Wage Differentials in Brazil: How far agglomeration explains?

ABSTRACT

The main purpose of this study was to identify the contribution of the worker, firm and region's effects, focusing on the density of employment as a measure of agglomeration, on the wage differential in the Brazilian regions in the period from 2010 to 2014. The analysis was made in panel, from RAIS data, which allows the monitoring of the individual over time. Initially, mincerian equations was applied to the OLS models, with analysis focused on the fixed effects results. Afterwards, the regression method was applied by instrumental variables in order to eliminate the possible endogeneity of the employment density. The results indicated that there is a wage differential, although small attributed to the density of employment, accordingly to the literature. It has been found that education, the sector which the worker is inserted, and the size of the firm have a strong influence on wage differentials, especially to, higher level education individuals, workers form the industry sector and those workers of big firms. Finally, it is conclude that the density of employment, characteristics of individuals, firms and region contribute to the existence of wage differentials.

KEYWORDS: Effects Fixed; Sorting Skills; wage Differentials

1. INTRODUCTION

The wage differential is an issue studied with some frequency in the economic literature and can be verified from analysis about characteristics of individual, of space, of the firm (FALCÃO AND NETO, 2007; Rocha et al, 2011 and SILVA, 2013). Most of these studies are based on the analysis of Mincer (1974), which proposes a wage equation from the perspective of human capital - including work experience and education - as an explanatory observable variable for the wages formation. In this same vein, Langoni (1973) shows sub-sector wage differentials in Brazilian industry, despite identifying relevant participation of individual characteristics in the configuration of the wages.

Moreover, empirical evidences (AZZONI, 1997; SERVO and AZZONI, 2002; FREGUGLIA, 2007) point out that industry characteristics are an important source of income differentials, since attributes such as the sector of activity, complexity of occupations or technological distribution keep important relationship with the productivity.

The fact is that wage differentials are persistent, even when controlling the observable and non-observable characteristics of the individuals and the firms, indicating the influence of a regional component in the determination of wages (SAVEDOFF, 1991; FREGUGLIA, 2007; ROCHA, 2011).

In this regard, Azzoni and Servo (2002) show that worker's or occupation's characteristics explain part of the regional income differences, but draw attention to the positive effect of the region size on wages. This same conclusion is indicated by Freguglia and Menezes-Filho (2011), according to them even though 63% of the wage differential of the Brazilian states between 1995 and 2002 is explained by observable and unobservable workers characteristics, there is still a wage differential. This suggests that location specific effects influence the wages.

Based on this, this paper aims to identify the contribution of the worker, firm and region effects, focusing on employment density as a measure of agglomeration, on the wage differential in the Brazilian regions, between 2010 and 2014. This analysis is important, first because there are few studies that analyze simultaneously how these effects contribute to the wage differentials (FREGUGLIA, 2007; ROCHA, 2011; SILVA, 2017). Second, in a country of continental dimensions like Brazil, workers and firms heterogeneity can have an enhanced effect on the productivity of workers as they

interact with agglomeration variables. Thus, the work contributes to the literature by considering interactions between regional characteristics, individual and employment.

The specific objectives of this study are: i) to estimate the relationship between agglomeration economies, from the density of employment, and wage differentials considering the relationship with the sorting; ii) estimate the contribution of observable and unobservable characteristics of the individuals, as well as regional and firms attributes, in the composition of the wage differentials; iii) describe socioeconomically of the Brazilian formal sector workers.

The model estimated in this study adopts the approach of income Mincer equations, using a model with panel data (MENEZES FILHO, MENDES AND ALMEIDA, 2004; FREGUGLIA, 2007; SILVA, 2007), considering workers in the Brazilian formal labor market, between 2010 and 2014. These data are from the Relatório Anual de Informações Socias (RAIS), of the Brazil's Labor and Employment Ministry (MTE).

Besides this introduction, this paper is organized as follows. Section 2 discusses and presents the theoretical model of sorting skills as a source of productive clusters in which this study is based on. Section 3 provides a literature review of wage differentials, which presents a brief discussion of the relationship between heterogeneity of workers, firms and the region and regional productivity and also brings some empirical evidences of agglomeration economies in Brazil. Section 4 presents the methodology, the data and the variables used in this research. Section 5 brings the main results. Section 6 presents the final considerations of this work.

2. LITERATURE REVIEW

This section presents a literature review about the worker, firm and region effects in wage determination as well as the economies of agglomeration, which cause effects that contribute to the wage gap.

2.1 Heterogeneity of workers, firms and the region as a source of regional productivity

More recent studies in Labor Economics consider the presence of unobserved skills of workers in wage differentiation (ARBACHE, 1999; FREGUGLIA, 2007 AND SILVA, 2017). According to Arbache (2001), the central idea in relation to the heterogeneity observed of worker, consists in the fact that workers with the same

observable characteristics receive unequal wages. This wage gap for these workers is given by the relevance of non-observable characteristics of workers.

Freguglia (2007) notes that the control of individual heterogeneity observed in the wage equation can reduce or even eliminate much of the apparent wage advantage associated with the denses urban areas. In addition, the author considers that including in the analysis the unobserved characteristics of the worker; it is possible to correct the selection bias linked to attracting more skilled workers to the big urban centers.

To Silva (2017) the observed characteristics of formal workers and the location effects in metropolitan areas were only able to explain about 40% of the real wages variation of formal Brazilians workers. While, by including the firm and workers' unobserved heterogeneity, the explanatory power of the real wage gap rose to approximately 91%, on the results. To the author, these results can be explained by the concentration of the workers and firm's unobserved heterogeneities.

Besides, Silva (2017) states that a significant part of the positive effects of location on real wages in local labor markets would be explained by sorting skills associated to the firms and workers' unobserved heterogeneities. Another observation made by the author is that the cities within the population arrangements have higher wages than those who aren't included in.

Therefore, to achieve robust results in surveys of wage differentials is essential that unobservable characteristics of workers, firms and region are considered.

2.2 Empirical evidences of agglomeration economies in Brazil

The literature shows that both the observable and the unobservable characteristics of worker influences to existence of a wage differential. Savedoff (1991) seeks to identify the existence of wage differences between the metropolitan areas of Brazil between 1976 and 1986 and found that there is indeed a wage difference between the metropolitan areas. This dispersion can be higher or lower depending on the year and the occupational category of workers.

Seeking to analyze the factors for the wage differential in the Brazilian states, Azzoni (1997) conducted a study, which had noted that observable personal characteristics of workers explain 60% of the income gap between the states. Same controlling for regional characteristics, including cost of living differentials, this effect is still positive, although with significantly reduced magnitude. It appeared then that the

productivity differences are potentially explained by both regional attributes and worker characteristics.

Arbache (2001) draws attention to the role of unobservable characteristics of workers in wage determination. He suggests that the hypothesis of unobserved heterogeneity, which takes into account characteristics that cannot be measured on the worker, such as personal rapport, the reasoning, the ability, the degree of adaptation to change, creativity, motivation, and other, can give a more robust explanation to the wage gap. Thus, the author states that it is important to take into consideration the explanations of wage gap, both personal characteristics of the individual-measurable and non-measurable.

In Brazil, wage differentials have been the subject of attention because of high income inequality that characterizes the country. The explanation for this question becomes relevant for the fact that, even when controlling for a number of observable characteristics such as education, age, region residence, etc. occupation, these differentials persist. One possibility is that these differences reflect unobservable productive characteristics (FREGUGLIA, E SOUSA MENEZES-FILHO, 2007).

To Silva (2017) the education and the unobservable characteristics of workers are largely responsible for explaining the existence of a wage differential in the Brazilian regions. But, even controlling these two variables, remains a small wage differential, which can be linked to specific effects of location on wages. In this context, there seems to be a combination of worker and location effects both in existence and in the composition of the wage differentials.

However, characteristics of the companies can also serve as an explanation for the existence of a wage gap. The sector to which the firm belongs, its size, its observable and unobservable characteristics, the type of capital (foreign or domestic) and level technology used in them can affect the firm in its determination of wages and, consequently, the wage difference paid between one firm and another. Thus, it is important to understand how the interaction between the firm's characteristics and worker characteristics happens and how this interaction influences the wage differential.

