *Economic Inquiry*, 26(1):23–41.

Rocha, R. H., Menezes Filho, N. A., Oliveira, A. P. d., and Komatsu, B. K. (2017). A relação entre o ensino superior público e privado e a renda e emprego nos municípios brasileiros. *Pesquisa e Planejamento Econômico*, 47(3):40–69.

Rocha, W. M., Ehrl, P., and Monasterio, L. (2020). Financiamento da educação superior no Brasil: o impacto do programa fies nos salários dos trabalhadores formais. *Pesquisa e Planejamento Econômico*, 50(2):7–29. http://dx.doi.org/10.38116/ppe50n2art1.

Saccaro, A. and França, M. T. A. (2021). Apoio financeiro ou tipo de escola de ensino médio? uma análise do Fies e do PROUNI sobre a sobrevivência de estudantes no ensino superior brasileiro. *Pesquisa e Planejamento Econômico*, 51(2):48–76. http://dx.doi.org/10.38116/ppe51n2art2.

Silva, D. and Azzoni, C. (2022). Worker and firm heterogeneity, agglomeration, and wages in Brazil. *Papers in Regional Science*, 101(1):107–133. https://doi.org/10.1111/pirs.12637.

Storper, M. and Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4):351–370.

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.

Wheaton, W. C. and Lewis, M. J. (2002). Urban wages and labor market agglomeration. *Journal of Urban Economics*, 51(3):542–562. https://doi.org/10.1006/juec.2001.2257.

Wheeler, C. H. (2006). Cities and the growth of wages among young workers: Evidence from the nlsy. *Journal of Urban Economics*, 60(2):162–184. https://doi.org/10.1016/j.jue.2006.02.004.

Wheeler, C. H. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38(2):101–118. https://doi.org/10.1016/j.regsciurbeco.2008.01.011.

Appendix A Additional Tables

Table 5. Sample percentiles for selected variables

Variable[a]	Sample percentile 10th	25th	50th	75th	90th
$\bar{\hat{\epsilon}}_c$	-0.395	-0.270	-0.156	-0.071	0.010
# college	54.857	207.929	659.714	1,507.857	3,606.143
College share	0	0.001	0.003	0.007	0.012
Population	115,394	129,822.1	207,089.8	349,394.1	674,654
Employment (E_c)	1500.71	4468	9655.07	17512	28725.14
# firms (F_c)	109.29	177.50	377.14	694.36	1258.14

Note: This table presents the sample percentiles for selected variables. [a] All variables represent the mean annual values over the period 2011-2017. # college is the number workers with college-or-more and # firms is the number of firms.

Table 6. TFP estimation

	(1)	(2)
h_{pjct}	0.298***	0.352***
	(0.004)	(0.004)
L_{pjct}	0.437***	0.468***
	(0.005)	(0.005)
K_{pjct}	0.585***	0.565***
	(0.114)	(0.068)
Observations	330,910	330,951

Table 6 shows the results of the estimation of TFP following Levinsohn and Petrin (2003). Column 1 present the results when the intermediary inputs is the consumption of raw materials. Column 2 present the results when the intermediary inputs is energy consumption. Significance level: *** $p < 0.01$.

Table 7. First step estimation - Dependent variable is the $\ln(\mathsf{TFP}_{pjct})$

	(1)	(2)
Age_{pt}	-0.0101***	-0.0100***
	(0.0006)	(0.0006)
D_{pc}^{size} [a]		
$[5, 9]$	-0.232***	-0.2598***
	(0.0372)	(0.0412)
$[10, 19]$	-0.407***	-0.466***
	(0.0315)	(0.0351)
$[20, 49]$	-0.433***	-0.521***
	(0.0308)	(0.0343)
$[50, 99]$	-0.484***	-0.613***
	(0.0310)	(0.0345)
$[100, 249]$	-0.672***	-0.842***
	(0.0318)	(0.0347)
$[250, 499]$	-0.904***	-1.131***
	(0.0318)	(0.0354)
$[500, 999]$	-1.128***	-1.384***
	(0.0329)	(0.0368)
[1000,+)	-1.457***	-1.773***
	(0.0342)	(0.0380)
Industry $\times$ Year FE (ε_{jt})	Yes	Yes
Year FE (ε_t)	Yes	Yes
City $\times$ Year FE (D_{ct})	Yes	Yes
Observations	330,910	330,951

Note: Table 7 shows the results of the estimation of equation 4 when TFP is estimated following Levinsohn and Petrin (2003). Column 1 present the results when the intermediary inputs is the consumption of raw materials. Column 2 present the results when the intermediary inputs is energy consumption. [a]: Omited category [-, 4]. Significance level: *** $p < 0.01$.