Department of Economics- FEA/USP

Regional labor markets in Brazil: the role of skills and agglomeration economies

ANA MARIA BONOMI BARUFI

WORKING PAPER SERIES Nº 2014-18

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2014-18

Regional labor markets in Brazil: the role of skills and agglomeration economies

Ana Maria Bonomi Barufi (ana.barufi@gmail.com)

Research Group: NEREUS

Abstract:

This paper aims to discuss how agglomerations economies are present in the equilibrium outcomes of the Brazilian formal labor market. There has been a wide discussion on how to correctly identify agglomeration economies given all the different types of endogeneity found in the labor market relationships, as well as taking into account all the relevant aspects that may affect the results. We make use of an individual-firm panel database from the Ministry of Labor (RAIS - Annual Report on Social Information) with information for six years (2003, 2004, 2005, 2008, 2009 and 2012). With the panel data setting, it is not only possible to account for individual unobserved characteristics constant in time, but also for sector and area effects. Moreover, by identifying skills according to the occupational position of the individuals in each firm, it is possible to control for the proximity to different skill levels (in the sector and municipality) to account for different levels of production knowledge externalities. Individual fixed effects control the potential endogeneity of the labor quality. In the case of labor quantity endogeneity, even if there is no consensus of how to best control for it, instruments based on long time lags are considered. The results show that there is a positive and significant effect of density over wages (Urban Economics literature), even when controlling for other relevant characteristics. Moreover, a measure of market potential, related to the New Economic Geography literature, does not capture this positive relationship with wages in the same way, changing sign in a specific setting. Finally, considering a quantile regression approach, there is an indication that agglomeration economies reinforce wage inequality, with a higher effect for the upper part of the wage distribution.

Keywords: agglomeration economies, regional labor markets, wage equation

JEL Codes: R23, E24, R30

Regional labor markets in Brazil: the role of skills and agglomeration economies

Ana Maria Bonomi Barufi
ana.barufi@gmail.com

1. Introduction

In the last decades, Brazil has shown an important decrease in personal and regional inequality, measured in different dimensions. In spite of that, its income inequality level is one of the highest in an international comparison (16th highest Gini index among 141 countries[1]), reaching 0,526 in 2012 from 0,603 in 1995. Regional income disparities are also considerably high, summarized by the fact that average personal income was 1.8 times higher in the South-Southeast than in the North-Northeast in 2012.

The Brazilian labor market is very important for this outcome, as it represents the largest share of total income obtained by the population (77.8% in 2012, according to the Brazilian National Household Sample Survey, PNAD, by the Brazilian Institute of Geography and Statistics, IBGE). More specifically, the formal labor market[2] has been having a significant role in the inequality reduction mentioned above, being responsible for around 43% of the whole Gini Index decline from 1995 and 2012. Therefore, understanding the dynamics of wages and the labor market in general is essential in the Brazilian context.

In a spatial perspective, there is a significant concentration of population and economic activity in the country. Cities shorten distances between economic agents and allow a reduction in the cost of the exchange of ideas, information and goods. In 2010, 84.4% of the total population was in urban areas, occupying 1.07% of the Brazilian territory. Regarding the economic concentration, in 2011, cities with more than 500,000 inhabitants hold a share of 41.2% of total GDP, while concentrating 29.3% of the population. Considering the correlation of the logarithm of wages and the logarithm of population density at the municipal level, it reached 0.06 in 2010 (for the whole labor market, and 0.05 for the formal sector).

[1] Considering the Brazilian Gini Index of 2012 and the most recent information for other countries, available at https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html and accessed in 27/06/2014.

[2] Formal jobs can be defined as contracts following CLT (Consolidation of Labor Laws), meaning that they are under restrictions concerning a required minimum wage, a maximum number of weekly working hours, firms can only fire workers under specific conditions and have to pay taxes and benefits, among other aspects. Because these restrictions imply higher hiring costs, a high percentage of workers is still hired under informal conditions in Brazil.

It is important to understand how the urban concentration of economic activity affects productivity. Taking wages as an indirect measure of the latter, this paper tries to assess whether there is a premium for the city size in Brazil, even after controlling for individual characteristics and other relevant information at the area and sector levels. The remaining of the text is organized as follows: section 2 presents a literature review; section 3 resumes the methodological approach; data is examined in section 4; estimation results are discussed in section 5; and section 6 concludes, proposing steps for future work.

2. Literature review

The urban wage premium is usually investigated by studies that try to identify how the density of economic activity affects the productivity of workers (Heuermann et al., 2010). The challenge is to isolate these effects from other explanatory factors of productivity differentials in space, while investigating the possibility of convergence and divergence between regions, sectors, educational levels, among others dimensions ((Lindley and Machin, 2014). Agglomeration economies are defined as external economies of scale, representing productivity gains generated by the concentration of economic agents in space. In this context, Glaeser and Gottlieb (2009) associate them to the reduction of transportation costs of goods (greater proximity between intermediate goods supply and demand), people (labor markets are more efficient in urban areas and service providers have greater access to their clients), and ideas (stronger knowledge exchange between people and firms).

The economic mechanisms that generate agglomeration economies are classified in different dimensions. Considering sectorial composition, there are localization economies (Marshall, 1890), defined by the idea that productivity gains of a firm can be related to the size of its sector of activity in the city, and urbanization economies (Jacobs, 1969), generated by the city scale, with the rational that the diversity of sectors can contribute to the exchange of ideas. Moreover, based on the Marshallian externalities (labor pooling, intermediate inputs and knowledge externalities), Duranton and Puga (2004) and Puga (2010) define three main microfoundations for agglomeration economies: (i) sharing (facilities, gains from individual specialization and variety, and risk – labor pool); (ii) matching (higher quality and quantity of matchings among workers and firms); and (iii) learning (knowledge generation, diffusion, and accumulation).

While there is a large literature investigating these mechanisms, there is also a great effort to measure the extent through which agglomeration economies really foster productivity. The main variation among these studies lays on the productivity measure, which can be directly derived

from the production function. However, there are other indirect strategies, as it can be hard to measure the contribution of capital and other factors in the production function.

