Should I stay or should I go? Dynamic agglomeration economies in Brazil

Ana Maria Bonomi Barufi[£]

[£] Doutoranda, Departamento de Economia, Universidade de São Paulo, ana.barufi@gmail.com

Área 10 – Economia Regional e Urbana

Resumo

O fenômeno da urbanização continua ocorrendo de maneira bastante intensa, especialmente em países em desenvolvimento, tornando necessária a compreensão do que faz as cidades terem tanto sucesso como centros produtivos. Nesse contexto, o tema das economias de aglomeração no mercado de trabalho tem sido explorado de maneira significativa no período recente em função do grande interesse em explorar novas bases de dados longitudinais e de propor políticas para o tamanho ótimo de cidades. A maioria dos estudos procura avaliar os ganhos estáticos da aglomeração, mas uma corrente mais recente da literatura tem buscado entender em que medida a aglomeração proporciona ganhos dinâmicos de produtividade (afetando o crescimento dos salários). Desse modo, utilizando dados da RAISMIGRA de 1995 a 2008, este estudo busca mensurar não apenas as externalidades estáticas como também as externalidades dinâmicas das aglomerações, procurando entender em que medida o movimento migratório dos indivíduos afeta os seus ganhos salariais futuros. Assim, calcula-se o impacto da experiência adquirida nos anos anteriores em diferentes tamanhos de cidade no salário atual, observando que o maior número de anos trabalhados em cidades mais densas resulta em maiores ganhos salariais.

Palavras-chave: economias de aglomeração, ganhos dinâmicos, prêmio salarial urbano

Abstract

The urbanization phenomenon is still going on, especially in developing countries, raising the question of why are cities this successful as productive centers. In this context, agglomeration economies in labor market outcomes have been explored significantly in the past years, even more because longitudinal individual micro-data were made available only recently in many countries. Most studies try to measure static agglomeration externalities, but more recently, there has been an increasing interest in dynamic agglomeration gains (which affect wage growth). Here we consider data for the Brazilian formal labor market from 1995 to 2008 (RAISMIGRA), trying to measure not only static agglomeration externalities, but also dynamic ones. In this sense, we try to understand how migratory movements affect the future wage level of individuals. We calculate the impact of acquired experience in different city sizes over the observed wage, finding that if individuals work more years in bigger cities, they will have higher wages in the future.

Keywords: agglomeration economies, dynamic gains, urban wage premium

JEL codes: J31, R32, R23

1. Introduction

Cities are more important than ever for economic growth and life in society. People living, working, and thinking together in dense areas is a sign of the success of cities (Glaeser, 2011). If in the United States 80.7% of the population live in 3.06% of the territory (Glaeser, 2008, and US Census Bureau[1]), in Brazil this spatial concentration is even higher (84.4% of the population was living in 1.07% of the territory in 2010[2]). In the United States, Glaeser and Maré (2001) find that in metropolitan areas workers earn 25% more than their non-urban counterparts do. In this perspective, increasing urbanization is interpreted as a sign of gains from agglomeration (net from costs).

In fact, the urban economics literature provides strong evidence of the existence of an urban wage premium. This sort of analysis can be found for example in Combes et al. (2008a), Glaeser and Maré (2001), Melo and Graham (2009), Mion and Naticchioni (2009), Ciccone and Hall (1996), Combes et al. (2010), Groot et al. (2014a), and Groot and de Groot (2014b). A literature review can be found in Puga (2010) and Rosenthal and Strange (2004). The urban wage premium can range from 1% to 11% depending on the sample and the country considered. This widespread effort is pushed by the fact that longitudinal individual-level microdata on the labor market has been recently made available in many different countries.

The studies aforementioned aim to measure the static effect of agglomeration economies on productivity on the regional level. Combes et al. (2008a) summarize the three main sources of regional wage disparities: (i) the composition of the local labor market; (ii) the availability of local non-human endowments that can increase productivity; and (iii) agglomeration economies. The third item is related to wage differences that follow from close proximity between firms and workers/consumers, thick labor markets and knowledge spillovers (Duranton and Puga, 2004).

In general, this type of work involves regressing nominal wages at the individual level (as a productivity outcome) on a measure of agglomeration at the local level (Duranton, 2014a). Wages are supposed to equalize the marginal productivity of labor under perfect competition, and in the case there is some market imperfection, they are supposed to be higher in places that are more productive. As pointed out by Duranton (2014b), wages should be in nominal instead of in real terms, because the former are more related to productivity differentials, while the latter represent life standard levels.

In addition to their static and instantaneous effect, agglomeration economies can have a long-lasting dynamic impact on productivity. There is no clear answer on the best strategy to follow to analyze dynamic externalities and separate them from static ones, due to limitations on data availability. De La Roca and Puga (2014) provide one of the most complete studies in this area, but relevant efforts were also made by Yankow (2006), Wheeler (2006), D'Costa and Overman (2014), Carlsen et al. (2013), and Gordon (2015), among others.

