External Returns to Human Capital: Evidence from Brazilian Cities

Klebson Moura[*], Raul Silveira Neto[†], Roberta Rocha[‡]

Resumo: No presente estudo procura-se identificar e medir os retornos externos ao capital humano no Brasil, utilizando informações do Relatório Anual de Informações Sociais (RAIS) do Ministério do Trabalho sobre todas as aglomerações urbanas do país para o período 2002-2014. Com uma estimativa de duas etapas e usando variáveis instrumentais para identificação, encontramos um efeito considerável da concentração de capital humano local nos salários locais. O conjunto de resultados indica que, no caso das áreas de mercado de trabalho brasileiro, os retornos externos ao capital humano são da ordem de 0,86% nos salários locais para um aumento de um ponto percentual nos graduados da faculdade. Em consonância com as expectativas teóricas os resultados são semelhantes aos encontrados na literatura pertinente, além disso, o impacto do capital humano local é mais forte para trabalhadores não qualificados do que para trabalhadores qualificados. Os resultados são robustos e não igualmente distribuídos entre os setores de trabalho

Palavras-Chave: capital humano; externalidades ; dados em painel ; variável instrumental.

Abstract: In this study we aim to identify and measure the external returns to human capital in Brazil using information from the Ministry of Labor's Annual Social Information Report (RAIS) on all urban agglomeration in the country for the period 2002-2014. With a two step estimation and using instrumental variables for identification we find a considerable effect of local human capital concentration on local wages. The set of results indicates that, in the case of brazilian labor market areas, the external returns to human capital are a 0.86% increase in local wages for an increase of one percentage point in college graduates. Consistent with theoretical expectations and similar to literature results, we also found that the impact of local human capital is stronger for unskilled than for skilled workers. Finally, results are robust and not equally distributed across sectors

Keywords: human capital; externalities ; panel data ; instrumental variable .

JEL Classification: R0 .

*Anpec: **Área 10**.*

[*]Department of Economics, Federal University of Pernambuco, Caruaru, Brazil. Email: *klebson.moura@gmail.com*.

[†]Department of Economics, Federal University of Pernambuco, Recife, Brazil. Email: *rau.silveira@uol.com.br*.

[‡]Department of Economics, Federal University of Pernambuco, Caruaru, Brazil. Email: *roberta_rocha_pe@yahoo.com.br*.

1 Introduction

The relation between individual human capital and earnings are quite known. Most of the empirical work confirms the intuitive notion that wage gains are in fact related to education attainment and not just to unobservable worker characteristics, leaving little doubt about the impact that education has on the private sphere. On the other hand, the relationship between concentration of local human capital in individual earnings are less known, despite human capital externalities being at the core of both theoretical models and justifications for education provision through public policies.

In this context, we can ask ourselfs what externalities would arise with the educational level of a given place and what would be their net effect. Answering this question matters to support the education subsidies and local development policies associated with the provision of education attaiment, since in the case of positive externalities the gains would not be limited to individuals but also present at the social level through educational spillovers. Despite the importance of this research agenda, empirical works with this specific goal are more recent and relatively scarce.

Among theoretical explanation for the existence of human capital externalities found in the literature, we have Rauch (1993)'s argument that knowledge and skills are shared between workers in formal and informal interactions, with higher level of human capital of individuals increasing probability that meaningful knowledge is shared when individual agents meet[1]. Additionally, Glaeser (1999) argues that larger cities, in addition to the possible existence of more formal education sources and specific training programs, would also offer unique opportunities through a non formal learning environment through social and business contacts, and therefore, would also benefit ordinary workers through the imitation of more skillful neighbors, leading to widespread impacts of human capital concentration and not just focused on innovation.

Acemoglu & Angrist (2001) divide theoretical arguments into two types, pecuniary and non-pecuniary. The first type are those that manifest themselves through interactions and sharing of knowledge and ideas in the lines of *Jacobs externalities*, due to Jacobs (1970) argument that cities are the main engine of growth because they facilitate exchange of ideas. The pecuniary ones, are in lines with *Marshalian externalities*, with the argument that increasing the geographic concentration of specialized inputs increases productivity, since the matching between factor inputs and industries is improved. Moretti (2004a) calls the human capital externalities of productive externalities, which occur when the presence of skilled workers makes other more productive workers so that an increase in aggregate human capital generates an effect on aggregate productivity.

Despite the clear mechanisms postulated by economic theory to explain external effects of human capital concentration, empirical estimation allways faced difficulties. The most important ones are mainly related to bias introduced by unobservable factors correlated with wages on one hand, and human capital concentration on the other. The first source of bias is the existence of sorting, the concentration of individuals of high levels of unobserved ability into cities. On the local level, a second problem is the bias introduced by city-specific unobserved characteristics that are correlated with human capital concentration. As suggested by Moretti (2004a), some examples are, unobserved differences in industrial mix, technology or natural resources. The empirical literature tries to control for these problems with, panel data for worker fixed effect, introduction of labor shock controls, and instrumental variables respectively.

However, only throughout recent years improvements in access to worker-level information fomented a rise in studies focused on wage disparities across regions which implements these estimation strategies. This allowed the emergence of research focused on corectly identify and measure of aglomeration and huma capital concentration effects (Moretti, 2004a,b; Combes *et al.* , 2008, 2010; Heuermann, 2011; Groot *et al.* , 2014; Broersma *et al.* , 2015; Barufi *et al.* , 2016). Empirical findings for develop countries shows that externalities due to human capital concentration are large and also affect unskilled workers. For the U.S estimates range from a 1% to 5% wage increase for a one year increase in local average scholling (Rauch, 1993; Acemoglu & Angrist, 2001). For West Germany, results are 1.8% and 0.6% (Heuermann, 2011). Unfortunately, research for developing countries is scarcer. Liu (2007), which following the empirical strategy of Moretti (2004a), assess the external returns to education in China. For Kenya, Manda *et al.* (2002) uses district-level average education attaiment to capture the effect of human capital on earnings, finding positive effects for male and female workers. Thus, there is a gap to be filled

[1]Moddeling approach used by Jovanovic & Rob (1989)

in the research program on the effects of human capital concentration, namely, increase evidence of human capital agglomeration in developing countries.

Specifically for Brazil, some facts can be considered stylized in the context of human capital. The first is the existence of high private return of education, ranging from 9.8 % to 18 % depending on the methodology used (Sachsida *et al.* , 2004; Barbosa Filho & Pessoa, 2008), secondly is the large regional wage inequality related to a inequalities in regional education attaiment (Barros *et al.* , 2007). In fact, the country has historically low human capital stock and presents a stagnation of labor productivity (Galeano & Feijó, 2013). In this setup the assessement of external effects of human capital is especially important because identify social returns to human capital concentration would serve as yet another argument to promote policies that aim human capital increase within the labor force. Despite those facts, studies with the specific goal of identifing external returns of human capital concentration are not numerous, some exceptions are Araujo Junior & Silveira Neto (2004) and Falcão & Silveira Neto (2007), however these initial studies do not fully address bias issues. Therefore, the main goal of this paper is to identify and measure external returns to human capital concentration on local wages in Brazil.

Specific contributions of the paper are twofold. First, despite the high known private returns of human capital in Brazil, the external effects are still unknown. We help to fill that gap in the national and developing countries literature by assessing evidence of human capital externalities. Second, although we focus on the effect of human capital concentration on local wages, we simultaneously include local employment density in our estimations. This inclusion is important to control for what Combes *et al.* (2010) calls "endogenous labor quantity", that is, more productive places and therefore with higher wages, can attract more skilled and unskilled workers becoming denser, which may affect the correct estimates of the effect of interest by, possibly, changing local workforce composition.