Among these, it is possible to consider variables that are indirectly related to labor productivity: job creation, new establishments, rents and wages (Rosenthal and Strange, 2004). The latter is most commonly used in empirical studies that measure agglomeration economies, especially because of its large availability. The main assumption here is that wages equalize the marginal productivity of labor (under perfect competition), or that at least they are higher in places that are more productive. An important drawback from this strategy is associated to the fact that the extent through which wages capture local productivity will depend on elasticities in the labor market.

2.1. Theoretical approaches

Departing from that, two main theoretical frameworks aim to establish the relationship between productivity and the size of the city. The first is associated to the Urban Economics (UE) literature and the relationship of wages and density, while the second is based on the New Economic Geography (NEG) approach, trying to understand the association among wages and market potential. Following Fingleton and Longhi (2013) and Combes, Mayer and Thisse (2008), the UE model is usually associated to Abdel-Rahman and Fujita (1990), and can be described as the following. A production function of a price-taking firm in region *r* and sector *i*, using l_j of labor and an amount k_j of other inputs is given by:

$$y_j = A_j\left(s_j l_j\right)^{\mu} k_j^{1-\mu} \tag{1}$$

In this case, A_j is the technology factor (Hicks-neutral), μ is the share of labor in the productive process and e_j is the efficiency level of workers. Moreover, w_j is the wage level and r_j is the price of other inputs. Then, firm *j* profit in all regions is:

$$\pi_j = \sum_v p_{jv} y_{jv} - w_j\, l_j - r_j\, k_j = p_j y_j - w_j\, l_j - r_j\, k_j \tag{2}$$

In this case, $p_j = \sum_v p_{jv} \frac{y_{jv}}{y_j}$ is the average unitary price, and y_{jv} is the exported amount of firm *j* to region *v*. Following first order conditions for the profit maximization, the following relations are obtained:

$$\frac{\partial \pi_j}{\partial l_j} = \mu p_j A_j e_j^{\mu} \left(\frac{k_j}{l_j}\right)^{1-\mu} - w_j = 0 \quad (3)$$

$$\frac{\partial \pi_j}{\partial k_j} = (1-\mu) p_j A_j e_j^{\mu} \left(\frac{k_j}{l_j}\right)^{-\mu} - r_j = 0 \quad (4)$$

Then, equations (3) and (4) are combined to isolate w_j:

$$w_j = \mu(1-\mu)^{(1-\mu)/\mu} e_j \left(\frac{p_j A_j}{r_j^{1-\mu}}\right)^{1/\mu} \quad (5)$$

In this case, information in the individual level are required (individuals that work for firm *j*), available only recently in some countries. When this is not the case, it is possible to estimate an equation based on a regional aggregation:

$$w_{rs} = \frac{\mu(1-\mu)^{(1-\mu)/\mu}}{n_{rs}} \sum_{j \in (r,s)} e_j \left(\frac{p_j A_j}{r_j^{1-\mu}}\right)^{1/\mu} \quad (6)$$

Here, n_{rs} is the number of firms in region *r* and sector *s*. It is possible to highlight the fact that wages are directly proportional to labor efficiency (e_j), while p_j and r_j capture agglomeration and dispersion forces respectively, and A_j is related to technological externalities (knowledge and learning spillovers, existing technology associated to workers abilities, among others).

The aggregated equation[3] to be estimated is the following:

$$ln(w_{rs}) = \alpha + \beta ln(den_r) + \varepsilon_{rs} \quad (7)$$

In which $ln(den_r) = ln \frac{1}{n_{rs}} \sum_{j \in (r,s)} e_j \left(\frac{p_j A_j}{r_j^{1-\mu}}\right)^{1/\mu}$. Usually, this equation captures agglomeration and dispersion forces, even if the microfoundations underlying the model are not very clear (Combes, Mayer and Thisse, 2008, state that consumer preferences and hypotheses on factor and goods mobility are ignored in this case). Density can affect wages through local technology level (A_j), final good price (p_j), price of other inputs (r_j) or local labor efficiency (e_j).

[3] In the individual level, this equation assumes the form $ln(w_{irs}) = \alpha + \beta ln(den_r) + \varepsilon_{irs}$

In the case of NEG models, short-run equilibrium is depicted by a set of simultaneous equations, among which there is a wage equation (Fujita et al., 1999). For sector *M* ("industry"), the basic wage equation equalizes wage in region *r* to a measure of market potential:

$$w_r^M = P_r^{\frac{1}{\sigma}} \tag{8}$$

Where:

$$P_r = \sum_v Y_v (G_v)^{\sigma-1} (T_{rv})^{1-\sigma} \tag{9}$$

In this context, Y is income, G is the price index of sector, T_{rv} is the transportation cost among regions *r* and *v*, and σ is the elasticity of substitution. Then, Y in region *r* and G in region *v* are obtained by:

$$G_r = \left[\sum_v \lambda_v (w_v^M T_{rv})^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{10}$$

$$Y_r = \theta \lambda_r w_r^M + (1-\theta) \phi_r w_r^C \tag{11}$$

In equation (11), the share of workers of sector *M* in region *r* is represented by λ_r, while ϕ_r is the share of workers of *C* in *r*. Finally, θ is the percentage of total workers in the country that are in sector *M*, and its counterpart $1-\theta$ is the percentage in *C*. Then, the NEG wage equation that can be compared to UE equation discussed above is:

$$ln(w) = \frac{1}{\sigma} ln(P) + \varepsilon \tag{12}$$

The model represented in equation (12) is more complex than the one in equation (7), as *P* has different components, and is endogenous because it depends on *w* (however, density in equation (7) presents endogeneity problems as well). Additionally, transportation costs are hard to measure.