There are only a few studies on agglomeration economies in Brazil, mostly based on the regional level or individual cross-sections (Simões and Freitas, 2014; Maciente, 2013; Fontes et al., 2010; Freguglia and Menezes-Filho, 2012; Silva and Silveira-Neto, 2009). According to the knowledge

[1] Available in http://www.census.gov/geo/reference/ua/urban-rural-2010.html and accessed in16/04/2014.
[2] IBGE, Mapa de uso da terra de 2010, available in ftp://geoftp.ibge.gov.br/mapas_tematicos/mapas_murais/shapes/uso_da_terra_2010/ and accessed in14/04/2014.

of the authors, so far there is no study covering dynamic agglomeration economies in the Brazilian context (or for an emerging economy).

Therefore, this study aims to fill this gap in the empirical research in Brazil, as well as to contribute to the methodological discussion of the literature. The first set of results covers the evaluation of static agglomeration economies using a large longitudinal database (covering the period from 1995 to 2008). In the second set of regression models, a first attempt of estimation of dynamic effects is provided, considering the effect of previous experience in different city sizes over wages.

The next sections are organised as follows: Section 2 contains a brief discussion of the conceptual differences between static and dynamic externalities; Section 3 discusses the methodological approach; Section 4 discusses the main studies on dynamic agglomeration economies; and then, Section 5 presents the data; Section 6 brings the results; and, finally, Section 7 draws the main conclusions from the analysis.

2. Static versus dynamic externalities

Static effects are based on the assumption that agglomeration economies have a once-and-for-all effect on productivity, with no further impact in the following periods. On the other hand, dynamic externalities aim to measure a long-lasting effect or a permanent impact of city size or urban density on productivity (Combes and Gobillon, 2015). Another way of distinguishing these concepts is the following: static productivity advantages refer to a higher productivity found in large cities when compared to smaller ones, while dynamic productivity advantages are associated to the fact that productivity can increase differently over time for each city size (Camagni et al., 2015).

Among the possible permanent impacts of agglomeration economies, local productivity growth can increase with technological spillovers, or individuals can learn more and faster in bigger cities (Lucas, 1988). In this perspective, individuals can learn by working in bigger cities (De La Roca and Puga, 2014). Glaeser (1999) defends that the rate of interaction between people is accelerated by urban density, leading to an increase in human capital accumulation due to human interactions. This effect is particularly high for highly skilled people, usually more concentrated in big cities. Yankow (2006) points out that this learning hypothesis requires some time in the city, meaning that there is no expectation of immediate wage gains when moving to a bigger urban concentration or any wage loss when workers leave cities.

3. Methodological approach

The literature on agglomeration economies largely explores the model developed within the Urban Economics framework, which aims to represent the relationship between local characteristics and productivity (nominal wage). The main assumption here is that worker's location choices are strictly exogenous, meaning that they do not take into account wage shocks due to localized job opportunities in their migration or job change decisions. Another possible strategy is to introduce static and dynamic agglomeration effects in a dynamic model of location choices (Baum-Snow and Pavan, 2012), but this sort of estimation requires a structural model behind it in order to achieve identification.

The following model can represent the static agglomeration approach, inspired in the Abdel-Rahman and Fujita (1990) tradition. A production function of a price-taking representative firm in region r and time t, using $l_{r,t}$ of labor and an amount $k_{r,t}$ of other inputs is given by (Combes et al., 2008b):

$$y_{r,t} = A_{r,t}\left(s_{r,t}l_{r,t}\right)^{\mu}k_{r,t}^{1-\mu} \tag{1}$$

In this case, $A_{r,t}$ is the technology factor (Hicks-neutral), μ is the share of labor in the productive process and $s_{r,t}$ is the efficiency level of workers. Moreover, $w_{r,t}$ is the wage level and $r_{r,t}$ is the price of other inputs. Then, the profit of the firm in region *r* is:

$$\pi_{r,t} = p_{r,t}y_{r,t} - w_{r,t}\, l_{r,t} - r_{r,t}\, k_{r,t} \tag{2}$$

In a competitive equilibrium, the first-order conditions for the optimal input mix generates a wage equation:

$$w_{r,t} = \mu(1-\mu)^{(1-\mu)/\mu}\left(\frac{p_{r,t}A_{r,t}}{r_{r,t}^{1-\mu}}\right)^{1/\mu} s_{r,t} \equiv B_{r,t}s_{r,t} \tag{3}$$

Here, $B_{r,t}$ represents the composite local productivity effect. This effect can be decomposed in pecuniary externalities, which appear through local markets (p_j and r_j capture agglomeration and dispersion forces respectively), and pure local externalities (do not appear through a market mechanism, A_j is related to technological externalities – knowledge and learning spillovers, workers and firms sharing indivisible goods, among others)[3]. Due to the great difficulty in estimating these effects separately, most studies in the literature quantify the overall local effect (Combes and Gobillon, 2015). Furthermore, the potential inclusion of measures of cost of living in the model to explain real wage inequalities between cities will be related to a completely different question, while the relation of local characteristics and nominal wages is enough to capture the net effect of agglomeration on productivity.