Our results indicate that, for Brazilian labor market areas, the effect of human capital concentration, defined by the fraction of individuals with at least a college degree, in local wages is about 0.839, that is, a one percentage point in the concentration o college graduates increases local wages by about 0.84%. Also the effect is larger than purely agglomeration effects arising from population density only. Furthermore, the effect is higher for unskilled workers then for skilled and are not equally distributed across sectors, with returns concentrated in the manufacturing and service sectors. Results were robust to different specifications and estimation strategies.

The remainder of the paper is organized as follows. The next section presents a theoretical framework. Section 4 describes the empirical strategy and the treatment of the underlying data. Section 5 presents and analyzes the results and robustness checks, with the conclusions being presented in the last section.

2 Previous Empirical Evidence

The formalization of the impacts of productivity gains from education (human capital) in the per capita production growth are laid out in Lucas (1988), with it's endogenous growth model based of heterogenous human capital distribution between workers as the main explanation from the cross-country incomes differences. However, The first paper to exploit differences in human capital across cities to identify externalities was Rauch (1993). Using the cross sectional data from the 1980 US Census he found that a one year increase in average education raises wages by 3% to 5% percent in 1980. As point out by Moretti (2004a), Rauch's paper suffers from two methodology limitation, first is that he does not directly account for the endogeneity of aggregate human capital and, second, he does not distinguish between externalities and complementarity between skilled and unskilled workers.

In order to circunvey these estimation problems, Acemoglu & Angrist (2001) uses variations in compulsory education laws and the birth quarter of individuals as an instrumental variable for secondary schooling in a panel fix effects of U.S. states, finding smaller external returns to education, between 1% and 3%. However, the authors themselves argue that variations in compulsory scholling laws would affect mainly secondary education. Ciccone & Peri (2002) argue that studies with a Mincerian approach that try to identify human capital externalites with a standard Micerian wage regression, such as the previous cited ones, may confound positive externalities with wage changes arising from a negatively inclined demand curve for human capital. The authors propose to use a new approach holding the labour force skill composition constant, not finding significant externalities for cities and states between 1970 and 1990.

Focusing on college level education, Moretti (2004a,b), using a panel data, estimate a model of non-random selection of workers among cities, accounting for unobservable city-specifc demand shocks by using two instrumental variables for human capital concentration, namely, lagged demographic scholling structure of the city and the existence of a land-grant college. He finds that a percentage point increase in the supply of college graduates raises high school drop-outs' wages by 1.9%, high school graduates' wages by 1.6%, and college graduates wages by 0.4%. Focusing on wage inequality across US metro regions, Florida & Mellander (2014) find that regional variation in wage inequality, on the one hand, is associated with human capital, skill levels and occupational structure, in line with previous studies of skill-biased change and job polarization.

Considering the evidence for European urban centers, Heuermann (2011) using panel data and employing instrumental variables estimate human capital externalities for highly and non-highly educational groups and different sectors. Although the authors finds high a positive and significant effect, on contrary to Moretti (2004a), the effect is larger for the highly educated group, a 1.8% increase in the high skilled wage group and 0.6% for the non-highly educated. Also, the effects are larger in the manufacturing sector than in the service sector. More recently, Broersma *et al.* (2015), analyzes the effect of human capital externalities specifically for low-educated workers and for jobs with low skill requirements, using a employer-employee matched dataset for Dutch workers. Through a mincerian estimation in a multilevel model approach, including regional level data, the authors find that, workers on low-skilled jobs earn higher wages when working in cooperation with workers in high-skilled jobs, while for low-educated workers such cooperation with high-educated workers is negative. In a cross-country analysis of productivity in five OECD countries (Germany, Mexico, Spain, United Kingdom, and United States), Ahrend *et al.* (2014) uses a two-step econometric approach that enable to capture the productivity clean for individual unobserved heterogenity and sorting of more productive individuals into cities. The autors find that city productivity premiums tend to increase with city size. Specifically, a twofold increase in city size is associated with a 2-5% increase in productivity.

For developing Countries the lirature is scarcer. Liu (2007), follows the empirical strategy of Moretti (2004a) for assessing the external returns to education in China. He finds that the external returns are at least as high as the private returns to education, ranging from a low of 4.9% to a high of 6.7%. Two-stage least squares estimates indicate that a one-year increase in city average education could increase individual earnings by between 11 and 13%, and social returns can be as high as 16%. A significantly higher return than those found in developed countries. For Kenya, Manda *et al.* (2002) uses a Mincerian equation approach with district-level average education attaiment to capture the effect of human capital on monthly earnings, finding positive effects for male and female workers. More interestings, an increase in average human capital for females has a positive impact on earnings of male workers relative to female workers.

Specifically for Brazil, most papers focus on the role of density through an agglomeration economies approach such as Rocha *et al.* (2011) and more recently, Barufi *et al.* (2016) analyzes the effect of sector-specific agglomeration economies. They find that urbanisation economies positively and significantly affect the manufacturing and servive sectors, with impact of populationd density in local wages ranging from 5.1% to 9.4%. Despite the scarce literature some exceptions can be found, early works have studied the impact of human capital concentration on wages (Araujo Junior & Silveira Neto, 2004; Falcão & Silveira Neto, 2007), however these initial studies do not address bias issues.

3 Theoretical Framework

In order to estimate the effect of human capital concentration on productivity (trought wages), we must first present a theorical model. We use a simple framework proposed by Moretti (2004a) based on the general equilibrium model of Roback (1980, 1982). In this proposed theoretical framework the identification of externalities generated by the concentration of human capital, considers the imperfect substitution by different types of workers, skilled and unskilled. Considering, like Heuermann (2011), that the production function of the Cobb-Douglas type depends on two inputs, skilled and unskilled workers, we have to:

$$Y_j = (\theta_{1j} N_{1j})^{\alpha} (\theta_{2j} N_{2j})^{1-\alpha}, \quad \alpha \in (0, 1) \tag{1}$$

where N_{1j} quantifies the total number of skilled workers in municipality *j* and N_{2j} the same measure for unskilled workers, with productivity mesured by θ_{1j} and θ_{2j} respectively. However, we use a more general and flexible functional that relates productivity and the share of skilled workers in region j (s_j), specifically:

$$\log(\theta_{ij}) = \phi_{ij} + f_i(s_j, \gamma), \quad i = 1, 2 \tag{2}$$

Where $f_i(s_j, \gamma)$ associates the productivity of each worker to the portion of skilled workers, and a positive parameter . Therefore, equation (2) indicates that the productivity of worker i in the municipality j depends on its own level of human capital, but also on the local stock of human capital of the locality j that must act by increasing the productivity of both workers positively. However, if the externalities generated by the concentration of human capital are not correlated with the productivity of the workers, it should be equal to zero. The generalization made by equation (2) is important because, unlike Moretti (2004a) and Heuermann (2011), the effects associated with human capital extrenalities does not necessarily act with equal intensity for both types of workers. Therefore, as seen bellow, with the generalization it is possible to have differentiated gains according to different levels of workers schooling.