In fact, the existence of productivity differentials is associated with competition between companies of various sectors of the industry with regard to the search for workers (NETO, FREGUGLIA AND FARJADO, 2012). Thus, according to the authors, the differential would be transitory until the firms could fit the market balance.

Arbache and De Negri (2004) examined the effects that industrial affiliation, employment bond in firms in the industrial sector, have on Brazilian wages. They controlled their analysis the following variables: plant size, nationality of capital, access to international markets and variables on workers. The results show that the pay gap can be explained, in part, by the worker industrial affiliation. Besides, the authors showed that other factors such as value added, profit margin and technology at the industry level also affect the wage differential between workers.

Freguglia, Menezes-Filho and Souza (2007) point out that the firm's effects explain very little the observed differential. For them it is important the role of unobserved heterogeneity of individuals in explaining wage differentials, because by including them in the estimates, the model adjustment rises from 21% to 81%, indicating that the characteristics observed and unobserved individuals explain most pay differentials. The rest of the explanations of the wage gap are found in the differences between regions and industries.

Silva (2015) found results that by adding the unobserved heterogeneity of firms and workers to other observable characteristics analyzed; the explanatory power in the variations in Brazil's real wages goes from 40.6% to 91% in the period from 1995 to 2008. By including only the unobservable characteristics of individuals, this explanatory power is 89%, when the unobserved effects of firm are added, in isolation; they provided explanation for 74% of the wage variation (SILVA, 2015). Thus, the fixed effects of individuals are more important in explaining the Brazilian wage differences than the firm's fixed effects.

However, Pessoa (1997) argues that the inherent characteristics of the region by interacting with firms and workers contribute to the explanation of the wage differentials. The author suggests that to evaluate if is the region that determines this differential is necessary to conduct studies what seek to analyze the differential, controlled by the characteristics of the workers. If the differential persists between individuals of the same features, the differential will be correlated with the region.

In this context, Azzoni and Servo (2002) found evidence that larger markets have a positive effect on wages, since even after controlling for worker characteristics and work and differences in cost of living, the wage gap persists in favor of Brazil's largest metropolitan areas. Similarly, Menezes and Azzoni (2006) show positive wage differentials in favor of workers in regions South / Southeast compared to Northern counterparts / Northeast.

Falcão and Neto (2007) suggest that the existing wage gap between regions can be explained by human capital externalities. According to the authors, these externalities can generate a different production structure for each region and affect the wages of each locality. The results found by the authors show that an increase in the share of workers with 15 years or more of study in the city tends to increase by 5.8% the wages of workers with incomplete 1st degree and 5.4% workers with up to the 2nd incomplete degree. As for workers between the 2nd degree complete and incomplete higher education, the increase would be 6.4% and for those with a college education or higher levels, the increase would be 7.4%. Therefore, the most skilled workers are the most benefited with a higher level of human capital local added.

This is consistent with the idea that the existence of human capital externalities located implies a different wage premium for each employee (HALFDANARSON, HEUERMANN and SUDEKUM, 2008). In other words, individuals have better pay in intensive areas in human capital and if the market is dynamic enough to promote better interactions between workers, there is better utilization of individual characteristics, resulting in a positive wage differential. Thus, the existence of a pay differential is explained in part by the characteristics of the region.

3. METHODOLOGY

This chapter presents the methodology adopted in this work. Section 3.1 will present the database and variables. Section 3.2 will present the econometric model; it is divided in sub-section 3.2.1, which addresses the endogeneity and instrumental variable.

3.1 Data and variables

The main data used in this study are from RAIS-MTE database, created in 1975, gathering information; compulsorily and annually form, of all establishments in the Brazilian formal labor market, covering more than 97% of the formal sector the economy. Its longitudinal format allows building panel data for workers and to monitor the geographical, occupational and sectoral trajectory of formal worker over time, and monitor selected features of the employer establishment.

The base has generated 3,946,910 observations, equivalent to 789 382 annual observations and follows the trajectory of formal workers of the Brazilian labor market between the years 2010 and 2014. In order to use a more homogeneous sample, in order

to obtain more comparisons relevant, the database considers only males of the private sector.

The dependent variable, natural logarithm the worker's wages, was built building on the nominal remuneration of the worker received in December of each year. The Índice Nacional de Preços ao Consumidor- IPCA (IBGE)[1], that measures the inflation in the metropolitan areas, it was used to deflate nominal wages, taking December 2014 as the base period.

The agglomeration variable - density employment - was built from the ratio population employed on site by the area in square kilometers of the locality in question. Following criteria IBGE (2015) were delimited were 373 (small, medium and big) population arrangements.

With regard to the observable characteristics of the worker, the explanatory variables are: education, age, age squared, gender, race, length of employment, job time squared and occupational group. In relation to this set of characteristics, the comparison category for binary variables are: i) education: illiterate workers; ii) occupational group: agricultural workers, forestry, hunting and fishing and iii) race: white.

Age and age squared are continuous variables and consider workers between 18 and 65 years. The use of time proxy for experience, and the squared employment are given time in months in the same bond.

The education variable is binary character and is classified into four levels of education: illiterate; Elementary (primary 1 incomplete; elementary school one full; elementary school incomplete 2; 2 full elementary school); High school (incomplete high school, finished high school) and Higher education (incomplete higher education and complete higher education).

With regard to occupational groups they are listed here from 1 to 9. They are: (1) senior members of the government, leaders, managers; (2) Professionals from the arts and sciences; (3) mid-level technicians; (4) Workers of administrative services; (5)Workers in services, trade sellers; (6)agricultural workers, forestry and fishing; (7)Workers of production of industrial goods and services (1); (8) Workers of production goods and industrial services (2) and (9) Workers repair and maintenance services.

[1] The IBGE is the Brazilian Institute of Geography and Statistics is the main provider of data and statistical information of the country.

The sectors of activity have binary character and are classified according to the IBGE classification: public administration (excluded), agriculture, livestock and forestry production (reference category); fishing and aquaculture; mining and quarrying; transformation industry; electricity, gas and water; construction; trade; vehicle repair; transport and mail; Housing and feeding; information and communication; financial activities; real estate activities; Professional activities; scientific and technical; administrative activities and complementary services; education; health and social services; arts, culture and sports; other collective services; domestic services and international organizations.

The firm size refers to the number of employees it has compared to the average employment in the sector, which it is inserted and is represented by dummies of size with the following criteria: Microenterprise (reference category) (1, if less than 20 employees, 0 otherwise); Small enterprise (1, 20-99 employees); medium-sized enterprise (1, 100-499 employees) and big companies (1, more than 500 employees).

With regard to the observable characteristics of the region we have: Brazilian macro-region that population arrangement is in (reference category, the North) and density of employment, as described in this subsection. In addition, this study used the population of the years 1910, 1920, 1930 and 1940 (IBGE), as instruments for the density of employment.

3.2 Econometric model

The methodology used in this study aims to identify the contribution of the worker, firm and region effects, focusing on employment density as a measure of agglomeration, on the wage differential in the Brazilian regions in the period 2010 to 2014. The model is estimated from Mincer income equations (Mincer 1974).

The model specification follows the functional form:

$$ln\, w_{ijt} = \beta X_{ijt} + \theta D_{ijt} + \varphi F_{jt} + \delta \gamma_{it} + \varepsilon_{ijt} \qquad (10)$$

Where $ln\, w_{ijat}$ is the log of the hourly wage of the individual *i* in the region *j* in time *t*, X_{ijt} is the vector of worker individual characteristics, which are included the variables of age, education, race, and experience; F_{jt} is the vector of firm characteristics - size and sector; γ_{it} is the observable characteristics of the region where the individual works. The variable of interest in this work isD_{ijt}, that represents the natural *ln* of employment density, with which it is possible to measure the influence of

agglomeration on regional wage differential (ABEL AND DEITZ, 2012; SILVA, 2017). The coefficientsβ,φ, δ and θ represent the parameters to be estimated, and ε_{ijat} is the stochastic error.