The component of local labor skills ($s_{r,t}$) should be included because higher skills may be overrepresented in larger cities (due to other factors than the city size itself, such as local amenities), meaning that its omission may overstate the impact of agglomeration economies (composition effect). On the other hand, people can also become more skilled in bigger cities (through a learning process, externality due to proximity to more skilled people), meaning that the inclusion of a local skill variable can capture part of the agglomeration effect. Combining these two elements, it is important to include skills, but not at the local level, and individual skills can do the job, by avoiding measuring this agglomeration effect. Therefore, equation 3 becomes:

$$w_{i,t} = B_{r,t}s_{i,t} \tag{4}$$

In this case, information at the individual level is required (individuals that work in region r), available only recently in some countries. This approach is based on the assumption that agglomeration economies embedded in $B_{r,t}$ have a once-and-for-all impact on productivity. However, as mentioned in the previous section, agglomeration economies can have a long-lasting or even permanent effect. Potential sources of dynamic gains can be identified in all components of $B_{r,t}$, but as in the static case, it is only possible to measure the overall impact of dynamic externalities (Combes and Gobillon, 2015).

The inclusion of dynamic effects on the aforementioned model is done by firstly simplifying equation 4 to consider a static-local-effects-only model for a local outcome, $y_{i,t} = u_i + \beta_{r(i,t),t} + \epsilon_{i,t}$. Then, the individual productivity growth is written as $y_{i,t} - y_{i,t-1} = \beta_{r(i,t),t} - \beta_{r(i,t-1),t-1} + \varepsilon_{i,t}$. The

[3] Scitovsky (1954).

inclusion of dynamic fixed effects is done by considering in this regression a fixed effect for region r in t (region r in $t-1$ has an impact on productivity growth between $t-1$ and t, for $t \geq 1$):[4]

$$y_{i,t} - y_{i,t-1} = \beta_{r(i,t),t} - \beta_{r(i,t-1),t-1} + \mu_{r(i,t-1),t-1} + \varepsilon_{i,t} \quad (5)$$

Equation 5 can be transformed to obtain:

$$y_{i,t} = y_{i,0} = \beta_{r(i,t),t} + \sum_{k=1}^{t} \mu_{r(i,t-k),t-k} + \zeta_{i,t} \quad (6)$$

This transformation shows that even when dynamic effects affect only the growth rate of the local outcome ($y_{i,t} - y_{i,t-1}$ in equation 5), they do have a permanent effect on its level. There are a few assumptions underlying the relationship considered in equation 5, namely the fact that dynamic effects are perfectly transferable over time, and that this specification makes more sense for stayers than movers. In the case of movers, it is relevant to consider not only the region of origin (in $t-1$), but also the region of destination (in t). Then, for movers it is more appropriate to re-write equation 6 in the following way:

$$y_{i,t} = y_{i,0} = \beta_{r(i,t),t} + \sum_{k=1}^{t} \mu_{r(i,t-k),r(i,t),t-k} + \zeta_{i,t} \quad (7)$$

The problem with this formulation is that the number of parameters for dynamic effects becomes very large ($R^2.T$, where R is the number of regions and T the number of time spams in the panel), and the identification of the model requires the imposition of restrictions on the parameters. This last aspect can be seen in the case of a re-written version of equation 5, for stayers:

$$y_{i,t} - y_{i,t-1} = \beta_{r(i,t),t} - \beta_{r(i,t-1),t-1} + \mu_{r(i,t-1),r(i,t)t-1} + \varepsilon_{i,t} \quad (8)$$

In which the variation of static and the level of dynamic effects cannot be estimated separately (and the same happens in the movers' case). The next section will present a description of the different strategies used in the literature so far, and it will be followed by a discussion of the empirical strategy applied here.

4. Strategies of estimation of dynamic agglomeration economies

There are only a few studies dealing with dynamic agglomeration economies and including all the issues discussed in the literature of static gains. In this section some of these studies are discussed, with their advantages and disadvantages.

One of the most complete empirical studies in dynamic agglomeration economies is done by De la Roca and Puga (2014). Their main goal is to separate the possible reasons why firms would pay higher salaries in bigger cities: (i) static advantages; (ii) sorting of more productive workers; (iii) dynamic advantages (cities facilitate learning and experimentation). Therefore, they estimate a mincerian wage equation including individual fixed effects to capture the sorting process, and the number of years of experience each individual has of working in different city sizes. They estimate this model to a very large Spanish panel data set (they follow workers from 1981 to 2009, and evaluate their wages from 2004 to 2009). Two restrictions applied to the model discussed in the

[4] These steps are discussed by Combes and Gobillon (2015).

previous sections are that the authors consider only a few groups of city sizes (instead of allowing a continuous range of city sizes) and that individual heterogeneity is equal in the dynamic and the static effects.

With a different strategy, D'Costa and Overman (2014) use a large panel (1998-2008) for British workers aiming to evaluate whether the sorting of high ability workers can explain urban wage premiums and if this wage premium is given to the workers immediately or over time through faster wage growth. Their conclusion is that both learning and sorting matter for understanding the effects of cities on wage growth. The main restriction imposed in the model discussed above is that even if they estimate the specification in first differences, allowing for a distinct heterogeneous effect in the static and the dynamic cases, they have to exclude movers from the analysis (therefore, this study does not measure between-city dynamic effects).

A few years earlier, Wheeler (2006) estimated the impact of density on annual wage growth, also calculating the within-job and between-job components of this growth, with data for young male workers in the United States. Their main finding is that there is a positive relationship between wage growth and city size, but when they control for individual fixed effects there is no evidence of this urban wage premium on growth.