2.2. Empirical literature

There is a large discussion on which is the best model to estimate the relationship among productivity and agglomeration. Evaluating this comparison with data for spatial units, Fingleton (2006) finds that there is indication that both NEG and UE models can explain the wage

distribution. Although, some authors defend that the UE model can deal better with spatial variation in smaller distances (Combes et al., 2005; Brakman et al., 2009; Fingleton, 2011). Fingleton e Longhi (2013) note that when an area effect is included in the regression, the potential market coefficient is no longer significant, while the density effect is still observed for women.

However, most of the studies focus on applying the UE model directly, without discussing further the theoretical approach. One of the first studies that aim to measure the urban wage premium while controlling for individual heterogeneity was conducted by Glaeser and Maré (2001). They consider longitudinal individual data for the United States and find evidence that there is an urban premium even when observed and unobserved individual characteristics are taken into account.

Combes, Duranton and Gobillon (2008) discuss three potential explanations for regional wage differentials in France: (i) spatial heterogeneity in the labor force composition (self-selection of workers and firms); (ii) local resources availability; (iii) agglomeration economies. In order to control for individual heterogeneity, the authors consider the longitudinal structure of the data. They apply a two-stage estimation, in which the first stage is based on a regression of individual wages against observed individual characteristics varying in time, individual fixed effects, area-time dummies, sector dummies, local characteristics of the sector. Then, in the second stage, the area-time dummies estimated in the first stage are regressed against time dummies and variables capturing local interactions among sectors and local characteristics (weighting for sample size in each area). Among these, they include a measure of density. Their results indicate that 40% to 50% of the urban wage premium is explained by the sorting of workers (captured by individual fixed effects).

However, as there is a strong concern that density can be endogenous, the authors consider different instruments, such as long temporal lags of population density (more than 100 years before). Combes et al. (2010) also consider soil formation, and find that both instruments give similar results, without changing significantly the main conclusions of the study. In the same direction, Mion and Naticchioni (2009) show that individual abilities and firm size explain a large part of wage spatial variation in Italy. In the case of Germany, Lehmer and Möller (2010) find that there are wage gains associated to rural-urban migratory movements (and losses with movements in the opposite direction), while urban-urban and rural-rural movements are also related to wage increases (with a greater effect in the former).

According to Groot et al. (2014), there are some important drawbacks of including individual fixed effects in the first stage of the estimation. Firstly, the sorting process becomes a black box, as there is no clear explanation of why and how it is happening. Moreover, considering the

identification of agglomeration economies, it will be mostly due to migrating workers, because those who stay in the same place will identify only the wage variation related to changes in agglomeration economies in that place, which are generally slow. Therefore, migratory movements, usually related to the sorting process, will capture most of this effect. Their solution for this problem is to estimate a pooled model, based on subsequent cross-sections, including as many observed individual characteristics as possible. In the Netherlands, they find results that are very close to other developed countries.

Finally, Matano and Naticchioni (2012) apply a quantile fixed effects regression for Italian data and find that the effects of agglomeration economies are higher for superior quantiles of the wage distribution. Even when considering similar instruments to the ones discussed above, this relation is still observed.

In Brazil, there are some studies that aim to measure the effect of agglomeration economies over productivity. In the area level, Amaral et al. (2010) estimate the NEG wage equation and find that market potential seems to be positively related to wages. Based on a NEG model, Chagas (2004) observes that the average wage responds positively to the size of the sector, local infrastructure and labor qualification in the municipality. In the case of a UE wage equation, Simões and Freitas (2014) apply areal data to find that urbanization economies are more relevant for high technological intensity sectors, while sectors with low and medium technological intensity are more benefited from mid-sized urban centers, relatively less diversified.

Fally et al. (2010) also apply the NEG model to study regional wage inequality in Brazil. They build measures of market access and supply access (intermediary goods). With individual data, the authors find that migratory movements have not promoted regional real wage equalization, even if there are not significant restrictions to worker displacement in the country, meaning that labor market frictions may be the cause for such a result. Moreover, wages seem to be affected positively by market and supply access.

Considering data from RAIS (Annual Report of Social Information, from the Ministry of Labor), Freguglia and Menezes-Filho (2012) control for individual heterogeneity to explain wage differentials in Brazil. They find that when this control is made, almost 63% of the total differential disappears, meaning that local policies should focus on human capital development in order to promote regional development. Silva (2012) also considers this database and finds that after controlling for individual heterogeneity, the urban wage premium reaches 3% from 1995 to 2008.

Finally, Maciente (2012) has developed a matching among the occupational classification in Brazil and the abilities profile of ONET (Occupational Information Network). Based on that, the author finds a weighting system for each ability required on the job, based on the worker qualification and the job complexity.

In the next section, the methodology of analysis will be discussed, alongside with the main challenges to be faced here.

3. Methodology

Following the theoretical discussion presented in Section 2, two main regression models will be estimated here, inspired by the New Economic Geography and the Urban Economics frameworks. Then, the general model will be the following:

$$\begin{aligned}\log(w_{i,a,r,s,t}) = {} & \beta_1 mod_{a,t} + \beta_2 \log(firms\_sector_{a,s,t}) \\ & + \beta_3 share\_emp\_sector_{a,s,t} + \beta_4 \log(diversity)_{a,t} \\ & + X_{i,t}\gamma + V_{s,a,t}\delta + K_{a,t}\theta + \tau_t + \delta_i + R_r + \varepsilon_{i,a,r,s,t}\end{aligned} \quad (13)$$

In which $mod_{a,t}$ can be alternatively labor density or market potential, when the model to be estimated refers to the UE or the NEG approach, respectively. Moreover, while $\log(firms\_sector_{a,s,t})$ (the logarithm of the number of firms in the sector, area and time) and $share\_emp\_sector_{a,s,t}$ (the share of employment in the sector, area, time, considering the total employment of the area in *t*) represent measures of specialization, $\log(diversity)_{a,t}$ (the logarithm of the inverse of a Herfindhal index in the area and time) is a measure of diversity. In this sense, the first ones are related to localization economies and the latter aims to represent urbanization economies. Observed individual characteristics varying in time are captured by $X_{i,t}$ (age, age squared and abilities associated to the occupation); other sector and areal observed characteristics are in $V_{s,a,t}$ and $K_{a,t}$ (average abilities). Finally, time effects (τ_t), individual unobserved effects invariant in time (δ_i) and region effects (R_r) can be controlled for.