Equating the wage of each type of worker to the marginal product of labor from the production function specified by equation (1), using equation (2) and $s_j = N_{1j}/N_{1j} + N_{2j}$ It follows:

$$\begin{aligned}\log(w_{1j}) = log(\alpha_1) + \alpha_1 \log(\theta_{1j}) + (\alpha - 1)\log(s_j) + (1 - \alpha_1)\log(\theta_{2j}) \\ + (1 - \alpha_1)\log(1 - s_j)\end{aligned} \tag{3}$$

$$\begin{aligned}\log(w_{2j}) = log(1 - \alpha_1) + \alpha_1 \log(\theta_{1j}) + (\alpha - 1)\log(s_j) + (1 - \alpha_1)\log(\theta_{2j}) \\ + (\alpha_1)\log(1 - s_j)\end{aligned} \tag{4}$$

By deriving equations (3) and (4) for region *j* as a function of the share of skilled workers, the impact of a marginal increase in the number of skilled workers on the salary of both workers is obtained as follows :

$$\frac{d\log(w_{1j})}{ds_j} = f_1^{'}(s_j, \gamma) + \frac{\alpha - 1}{s_j - s_j^2}, \tag{5}$$

$$\frac{d\log(w_{2j})}{ds_j} = f_2^{'}(s_j, \gamma) + \frac{\alpha}{s_j - s_j^2}. \tag{6}$$

Skilled workers will benefit the most by higher wages, by taking advantage of the externalities generated by the concentration of human capital if the following condition is observed:

$$f_1^{'}(s_j, \gamma) > f_2^{'}(s_j, \gamma) + \frac{1}{s_j - s_j^2} \tag{7}$$

Here two consideration must be made. According to expression (7), a marginal increase in the stock of human capital in a given region will have a greater positive effect on skilled workers wages than of unskilled workers if $f_1^{'}(s_j, \gamma)$ the "externality effect"overcome the wage gains of unskilled workers $f_2^{'}(s_j, \gamma)$ added to the "neoclassical net effect"$1/s_j - s_j^2$. Also, the neoclassical "net effect"is minimum when $s_j = 1/2$, taking te low percentage of skill workers in Brazil we can assume that $s_j < 1/2$, implying a large “neoclassical effect” on regions. Therefore, for an empirical confirmation of equation (7), externalities must be stronger than the “neoclassic net effect” of the region. Analogously, under these conditions, in the case when $f_1^{'}(s_j, \gamma) = f_2^{'}(s_j, \gamma) = \gamma s_j$, external returns to education for unskilled workers will be higher for low human capital stocks in "j". According to this model, it is possible that empirical evidence for countries or even regions of the same country with different levels of human capital presents different results regarding the magnitude of the relation between concentration of human capital and productivity gains, not being directly comparable.

In summary, the main result of the model for equation (2), suggests that human capital concentration influences skilled and unskilled workers wages differently through two forces, “spillovers effect” and the “neoclassical effect”. The former operates through productive externalities generated by the concentration of skilled workers - which makes access to information and learning easier and faster, among other reasons - increasing the productivity of both workers. The later derives from imperfect substitution between skilled and unskilled workers - an increase in the supply of skilled workers reduces the productivity of these individuals and thus their wages, but raises the productivity of the unskilled.

4 Empirical Strategy and Descriptive Statistics

To achieve the goal of correctly estimate human capital externalities in local productivity, measure trought real wages, we must consider the possible sources of omitted variable bias and the subsequent difficulties that its existence imposes the estimation

The first source of bias, on the individual level, is the workers unobserved heterogeneity may affect estimation. Individuals observed in cities with a higher concentration of human capital may have more ability than individuals with the same formal educational level residing in a location with a lower concentration of human capital. As stated by Moretti (2004a), this type of sorting can be problematic if cities with a higher share of college are associated with a high return for unobserved abilities, attracting the most skillful and affecting the labor force composition.

Second, on the local level, another potential source of bias is related to shocks in the local labor market, which may be correlated with the concentration of skilled workers in the locality. Even with cities fixed effects, time-varying factors may affect the stock of human capital or wages between cities. For example, shocks can attract skilled workers to a particular region and to raise wages (Moretti, 2004a; Falcão & Silveira Neto, 2007). This problem is adressed by constructing a demand shock index for each educational group and area, which represents the expected change in employment for an educational group in the area.

Finally, as explained in Combes *et al.* (2010), density and measures of productivity (wages) may be simultaneously determined, because more productive places tend to attract more workers and as a result become denser. This issue is refered as the "endogenous quantity of labor" problem, a higher employment density can exacerbate the knowledge spillover, despite of human capital concentration. Including employment density as a control variable, seeks to obtain a effect of human capital externalities "clean" of density effect.

4.1 Model Specification

To simultaneously control all of these bias sources, we use a two step estimation approach[2] adapting from Heuermann (2011) and Groot *et al.* (2014). The first step consists in estimate a Mincerian type equation for individual-level data given by equation (8):

$$\begin{aligned} \ln(w_{i,j,t}) &= \beta_0 + \theta_i + \beta_1 age_{i,j,t} + \beta_2 age^2_{i,j,t} + \beta_3 tenure_{i,j,t} + \textstyle\sum_{edu} \beta_{edu} D^{edu}_{i,j} \\ &+ \textstyle\sum_{sector} \beta_{sector} D^{sector}_{i,j} + \textstyle\sum_{size} \beta_{size} D^{size}_{i,j} + \textstyle\sum_j \sum_t \lambda_{j,t} D_{j,t} + \epsilon_{i,t}, \end{aligned} \quad (8)$$

where $w_{i,j,t}$ is the wage from individual i in local j on year t, θ_i is the individual fix effect, $D^{edu}_{i,j}$ are a series of educational dummies, $D^{sector}_{i,j}$ are sector dummies, include to control for sector-specific effects, $D^{size}_{i,j}$ are firm size dummies, measure by number of employees. Most importantly, $D_{j,t}$ are dummies for each local and year, providing a local-year specific wage index, given by $\lambda_{j,t}$. Ideally, estimating equation (8) using longitudinal data, allows to control for an important source of bias, the unobserved heterogeneity in workers abilities.

The second step regression uses first step estimates of local wage indexes, $\hat{\lambda}_{j,t}$, as dependent variable in a regression on local human capital concentration, employment density. Also, as we know from economic theory, market structure can affect prices and wages, therefore, we include the share of jobs in industry, a sector with historically higher wages and three measures for, specialization, diversity and competition[3] (that will be define below) as shown in equation (9):

[2] A similiar approach was used by Ahrend *et al.* (2014) for estimate agglomeration benefits based on city productivity differentials across five OECD countries (Germany, Mexico, Spain, United Kingdom, and United States).

[3] see Barufi *et al.* (2016).

$$
\begin{aligned}
\hat{\lambda}_{j,t} \quad &= \alpha_0 + \alpha_1 humancapital_{j,t} + \alpha_2 \ln(employ.density_{j,t}) + \alpha_3 \ln(area_j) \\
&+\alpha_4 industryshare_{j,t} + \alpha_5 specialization_{j,t} + \alpha_6 diversity_{j,t} \\
&+\alpha_7 competition_{j,t} + \alpha_8 demandshocks_j + \varepsilon_{j,t}
\end{aligned} \tag{9}
$$

In this set up we are interested in correctly estimate α_1, the effect of human capital concentration on the local wage index. However, as discussed earlier, in the second step we also should use strategies to control for possible sources of bias.