To avoid the correlation problem between ε_i and the explanatory variables of the model, it is necessary that the assumptions in the equation below are maintained:

$$\mathrm{E}(\varepsilon_i|X_i) = 0;\ \mathrm{E}(\varepsilon_i|F_i) = 0;\ \mathrm{E}(\varepsilon_i|\gamma_i) = 0\ e\ \ E(\varepsilon_i|D_i) = 0 \qquad (11)$$

Without these assumptions, there is no way to keep the causal relationship. And this can cause endogeneity problems in the model, generating estimates of parameters biased and inconsistent. (Wooldridge, 2011).

Freguglia (2007) argues that the existence of individual's unobserved heterogeneities can cause endogeneity, making inconsistent estimators, through the omission of variables or selection bias. If the individual is endowed with non-measurable attributes and there is interference of these on productivity, wage differential in favor of large urban centers probably reflects this by requiring appropriate mechanisms for correction of estimates (SILVA, 2013).

To control this problem, the previous model is estimated with fixed effects. As a result, individual unobservable attributes that are fixed in time are captured separately, avoiding the correlation between the error term and the explanatory variables. Thus, the endogeneity attributed to the invariant features in time is controlled, allowing more credible estimation of agglomeration effect on individual wages (SILVA, 2013).

Thus, the model specification follows the functional form below:

$$ln\ w_{ijt} = \ \beta X_{ijt} + \theta D_{ijt} + \varphi F_{jt} + \ \delta\gamma_{it} + T_t + \ \emptyset_i + \ \varepsilon_{ijt} \qquad (12)$$

With $\emptyset_i$ being the fixed effect of individual. The identification hypothesis of this model requires that$\mathrm{E}(\varepsilon_i|\emptyset_i) = 0$. This expression denotes that the correlation between X_i and ε_i is captured by fixed explanatory variables ($\emptyset_i$), that is, invariant over time. With the inclusion of this effect the model becomes more consistent and non-biased.

3.2.1 Endogeneity and Instrumental Variable (IV)

The inclusion of fixed effects in the estimation does not guarantee total correction of endogeneity, especially if there is evidence of simultaneity in the model. According to Wooldridge (2011), concurrency occurs when one of the explanatory variables is determined by the model dependent variable.

This work uses the employment density variable to identify the contribution of agglomeration effects in the local wage determination. This variable is endogenous to

the model of this research, because according to Ciccone and Hall (1996); Combes, Duranton and Gobillon (2008) and Silva (2017) contemporaneous shocks in a region can increase local wages, generating reverse causality of employment density with wages from this location. That is, the density of employment can influence local wages, and it can also be influenced by them. If this occurs, the hypothesis of exogeneity($Cov\ (D_{at}, \varepsilon_{at}) = 0$) do not hold and, therefore, the estimator becomes inefficient.

In situations like this, Wooldridge (2011) proposes the use of (IV). For the case in question, must find an instrument that determines the current density, but not wages, except for the effect of density on wages. This method provides consistent estimates of the parameters of interest through the use of an additional variable (C_{at}), that serves as an instrument and that is not included in the equation. The adequacy of the instrument depends on the following conditions:

$$Cov\left(c_{jt}, \varepsilon_{at}\right) = 0 \quad (14)$$

$$Cov\left(c_{jt}, D_{jt}\right) \neq 0 \quad (15)$$

The condition 1 requires that the instrument used is not correlated with the stochastic error. The second condition requires correlation between the instrument and the endogenous variable. That is, the instrumental variable, c_{jt}, cannot be correlated with the error of equation 1, but must be correlated with the employment density,D_{jt}.

In urban economics studies (CICCONE AND HALL, 1996; COMBES ET AL, 2007 AND MELO AND GRAHAM, 2009), the variable logarithm of lagged population proved to be a valuable tool, since the current density should be correlated with the pattern of population concentration from the past, but current productivity levels do not keep relations with the past distribution of the population. Thus, the population time lag reduces the endogeneity in the density of employment, generating consistent estimates of the effects of urban agglomeration on local wages.

Following therefore the literature consensus the instrument of this work to the density of employment is the natural logarithm of the population of the regions analyzed (C_{jt}). Once defined the instrumental variable proceeded to the estimation by the method of least squares in two stages (2SLS).

According Gujarati and Porter (2011) to mitigate the likely correlation between the endogenous explanatory variable and the stochastic error applies the 2SLS,

regressing in the first stage, the endogenous explanatory variable on all the predetermined variables throughout the system, including the instrumental variable:

$$D_{ijt} = \alpha + \beta X_{ijt} + \varphi F_{ijt} + \delta\gamma_{it} + \sigma C_{jt} + T_t + \emptyset_i + \varepsilon_{ijat} \quad (16)$$

Where $\varepsilon_{ijat}$are the usual waste OLS. Thus we have:

$$\widehat{D}_{ijt} = \alpha^{+\beta^{X}{}_{it}} + \varphi^{F}{}_{jt} + \delta\gamma_{it} + \sigma C_{jt} + T_t + \emptyset_i + \varepsilon_{ijt} \quad (17)$$

Where $\widehat{D}_{ijt}$ is an estimate of the average value of D_{ijt} conditioned to the independent model variables. According Gujarati and Porter (2011), using the IV and regressing the model by OLS method, in the first stage, the correlation between the variable $\widehat{D}_{ijt}$ and the stochastic error ceases to exist.

After estimates the model using in place of the endogenous variable, the parameter ($\widehat{D}_{ijt}$) found in the first stage:

$$ln\, w_{ijat} = \alpha + \theta\widehat{D}_{ijt} + \beta X_{ijt} + \varphi F_{ijt} + \delta\gamma_{jt} + T_t + \emptyset_i + \varepsilon_{ijat} \quad (18)$$

Once done this procedure, are obtained consistent estimates of the effect of the employment density on the wage differentials (GUJARATI AND PORTER, 2011).

Silva (2017) presents two main reasons for the use of model estimation in two stages. The first reason is that the model estimation in a single stage does not allow calculating the variance of local shocks, generating bias in the standard errors of estimates of aggregate variables. The second reason is that it is not possible to distinguish local shocks of idiosyncratic shocks in the worker level.

4. RESULTS

This section is organized as follows. First up is a brief profile of the Brazilian worker in the analysis period. After, models are estimated on panel data to identify the effect of agglomeration, as well as feature vectors of firms and workers on wage differentials. Finally, it is estimated a model with instrumental variables in order to deal with the endogeneity of employment density.

4.1 Profile of the Brazilian worker in the period from 2010 to 2014: basic statistics

The selected sample has 3,946,910 observations equally divided for each year of observation ranging from 2010 to 2014. Table 1 and Table 2 bring the main features related to the employee.

Table 1 Workers' characteristics based on the RAIS (2010-2014) (to be continued)

Variable	Observations	Average	Standard Deviation	Min.	Max.
LNW real-time	3946910	7.53	0.780	5,030	11.59
Age	3946910	37.87	10.30	18	65

Table 1 Workers' characteristics based on the RAIS (2010-2014) (conclusion)

age2	3946910	1.54	824.8	324	4225
Exp.	3946910	70.19	75.11	0	597.9
Exp.2	3946910	10.57	23.53	0	357 484

Source: Elaborated by the own author based on RAIS data (2010-2014).

The variable of interest in this study, the wage natural logarithm, had an average of 7,534, with a wage dispersion of 0.78, indicating reasonable distortion of wages, although they are considered only males. In part, this data may reflect the existence of segmentation in the labor market, so that variables such as race may be behind this result. In addition, the formal stock of human capital (education and experience) and attributes inherent in the occupation and region may be important explanatory factors of the trajectory of wages in Brazil. As the use of time proxy for experience in this study, the observed mean is 70.19 in the period, with a deviation of more than 75. This signals to workers with different levels of experience in the labor market,

Regarding the age of the individuals observed, the average is 37.87 years, with a deviation of 10.3 years. This points to two important conditions: the average Brazilian worker is not so young, and this has effects on individual productivity and therefore wages; wage distortions partially reflect the diversity in the age composition of the labor force in Brazil. Regarding race / color, the number of white workers represents 60% of the sample, followed by brown (28.7%) and black, on a much smaller percentage, 5.5%. This suggests that there is a predominance of white men in the Brazilian labor market, which can lead to wage differences. This, however, must be verified in the estimations in the following results.