Following Combes, Duranton and Gobillon (2008), individual fixed effects are used to capture at least part of the sorting process related to possible endogeneity of the quality of labor. More specifically, cities paying higher wages attract qualified workers, who may increase overall productivity in the city, leading to a further increase in wages. There is also a strong concern of potential endogeneity of the quantity of labor (density and market potential variables). Here only one kind of instrument is proposed, based on long temporal lags of these variables (following

Combes, Duranton and Gobillon, 2008, Combes et al., 2010, and Groot et al., 2014, among others).

In this first effort, all the results will be based on a single-stage estimation. Future studies will consider different approaches. Additionally, a quantile regression version of equation (13) is estimated, in order to assess the effects of agglomeration economies along the wage distribution. Only the results of a pooled cross section for the quantile regression are discussed here.

4. Data sources and descriptive analysis

The wage equation discussed in section 3 will be estimated for the Brazilian labor market considering registration data from RAIS (Annual Report of Social Information, from the Ministry of Labor), informed annually by all firms to the federal government. In this sense, it is important to notice that this database refers only to the formal labor market, meaning that the informal sector, which is very significant in Brazil, is not represented. Another potential problem of this data comes from the fact that larger firms have a more accurate reporting process, what implies that smaller firms and smaller municipalities will have more missing data. On the other side, because RAIS is based on registration data, there is less risk of wage under-reporting.

RAIS database has increased significantly in the last few decades, due to the formalization process that happened in the period (if in 2004, 30.2% of total jobs[4] were based on a formal contract, in 2012 this percentage increased to 39.3%). Therefore, the database size has also expanded. In this first effort, the individual-firm data will be considered for years 2003, 2004, 2005, 2008, 2009 and 2012[5], keeping only individuals that are present in every year (what induces a potential bias in the analysis, to be discussed in future works). A few steps[6] were taken to get a balanced panel

[4] Total jobs include public sector and military, formal and informal employees, self-employment, entrepreneurs, and unpaid work.

[5] Considering current access to the microdata.

[6] The original database contains individual information for each contract, meaning that individuals can appear more than once if they work for more than one firm. A few steps were taken in order to select the desired observations (ending up with only one contract for worker). Firstly, a filter selected only active contracts in December of the year of individuals working for private companies in permanent jobs. Then, contracts with missing individual ids, wages equal to zero or less than 20 weekly hours were excluded. The next steps for individuals with multiple contracts included keeping those with 5 or less contracts, dropping those with different gender in each contract, and keeping the contracts with the highest number of weekly hours and with the oldest hiring date. Finally, for the remaining cases of multiple contracts, only one of them was randomly selected, leading to one contract per individual for the whole database. After all these procedures, the database size ranged from 13.1 to 20.7 million between 2003 and 2012 (related to the formal sector expansion previously mentioned). In this first effort, only individuals observed in every available year were kept, generating a balanced panel with 4.04 million of individuals. A last adjustment was necessary in order to guarantee that all remaining information were trustable involved keeping individuals with the same information for gender and birth date in every year. Finally, following the literature, the

comprising 1,126,908 individuals observed in six years. In order to increase data tractability, a sample based on individuals born in January was selected, resulting in 553,740 observations for 92,990 individuals.

The RAIS microdata mentioned above provides information on the individual level ('selected microdata from RAIS'). For the sector and area levels, the main sources are the RAIS aggregated information provided by the Ministry of Labor[7] (hereby referred as 'full aggregated RAIS') and the full microdata (before the cleaning process described in footnote 4, referred as 'full microdata from RAIS'). Apart from the variables obtained directly from RAIS, the abilities of workers are captured according to their occupations, following Maciente (2012, 2013). The author considers the matching proposed by ONET (Occupational Information Network) between occupations and skills in the United States to build the same comparison for the Brazilian formal labor market, weighting each of the final 21 skills factors according to the job requirements (cognitive skills, physical strength, managerial skills, among others).

Apart from the individual level, there is information for sectors and areas. The former is based on the two-digit division of CNAE 1.0 (National Classification of Economic Activities), while the latter considers minimum comparable areas (MCAs) 2000-2010[8], totalizing 5,479 areas. Finally, in one of the specifications regional cluster effects are considered. These effects are based on urban regions of immediate articulation (482 areas containing all municipalities), proposed by the Brazilian Institute of Geography and Statistics for 2007 (IBGE, 2013), taking into account all daily commuting and transportation connections among municipalities. Table 1 provides a detailed description of all variables considered in section 3, including methods of calculation and data sources. Moreover, table 2 presents the main descriptive statistics of the variables considered in the model.

analysis was conducted for men with 18 to 56 years old in 2003, working in manufacture and service sectors.

[7] www.mte.gov.br

[8] MCAs harmonize municipal borders taking into account the creation of new municipalities (equivalent to counties) in the period, meaning that MCAs 2000-2010 consider all municipalities divided and merged from 2000 to 2010 as being part of a bigger area, that can be aggregated in both periods.

Table 1. Definition of the variables and data sources.