First, in order to control for the "endogenous quantity of labor" we first use lagged population density as instrument for current employment density. As argued by Combes *et al.* (2010) for the french case, considering a relatively constant urban hierarchy, highly dense regions today were highly dense in the past, however, in those days the most important sector was agriculture and density was affect only by the capability of sustain population. In a relative young country as Brazil, the urban hierarchy may not be highly stable and the drivers of urban center development in Brazil can be potentialy related to main historical economic cycles from the colonial period, such as the sugar-cane, gold mining and coffee production as discussed by Naritomi *et al.* (2012). The autors construct a variable which indicates the influence of each of those economic cycles in the brazilian municipalites with value 1 for those directly affected and defined by

$$
I_i = \begin{cases} \left(\dfrac{200 - d_i}{200}\right) & \text{se } d_i \leq 200\text{km}, \\ 0 & \text{otherwise}, \end{cases} \tag{10}
$$

where d_i is the distance from municipality i to the closest municipality directly involved in the respective resource boom. The assumption is that these economic episodes help define the brazilian urban structure by partialy deteming urban centers location and development and no longer are related to the current productive structure, making then a plausable instrument for employment density.

Also, we must include other market area variables related to local wages. For market measures we first use the specialization index given by equation (11), in which a value close to 0 indicates industrial composition in the region is similar to the national one, for values closed to 1 the labor market area is completely specialized. As Henderson (2010) affirms, standardized manufacturing activity benefits from agglomeration generating productivity improvents that can affect local wages.

$$
Specialization_{region} = \sum_i \left(\frac{E_{i,region}}{E_{i,region}} - \frac{E_{industry}}{E_{country}} \right)^2 \tag{11}
$$

The second market measures included as control is for diversity, since Jacobs (1970) arguments that a city with larger sector diversity, because in diverse cities there is more interchange of diferent ideas (Glaeser *et al.* , 1992). Following Combes *et al.* (2011), we use a Inverse Herfindahl Index given by equation (12), in which higher values represents higher diversity within labor market area.

$$
IHI = \left(\frac{E^2_{region}}{\sum_{ind} E^2_{industry,region}} \right) \tag{12}
$$

For the last market measure, we include a degree of competition measure given by equation (13), where values larger then one indicates larger competition in the industry within a particular labor market area. Competition can have a dual effect on productivity improvements, which can affect local wages. A greater competition accelerate imitation and improvement of inovative ideias on one hand and reduces returns to inovators (Glaeser *et al.* , 1992).

$$C_{industry,region} = \frac{F_{industry,region}/E_{industry,region}}{F_{industry,country}/E_{industry,country}} \tag{13}$$

Finally, we use an instrumental variable approach proposed by Moretti (2004a) to control for possible demand shocks that attract skilled workers to a particular region and raise wages. The proposed instrument seeks to capture the generational change in the schooling of the regions considered in the analysis, such instrument is supposedly correlated with the human capital stock, but is not influenced by current shocks in the local labor market. Moretti (2004a) argues the existence of a long-run trend of incresing education, as younger and more educated cohorts enters the labor force, as relative population shares of diferent cohorts vary across cities, this will lead to diferential trends in college share across cities. The instrument is defined as:

$$IV = \sum_{m} \omega_{mj} \Delta P_m, \tag{14}$$

where m identifies age groups, j the geographical unit, ω_{mj} the proportion of individuals living in the region j in 1980 and who, in 2000's, belonged to the m age group, finally ΔP_m captures national change in the share of skilled workers by group m between the various years of the study (2002 to 2014). The authors note that, if the age distribution of cities may reflects expected changes in the local economy the estimates would still be biased, usage of lagged values circunvent this issue.

4.2 Data

As previously discussed, in order to achieve proposed objective, it's necessary to initially estimate equation (8), generating a local wage index . For this, we must have individual-level wage information, socioeconomic characteristics, labor relations type and individual location. For the brazilian case, at the national level, the main database that meets these requirements is given by the Ministry of Labor's Annual Social Information Report (RAIS).

The RAIS dataset consists of information on all formal workers in the country, whose advantages are its scope, longitudinal structure of the information and geographical breakdown. These advantages allow include worker fixed effects in order to correct the unobserved heteroneneity bias. However, because it is restricted to formal workers, it leaves aside a large part of the individuals employed in the country. Fortunately, as emphasized by Baruffi et al. (2016), during the first decades of economic stabilization (first and second decades of the current century), the process of formalization guarantees a greater representativeness of the RAIS dataset. To put into perspective, for 2002 the RAIS dataset contains about 46 million workers, increasing in 2014 to more then 76 million workers.

In order to avoid computational issues we decide to keep only seven time periods, all the even year from 2002 to 2014. Also, we keep only the workes who had a permanent job contract with at least 20 weekly hours that was active at the end of the year. In order to create a more homogeneus sample we also removed the public administration workers, since wages in this sector do not necessarily reflect workers productivity and presents a specific contract relation. Finally, we considered only men between 18 and 56 years of age. With the selections discussed above, we built a panel at individual level for the workers present in every considered year. With this in mind our final sample consists of 1,327,411 workers for each year, for a total of 9,291,877 observations.

For the Brazilian case the most disaggregated administative unit is the municipality, however, labor market interactions often cross administratively defined boundaries. For this reason, Combes *et al.* (2008, 2010) for the french case, use employment zones built by the French Ministry of Labor in their analisys of agglomeration externalities and Ahrend *et al.* (2014) uses *functional urban areas*, built by for OECD countries, in their analysis cross-country analysis of productivity. A direct equivalent for Brazil are the so-called *Regiões de Influência* (Regions of Influence) which are areas of influence of local urban centers that takes into account the brazilian urban network and all daily commuting and transportation connections among the municipalities, created by the Brazilian Statistics Bureau IBGE (2013). In it's most desagregates form, the country is divided into 482 Regions of Influence (REGIC) that can be thought as labor market areas and will be reffered to as such from now on. Before presenting results, we perform a brief descriptive analysis of the data used, in order to identify the main characteristics of the dataset and to substantiate the later discussion of results.

Tabela 1: Descriptive Statistics for Individual Workers Sample (First Step)

	2002		2014		Whole Period	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Wage (R$ per month)	741.07	26,677.3	2,304.4	4,762.7	1,370.94	11,592.8
Age (years)	33.89	9.10	45.9	9.11	39.63	10.03
Tenure (months)	61.6	69.5	128.6	109.6	97.53	91.1
Educational Levels						
Illiterate (%)	0.02		0.01		0.01	
Incomplete Middle School (%)	0.36		0.24		0.29	
Complete Middle School (%)	0.28		0.22		0.25	
Complete High School (%)	0.26		0.38		0.32	
Incomplet College (%)	0.03		0.02		0.03	
College (%)	0.06		0.13		0.10	
Firm Size (Employee Number)						
(0;4)	0.095		0.076		0.059	
(5;9)	0.099		0.087		0.087	
(10;19)	0.109		0.095		0.097	
(20;49)	0.138		0.131		0.122	
(50;99)	0.101		0.097		0.106	
(100;249)	0.128		0.119		0.115	
(250;499)	0.103		0.099		0.108	
(500;999)	0.091		0.097		0.096	
1000+	0.134		0.198		0.147	

Source: author's own calculation based on the RAIS dataset.

In Table 1, we present the descriptive statistics for the first step variables, that is, in equation (8) for 2002, 2014 and the role period. As can be seen, the average monthly wage for the period is R$ 1,370 with a high standard deviation indicating great wage inequality among sample workers. Also, the average age is almost 40 years and a 98 months tenure, this can be reflected in the restriction of the sample to individuals that are present in the sample in all seven periods considered between 2002 and 2014. The low percentage educational level obtaneid by the workforce is striking, Table 1 shows that only 10% of them have at least a college degree, the majority finishes high school. By this metric it is easy to observe the low overall stock of human capital in brazilian cities.