Table 2 Workers' characteristics based on the RAIS (2010-2014) (to be continued)

Variable	**Observations**	**Frequency**
Education		
Illiterate	19,769	0.5%
Elementary	1265995	32.1%
High school	2006316	50.8%
Higher education	654 830	16.6%
Total	3946910	100.0%
Race / Color		
Indigenous	8505	0.22%
White	2368140	60%
Black	217 270	5.5%
Yellow	30,282	0.77%
Brown	1132209	28.7%
Not identified	190 504	4.83%
Total	3946910	100.0%

Group occupancy	**Observations**	**Frequency**
Military, police ...	161	0.00%
Members of the government ...	187 645	4.76%
Professionals from the arts and sciences	227 081	5.75%
Mid-level technicians	383 864	9.73%
Workers of administration	539 681	13.68%

Table 2 Workers' characteristics based on the RAIS (2010-2014) (conclusion)

Service workers ...	813 894	20.62%
Agricultural workers ...	191 153	4.84%
Production workers (1) ...	1155829	29.29%
Production workers (2) ...	250 408	6.35%
Maintenance workers ...	196 735	4.99%
Total	3946910	100.0%

Source: Elaborated by the own author based on RAIS data (2010-2014).

The variable of interest in most urban economy studies and work is the education level of workers. Authors like Mincer (1974) argue that it is the main approach to the formal human capital of individuals and should have a major effect on wages. Observing the data in Table 2, it appears that the majority of Brazilian workers have high school (50.8%), but a considerable portion has higher education (16.6%). If the most skilled workforce is heterogeneously distributed in Brazil, this may signal to regional imbalances in the distribution of opportunities (inefficient matching). Thus, regional attribute can act as a selection mechanism of most skilled individuals (sorting) and, with it, regional differences in wages would reflect partial contribution of characteristics of the region itself.

As for occupation, occupational groups 7 (service workers, trade vendors in shops and market) and 5 (workers of production of industrial goods and services (1)) represent 29.29% and 20.62% of the total sample workers, respectively.

Table 3 Description of the firm's characteristics based on the RAIS (2010-2014) (to be continued)

Activity Sector	**Observations**	**Frequency (%)**
Agriculture, livestock and crop production	182 726	4.64%
Fisheries and aquaculture	1762	0.04%
Mining and quarrying	57,391	1.46%
Transformation industry	1129063	28.67%
Electricity, gas & water	91,829	2.33%
Construction	244 295	6.20%
Trade; vehicle repair	794 611	20.17%
Transport and Post	393 278	9.98%
Accommodation and feeding	94,288	2.39%
Information and communication	103 067	2.62%
Financial activities	127 029	3.23%
Real Estate activities	9550	0.24%
Professional activities	89,963	2.28%
Administrative activities and serv.	412 150	10.46%
Education	56,286	1.43%
Health and social services	60,935	1.55%
Arts, culture and sports	18,245	0.46%
Other community	71,264	1.81%
Domestic services	278	0.01%
International organizations	732	0.02%
Total	3938742	100%
Firm size		

Table 3 Description of the firm's characteristics based on the RAIS (2010-2014) (conclusion)

Micro	678 755	17.2%
Small	926 252	23.5%
Medium-sized enterprises	924 306	23.4%
Big	1417577	35.9%
Total	3938742	100%

Source: Elaborated by the own author based on RAIS data(2010-2014).

Table 3 shows the industry sectors in which individuals of the sample are allocated. According to these data, the sectors with the highest number of persons employed in the study period were the sectors of processing and trade industry; repair of vehicles, respectively, 28.67% and 20.17% of the total number of observations. It is important to note that these two sectors are linked to the industry.

With regard to firm size, most workers are distributed in large firms, 35.9%. The medium and small firms employ almost the same number of workers, the difference is 0.1%. Micro-firms employ 17.2% of subjects, showing that the general employees are well distributed, according the size of firms, the question is whether this also implies a better distribution of wages.

Table 4 shows how the sample is divided according to the Brazilian macro-region.

Table 4 Description of the characteristics of the region (2010-2014)

Region	**Observations**	**Frequency (%)**
Midwest	264 102	6.69%
North	145 034	3.67%
Northeast	592 859	15.02%
Southeast	2052645	52.01%
South	717 553	18.18%
Total	3938742	100%

Source: Elaborated by the own author based on RAIS data (2010-2014).

Table 4 it can be seen that most Brazilian workers are concentrated in the Southeast (52.01%), while the lower portion is in the North (3.67%). Even compared with the South (18.18%) and Northeast (15.02%), second and third largest labor markets, respectively, the difference in the allocation of workers is significant in favor of the Southeast. This should be a reflection of the dynamism of this market, which implies greater diversity of opportunities and the best production structure, which has significant effects on wages. Moreover, agglomeration economies must be formed on the basis of regional attributes, which has a direct effect on the productivity of both workers and firms.

Table 5 shows the average wage of sample according regions of Brazil.

Table 5 Income average per area during the analysis period

Region	Observations	Average (R$)	Des. Pad. (R$)	Min. (R$)	Max. (R$)
Northeast	592 859	1939.98	2855.327	153	104,881.6
North	145 034	2412.57	3149.096	156	10069.9
Southeast	2052645	3017.67	3757.746	154	108,468.8
South	717 553	2506.55	2815.405	154.58	106 103
Midwest	264 102	2606.85	3366.629	156.3	83977.22
Total	3772193	-	-	-	-
Frequency (%)	95.77%				

Source: Elaborated by the own author based on RAIS data(2010-2014).

The results of Table 5 above show that the highest average wages in the period was in the Southeast, R$3017.67 for the observed individuals, followed by the Midwest region with R$2,606, 85 and South, with R$2506.55. Individuals in the Northeast region were those who had the lowest average wage, R$1939.981, followed by workers in the northern region, whose average wage was R$2412.575. The wages standard deviations in all regions were higher than average; it shows that there are large differences in earnings between individuals of the Brazilian regions, but also within the regions themselves.

4.2 Results of the panel data estimates

Empirical regularities point to the existence of a wage premium in favor of large cities (GLAESER AND MARÉ, 2001; FREGUGLIA, 2007; ROCHA, 2011). With this based, this section presents the estimation results of the models with panel data: pooled data (POLS), random effects (RE) and fixed effects (FE). It is important, first of all, to point out that the choice of the most consistent models is given considering the Breusch-Pagan and Hausman test, by which it was evident that the fixed effects model is the most appropriate, since there are indications that the individual unobserved component is invariant in time.