Variable	Definition	Level	Data source
Real hourly wage	Monthly wage received in December, deflated by INPC (National Index of Consumer Prices, by IBGE) and divided by 4 times the numbeer of weely hours in the contract.	Individual	Selected microdata from RAIS
Labor density in the formal sector	Total employment in the MCA divided by total MCA area (in km2).	MCA	Full aggregated RAIS
Population density in 1920	Population in 1920 for MCAs 1920-2000 is redistributed for MCAs 2000-2010 based on the populational share of the latter on the former in 2000, and divided by MCA 2000-2010 area in km2.	MCA	IPEADATA
Population density in 1940	Population in 1940 for MCAs 1940-2000 is redistributed for MCAs 2000-2010 based on the populational share of the latter on the former in 2000, and divided by MCA 2000-2010 area in km2.	MCA	IPEADATA
Market potential	Average density (employment in the formal sector) of the neighbors (weights matrix for the k=8 nearest neighbors).	MCA	Full aggregated RAIS
Market potential in 1920	Average density (population density in 1920) of the neighbors (weights matrix for the k=8 nearest neighbors).	MCA	IPEADATA
Market potential in 1940	Average density (population density in 1940) of the neighbors (weights matrix for the k=8 nearest neighbors).	MCA	IPEADATA
Age	Age at the end of the year	Individual	Selected microdata from
Number of establishments in the sector in the area	Number of establishments	Sector-MCA	Full aggregated RAIS
Share of sector employment in total employment of the area	$Specialization_{sector,MCA,t} = \frac{E_{sector,MCA,t}}{E_{MCA,t}}$	Sector-MCA	Full aggregated RAIS
Diversity	Inverse of Herfindhal index: $Diversity_{mca,t} = \frac{Emp^2_{mca,t}}{\sum_{sector} Emp^2_{mca,sector,t}}$	MCA	Full aggregated RAIS
Individual abilities	Matching between occupation and skills required for the job; for sector-MCA and MCA levels, the variables are averages of individual values.	Individual, sector-MCA, MCA	Full microdata from RAIS, Maciente (2012)

Source: Elaboration of the author.

Table 2. Descriptive statistics of the main variables.

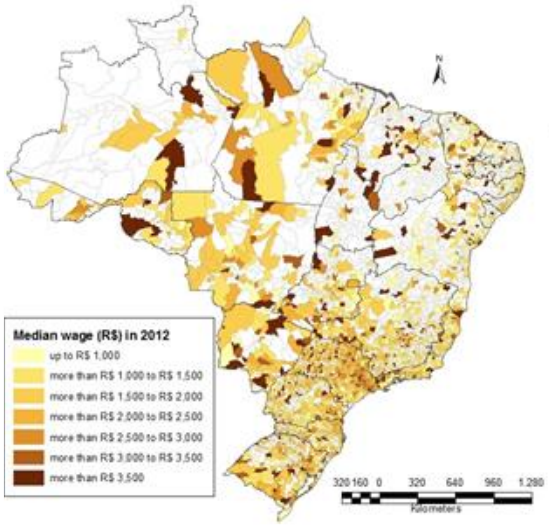
	Mean	Standard deviation	Minimum	Maximum	p50
ln(real hourly wage)	2.41	0.84	-0.36	6.34	2.28
ln(labor density in the formal sector)	4.23	1.77	-5.01	6.61	4.17
ln(population density in 1920)	3.27	1.40	-3.83	4.84	3.55
ln(population density in 1940)	3.71	1.42	-3.75	5.23	3.63
ln(market potential)	3.09	1.29	-2.05	5.15	3.20
ln(market potential in 1920)	2.96	1.11	-2.86	4.31	3.36
ln(market potential in 1940)	3.19	0.98	-2.83	4.49	3.35
Age	37.02	9.39	18.00	64.00	36.00
ln(umber of establishments in the sector in the area)	6.47	2.20	0.00	10.59	6.48
Share of sector employment in total employment of	4.68%	5.80%	0.00%	76.11%	2.66%
ln(diversity)	2.45	0.33	0.20	3.07	2.51
Individual abilities - individual level					
Factor 1 - Cognitive skills	-0.57	0.82	-2.16	2.65	-0.81
Factor 2 - Maintenance and operation skills	0.04	1.00	-1.87	2.41	-0.06
Factor 3 - Assistance skills	-0.17	0.68	-2.29	3.57	-0.23
Factor 4 - Management skills	-0.28	0.92	-1.76	3.53	-0.53
Factor 5 - Design and engineering skills	-0.14	0.87	-2.18	3.85	-0.23
Factor 6 - Transportation skills	0.21	1.30	-2.00	7.75	-0.13
Factor 7 - Artistic skills	-0.34	0.76	-3.20	5.21	-0.41
Factor 8 - Accuracy and automation skills	0.05	0.75	-3.68	3.74	0.01
Factor 9 - Supervised work skills	0.12	0.86	-3.16	3.19	0.16
Factor 10 - Teaching and social science skills	-0.23	0.56	-1.88	6.09	-0.20
Factor 11 - Physical strength	0.20	0.85	-2.32	7.41	0.21
Factor 12 - Telecommunication skills	-0.12	0.90	-2.13	4.38	-0.31
Factor 13 - Independence skills	-0.29	1.21	-3.78	3.05	-0.03
Factor 14 - Natural science skills	-0.16	0.78	-2.62	2.84	-0.24
Factor 15 - Attention skills	-0.17	0.90	-3.23	5.18	-0.19
Factor 16 - On-the-job experience	-0.32	0.88	-2.54	5.39	-0.49
Factor 17 - Conflict management skills	0.01	0.90	-2.47	5.04	-0.04
Factor 18 - Team-work skills	-0.13	0.83	-4.31	3.12	-0.08
Factor 19 - Sales skills	-0.21	0.89	-2.16	4.29	-0.19
Factor 20 - Monitoring and compliance skills	-0.15	0.95	-3.20	2.94	-0.34
Factor 21 - Clerical skills	-0.09	1.08	-3.91	3.23	-0.19

Source: Elaboration of the author.