The descriptive statistics of second step (labor market area level) information are show in Table 2. As previously discussed, local wages indexes where estimated by including dummies for each labor market area and year combination, thus it becomes a measure of wage for each labor area. On average, 6% of a regic workers have college degrees, our chosen metric for human capital concentration. The high standard deviation (0.12) imply a high human capital inequality between labor market areas. Indeed, many recent articles in Brazil discuss the importance of human capital and wage differentials in income inequality and its recent decline (Barros *et al.* , 2007; Tavares & Menezes Filho, 2011; Silva *et al.* , 2016).

As can be seen, on average labor areas have similar composition to national one. The index shows a larger variation with an average of 0.31 and 0.36 as standard deviation. As we can see, on average, labor market areas presents less competition compared to the country. However, there is cases of high competition as the maximum copetition index approachs 5.

Finally, in Table 2 we have the measure of proximity of each labor market area to major colonial economic cicles. As previously explained, values close to 1 shows greater proximity and therefore greater influence. As shown by Naritomi *et al.* (2012), these economic episodes have a long standing impact on local institutional framework, measure by access to justice, land inequality and governance. For our purpuses, these institutional outcomes can affect human capital concentration and density today, and must be considered.

Tabela 2: Descriptive Statistics for Local Variables (Second Step)

Variable	Mean	Std. Dev	Min.	Max
Local Wage Index (estimated coefficients)	4.34	0.17	3.17	4.90
Local Human Capital (share of college graduates)	0.06	0.12	0.01	0.22
ln *Employment Density* ($jobs/km^2$)	0.32	1.64	-5.44	4.31
ln *populational density in 1940* (hab/km^2)	2.15	1.54	-3.63	5.24
ln *area* (km^2)	8.85	1.28	5.70	13.6
Jobs in Industry (%)	0.22	0.09	0.01	0.49
Skilled Demand Shock	1.07	0.21	0.00	1.41
Unskilled Demand Shock	0.39	0.15	-1.15	0.89
Market Measures[b]				
Specialization Index	0.02	0.03	0.00	0.37
Diversity Index	0.31	0.12	0.08	0.76
Competition Index	0.84	0.26	0.36	4.88
Historical Variables[c]				
Sugar Cane Boom	0.08	0.21	0.00	1.00
Gold Boom	0.11	0.26	0.00	1.00
Coffee Boom (Colonial Era)	0.09	0.26	0.00	1.00
Coffee Boom (Post-Colonial Era)	0.15	0.33	0.00	1.00

Source: author's own calculation based on the RAIS dataset.
[a] All calculations based on the Region of Influence (REGIC) geographical unity.
[b] Following Barufi *et al.* (2016) the respective measures are: degree of specialization, inverse herfindahl and degree of competition.
[c] Refers to the value of the index created by Naritomi *et al.* (2012) for mesuring the proximity to each of the refered colonial economic booms.

5 Results

In Table 3, we show the results for the first step estimation, following equation (8). As we can see, the overall fit is satisfactory with a R^2 of 0.45 with all control variables presenting highly significant coefficients. Wages are higher for older and longer tenured workers, although there is a nonlinearity in age presente by the negative coefficiente of age^2. Higher educational levels and working on larger firms positively correlate with higher wages. The inclusion of dummies for REGIC and year interactions, allows the estimation of $\lambda_{j,t}$ to be use in the second step.

As previously discussed, the unobserved worker heterogeneities can turn OLS estimations inconsistent, therefore, we perform a panel with individual fixed-effect estimation, the last two columns of Table 3 presentes the results. As we can see, effects of educational levels an firm size on individual wages, although still highly significant, are strongly reduced, indicating that a considerable share of those returns were from unobserved ability. This highlights the importance of controls for unobserve ability. Given that fixed-effect estimation appears to be the most adequate[4], in the second step the wage indexes from this estimation will be used.

In the second step, with the estimates of local wage indexes and control variables aggregated in labor market area level, we estimate equation (9). Again, to provide stronger conclusions, the existence of sorting and consequent endogeneity of human capital concentration and local wage indexes requires an estimation strategy with instrumental variables, therefore the second step was estimated using a 2SLS approach. In order to assess their relevance, in Table 4 we present the first stage estimation. We can see that as expected the proposed instruments are relevant only for it's endogenous counterparts, that is, the demographic schooling structure instrument is relevant for Human Capital and population density, sugar, gold and coffee booms are relevant for employment density.

Second step results are presented in Table 5. For comparison, we present, in the first column, OLS estimation of the equation (9), we can see that the concentration of human capital has a positive

[4] As in Combes *et al.* (2010) indetification of parameters come from worker movement between labor market areas and from time variation for worker ho didn't move

Tabela 3: Human Capital Concentration and Local Wages (Mincerian Equation)

	Dependent Variable.: Individual Hourly Wage			
	OLS		Panel	
Explanatory Variables	Coef.	Stand. Dev.	Coef.	Stand. Dev.
age (years)	0.044***	(0.000)	0.097***	(0.001)
age^2	-0.001***	(0.000)	-0.002***	(0.000)
tenure (months)	0.002***	(0.000)	0.002***	(0.000)
tenure^2	0.000***	(0.000)	0.000***	(0.000)
Educational Levels				
< Middle School	0.177***	(0.002)	0.007***	(0.002)
Middle School	0.296***	(0.002)	0.001***	(0.002)
High School	0.581***	(0.002)	0.017***	(0.002)
College	1.397***	(0.002)	0.131***	(0.002)
Firm Size (Employee Number)				
(0;4)	-0.192***	(0.012)	-0.124***	(0.012)
(5;9)	-0.123***	(0.012)	-0.090***	(0.012)
(10;19)	-0.041***	(0.012)	-0.045***	(0.012)
(20;49)	0.051***	(0.012)	-0.002***	(0.012)
(50;99)	0.136***	(0.012)	0.046***	(0.012)
(100;249)	0.229***	(0.012)	0.094***	(0.012)
(250;499)	0.275***	(0.012)	0.130***	(0.012)
(500;999)	0.298***	(0.012)	0.151***	(0.012)
1000+	0.326***	(0.012)	0.167***	(0.012)
Ignored	0.305***	(0.012)	0.180***	(0.012)
Intercept	0.250***	(0.072)	0.356***	(0.075)
Sector Dummies	Yes		Yes	
$\lambda_{j,t}$ *Dummies*	Yes		Yes	
R^2	0.452		0.455	
N	8,928,383		8,928,383	

Significance levels: *** 1% , ** 5% and * 10% .

and significant effect on the hourly wages of workers, a one point increase in the share of workers with college increases labor market areas wages by 0.70%. However, employment density presents a negative impact of wages, contradicting the aglomeration literature when we don't control for the "endogenous labor quantitiy".

The 2SLS estimation results are presented in the third column of Table 5. As we can see, the effect of interest grows to 0.859, a 22% increase, indicating that a one point increase in the share of workers with college increases labor market areas wages by 0.86%. This effect in similar to Moretti (2004a) findings of positive externalities for human capital concentration (around 1.14%). With inclusion of instruments for employment density it's estimate becomes positive and significant, implying that denser labor market areas presents higher wages as in Barufi *et al.* (2016). Market measures introduced as controls are not statistically significant but their inclusion in important for an obtaining an unbiased estimate of human capital concentration on local wages. Finally, although skilled demand shock no longer plays a significant role, the impact of unskilled demand shocks remains important for the wages in the period.