Table 6 Effect of Agglomeration, with controls for individual characteristics, on wage differentials of Brazilian workers between 2010 and 2014

Dependent Variable *ln* of real average income (to be continued)

	POLS without controls	POLS	RE	EF
ln employment dens.	0.0565 ***	0.0298 ***	0.0190 ***	0.00388 ***
	(420.40)	(272.80)	(89.28)	(13.43)
Labor Features				
Age		0.0548 ***	0.0677 ***	0.0607 ***
		(268.47)	(226.74)	(108.25)
Controls for Education				
Elementary		0.216 ***	0.0596 ***	-0.00805
		(53.05)	(13.80)	(-1.78)
High school		0.428 ***	0.145 ***	-0.0107 *
		(104.84)	(33.11)	(-2.33)

Table 6 Effect of Agglomeration, with controls for individual characteristics, on wage differentials of Brazilian workers between 2010 and 2014

Dependent Variable *ln* of real average income (conclusion)

	POLS	RE	EF
Higher education	1,093 ***	0.421 ***	0.0383 ***
	(261,63)	(90,21)	(7,83)
Controls for Race / Color			
Indigenous	-0.108 ***	-0.0295 ***	
	(-17.71)	(-4.77)	
Black	-0.109 ***	-0.0422 ***	
	(-86.05)	(-23.84)	
Yellow	0.000536	0.00824 *	
	(0.16)	(1.98)	
Brown	-0.121 ***	-0.0428 ***	
	(-185.67)	(-44.48)	
Controls for Occupations			
Science / Arts	0.595 ***	0.376 ***	0.0991 ***
	(291.59)	(117.45)	(27.79)
Middle level. Tec	0.420 ***	0.244 ***	0.0506 ***
	(244.37)	(89.57)	(16.00)
Admin service	0.119 ***	0.145 ***	0.000716
	(73.22)	(55.50)	(0.23)
Workers. services	-0.0974 ***	0.0234 ***	-0.0220 ***
	(-63.61)	(9.24)	(-7.28)
Workers production 1	0.218 ***	0.129 ***	0.0320 ***
	(149.02)	(52.97)	(10.96)
Workers production 2	0.253 ***	0.168 ***	0.0580 ***
	(143.15)	(61.57)	(18.16)
Maintenance workers	0.183 ***	0.0960 ***	0.0256 ***
	(195,75)	(65.41)	(15.02)
Dummies of year			
year 2011	0.101 ***	0.112 ***	0.134 ***
	(113.17)	(331.83)	(202.64)
year 2012	0.197 ***	0.222 ***	0.265 ***
	(218.97)	(524.13)	(245.74)
year 2013	0.289 ***	0.324 ***	0.386 ***
	(319.24)	(645.71)	(256.75)
year 2014	0.352 ***	0.400 ***	0.483 ***
	(385.68)	(686.95)	(249.63)
Constant	5,371 ***	5,510 ***	6,138 ***
	(954.77)	(745.26)	(355.33)
N	3,947e + 06	3,947e + 06	3,947e + 06
R2	.4806		
R (overall)		.4175	0.1384
Prob> Chi		0.0000	

Source: Elaborate by the own author based on RAIS data (2010-2014)

Note: * p <0.05, ** p <0.01, *** p <0.001

Table 6 shows the results for the estimates of the employment density on wages. The goal, a priori, is to capture the net effect of agglomeration of the proposed measure on the wage differential. The values show a positive effect, though small in magnitude,

employment density on wage differentials. In the OLS model, without control, the density effect of employment on the employee's remuneration is 5.65%. This corroborates the findings of Abel and Deitz (2012) and Silva (2017) and is consistent with the idea that agglomeration economies are potentially explanatory factors for regional differences in productivity. Still, it should be noted that workers', firms' or even region's attributes may affect the behavior of wages and, accordingly, in this sense, it is fundamental to implement controls in order to discount the effect of these characteristics on individual productivity.

The next step was to estimate the agglomeration effect on the wage differential considering a vector of individual characteristics (Table 6). In this first period are discounted only the individual effects, since the idea is to identify how these affect the wage structure between regions and after adopting them as the associated control the agglomeration differential persists. They are estimated OLS models, random effects and fixed effects. Considering, however, that the unobserved heterogeneity is invariant in time, and there are empirical evidences that attest to this, as the Hausman test, the analysis focuses on the comparison between OLS and FE models.

By adding individual controls, the estimated wage differential associated with the density of employment falls to 2.98% against 5.65% in the model without control. It is still a considerable effect and reflects the relevant character of the agglomeration as an explanatory factor for differences in productivity between locations. In practice, it is expected that the larger agglomerations resides a wider range of productive features, which generate benefits to worker level, via more robust interactions with firms, which may have an effect on wages. Moreover, if these regions there are advantages such as economies of scale, better inputs sharing structure or learning purposes, more productive firms to be attracted to these markets and, therefore, more productive workers should also be. In this sense,

On the other hand, the reduction coefficient is unquestionable when the implementation of controls by the observable characteristics of individuals. In contrast, most of the estimated coefficients for the individual characteristics have statistically significant and positive effect on wages. Draws attention the education variable, where individuals gain on average 2 times more than in comparison category (1.98). This draws attention to some issues. The first is human capital is still a potential source of wage differentials and, in fact, education is a major factor in the composition of observable skills of workers. Second, if education extends the information set of

individuals is to be expected that the more educated to focus on regions with the greatest diversity of opportunities and features potentiating productivity effect. In these terms, the wage differential associated with more schooling reflects not only the set of productive skills inherent in them, but also the ability to obtain information to enable them to allocate more compatible in posts with stock your skill. Thus, behind the agglomerative forces, there may be a sorting effect skills, acting as a centripetal force on the most skilled workers. The wage differential associated with more schooling reflects not only the set of productive skills inherent in them, but also the ability to obtain information to enable them to allocate more compatible in posts with stock your skill. Thus, behind the agglomerative forces, there may be a sorting effect skills, acting as a centripetal force on the most skilled workers. The wage differential associated with more schooling reflects not only the set of productive skills inherent in them, but also the ability to obtain information to enable them to allocate more compatible in posts with stock your skill.

As for the occupational groups, the highest differences are for the professional workers of science and arts' group, even with fixed effect of individual (around 10%), compared to agricultural workers. It makes perfect sense, because in theory, this occupational group are placed individuals with best qualifications, since the set of occupations inherent to it include those of more complex nature. In this sense, it is expected best combinations of workers and occupations, that is, productive characteristics of individuals and occupational requirements of firms, reflecting higher salaries (SILVA, 2017). In other words, groups linked to agricultural activity they tend to receive lower wages compared to those linked to the sectors of industry and high demand for professional qualification. This is in line with the findings by Gatica et al. (1995) for which activities related to complex sectors generate higher wage returns.

As expected, the wage gap is still favorable to whites, reflecting racial segregation in the Brazilian labor market. As for age, observed a growing relationship with wages (0.55%). This same behavior persists in the estimations with fixed effects. In fact, empirical evidences (TOPEL and Ward, 1994; Balassiano et al., 2005) show that at a certain age wages are expected to grow in response to a dynamic process of acquiring skills.

The point is that even controlling for observable and unobservable characteristics of individuals, the wage differential associated with the labor market persists size, however small (only 0.383%). In this sense, some evidence can be drawn

from these results. In practice, the effect of employment density is very expressive on wage differentials and this reflects the diversity of characteristics of workers in the various labor markets in Brazil. Thus, once again the results point to a favor skills of sorting effect of large urban centers, so that the productivity differences between regions may be more related to the set of individual characteristics than agglomeration.

Thus, in order to get more information on the influence of the firm's characteristics and the region on wages, the next analysis is to implement controls for these two units. This procedure was adopted, plus the assumption inherent in firms, for a few reasons: first because despite the density of employment is a measure of crowding, lack of controls by other regional characteristics can make the skewed results; second, there is empirical evidence pointing to a macro-regional component in Brazil (BARROS, 2011; ROCHA, 2011).