It is important to visualize the relationship between wages and density in space. Moreover, comparing the sample with the whole database is also necessary in order to understand how relevant is the regression analysis conducted in the next section. With that in mind, figures 1 and 2 present wages distribution along MCAs, the latter with the whole database for manufacture and service sectors, and the former with the selected data for the regression analysis. In the direct comparison of median wages, the selected sample has some specific characteristics, specially the

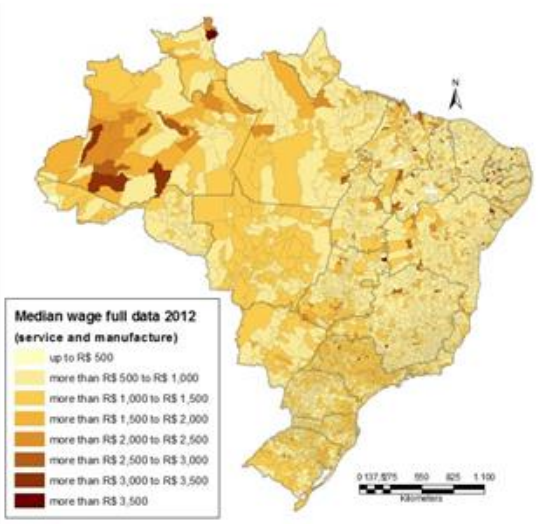
fact that there are only men, aged 18 to 56 in 2003, working at least 20 hours per week. On the other side, in the whole database there are women and older and younger workers, meaning that wages in general will be lower in Figure 2. The second important aspect is that not all MCAs have observations in the selected database (around 90,000 individuals). Even tough, there are similarities among them regarding high median wages in the North and Center-West regions of the country.

Figure 1. Median wages in 2012 for the selected database (born in January, in every year of the studied period, only men, 18 to 56 years old in 2003), services and manufacture.



Source: RAIS microdata.

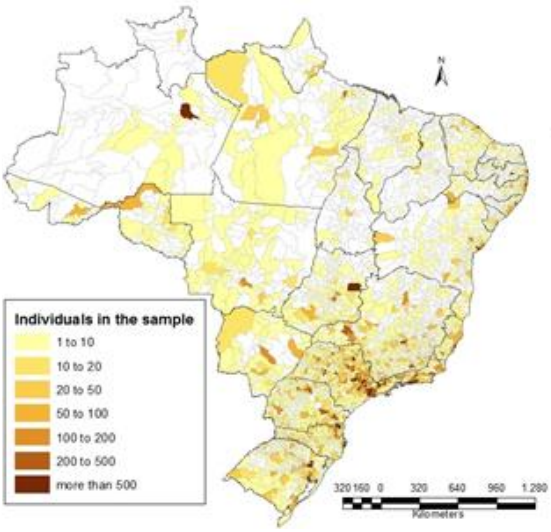
Figure 2. Median wages in 2012 for the whole database, with services and manufacture.



Source: RAIS.

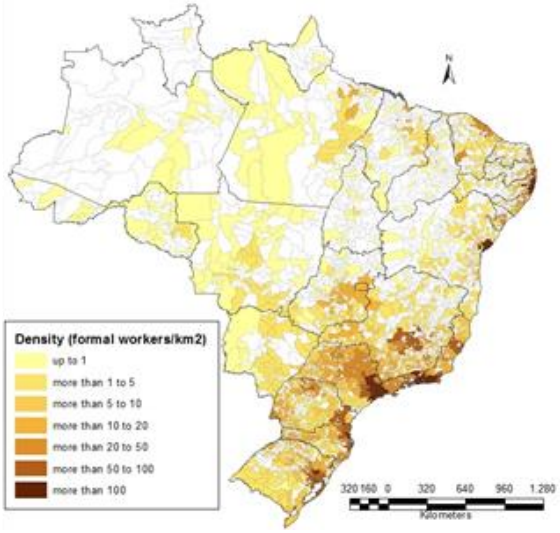
The regional distribution of the individuals analyzed in the sample can be found in figure 3, while figure 4 shows the density of formal workers in 2012. The high values of median wage observed for some MCAs in the North, Northeast and Center-West of the country (figure 1) are usually associated to a small number of individuals in the sample (figure 3), meaning that the formal labor market is not significant in these areas. However, in the regression analysis these outliers are not strongly affecting the relationship between wages and density, because they are associated to a small number of individuals (and the regressions are on the individual level).

Figure 3. Regional distribution of the individuals in the sample.



Source: RAIS microdata.

Figure 4. Density of employment in the formal sector (employees/km^2).



Source: RAIS.

Finally, comparing figures 3 and 4, denser areas also have more observations in the sample, meaning that at least in the regional level, the selected database is representative of the population.

5. Results

As mentioned in section 3, there are two main groups of models being estimated here, aiming to compare Urban Economics (UE) and New Economic Geography (MEG) approaches to explain the relationship of agglomeration economies and productivity. The difference between these two frameworks in the estimated models is the main variable of interest: in the UE case, it is the employment density at the MCA of work; for the NEG framework, it is the employment density of the neighboring MCAs (market potential). The estimation strategies adopted here are based on pooled OLS (POLS – Models 1, 2, 9 and 10) and fixed effects (remaining models), considering alternative models with instrumental variables to deal with the endogeneity of density and market potential. Table 3 presents the main results, considering different controls and estimation structures. As already mentioned in section 4, the analysis is conducted for a sample of individuals working in the formal sector in 2003, 2004, 2005, 2008, 2009 and 2012.

Evaluating the results presented in table 3, Models 1 and 9 refer to the simple correlation between wages and the variables of interest (density and market potential). Both these effects are positive and significant, and their values are higher than the ones for the other specifications proposed in table 3, which consider more controls. The biggest drop in this coefficient happens when individual fixed effects are included in the regression. This result is similar to those found in other

countries, especially France (Combes, Duranton and Gobillon, 2008) and the United States (Glaeser and Maré, 2001).

Adding region fixed effects in Model 13 leads to a sign change in the market potential coefficient, while the density coefficient is still positive and significant in Model 5. Therefore, comparing UE and NEG results, the former seems to be more stable to different specifications, seeming more suitable for this kind of analysis. Regarding urbanization and localization economies, the results presented in table 3 give an indication that both phenomena are present, as diversity (urbanization) and share of sector in the MCA employment are both significant and positive.

Models 6 to 8 and 14 to 16 consider an instrumental variables approach, instrumenting density and market potential variables with their long temporal lags counterparts. Their first stage results lead to a correlation around 0.70 between the endogenous variable and the instrument for all cases. However, the lack of significance of the instrumented variable coefficient in Models 8 and 16, when including region fixed effects, may be an indication of inadequacy of this instrument. Further studies should consider other possibilities of instruments, even more because there is a large discussion in the literature regarding the real exogeneity of long temporal lags for density and market potential.