These results offer evidence on the existence of human capital externalities, however, in order to empiricaly test the assumption of a distinct effect between educational groups as proposed by the theoretical model of Section 3, we must perform separate estimates for different educational groups.

Tabela 4: Human Capital Concentration and Local Wage Index (2SLS First Stage)

	Human Capital[a]		ln *Employ. Density*	
Explanatory Variable	Coef.	S.D	Coef	S.D
Age Structure	2.36***	(0.104)	-5.240	(3.604)
Pop. Density 1940	0.000	(0.000)	0.325***	(0.062)
Sugar Boom	-0.003	(0.004)	2.63***	(0.255)
Gold Boom	-0.003	(0.003)	0.544**	(0.242)
Coffee Boom (Colonial)	0.003	(0.005)	1.453***	(0.193)
Coffee Boom (Post-Colonial)	0.016**	(0.006)	0.603***	(0.235)
Constant	0.035***	(0.010)	-3.902***	(0.716)
Exogenous Regressors	Yes		Yes	
F-test	453.5***		35.40***	
Sanderson-Widjmeier[a]	680.09***		243.95***	
N	390		390	

Significance levels: *** 1% , ** 5% and * 10% .
[a] Human Capital Concentration is measure as the share of workers with at least a college degree;
[b] Consists of a weak identification test of individual endogenous regressors. They are constructed by "partialling-out"linear projections of the remaining endogenous regressors.

5.1 Effect on Different Educational Groups

As previously discussed, human capital concentration influences skilled and unskilled workers wages differently through two forces, “spillovers effect” and the “neoclassical effect”. The former operates through productive externalities generated by the concentration of skilled workers - which makes access to information and learning easier and faster, among other reasons - increasing the productivity of both workers. Therefore we must reestimate the equation (8) considering this differentiation.

Considering that both groups are present in the regions labor markets and wage determination is simultaneos, in this set up separately estimate the first step for each group is not a good strategy. Therefore, we included a interaction dummy between $D_{j,t}$ the dummies for each local and year, with a skilled worker dummy. Implicitly considering the hypothesis that there is an effect of the concentration of human capital for all the workers, with a specific return added for skilled workers trough education.

The second step regression uses first step estimates of local wage indexes $\hat{\lambda}_{j,t}$ only, for the unskilled group and $\hat{\lambda}_{j,t}$ added with a dummy for skilled group, estimation results are presented in Table 6. Theorical prediction, according to equation (7) is that a marginal increase in the stock of human capital in a given region will have a larger positive effect on skilled workers wages than of unskilled if the "externality effect"overcome the neoclassical "net effect". As we can see, our results show that wages of both groups are affected by the externalities resulting from the concentration of human capital. Moreover, as the effect found is greater for non-skilled individuals (0.73%) then for skilled (0.25%), we can imply that these externalities for the skilled group are not strong enough to outweigh the gains in the wages of the unskilled when we also consider the neoclassical effect. These results are similar to Moretti (2004a), which finds a 1.6% for highschool graduates and 0.4% for college graduates.

5.2 Effect on Different Sectors

As stated by Heuermann (2011), labour market institutions and the relative importance of knowledge and physical capital as factors of production can be quite diverse across industries. This fact together with findings of wage differences across sectors makes interesting to know the specific impacts of human capital concentration across industries in order to promote strategies and improve their efficiency. Therefore, in order to evaluate these differences, we repeat our baseline estimation separately considering a agregation of two-digit CNAE divisions into four large groups, namelly, manufacturing industry, extractive industry, construction and services. Results are presented in Table (7)

Tabela 5: Effects of Human Capital Externalites in Local Wage - Reduced Form

	OLS		2SLS	
	Coef.	Std. Dev	Coef.	Std. Dev
Human Capital[a]	0.699***	(0.130)	0.839***	(0.129)
ln *Employment Density*	-0.004	(0.007)	0.024**	(0.010)
ln *REGIC Area*	0.002	(0.007)	0.002	(0.006)
Jobs in Industry	-0.127	(0.097)	-0.134	(0.095)
Diversity	0.049	(0.048)	-0.005	(0.048)
Specialization	0.115	(0.307)	-0.580	(0.428)
Competition	0.043	(0.027)	0.001	(0.018)
Centrality	-0.098**	(0.042)	-0.085*	(0.045)
Foundation (year)	0.000	(0.000)	0.000	(0.000)
Major Port Dummy	0.024	(0.047)	-0.052	(0.040)
Skilled Demand Shock	-0.053**	(0.024)	-0.024	(0.021)
Unskilled Demand Shock	-0.075**	(0.024)	-0.106***	(0.035)
Constant	4.46***	(0.016)	4.39***	(0.145)
Regional Dummies	Yes		Yes	
Kleibergen-Paap[b]			78.30	
Cragg-Donald Wald F[c]			33.2	
Hansen-J[d]			8.20	
R^2	0.311			
N	482		390	

Note: Dependent variable in the local wage index generated at the first step mincerian wage equations given by equation (8). Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.

[a] Human Capital Concentration is measure as the share of workers with at least a college degree;

[b] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.

[c] Should be compared with Stock & Yogo (2002) critical values.

[d] Hansen-J test is for overidentified restrictions. The null hipothesis is that all instruments are valid.

Significance levels: *** 1% , ** 5% and * 10% .

As we can see, the results confirm a variety in human capital concentration effects the effects between different sectors, implying that human capital concentration benefits are unevenly distributed. For example, we find larger effects for manufacturing with (0.52%) and services (0.36%), with no significant effect for the remaining two groups, extractive industry and construction. However, these results are intuitive given their underlying caracteristics of low gains of exchange of information and social enviorment. On the other hand, concentration externalities play a important role for extractive, as the coeficient of employment density was positive and highly significant for these sectors.

Our results are in line with (Heuermann, 2011), which finds for West Germany positive effects for the manufacturing industry and services, the former with highest effects than the later. As an explanation, the author uses the notion of pecuniary externalities versus tecnological externalities as microeconomic sources for human capital externalities. The first arises when the firms relate their investments in physical capital to local human capital endowments, which could be true for physical capital intensive industries like the manufacturing. while the second arises when externalities come from the exchange of knowledge between firms and workers, more probable in services. Specifically for Brazil, we have no knowledge of directly comparable results, since Barufi *et al.* (2016) focus on the employment density effect.

Tabela 6: Human Capital Concentration and Local Wage Index (2SLS Reduced Form Educational Groups)

	Depedent Variable: Local Wage Index			
	Skilled[a]		*Unskilled*[b]	
	Coef.	Stand. Dev.	Coef.	Stand. Dev.
Human Capital	0.254**	(0.134)	0.737***	(0.136)
ln *Employment Density*	0.006***	(0.005)	0.014***	(0.008)
ln *REGIC Area*	0.001	(0.004)	-0.012	(0.009)
Jobs in Industry	0.010	(0.080)	-0.117	(0.095)
Diversity	0.001	(0.043)	0.001	(0.047)
Specialization	-0.628	(0.347)	-0.279	(0.439)
Competition	-0.005	(0.014)	0.004	(0.018)
Skilled Demand Shock	-0.020	(0.024)	-0.005	(0.025)
Unskilled Demand Shock	-0.126	(0.036)	-0.107***	(0.040)
Constant	-0.804***	(0.058)	4.70***	(0.000)
Macro Dummies	Yes		Yes	
Kleibergen-Paap[c]	72.5		72.5	
Cragg-Donald Wald F[d]	37.5		37.5	
Hansen-J	2.55		9.14	
N	390		390	

Note: Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.
[a] Dependent variable is $\hat{\lambda}_{j,t} + \hat{\gamma}_{j,t}$, the local wage index generated at the first step mincerian added with a local wage index of only the skilled workers.
[b] Dependent variable is $\hat{\lambda}_{j,t}$ the local wage index generated at the first step mincerian equations given by equation (9)
[c] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.
[d] Should be compared with Stock & Yogo (2002) critical values.
[e] Hansen-J test is for overidentified restrictions. The null hipothesis is that all instruments are valid.
Significance levels: *** 1% , ** 5% and * 10% .