Table 7 Effect of Agglomeration, with controls for individual characteristics, region and firm on the wage differentials of Brazilian workers between 2010 and 2014
Dependent Variable ln the real average income (to be continued)

	POLS	RE	FE
ln employment dens.	0.0201 ***	0.0133 ***	0.00224 ***
	(177.25)	(61.69)	(7.68)
Worker caractheristics			
Age	0.0517 ***	0.0649 ***	0.0595 ***
	(275.81)	(227.95)	(107.48)
Controls for Education			
Elementary	0.177 ***	0.0567 ***	-0.00672
	(47.32)	(13.67)	(-1.50)
High school	0.366 ***	0.148 ***	-0.00688
	(97.59)	(35.16)	(-1.52)
Higher education	0.912 ***	0.414 ***	0.0384 ***
	(237.12)	(92.10)	(7.96)
Controls for Race / Color			
Indigenous	-0.0686 ***	-0.0208 ***	
	(-12.26)	(-3.41)	
black	-0.0921 ***	-0.0432 ***	
	(-78.54)	(-25.20)	
Yellow	0.00976 **	0.0118 **	
	(3.27)	(2.89)	
Brown	-0.0632 ***	-0.0332 ***	
	(-96.62)	(-34.80)	
Not identified	-0.0302 ***	-0.0228 ***	
	(-24.39)	(-16.48)	
Controls for Occupations			
Science / Arts	0.647 ***	0.375 ***	0.0995 ***
	(299.39)	(115.00)	(28.19)
Middle level. Tec	0.412 ***	0.232 ***	0.0488 ***
	(216.77)	(83.16)	(15.56)
Admin service.	0.158 ***	0.133 ***	0.000557
	(85.16)	(49.74)	(0.18)
Workers. services	0.0464 ***	0.0418 ***	-0.0133 ***
	(25.69)	(15.90)	(-4.44)
Workers. production	0.149 ***	0.0920 ***	0.0219 ***

Table 7 Effect of Agglomeration, with controls for individual characteristics, region and firm on the wage differentials of Brazilian workers between 2010 and 2014
Dependent Variable ln the real average income (to be continued)

	POLS	RE	FE
	(85.92)	(36.46)	(7.58)
Workers. production 2	0.157 ***	0.123 ***	0.0452 ***
	(80.00)	(43.86)	(14.31)
Workers. Maintenance	0.165 ***	0.0861 ***	0.0230 ***
	(163.46)	(58.03)	(13.63)
Controls for the size firm			
Small business	0.157 ***	0.0755 ***	0.0438 ***
	(186.46)	(97.86)	(54.46)
Medium-sized enterprises	0.283 ***	0.154 ***	0.0919 ***
	(326.28)	(151.55)	(82.59)
Big companies	0.382 ***	0.217 ***	0.128 ***
	(448.30)	(181,86)	(94.39)
Controls by Sector Firm Activity			
Fisheries and aquaculture	-0.0558 ***	-0.0181	-0.0300
	(-4.52)	(-0.91)	(-1.25)
Mining and quarrying	0.506 ***	0.311 ***	0.0994 ***
	(184.96)	(59.74)	(16.38)
Transformation. Ind	0.0640 ***	0.109 ***	0.0436 ***
	(37.23)	(39.92)	(12.66)
Electric., Gas and water	0.159 ***	0.185 ***	0.0452 ***
	(65.90)	(41.53)	(8.25)
Construction	0.0450 ***	0.0628 ***	0.0181 ***
	(22.50)	(19.83)	(4.70)
Trade	-0.103 ***	0,000472	-0.0134 ***
	(-58.25)	(0.17)	(-3.80)
Transport and Post	0.000113	0.0357 ***	-0.00877 *
	(0.06)	(11.79)	(-2.30)
Accomodation and food	-0.265 ***	-0.119 ***	-0.0557 ***
	(-109.91)	(-29.66)	(-10.71)
Inf. And Communication	-0.0953 ***	0.0572 ***	-0.0284 ***
	(-40.15)	(14.19)	(-6.05)
Financial. Act	0.305 ***	0.373 ***	0.102 ***
	(133.11)	(75.49)	(15.50)
Real estate. Act	-0.142 ***	-0.00572	-0.0185 *
	(-25.61)	(-0.77)	(-2.33)
Technician	-0.00637 **	0.0626 ***	-0.0152 ***
	(-2.64)	(16.24)	(-3.46)
Administrative. Act	-0.163 ***	-0.0446 ***	-0.0368 ***
	(-85.00)	(-14.76)	(-9.86)
Education	-0.330 ***	-0.0613 ***	-0.0541 ***
	(-119.59)	(-11.17)	(-7.55)
Health and Social service	-0.306 ***	-0.0464 ***	-0.0181 **
	(-113.23)	(-9.79)	(-2.94)
Arts, cult. and sports	-0.346 ***	-0.110 ***	-0.0324 **
	(-83.31)	(-14.10)	(-3.22)
Other Being. Collective	-0.206 ***	-0.0356 ***	-0.0296 ***

Table 7 Effect of Agglomeration, with controls for individual characteristics, region and firm on the wage differentials of Brazilian workers between 2010 and 2014
Dependent Variable ln the real average income (conclusion)

	POLS	RE	FE
	(-80.57)	(-8.88)	(-6.14)
Domestic services	-0.261 ***	-0.0881 **	-0.0392
	(-8.45)	(-2.93)	(-1.32)
International. Org	0.425 ***	0.0905 ***	-0.0591 ***
	(22.27)	(5.51)	(-4.07)
Macrolocation control			
Northeast	-0.210 ***	-0.192 ***	-0.0495 ***
	(-245.58)	(-87.12)	(-10.85)
Southeast	0.146 ***	0.126 ***	0.0191 ***
	(204.11)	(74.43)	(5.76)
Midwest	0.0919 ***	0.0675 ***	0.0338 ***
	(80.38)	(25.27)	(7.24)
South	-0.0904 ***	-0.0843 ***	-0.0399 ***
	(-114.22)	(-46.63)	(-8.77)
Dummies year (Yes)			
N	3,947e + 06	3,947e + 06	3,947e + 06
R2	.5640		
R (overall)		.5233	.2235

Source: Elaborate by the own author based on RAIS data (2010-2014)
Note: * p <0.05, ** p <0.01, *** p <0.001

Table 7 shows that after including all controls, employment density effect on wages was significantly reduced (0.224%). This effect is expected, because the controls are implemented by firm size and sector of activity, two potential sources of productivity were incorporated (Freguglia, 2007). In this sense, there is apparently a combined effect of worker - firm acting in the determination of wages in Brazil.

With regard to firm size, in all estimation methods used, the larger firms had higher wage differentials. In the estimation of fixed effects, they had a positive differential of 19.72% compared to their counterparts in microfirms. This result is consistent with the findings of Topel and Ward (1994) and Silva (2017), in which had been found that the establishment of the size affects the productivity and therefore is a potential source of wage differentials.

With respect to sectors, the results corroborate with Hoffman (2001) and Freguglia (2007) who claim that the industrial sectors have greater wage differentials in relation to other sectors. When compared to the industry sectors of agriculture, livestock and crop production are the financial activities sector (10.2%), the extractive industry (9.95%), the manufacturing industry (4.9%) and electricity, gas and water (5.5%). Thus, it can be suggested that more complex sectors of the economy keep fundamental relationship with nature of the gains. This makes sense and is expected to be considered the equivalence between sectors and occupations. In other words, most dynamic sectors

require more complex occupations, which support the demand for more skilled workers. In this context, it can be expected that the firms in the more technological sectors with higher qualification requirements will be those that pay higher wages.

In this sense, the evidence presented here is supported by the literature and evidence relevant explanatory power of the characteristics of the firms on the productivity of the Brazilian regions. Anyway, inter-regional production distortions may reflect weaknesses in the local productive matrix and this may be related to differences in the regional distribution of firms. In other words, some regions concentrate the most productive firms, while others focus more rudimentary activities.

Finally, the controls for macrolocation show positive wage gains in favor of Southeast workers (1.9%) and Midwest (3.4) compared to the North. Already the Northeast region had negative differentials in all analyzes of the work. These results reflect the differences in production structure between regions in Brazil and those regional differences incorporate structural elements, which are behind the generation of productive opportunities.

The estimated models show here, however, that despite the agglomeration influence the distribution of wages in Brazil has modest role in compensation differential. Characteristics of firms and workers are more relevant and reflect apparently better interactions between individual skills and sector skills. In fact, a higher density employment may be contaminated by a poor relationship between firm's expertise and skill of the workers, so as not to get excellent results on productivity, which may be a factor behind the poor relationship between agglomeration and wages...

4.2.1 Model estimated by instrumental variable

As mentioned in the methodology of this study, the final step of estimation of the proposed model is to estimate the model, through an instrumental variable, through 2SLS. In the first stage regression, as stated by Gujarati and Porter (2011), the effect of employment density endogeneity will be eliminated once used the instrumental variable. In this case, the instrument of this work is the logarithm of the population lagged to decades from 1910 to 1950.