Table 3. Estimation results for the logarithm of the real hourly wage, considering POLS, fixed effects and instrumental variables for alternative models related to the UE and NEG frameworks.

	Urban Economics								New Economic Geography							
	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6	Mod. 7	Mod. 8	Mod. 9	Mod. 10	Mod. 11	Mod. 12	Mod. 13	Mod. 14	Mod. 15	Mod. 16
ln(labor density in the formal sector)	0.103***	0.084***	0.043***	0.031***	0.117***	0.016***	0.012***	-0.074								
ln(market potential)									0.163***	0.134***	0.073***	0.049***	-0.028*	0.031***	0.029***	0.008
ln(umber of establishments in the sector in the MCA)				0.005***	0.007***	0.012***	0.014***	0.010***				0.015***	0.009***	0.016***	0.016***	0.263
Share of sector employment in total employment of the MCA				0.311***	0.280***	0.294***	0.290***	0.258***				0.247***	0.268***	0.258***	0.259***	-0.003
ln(diversity)				0.040***	0.025**	0.049***	0.052***	-0.008				0.011*	0.007	0.028***	0.030***	0.102
Instrument: ln(pop. density 1920)						x		x								
Instrument: ln(market potential 1920)														x		x
Instrument: ln(pop. density 1940)							x									
Instrument: ln(market potential 1940)															x	
Individual observed characteristics		x	x	x	x	x	x	x		x	x	x	x	x	x	x
Individual fixed effects			x	x	x	x	x	x			x	x	x	x	x	x
Year dummy			x	x	x	x	x	x			x	x	x	x	x	x
MCA and sector characteristics				x	x	x	x	x				x	x	x	x	x
Region dummy					x			x					x			x
N	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740	553,740

Controls: i) individual observed characteristics: age, age^2, skills required for the occupation; ii) MCA-sector characteristics: sector dummies, average skills required for the occupation, in the sector and MCA, average skills in the MCA.
T = 6 (2003, 2004, 2005, 2008, 2009, 2012), individuals = 92,990.
*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$;
Source: Elaboration of the author.

The second set of equations discussed here is based on quantile regressions that allow in a certain way the comparison of coefficients related to different parts of the distribution of real hourly wage. As previously mentioned, this first effort will involve only a quantile regression estimation based on the pooled database. Table 4 presents the results for percentiles 10, 50 (median) and 90 of the wage distribution. Overall, agglomeration effects (density and market potential) are higher for the upper part of the distribution, an indication of their contribution for an increase in wage inequality.

Regarding urbanization economies, they seem to favor the reduction of wage inequality, as the diversity coefficient is smaller for higher percentiles. On the other hand, there is a suggestion that localization economies act in the opposite direction, as the coefficient for the share of the sector in total employment in the MCA increases for higher percentiles.

Table 4. Quantile regressions for the logarithm of the real hourly wage, considering pooled cross-sections.

	Urban Economics			**New Economic Geography**		
	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
	Percentile 10	Percentile 50	Percentile 90	Percentile 10	Percentile 50	Percentile 90
ln(labor density in the formal sector)	0.022***	0.030***	0.043***			
ln(market potential)				0.051***	0.056***	0.058***
ln(umber of establishments in the sector in the MCA)	-0.042***	-0.058***	-0.064***	-0.039***	-0.053***	-0.055***
Share of sector employment in total employment of the MCA	0.462***	0.431***	0.707***	0.444***	0.373***	0.608***
ln(diversity)	0.224***	0.173***	0.111***	0.179***	0.128***	0.071***
Constant	0.455***	1.134***	1.800***	0.399***	1.110***	1.838***
Individual observed characteristics	x	x	x	x	x	x
MCA and sector characteristics	x	x	x	x	x	x
N	553,740	553,740	553,740	553,740	553,740	553,740

Controls: i) individual observed characteristics: age, age^2, skills required for the occupation; ii) MCA-sector characteristics: sector dummies, average skills required for the occupation, in the sector and MCA, average skills in the MCA.
*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$;
Source: Elaboration of the author.

The following can summarize the main results discussed in this section: firstly, density and market potential seem to have a positive effect over wages. Furthermore, when controlling for individual characteristics and other sector-MCA variables, there is an important reduction of density and market potential coefficients (especially when individual fixed effects are included in the regression). Thirdly, even if the instruments used here are not the most appropriate, they contribute to a further decline of density and market potential coefficients. Finally, agglomeration economies seem to contribute to an increase in wage inequality in Brazil.

6. Conclusions

The first part of this paper described the important role of the Brazilian labor market for income dynamics in Brazil, highlighting its relevance for the significant income inequality reduction observed in the last decade. In addition, it discussed the possibility of a positive relationship between the spatial concentration of economic activity and local productivity.

Then, a literature review presented the main concepts behind agglomeration economies (external economies of scale related to city size), discussing Urban Economics and New Economic Geography theoretical approaches to understand the relationship between agglomeration and productivity. Moreover, a brief review of the empirical literature is presented, showing that there is space for further studies based on the wage equation in the Brazilian context.

The results show that density and market potential seem to have a positive effect over wages. When individual characteristics and other sector-MCA variables are controlled for, there is an important reduction of the coefficients of density and market potential (especially when individual fixed effects are included in the regression). Finally, agglomeration economies seem to contribute to an increase in wage inequality in Brazil, meaning that concentrating the economic activity in bigger cities may strengthen regional inequalities (and a deconcentration process can work in the opposite direction).

The contributions of this paper can be summarized as follows. Firstly, an empirical comparison among UE and NEG models in the Brazilian case is proposed, showing a slight advantage for the former. Furthermore, quantile regressions try to assess the inequality impacts of agglomeration economies. Finally, as RAIS database provides sufficiently detailed information on labor contracts, it allows the control for skills in different levels (individual, sector, area).