5.3 Robustness Checks

In order to assess the validity of human capital agglomeration effects, in this section we perform a series of robustness checks. First we include as controls variables related to amenities and geographical characteristics possibly correlated to wages. With the notion of compensatory wages[5], part of the existing wage variations compensate the accessibility to amenities. Without including controls for this compensation could lead to a overestimation of the true effect, therefore, in order to correctly evaluate the effect of human capital concentration on local wage indexes, inclusion of amenities and geographic characteristics controls is necessary. Therefore, we include altitude, number of sunny days, number of rainy days, distance to equator and distance to sea as controls.

In the remainder robustness checks, we repeat the baseline estimation using subsets of labor market areas. First we exclude Rio de Janeiro e São Paulo labor makert areas from the estimation based on concerns that, given Brazil's urban hierarchy with two disproportionately big metropolitan areas, São Paulo and Rio de Janeiro influence on the overall results may be overwhelming. We then restrict the estimation for only non-Metropolitan labor makert areas is order to evaluate if results hold outside metropolitan areas. For this we considered only the areas were none of its composing municipalities lay in metropolitan regions, ending up with 357 non-metropolitan labor market areas.[6]. Finally, as a way to evaluate if the results are

[5] See Roback (1982).

[6] Unfortunately, it is not possible to perform an estimation considering only the metropolitan regions by their small number, creating problems on the estimation matrix rank.

Tabela 7: Human Capital Externalities Effects by Economic Sectors

	Depedent Variable: Local Wage Index			
	Manufacturing	Extractive	Construction	Services
Human Capital^{b}	0.524**	-1.687	-1.141	0.365***
	(0.244)	(2.749)	(1.142)	(0.166)
ln *Employment Density*	0.031	0.230**	0.119	0.025
	(0.029)	(0.106)	(0.082)	(0.017)
Constant	5.73***	5.57***	5.11***	5.75***
	(0.270)	(1.062)	(0.764)	(0.186)
Controls	Yes	Yes	Yes	Yes
Macrorregion Dummies	Yes	Yes	Yes	Yes
Fist Stage F-*test*	75.2***	71.8***	71.8***	72.5***
Kleibergen-Paap^{c}	33.2	36.1	34.8	35.6
Hansen-J	2.96	3.06	3.28	4.36
N	390	349	387	390

Note: Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.

[a] Dependent variable is $\hat{\lambda}_{j,t} + \hat{\gamma}_{j,t}$, the local wage index generated at the first step mincerian added with a local wage index of workers in that sector. This is similar with the strategy used for skilled and unskilled estimations.

[b] Human Capital Concentration is measure as the share of workers with at least a college degree;

[c] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.

[d] Should be compared with Stock & Yogo (2002) critical values.

Significance levels: *** 1% , ** 5% and * 10% .

generalized even in the smaller labor market areas of the sample, we separatly estimate the model for the 50% smaller and 50% larger areas, which produces a population cutpoint of about 250,000.

Results for all robustness checks are presented in Table (8). Column (1) shows the effect of interest on a regression including the aforementioned variables. This result should be compared with the coeficient of the 2SLS estimation of table (5). The inclusion of amenities and geografical control variables reduces the human capital concentration effect on local wage indexes from 0.86% to 0.77%, implying that 11% of the effect comes from compensation for amenities accessibility and suggesting that a high parcel of college degree ocurrs in less ameable places. More importantly, the effect of interest remains highly significant. On column (2) we have results from estimation without Rio de Janeiro and São Paulo, as we can see, they remain unchanged relative to the baseline estimate, indicating that the impact of the concentration of human capital on wage rates is not by the two largest brazilian labor market areas. Results are show in the third column of Table (8) and in this case, we found that human capital externalities are slightly larger (0.89%) and still highly significant. Results from the last two column of Table (8) shown a slightly higher point estimate of human capital concentration effect on local wages in the smaller regics, which could be associated implying with the relative scarcity of human capital in those areas.

6 Concluding Remarks

As recent research about the external returns to human capital shows, these gains can be bigger than individual returns and, thus, can play an important role in explaining spatial urban inequalities. Recent research about urban agglomeration gains in Brazil indicates that they are significant , but these researches have not explored the importance of local concentration of human capital. The present paper, fills that gap, with the objective of identify and to measure the external returns to human capital concentration, defined by the share of local workes with at least college degree, in Brazil. In agreement with the avaible national and international literature, we find strong evidence of the existence of large external effects of concentration of human capital in Brazil, with a 0.86% baseline effect of human capital concentration on local wages.

Tabela 8: Human Capital Externalities Effects - Robustness Checks

	Depedent Variable: Local Wage Index				
	(1)	(2)	(3)	(4)	(5)
Human Capital	0.770***	0.856***	0.887***	0.853***	0.825***
	(0.138)	(0.128)	(0.130)	(0.213)	(0.169)
ln *Employment Density*	0.031**	0.016***	0.017**	0.014	0.017
	(0.015)	(0.007)	(0.007)	(0.012)	(0.013)
Constant	4.59***	4.55***	4.54***	4.57***	4.54***
	(0.121)	(0.080)	(0.086)	(0.111)	(0.139)
Controls	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes
Amenities[a]	Yes	No	No	No	No
Excluding São Paulo and Rio	No	Yes	No	No	No
Only Non Metropolitan	No	No	Yes	No	No
Only 50% Smaller Regics	No	No	No	Yes	No
Only 50% Bigger Regics	No	No	No	No	Yes
Kleibergen-Paap[b]	53.3***	75.6***	71.6***	36.9***	39.5***
Cragg-Donald Wald F[c]	16.5	38.6	35.7	17.7	17.9
Hansen-J[d]	6.47	8.57	9.29	8.43	5.95
N	390	388	357	195	195

Note: Dependent variable in the local wage index generated at the first step mincerian wage equations given by equation (8) for each educational group. Excluded instruments are: Lagged age structure, 1940's population density, Sugar, Gold and Coffee boom proximity.

[a] The amenity variables are altitude, number or sunny days, number of rainny days, distance to the equador and distance to sea.;

[b] test of whether the equation is identified, i.e., that the excluded instruments are "relevant", meaning correlated with the endogenous regressors.

[c] Should be compared with Stock & Yogo (2002) critical values..

Significance levels: *** 1% , ** 5% and * 10% .

Moreover, theoretical prediction is that marginal increase in the stock of human capital in a given region will have a greater positive effect on skilled workers wages than of unskilled workers only if the "externality effect"overcome the wage gains of unskilled workers added to the neoclassical "net effect". We find larger effect for the unskilled workers so we may infer that human capital externalities for the skilled group, although existing, are not strong enough to outweigh the gains in the wages of the unskilled when we also consider the neoclassical effect. We also investigated the differentiated effects among sectors, finding large and significant effects for manufacturing and commercer, with no effect for extractive industry and construction, in line with available international results.