Yankow (2006) also regress the wage growth against changes in the location of work, separating the mobility effect from the growth effect. Their conclusions for United States male young workers is that when they move into cities, they have a significant wage growth (6 percentage points) in comparison to the individuals that stayed in a rural area (with a symmetric effect for out-of-the-city migrants). The main drawback in their strategy is that they cannot control for sorting on unobservables, a limitation that D'Costa and Overman (2014) try to overcome.

5. Empirical strategy

Equation 4 discussed in section 3 can be estimated in the following way:

$$y_{i,t} = \mu_i + X_{i,t}\theta + Z_{r(i,t),t}\gamma + \eta_{r(i,t),t} + \varepsilon_{i,t} \quad (9)$$

In which $y_{i,t}$ represents the logarithm of the hourly wage and $Z_{r(i,t),i}$ contains the logarithm of employment density. Most empirical studies consider employment instead of population as a measure of the agglomeration because it is as better measure of local economic activity and it is easily available for many years (Combes and Gobillon, 2015). Ciccone and Hall (1996) defend that density should be included instead of total employment in this regression, because it deals better with the heterogeneity of the spatial extent of the geographical units considered. The estimated parameter for density will be an elasticity, whose meaning is easily understood.

Following the discussion widely explored in Combes et al. (2008a, 2010 and 2011) and De la Roca and Puga (2014), in order to avoid potential endogeneity in the estimation of equation 9, it is important to control for individuals unobserved heterogeneity (fixed effects) to deal with the sorting process (endogenous quality of labor). Moreover, there is a high risk of endogeneity of the density measure (endogenous quantity of labor), what requires an instrumental variables estimation. Following the argument defended by Groot et al. (2014a) that fixed effects create a black box in the explanation of urban wage differentials, the estimations presented here will also allow for a POLS strategy.

Then, proceeding to the estimation of a dynamic model, an equation inspired in the work of De la Roca and Puga (2014) is considered:

$$y_{i,t} = \mu_i + X_{i,t}\theta + Z_{r(i,t),t}\gamma + \eta_{r(i,t),t} + \sum_{j=1}^{R} \delta_{j,r} exper_{i,j,t} + \varepsilon_{i,t} \tag{10}$$

Here, once again $y_{i,t}$ is the outcome variable, μ_i is the individual fixed effect, $X_{i,t}$ is the set of time-varying individual characteristics, $Z_{r(i,t),t}$ contains employment density of region r in t, $\eta_{c(i,t),t}$ represents a region fixed effect, $exper_{i,j,t}$ is the experience acquired by worker i in region j up until time t and $\varepsilon_{i,t}$ is the error term.

In order to reduce computational complexity[5], all these estimations will be conducted in one stage only (usually in the literature there is an effort to run a two-stage regression with the second stage being the regression of the spatial wage residual against local characteristics – Groot et al., 2014a; Combes et al., 2008a; among others). One additional aspect is that $\eta_{r(i,t),t}$ will be replaced by macro-regions dummies, and $Z_{r(i,t),t}$ may contain either the logarithm of density or a categorical variable grouping REGIC areas in four groups, defined by the quartiles of density for all individuals.

6. Data

The main database considered here is RAISMIGRA (Annual Report of Social Information - Migration, from the Ministry of Labour), which consists of identified registration data of all formal firms and their employees in the Brazilian labor market, focusing on the characteristics of the work contract. This database provides longitudinal data for all formally employed individuals in the private sector (or part of the public sector, depending on the type of contract), with a significant regional disaggregation (municipal level), from 1995 to 2008.

One of the main advantages of this database is that it is a mandatory report, covering the entire formal sector. Because of this, differently from individual self-reporting surveys, there is a smaller risk of wage under-reporting. Nonetheless, the fact that it only covers the formal sector presents a potential drawback in the analysis, as the informal sector in Brazil is very relevant (RAIS represents less than 30% of the workforce in the initial years of the period analyzed, reaching around 35% in 2008. Moreover, the reporting process is supposed to be more accurate in the case of bigger firms, which are usually located in larger cities.

The regional unit of analysis is the REGIC area[6], which is a better measure of a labor market area. In fact, this level of analysis is more suitable for the empirical analysis of agglomeration economies, as the estimation will not be affected by daily commuting (that is very different from

[5] This estimation in two stages would mean to include 14*482 dummies in the first stage for REGIC area and year combinations, apart from individual fixed effects.

[6] REGIC areas (482) aggregate municipalities based on their interaction, mostly associated to daily commuting flows, transportation links and interconnectedness in general terms (IBGE, 2013). From 1995 to 2008 some new municipalities appeared, requiring an adaptation of the original REGIC areas to encompass these new municipalities according to the municipalities they were originated from. The authors can provide details of this aggregation process.

moving permanently to another area to look for another job). A few steps were conducted[7] in order to achieve a database representing the dynamics of a competitive labor market, in manufacture and service sectors. Only individuals observed in all years (from 1995 to 2008, and separately from 1995 to 2002 and 2003 to 2008) Finally, due to limitations on computer processing, it was necessary to select a 10% sample[8] for each database.