With the use of instrumental variables, the likely endogeneity of employment density will be treated. Below are the main results of this estimation are shown.

The results of the estimation by instrumental variable are arranged in table 8.

Table 8 Estimation by instrumental variable of the Agglomeration Effect, with controls by individual characteristics, region and firm, on the Wage Differentials of Brazilian Workers between 2010 and 2014

Dependent Variable ln the real average income (to be continued)

	(IV and RE)	(VI and FE)
ln employment dens.	0.0141***	0.00404***
	(45.42)	(8.92)
Workers characteristics		
Age	0.0631***	0.0599***
	(223.88)	(133.35)
Controls for education		
Elementary	0.2563	-0.01892
	(54.7)	(-3.76)
High school	0.32***	-0.0176*
	(68.84)	(-3.46)
Higher education	2.414***	0.2001
	(357.18)	(0.41)
Controls for Occupation		
Science / Arts	0.108***	0.0442
	(3.76)	(1.58)
Middle level Tec	0.00343	-0.00270
	(0.12)	(-0.10)
Administration Services	-0.0937	-0.0154
	(-3.28)	(-0.55)
Workers. services	-0.162	-0.00716
	(-5.63)	(-0.25)
Workers. production 1	-0.0444	0.0190
	(-1.55)	(0.68)
Workers. production 1	-0.0130	0.0426
	(-0.45)	(1.52)
Maintenance worker	0.0426	0.0406
	(1.49)	(1.45)
Controls by Sector Firm Activity		
Fisheries and aquaculture	0.00702	-0.00620
	(0.40)	(-0.30)
Mining and quarrying	0.357***	0.104***
	(90.02)	(22.24)
Transformation. Ind	0.141***	0.0567***
	(52.99)	(17.55)
Electric., Gas and water	0.209***	0.0531***
	(56.79)	(11.83)
Construction	0.0864***	0.0264***
	(29.37)	(7.60)
Trade	0.0137	-0.00521
	(5.00)	(-1.59)
Transport and Post	0.0550	-0.000463
	(19.10)	(-0.13)
Accomodation and food	-0.111***	-0.0486***
	(-31.12)	(-11.36)
Inf. And Communication	0.0829***	-0.0173***
	(25.22)	(-4.54)
Financial. Act	0.400***	0.107***
	(112.12)	(23.74)
Real estate. Act	0.00913	-0.0118
	(1.56)	(-1.88)
Technician	0.0905***	-0.00476
	(28.09)	(-1.28)
Administrative. Act	-0.0332***	-0.0292***
	(-11.54)	(-8.56)

Table 8 Estimation by instrumental variable of the Agglomeration Effect, with controls by individual characteristics, region and firm, on the Wage Differentials of Brazilian Workers between 2010 and 2014

Dependent Variable ln the real average income (conclusion)

	(IV and RE)	(IV and FE)
Education	-0.0549***	-0.0395***
	(-13.51)	(-8.04)
Health and Social service	-0.0401***	-0.00871
	(-10.24)	(-1.86)
Arts, cult. and sports	-0.0973***	-0.0140*
	(-16.41)	(-1.96)
Other Being. Collective	-0.0257***	-0.0219***
	(-7.35)	(-5.41)
Domestic services	-0.0739*	-0.0363
	(-2.52)	(-1.22)
International. Org	-0.0212*	-0.0595*
	(-0.93)	(-2.56)
Controls for the size firm		
Small business	0.0775***	0.0394***
	(103.84)	(52.19)
Medium-sized enterprises	0.157***	0.0842***
	(176.75)	(90.34)
Big companies	0.220***	0.116***
	(222.17)	(108.41)
Macrolocation controls		
Northeast	0.00399	0.0846***
	(1.27)	(14.80)
Southeast	-0.204***	-0.0359***
	(-112.35)	(-11.84)
Midwest	0.118***	0.0208***
	(94.77)	(9.58)
South	0.0512***	0.0354***
	(21.30)	(8.35)
Dummies of year (yes)		
N	3052701	3052701
R(overall)	0.5426	0.1973

Source: Elaborate by the own author based on RAIS data (2010-2014)

Note: * p <0.05, ** p <0.01, *** p <0.001

The results of instrumental variable estimation for fixed effects show that there was a slight increase of 0.3% in the model with fixed effects model to 0.4% in fixed combination and instrumental variable effect. The signal that this estimate presents is going against what Silva (2017) and Combes, Duranton and Gobillon (2008) found in their respective studies, where the explanation percentage of the density of employment on wages, is reduced. However, this result corroborates with what Barufi, Haddad and Nijkamp (2016) observed in their study. The authors instrumentalized the density of employment, using 1940 data to understand the explanatory power of agglomeration economies. So there is in both national and international literature, as shown, a divergence in the results found with the use of instrumental variables, which may be connected to the instrument used.

Interestingly, the instrument used in this study, was also adopted by Combes, Duranton and Gobillion (2008) and Barufi, Haddad and Nijkamp (2016), the logarithm of the lagged population. Therefore, the result here found for the density of employment is another result that literature will have available for comparisons.

With regard to the other variables it was found that, by the method of instrumental variables and fixed effects, which changes the parameters are the values of the estimates, but the meaning remains the same. As in the model of fixed effects, the larger firms, the top-level workers, who are part of groups of professionals in the sciences and the arts; the average technicians and workers in the production of industrial goods and services (2) were the ones that showed positive wage differentials compared their respective categories of analysis.

The firm's characteristics showed the same behavior in two models, with minor differences in the parameters. The sectors of financial activities, mining and quarrying and manufacturing and electricity, gas and water showed the biggest wage differentials.

The Midwest and Southeast showed the highest positive differential, respectively. While the Northeast the largest negative differential.

In general, the estimates varied, but did not exceed the margin of 1.5% more or less. According to Silva (2017) evidence suggests that some of the endogeneity density seems to be originated from individual unobserved components and firms. Thus, to author the model including such components, endogeny density becomes just a residual problem.

5. FINAL CONSIDERATIONS

This work was the main objective to identify the contribution of the worker, firm and region effects, focusing on employment density as a measure of agglomeration, on the wage differential in the Brazilian regions in the 2010 to 2014. Econometric models based on mincerian equation wages were estimated. The main strategy was the regression estimation by panel data, for monitoring the individual over time.

The first OLS regressions were designed to capture the pure effects of agglomeration economies as a measure of sorting skills on wages. This effect ranged from 1.82% to 5.65%, depending on the suitable estimation technique. Thereafter, other controls manner being added until it reached the maximum level controls. The panel data with fixed effects was the most appropriate strategy for this model, with a view to the possibility of existence of unobserved heterogeneity.

The results showed that there is indeed a wage difference correlated with economies of agglomeration, by sorting kills. Even after all the controls have been added, and the estimation has been made with panel data, the sorting was able to explain the wage differential of 0.4%. Therefore, the sorting influences the local productivity, contributing to the density of employment and hence to a pay differential.

With regard to controls of individuals, education and occupational group in which the worker is inserted have a strong influence on wage differentials. Individuals with higher education have a wage differential in their favor in the order of 3.84% when the strategy adopted is FE. The professionals of science and arts (9.95%); mid-level technicians (4.9%) and workers in the production of industrial goods and services 2 (4.5%) are the groups that hold more explanation on wage differences. Note that these groups require a higher educational level than the groups linked to the rural sector.

Regarding firms, the size of it and the sectors in which it is inserted have a strong influence on the actual variations of wages. The bigger the firm, the higher wage income. With regard to sectors, that offers higher wage differentials are those most industrialized and require higher technical level. The sector of financial activities, the mining industry, the electricity sector, gas and water processing industry are that offer higher wages differentials, in the order of 10.2%; 9.95%; 4.5% and 4.3%. We conclude therefore that sectors related industries when compared to non-industrial proffer higher wages.

With regard to regions, the Northeast region presented on all models tested negative differentials, being the leader in this regard. Already the Midwest and Southeast regions are those with the highest positive differential, respectively. The Southeast has a salary greater than 50% difference compared to the Northeast.