The future steps of this work include estimating different wage equations changing the period of analysis (2004, 2008, 2012), in order to capture the effect of formalization over the results (as this process occurred during that period). Another important advance involves pursuing a two-stage estimation, similar to the one proposed by Combes, Duranton and Gobillon (2008). Finally, for the quantile regression, a fixed effects approach will also be considered.

References

Abdel-Rahman, H.; Fujita, M. Product variety, marshallian externalities, and city sizes. *Journal of Regional Science*, Vol. 30, No. 2, pp. 165-183, 1990.

Amaral, P.V.; Lemos, M.; Simões, R.; Chein, F. Regional imbalances and market potential in Brazil. *Spatial Economic Analysis*, Vol. 5, Issue 4, pp. 463-482, 2010.

Brakman, S.; Garretsen, H.; van Marrewijk, C. Economic geography within and between European nations: the role of market potential and density across space and time. *Journal of Regional Science*. 49(1), pp. 777-800, 2009.

Chagas, A.L.S. Externalidades da aglomeração: microfundamentação e evidências empíricas. *Dissertação de Mestrado*, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, 2004, em http://www.teses.usp.br/teses/disponiveis/12/12140/tde-07102009-105106/pt-br.php e acessada em 14/05/2014.

Combes, P.P.; Duranton, G.; Gobillon, L. Spatial wage disparities: sorting matters! *Journal of Urban Economics*, 63, pp. 723-742, 2008.

Combes, P.P.; Duranton, G.; Gobillon, L.; Roux, S. Chapter 1. Estimating agglomeration economies with history, geology, and worker effects. In: Glaeser, E.L. *Agglomeration Economies*. The University of Chicago Press, pp. 15-66, 2010.

Combes, P.P.; Duranton, G.; Overman, H.G. Agglomeration and the adjustment of the spatial economy. *Papers in Regional Science*, Vol. 84, Issue 3, pp. 311-349, 2005.

Combes, P.P.; Mayer, T.; Thisse, J.F. *Economic Geography*. Princeton University Press, Princeton, USA, 2008.

De Groot, S.P.T.; De Groot, H.L.F.; Smit, M.J. Regional wage differences in the Netherlands: micro evidence on agglomeration externalities. *Journal of Regional Science*, Vol. 54, Issue 3, pp. 503-523, 2014.

Duranton, G.; Puga, D. Chapter 48 – Micro-foundations of urban agglomeration economies. In: Henderson, J.V.; Thisse, J.F. (Eds.) *Handbook of Regional and Urban Economics*. Volume 4. Elsevier, pp. 2063-2117, 2004.

Fally, T.; Paillacar, R.; Terra, C. Economic geography and wages in Brazil: evidence from micro-data. *Journal of Development Economics*, 91, pp. 155-168, 2010.

Freguglia, R.S.; Menezes-Filho, N.A. Inter-regional wage differentials with wage heterogeneity: evidence from Brazil. *The Annals of Regional Science*, Vol. 49, Issue 1, pp. 17-34, 2012.

Fingleton, B. The empirical performance of the NEG with reference to small areas. *Journal of Economic Geography*, Vol. 11, pp. 267-279, 2011.

Fingleton, B.; Longhi, S. The effects of agglomeration economies on wages: evidence from the micro-level. *Journal of Regional Science*, Vol. 53, No. 3, pp. 443-463, 2013.

Fujita, M.; Krugman, P.; Venables, A.J. *The Spatial Economy: cities, regions and international trade*. The MIT Press, Cambridge, MA, United States, 1999.

Glaeser, E.L.; Gottlieb, J.D. The wealth of cities: agglomeration economies and spatial equilibrium in the United States. *NBER Working Paper* 14806, 2009.

Glaeser, E.L.; Maré, D.C. Cities and skills. *Journal of Labour Economics*, 19 (2), pp. 316-342, 2001.

Heuermann, D.; Halfdanarson, B.; Suedekum, J. Human capital externalities and the urbang wage premium: two literatures and their interrelations. *Urban Studies*, Vol. 47, No. 4,pp. 749-767, 2010.

IBGE – Instituto Brasileiro de Geografia e Estatística. *Divisão Urbano Regional*. 2013.

Jacobs, J. *The Economy of Cities*. New York, Vintage, 1969.

Lehmer, F.; Möller, J. Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for Germany. *The Annals of Regional Science*, Vol. 45, pp. 31-53, 2010.

Lindley, J.; Machin, S. Spatial changes in labour market inequality. *Journal of Urban Economics*, 79, pp. 121-138, 2014.

Maciente, A.N. A mensuração das competências cognitivas e técnicas das ocupações brasileiras. *IPEA – Radar Tecnologia, Produção e Comércio Exterior*, Nº 23, pp. 17-25, 2012.

_____. The determinants of agglomeration in Brazil: input-output, labor and knowledge externalities. *PhD Dissertation*, University of Illinois at Urbana-Champaign, 2013.

Marshall, A. *The Principles of Economics*. London, Macmillan, 1890.

Matano, A.; Naticchioni, P. Wage distribution and the spatial sorting of workers. *Journal of Economic Geography*, Vol. 12, pp. 379-408. 2012.

Mion, G.; Naticchioni, P. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, Vol. 42, No. 1, pp. 28-55, 2009.

Puga, D. The magnitude and causes of agglomeration economies. *Journal of Regional Science*, Vol. 50, N. 1, pp. 203-219, 2010.

Rosenthal, S.S.; Strange, W.C. Chapter 49 – Evidence on the nature and sources of agglomeration economies. In Henderson, J.V.; Thisse, J.F. (Eds.) *Handbook of Regional and Urban Economics. Volume 4*. Elsevier, pp. 2119-2171, 2004.

Silva, D.L.G. Distribuição especial dos efeitos dos ganhos de aglomeração sobre os retornos à educação no Brasil. *Dissertação de Mestrado*, Universidade Federal da Bahia, 2012.

Simões, R.F.; Freitas, E.E. Urban attributes and regional differences in productivity: evidence from the external economics of Brazilian micro-regions from 2000 – 2010. *International Journal of Economics*, Vol. 01, No. 2, pp. 30-44, 2014.