There are other data-sources considered in this study. For instance, skills are obtained from the original study of Maciente (2012), who converted the ONET classification to the Brazilian occupational classification (of 2002). However, the RAISMIGRA database contained two different occupational classifications along the whole period (from 1995 to 2002, the available classification is the occupational structure of 1994, while the period from 2003 to 2008 has the occupational structure of 2002). Moreover, Maciente's study considers occupations disaggregated at the five-digit level. As the RAISMIGRA database allows a three-digit disaggregation only, skills classified at the five-digit classification were aggregated through a simple average (and a conversion to the 1994 classification was applied for the 1995-2002 period). Therefore, as there are some methodological differences in the construction of the data from 1995 to 2002 and 2003 to 2008, whenever skills are considered, it is not possible to analyze the whole time period at once (1995-2008). In order to do this more comprehensive analysis, educational level dummies were included.

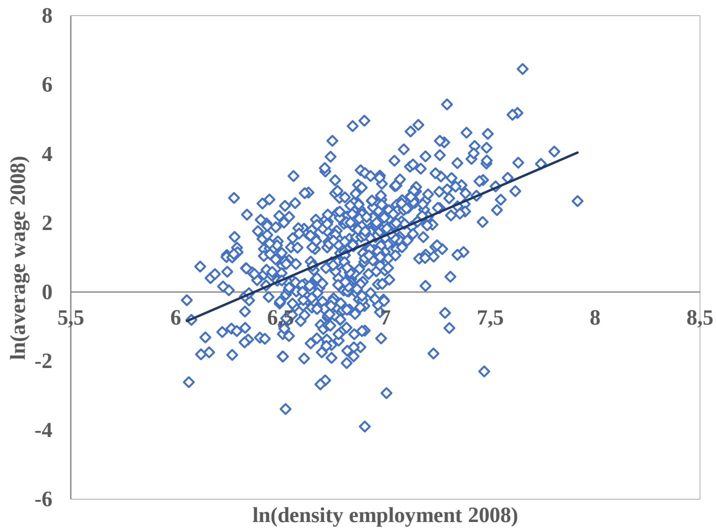
Finally, instruments related to colonial economic activities and geographic characteristics were provided by Naritomi et al. (2012) and aggregated to the REGIC area level. Population density in 1940 was obtained at the IPEADATA website. Further details are presented in Table A1 in annex.

A first analysis based on aggregated data indicates that there is a positive relationship between the logarithm of wages and the logarithm of local density (0.52). However, it is necessary to evaluate this correlation net of other confounding factors.

[7] There was an initial selection of active contracts in December of each year for male individuals working for private companies in permanent jobs. Then, contracts with wages equal to zero, or with less than 20 weekly hours were excluded, as well as contracts with missing information on educational level and occupation. The next step conducted was the selection of individuals aged 18 to 36 years in 1995 (for databases starting in 1995) and 18 to 36 in 2003 (for databases starting in 2003). Finally, only individuals working in manufacture and service sectors were kept in the database (sectors 15 to 36, 55 to 74 and 76 to 99, at the 2 digits-CNAE 1.0 classification).

[8] These samples are representative for the following characteristics in the initial year of each database: age group (less than 25 years old, 25 to 29 years old, 30 to 36); employment level in the REGIC area (less than 50 thousand, 50 thousand to less than 100 thousand, 100 thousand to less than 500 thousand, 500 thousand to 1 million, 1 million or more); firm size (up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more); macro-region; educational level (illiterate, incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more); and sector (CNAE 1.0 at the 2 digits level).

Figure 1. Correlation between the logarithm of employment density and the logarithm of wages at the REGIC area level, 2008.



Source: RAIS/MTE and IBGE.

The main descriptive statistics are presented in Table 1 and show that there is a higher concentration of workers in manufacture sectors, in the age group from 30 to 36 years old in 1995. Furthermore, higher percentages of workers are found in firms with 250 or more employees, in the Southeast region of the country (North and Center-West are underrepresented, as a sign of their higher concentration in extractive activities and agriculture). In terms of educational attainment, there is a high concentration of workers with incomplete middle school or with incomplete college level. Finally, comparing wages in 1995 and 2008 (deflated by INPC-IBGE), they seem almost stable. However, density of employment seems to have increased in the period. Some of these characteristics of the database are due to the small size of the sample (only 10% of the original database comprising workers observed in the whole period).

7. Results

As mentioned in Section 4, the first set of results explore static agglomeration economies in Brazil. In order to provide a robust result, eight different specifications were estimated, with different compositions of controls and estimation methods. Comparing Models 1 and 2 of Table 1, it is possible to note that there is a positive partial correlation between density and wages, around 0.08, and when individual fixed effects are taken into account, this correlation drops to 0.01. This is an indication that the sorting process is very relevant to determine the wage level in Brazil. Another important consequence is that the effect measure by movers (which is calculated when individual fixed effects are included) is much lower than the one measured with the whole database.

Table 1. Descriptive statistics of the main variables in the first and the last year of the database.