Moreover, as the literature considers the employment of an endogenous variable density, we attempted to alleviate this problem by means of instrumental variable, in this case, the natural logarithm of the lagged population. Made this correction, the results show that density employment had its percentage of explanation of the high salaries by 13.3%. As for the observable characteristics of the worker, the firm and the region fluctuations did not reach 1.5 points percentages are down or up, when comparing the results of estimations by IV and FE with FE only.

We conclude, therefore, that the work reached the desired objectives to measure how the sorting contributes to the wage gap and also to present the main observable characteristics of individuals, firms and regions, all this through a robust methodology,

backed by international and national literature. Moreover, this work is one of the few papers to adopt so robust model of simultaneous effects. Finally, this study makes important contributions to national and international literature on the wage differential considering the sorting skills as most important agglomeration factor in these differentials, also leaves results how to correcting the endogeneity of employment density by instrumental variables.

REFERENCES

ABEL, J.R.; ISHITA, D.; GABE, T. M. Productivity and the density of human capital. **Journal of Regional Science**, v.5, n. 4, p.562-586, 2012.

ARBACHE, J. S. Do unions always decrease wage dispersion? The case of Brazilian manufacturing. **Journal of Labor Research**. 1999.

ARBACHE, J. S. Wage differentials in Brazil: theory and evidence. **Journal of Development Studies**, Londres, v. 38, n. 2, p. 691-714, 2001.

ARBACHE, J.S; DE NEGRI, J.A. Filiação industrial e diferencial de salários no Brasil. **RBE**. 2004.

AZZONI, C. "Personal income distribution within states and income inequalities between states in Brasil: 1960, 1970, 1980 e 1991".**Latin American Economics Abstracts**, EstadosUnidos, v. 1, n. 12, 1997.

AZZONI, C.R; SERVO, L.M. **Education, cost of living and regional wage inequality in Brazil**. 2002.

BALASSIANO, M. et. al. **Escolaridade, salários e empregabilidade: tem razão a teoria do capital humano?**. 2005.

BARROS, A. R. **Desigualdades regionais no Brasil: naturezas, causas, origens e soluções**. Rio de Janeiro: Elsevier, 2011.

BARUFI, A.M.B.; HADDAD, E.A.; NIJKAMP, P. Industrial scope of agglomeration economies in Brazil. **The Annals of Regional Science**, v. 56, n. 3, p. 707–755, 2016.

CICCONE, A; HALL, R. E. "Productivity and the Density of Economic Activity", **American Economic Review**. v. 86: 54-70. 1996.

COMBES, P. P.; DURANTON, G.; GOBILLON, L. Spatial wage disparities: Sorting matters. **Journal of Urban Economics**, n. 63, p. 723–742, 2008.

COMBES, P.P., DURANTON, G.; GOBILLON, L. "Spatial Wage Disparities: Sorting Matters!". **Journal of Urban Economics**. 2007.

FALCÃO, N.; NETO, R. S. Concentração espacial de capital humano e externalidades: o caso das cidades brasileiras. In: 35º Encontro Nacional de Economia, Recife, 2007. **Anais**... Recife: ANPEC. 2007.

FREGUGLIA, R.S. **Efeitos da Migração sobre os salários no Brasil**. Tese de Doutorado. Universidade de São Paulo, 2007.

FREGUGLIA, R.S; MENEZES-FILHO, N.A; SOUZA, D.B. Diferenciais salariais inter-regionais, interindustriais e efeitos fixos individuais: uma análise a partir de Minas Gerais. **Estudos Econômicos**. 2007.

GATICA, J., MIZALA, A., ROMANGUERA, P. Interindustry wage differentials in Brazil. Economic Development and Cultural Change. 1995.

GLAESER, E; MARÉ, D. Cities and Skills. **Journal of Labor Economics**, v. 19, n. 2, p. 316-342, 2001.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. São Paulo: Mcgrawhill Bookman, 2011.

HALFDANARSON, B.; HEUERMANN, D.F.; SUDEKUM, J. H. **Capital Externalities and the Urban Wage Premium: Two Literatures and their Interrelations... 2008**. 2008.
HOFFMANN, R. **Distribuição de renda: medidas de desigualdade e pobreza.** 2001.
INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Arranjos Populacionais e Concentrações Urbanas do Brasil**. Rio de Janeiro, 2015. Available in: <https://www.ibge.gov.br/apps/arranjos_populacionais/2015/pdf/publicacao.pdf>. Accessed on aug. 13rd ago, 2018.
INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Classificação Nacional de Atividades Econômicas – CNAE (versão 1.0)**. Available in: <http://cnae. ibge.gov.br/busca-online-cnae.html?view=estrutura>. Accessed on aug. 13rd, 2018.
INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Downloads: Áreas Territoriais**. Availabel in: <https://www.ibge.gov.br/geociencias-novoportal/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=downloads>. Accessed on aug. 13rd aug, 2018.
INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Índice Nacional de Preços ao Consumidor Amplo - IPCA.** Available: http://www.sidra.ibge. gov.br/snipc/ tabelaIPCA.asp?o=3&i=P . Accessed on aug. 15th aug, 2018
LANGONI, C. G. **Distribuição de Renda e Desenvolvimento Econômico no Brasil**. 3. ed. Rio de Janeiro: FGV, 280p. 1973.
MARSHALL, A. **Principles of Economics**. 8th ed. London: Macmillan, 1920.
MELO, P. C.; GRAHAM, D. J.; NOLAND, R. B. A meta-analysis of estimates of urban agglomeration economies. **Regional Science and Urban Economics**, v. 39, n. 3, p. 332-342, 2009.
MENEZES, T. A; AZZONI, C.R. Convergência de salários entre as regiões metropolitanas brasileiras: custo de vida e aspectos de demanda e oferta de trabalho. **Pesquisa e Planejamento Econômico**. 2006.
MENEZES-FILHO, N.A; MENDES, M; ALMEIDA, E.S. O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção? **RBE**. 2004.
MINCER, J. Schooling Experience and Earnings. **National Bureau for Economic Research**. New York, 1974.
NETO, A.B.F; FREGUGLIA, R.F; FARJADO, B. G. Diferenciais salariais para o setor cultural e ocupações artísticas no Brasil. **Economia Aplica**. 2012.
PESSOA, S. A. Existe um problema de desigualdade regional no Brasil. **Anpec**. 1997.
ROCHA, R. M.; NETO, R. M. S; GOMES, S. M. F. P. O. Maiores cidades, maiores habilidades produtivas: ganhos de aglomeração ou atração de habilidosos? Uma análise para as cidades brasileiras. In: ENCONTRO REGIONAL DE ECONOMIA – ANPEC NORDESTE, 16, 2011, Fortaleza. **Anais...** Fortaleza: ANPEC, 2011. Available in: http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_maiores_cidade s.pdf . Accessed on dec 15th, 2011.
SAVEDOFF, W. D., Os diferenciais regionais de salários no Brasil: segmentação versus dinamismo da demanda. **Pesquisa e Planejamento Econômico**. 1991.
SILVA, A. S**. Migração e Prêmio Salarial: evidências para os trabalhadores do agreste pernambucano**, Dissertação (Mestrado) – Universidade Federal de Pernambuco, CAA. Programa de Pós-Graduação em Economia, - Caruaru, 2013.
SILVA, D. L.G. **Contribuição dos efeitos de Firma e de indivíduo para os efeitos de localização sobre os salários e para a variação salarial do trabalhador formal do Brasil**. 2015.

SILVA, D. L.G. **Economias de Aglomeração e heterogeneidade de trabalhador e firma na determinação de salários no Brasil.** Tese de Doutorado. Universidade de São Paulo, 2017.
TOPEL, R. H & WARD, M. P. Job Mobility and the Careers of Young Men. **The Quarterly Journal of Economics**, MIT Press, vol. 107, nº2, 1992.
WOOLDRIDGE, J. M. Econometric analysis of cross section and panel data. **Cambridge: Massachusetts Institute of Technology – MIT Press**, 2011. 2º ed. 741p.