Results were robust to a series of checks, such as inclusion of amenities related and geographical characteristics controls variables possibly correlated to wages, we still find a highly signicant human capital concentration effect 0.77%, a 11% reduction from baseline estimates. We repeat the baseline estimation using subsets of labor market areas, excluding Rio de Janeiro e São Paulo, consider only non-Metropolitan, smaller and larger areas. Results ranged from 0.88% in non-metropolitan areas to 0.83% for the smaller population areas.

The existence of external returns, as found in the present study, corroborates the importance of human capital investiment as a police for reducing regional inequality as a development strategy. By presenting evidence of human capital concentration on local wages, identifing another powerful explanation for wages differences across brazilian labor markets, while at the same time contributing to evidences of human capital externalities in developing countries

Referências

ACEMOGLU, DARON, & ANGRIST, JOSHUA. 2001. How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws. *Pages 9–74 of: NBER Macroeconomics Annual 2000, Volume 15*. National Bureau of Economic Research, Inc.

AHREND, RUDIGER, FARCHY, EMILY, KAPLANIS, IOANNIS, & LEMBCKE, ALEXANDER C. 2014. What Makes Cities More Productive? Evidence on the Role of Urban Governance from Frive OECD Countries. *OECD Regional Development Working Papers, 2014/05, OECD Publishing, Paris.*

ARAUJO JUNIOR, IGNACIO, & SILVEIRA NETO, RAUL. 2004. Concentração Geográfica de Capital Humano, Ganhos de Produtividade e Disparidades Regionais de Renda: Evidências para o Brasil Metropolitano. 297–314.

BARBOSA FILHO, FERNANDO DE HOLANDA, & PESSOA, SAMUEL. 2008. Retorno da Educação no Brasil. *pesquisa e planejamento econômico*, **38**(1), 97–125.

BARROS, RICARDO PAES, FRANCO, SAMUEL, & MENDONÇA, ROSABE. 2007. A Recente Queda da Desigualdade de Renda e o Acelerado Progresso Educacional Brasileiro da Última Década. *Texo para Discussão - IPEA*, **1304**, 40.

BARUFI, ANA MARIA BONOMI, HADDAD, EDUARDO AMARAL, & NIJKAMP, PETER. 2016. Industrial scope of agglomeration economies in Brazil. *Annals of Regional Science*, **56**(3), 707–755.

BROERSMA, LOURENS, EDZES, ARJEN J. E., & VAN DIJK, JOUKE. 2015. Human Capital Externalities: Effects for Low-Educated Workers and Low-Skilled Jobs. *Regional Studies*, **3404**(October), 1–13.

CICCONE, ANTONIO, & PERI, GIOVANNI. 2002. *Identifying Human Capital Externalities: Theory with an Application to US Cities*. IZA Discussion Papers 488. Institute for the Study of Labor (IZA).

COMBES, PIERRE PHILIPPE, DURANTON, GILLES, & GOBILLON, LAURENT. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63**(2), 723–742.

COMBES, PIERRE-PHILIPPE, DURANTON, GILLES, GOBILLON, LAURENT, & ROUX, SÉBASTIEN. 2010. *Estimating agglomeration economies with history, geology, and worker effects*. Vol. I.

COMBES, PIERRE PHILIPPE, DURANTON, GILLES, & GOBILLON, LAURENT. 2011. The identification of agglomeration economies. *Journal of Economic Geography*, **11**(2), 253–266.

FALCÃO, NATASHA, & SILVEIRA NETO, RAUL. 2007. *Concentração Espacil de Capital Humano e Externalidades: o caso das cidades brasilerias*. Anais do XXXV Encontro Nacional de Economia [Proceedings of the 35th Brazilian Economics Meeting]. ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics].

FLORIDA, RICHARD, & MELLANDER, CHARLOTTA. 2014. The Geography of Inequality: Difference and Determinants of Wage and Income Inequality across US Metros. *Regional Studies*, **0**(February 2015), 1–14.

GALEANO, EDILEUZA, & FEIJÓ, CARMEN. 2013. A estagnação da produtividade do trabalho na indústria brasileira nos anos 1996-2007 :. *Nova Economia - Belo Horizonte*, **23**(1), 9–50.

GLAESER, EDWARD L. 1999. Learning in Cities. *Journal of Urban Economics*, **46**(2), 254–277.

GLAESER, EDWARD L., KALLAL, HEDI D., SCHEINKMAN, JOSÉ A., & SHLEIFER, ANDREI. 1992. Growth in Cities. *Journal of Political Economy*, **100**(6), 1126–1152.

GROOT, STEFAN P T, DE GROOT, HENRI L F, & SMIT, MARTIJN J. 2014. Regional wage differences in the netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, **54**(3), 503–523.

HENDERSON, J. VERNON. 2010. Cities and development. *Journal of Regional Science*, **50**(1), 515–540.

HEUERMANN, DANIEL. 2011. Human Capital Externalities in Western Germany. *Spatial Economic Analysis*, **6**(2), 139–165.

IBGE. 2013. *Regiões de Influência das CIDADES*.

JACOBS, JANE. 1970. *The Economy of Cities*. Randon House Inc.

JOVANOVIC, BOYAN, & ROB, RAFAEL. 1989. The Growth and Difussion of Knowledge. *Review of Economic Studies*.

LIU, ZHIQIANG. 2007. The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics*, **61**(3), 542–564.

LUCAS, ROBERT E. 1988. On the mechanics of economic development. *Journal of Monetary Economics*, **22**(1), 3–42.

MANDA, D, GERMANO, M, & MWANGI, M. 2002. Human Capital Externalities and Returns to Education in Kenya. *KIPPRA Discussion Paper No. 13*, 26.

MORETTI, ENRICO. 2004a. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, **121**(1-2), 175–212.

MORETTI, ENRICO. 2004b. Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, **94**(3), 656–690.

NARITOMI, JOANA, SOARES, RODRIGO R, & ASSUNÇÃO, JULIANO J. 2012. Institutional Development and Colonial Heritage within Brazil. *The Journal of Economic History*, **72**(2), 393–422.

RAUCH, JAMES. 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, **34**(3), 380–400.

ROBACK, J.A. 1980. *The Value of Local Urban Amenities: Theory and Measurement*.

ROBACK, JENNIFER. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy*, **90**(6), 1257–1278.

ROCHA, ROBERTA, SILVEIRA NETO, RAUL, & GOMES, SÓNIA MARIA. 2011. Maiores Cidades, Maiores Habilidades Produtivas: Ganhos de Aglomeração ou Atração de Habilidosos? Uma Análise para as Cidades Brasileiras. *Revista Economica do Nordeste*, 675–695.

SACHSIDA, ADOLFO, LOUREIRO, PAULO ROBERTO AMORIM, & MENDONÇA, MÁRIO JORGE CARDOSO. 2004. Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia*, **58**(2), 249–265.

SILVA, VITOR HUGO MIRO COUTO, DE FRANÇA, JOÃO MÁRIO SANTOS, & NETO, VALDEMAR RODRIGUES DE PINHO. 2016. Capital Humano E Desigualdade Salarial No Brasil : Uma Análise De Decomposição Para O Período 2001-2012 . *Estudos Econômicos*, **46**(3), 579–608.

STOCK, JAMES H., & YOGO, MOTOHIRO. 2002 (November). *Testing for Weak Instruments in Linear IV Regression*. Working Paper 284. National Bureau of Economic Research.

TAVARES, PRISCILLA ALBUQUERQUE, & MENEZES FILHO, NAÉRCIO AQUINO. 2011. Human Capital and the Recent Decline of Earnings Inequality in Brazil. *Brazilian Review of Econometrics*, **31**(2), 231–257.