	1995		2008	
	mean	**s.d.**	**mean**	**s.d.**
ln(real hourly wage)	2.88	0.86	2.80	0.87
ln(density)	3.85	1.85	4.32	1.74
Sector				
Manufacture	55.7%		54.9%	
Services	44.3%		45.1%	
Age group (in 1995)				
18 to 24	14.0%			
25 to 29	17.2%			
30 to 36	68.8%			
Firm size				
Up to 4 employees	4.2%		4.0%	
5 to 9	5.4%		5.8%	
10 to 19	5.9%		5.8%	
20 to 49	9.7%		8.9%	
50 to 99	8.5%		8.0%	
100 to 249	14.5%		13.0%	
250 to 499	14.4%		13.9%	
500 to 999	13.4%		13.4%	
1,000 or more employees	24.1%		27.2%	
Macro-region				
North	1.8%		1.9%	
North-east	10.3%		10.3%	
South-east	65.1%		65.1%	
South	19.3%		19.2%	
Centre-West	3.5%		3.6%	
Education attainment				
Illiterate	1.0%		0.4%	
Incomplete primary school	44.0%		27.5%	
Comp. primary school - incompl. middle school	27.1%		23.5%	
Complete middle school - incomplete college	22.4%		33.7%	
College degree or more	5.3%		15.0%	
N	56,950		56,950	

Source: Authors' own calculations.

The second potential endogeneity issue is dealt with the inclusion of instruments for employment density (ln(population density in 1940), sugar cane, gold and coffee exploration in the colonial period, geographical characteristics). As Models 3 and 4 show, there is a slight change in the size of the estimated coefficient for density, but the main conclusion still hold.

Due to the restrictions in the skills variables discussed in the previous section, it is not possible to estimate a model including them for the whole period. Therefore, Models 5 to 8 divide the database in two, from 1995 to 2002 and from 2003 to 2008. In all these models, the estimation is already including individual fixed effects and instrumental variables. The comparison of Models 5 and 6, and 7 and 8, shows that controlling for skills or for education attainment generates very similar results. Finally, it is also possible to see that the density elasticity for the second time period is much smaller than the one for the first time period. A possible explanation for this result is that there was a concentration of workers in bigger cities over the years, increasing the competition for jobs in those places. Another potential reason is that Brazil saw a significant formalization process during that decade, and that can be affecting the results over the years.

Table 2. Regression models for the logarithm of the hourly wage, with static agglomeration economies and density as a continuous variable.

	Model 1	**Model2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Period	1995-2008	1995-2008	1995-2008	1995-2008	1995-2002	1995-2002	2003-2008	2003-2008
Estimation strategy	OLS	FE	IV	IV - FE	IV - FE	IV - FE	IV - FE	IV - FE
ln(density)	0.0806***	0.0113***	0.0699***	0.00588***	0.0143***	0.0145***	0.00417***	0.00411***
Controls included								
Macro-region dummies	x	x	x	x	x	x	x	x
Age, age squared	x	x	x	x	x	x	x	x
Education level dummies	x	x	x	x	x		x	
Skills						x		x
Firm size dummies	x	x	x	x	x	x	x	x
2-digit sector dummies	x	x	x	x	x	x	x	x
Year dummies	x	x	x	x	x	x	x	x
Instruments			x	x	x	x	x	x
R squared	0.559		0.558					
N	797,251	797,251	797,251	797,251	748,262	738,527	1,033,494	930,414
T	14	14	14	14	8	8	6	6
Individuals	56,947	56,947	56,947	56,947	93,533	92,316	172,249	155,069

Source: Authors' own calculations.

Table 3 brings very similar specifications to Table 2. The only difference is that instead of considering density as a running variable, four categories of dummies are included. Apart from the fact that the main conclusions discussed above hold, it is possible to note that people working in REGIC areas of the upper quartile will have higher salaries. Therefore, it seems that the relationship between wages and density is not linear.

Table 3. Regression models for the logarithm of the hourly wage, with static agglomeration economies and density as a categorical variable (based on quartiles)

	Model 1	**Model2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Period	1995-2008	1995-2008	1995-2008	1995-2008	1995-2002	1995-2002	2003-2008	2003-2008
Estimation strategy	OLS	FE	IV	IV - FE	IV - FE	IV - FE	IV - FE	IV - FE
ln(density) - quartile 1 (omitted)								
ln(density) - quartile 2	0.171***	0.0193***	0.222***	0.0263***	0.0523***	0.0504***	0.0288***	0.0224***
ln(density) - quartile 3	0.202***	0.0250***	0.107***	0.0242***	0.0507***	0.0498***	0.0569***	0.0526***
ln(density) - quartile 4	0.413***	0.0152***	0.563***	0.0496***	0.103***	0.104***	0.0393***	0.0295***
Controls included								
Macro-region dummies	x	x	x	x	x	x	x	x
Age, age squared	x	x	x	x	x	x	x	x
Education level dummies	x	x	x	x	x		x	
Skills						x		x
Firm size dummies	x	x	x	x	x	x	x	x
2-digit sector dummies	x	x	x	x	x	x	x	x
Year dummies	x	x	x	x	x	x	x	x
Instruments			x	x	x	x	x	x
R squared	0.562		0.552					
N	797,251	797,251	797,251	797,251	748,262	738,527	1,033,494	930,414
T	14	14	14	14	8	8	6	6
Individuals	56,947	56,947	56,947	56,947	93,533	92,316	172,249	155,069

Source: Authors' own calculations.

Finally, Table 4 presents the results of a first attempt to estimate dynamic and static agglomeration economies for the Brazilian labor market. Analyzing Model 1, it is noticeable that previous experience is not significant, while static gains are significant. When interactions between experience and the quartile of density at the present location are included, some of them become significant, but with non-intuitive signs.

Table 4. Regression models for the logarithm of the hourly wage, with static and dynamic agglomeration economies and density as a categorical variable (based on quartiles)

	Model 1	**Model2**	**Model 3**	**Model 4**
Estimation strategy	POLS	POLS	FE	FE
ln(density) - quartile 1 (omitted)				
ln(density) - quartile 2	0.0162***	0.241***	-0.000270	0.0233
ln(density) - quartile 3	0.128***	0.370***	0.00393	0.0259
ln(density) - quartile 4	0.197***	0.426***	-0.00198	0.0480*
Experience in REGIC areas with density on quartile 1	-0.119	-0.149	-0.000599	-0.000425
Experience in REGIC areas with density on quartile 2	-0.104	-0.125	-0.00351***	0.00901***
Experience in REGIC areas with density on quartile 3	-0.106	-0.0865	-0.00691***	-0.00972***
Experience in REGIC areas with density on quartile 4	-0.102	-0.0650	-0.00858***	-0.00899***
Interaction variables				
density in t in quartile 1 & experience in quartile 1		0.0312***		-0.000115
density in t in quartile 2 & experience in quartile 1		0.00175		-0.00277
density in t in quartile 3 & experience in quartile 1		(omitted)		(omitted)
density in t in quartile 4 & experience in quartile 1		0.00698*		-0.00142
density in t in quartile 1 & experience in quartile 2		0.0669***		(omitted)
density in t in quartile 2 & experience in quartile 2		-0.00153		-0.0121***
density in t in quartile 3 & experience in quartile 2		-0.00400		-0.0133***
density in t in quartile 4 & experience in quartile 2		(omitted)		-0.0122***
density in t in quartile 1 & experience in quartile 3		(omitted)		0.00630*
density in t in quartile 2 & experience in quartile 3		-0.0297***		(omitted)
density in t in quartile 3 & experience in quartile 3		-0.0452***		0.00322
density in t in quartile 4 & experience in quartile 3		-0.0375***		-0.000205
density in t in quartile 1 & experience in quartile 4		(omitted)		-0.00326
density in t in quartile 2 & experience in quartile 4		-0.0342***		(omitted)
density in t in quartile 3 & experience in quartile 4		-0.0291***		0.00660***
density in t in quartile 4 & experience in quartile 4		-0.0604***		0.000438
Controls included				
Macro-region dummies	x	x	x	x
Age, age squared	x	x	x	x
Education level dummies	x	x	x	x
Firm size dummies	x	x	x	x
2-digit sector dummies	x	x	x	x
Year dummies	x	x	x	x
Individual fixed effects			x	x
N	512,532	512,532	512,532	512,532
T	9	9	9	9
Individuals	512,532	512,532	512,532	512,532

Source: Authors' own calculations.

However, when individual fixed effects are included, all signs change significantly. Model 4, the preferred composition, shows that almost all static effects loose their significance, except for the higher quartile, whose coefficient is positive. Therefore, only individuals working in denser cities will get a significant advantage from that. Experience in different types of groups of cities is negative and significant for quartiles 3 and 4.

A result that is more interesting is that individuals who have experience in cities from quartile 2 working in any city size have a lower wage level nowadays, and individuals with experience in cities of quartile 4 working in cities of quartile 3 have a positive and significant gain from that.

This type of analysis presents many possibilities which still have to be explored in future studies.

8. Conclusion

The main conclusions from the analysis presented above are that there is a positive partial correlation between density and wages, around 0.08, and when individual fixed effects are taken into account, this correlation drops to 0.01. This is an indication that the sorting process is very relevant to determine the wage level in Brazil. However, it seems that the relationship between wages and density is not linear.

Considering dynamic and static externalities, when they are estimated simultaneously and individual heterogeneity is taken into account, there is indication that static effects loose importance (but are still relevant for individuals working in the group of denser cities). Furthermore, when experience is combined with the present location, there is a positive gain only for workers in cities with medium density, that had worked in high density cities.

Future studies should consider other estimation strategies to try to assess the relative importance of static and dynamic agglomeration externalities. There is still a long road to go.

Annex

Table A1. Definition and source of the main variables considered in the model.

Variables	**Definition**	**Level**	**Data source**
Hourly wage	Monthly wage received in December divided by 4.3 times the number of weekly hours in the contract.	Individual	RAISMIGRA microdata
Age	Age at the end of the year.	Individual	RAISMIGRA microdata
Education attainment	Illiterate, incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more.	Individual	RAISMIGRA microdata
Skills - factors	Methodological aspects are better described in Maciente (2012) and Maciente (2013)	Individual	RAISMIGRA microdata and Maciente (2012)
Firm size	Size of the firm in which the individual is working: up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more.	Individual	RAISMIGRA microdata
Labour density in the formal sector	Total employment divided by the area (in km^2).	REGIC	RAIS - aggregated data
Area	Area in km^2.	REGIC	IPEADATA
Instruments	**Definition**	**Level**	**Data source**
Population density in 1940	Population in 1940 for minimum comparable areas 1940-2000 is redistributed for minimum comparable areas 2000-2010 based on the share of the population of the latter in the former in 2000, aggregated by REGIC and divided by the area in km^2.	REGIC	IPEADATA
Distance to the coast	Distance to the coast (in 100 km) - obtained as a weighted average of the distance to the sea of the centroids of all municipalities that compose the REGIC (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Altitude	Average altitude of the REGIC area obtained from the weighted mean of the altitude of each municipality (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Distance to the Equator line	Distance to the Equator measured as the absolute value of the latitude coordinate - obtained as a weighted average of the distance to the Equator line of the centroids of all municipalities that compose the REGIC area (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Sugar	Proximity to the sugar cane boom (until 1760), calculated as the weighted average of the municipal index for all municipalities belonging to the REGIC area, ranging from 0 (more than 200 km) to 1, according to the proximity to the nearest municipality in the sugar cane areas (Naritomi et al., 2012) - weight = area.	REGIC	Adapted from Naritomi et al. (2012)
Gold	Proximity to the gold boom, calculated as the weighted average of the municipal index for all municipalities belonging to the REGIC area, ranging from 0 (more than 200 km) to 1 according to proximity to the nearest municipality in gold exploration areas (Naritomi et al., 2012) - weight = area.	REGIC	Adapted from Naritomi et al. (2012)
Coffee	Proximity to the coffee boom until 1886, calculated as the weighted average of the municipal index for all municipalities pertaining to the REGIC area, ranging from 0 (more than 200 km) to 1, according to the proximity to the nearest municipality in coffee areas (Naritomi et al., 2012) - weight = area.	REGIC	Adapted from Naritomi et al. (2012)

Source: Authors' own calculations.

References

Andersson, M.; Klaesson, J.; Larsson, J.P. (2014). The sources of the urban wage premium by worker skills: spatial sorting or agglomeration economies? *Papers in Regional Science*, Vol. 93, No. 4, pp. 727-747.

Baum-Snow, N.; Pavan, R. (2012). Understanding the city size wage gap. *Review of Economic Studies*. 79, pp. 88-127.

Carlsen, F.; Rattso, J.; Stokke, H. (2013). Education, experience and dynamic urban wage premium. *Norwegian University of Science and Technology Working Paper*.

Camagni, R.; Capello, R. Caragliu, A. (2015) Agglomeration economies and urban growth. *AISRe – ABC Workshop, The Frontier of the debate in Regional and Urban Economics*, available in http://www.aisre.it/images/call_for_paper/Workshop_dottorandi_ABC/03_Capello.pdf and accessed in 15/07/2015.

Ciccone, A., Hall, R.E. (1996), Productivity and the density of economic activity. *The American Economic Review*, Vol. 86, No. 1, pp. 54-70.

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

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

__________________________________ (2011), The identification of agglomeration economies. *Journal of Economic Geography*. Vol. 11, pp. 253-266.

Combes, P.P., Duranton, G., Gobillon, L., Roux, S. (2010), 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.

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

Combes, P.P.; Gobillon, L. The empirics of agglomeration economies. In: *Handbook of Regional and Urban Economics Vol 5A*, North-Holland, Amsterdam, 1st Edition, pp. 247-348, 2015.

D'Costa, S.; Overman, H.G. (2014). The urban wage growth premium: sorting or learning? *Regional Science and Urban Economics*. Vol. 48, pp. 168-179.

De la Roca, J.; Puga, D. (2014). Learning by working in cities. Available in http://diegopuga.org/papers/esurban.pdf and accessed in 01/07/2015.

Duranton, G. Agglomeration and jobs. In: Fischer, M.M.; Nijkamp, P. (eds.) *Handbook of Regional Science*. Spring-Verlag Berlin Heidelberg, pp. 631-648, 2014a.

__________. Growing through cities in developing countries. *The World Bank Research Observer Advance Access*, pp. 1-35, Abril, 2014b.

Duranton, G., Puga, D. (2004), Micro-foundations of urban agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.) *Handbook of Regional and Urban Economics*. Vol. 4. Amsterdam: North-Holland, pp. 2063-2117.

Glaeser, E.L. *Cities, Agglomeration and Spatial Equilibrium*. Oxford ; New York : Oxford University Press, 2008.

__________. (1999) Learning in cities. *Journal of Urban Economics*, Vol. 46, pp. 254-277.

__________. (2011), *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. London, UK: Macmillan.

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

Gordon, I.R. (2015). Ambition, human capital acquisition and the metropolitan escalator. *Regional Studies*, Vol. 49, No. 6, pp. 1042-1055.

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

Groot, S.P.T., De Groot, H.L.F. (2014b). Estimating the skill bias in agglomeration economies externalities and social returns to education: evidence from Dutch matched worker-firm micro-data. *Tinbergen Institute Discussion Paper TI 2014-088/VII*.

Hanlon, W.; Miscio, A. (2014). Agglomeration: dynamic approach. *NBER Working Paper 20728*.

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

Maciente, A.N. (2013), The determinants of agglomeration in Brazil: input-output, labor and knowledge externalities. PhD Dissertation, University of Illinois at Urbana-Champaign.

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

Melo, P.C.; Graham, D.J. Agglomeration economies and labour productivity: evidence from longitudinal worker data for GB's travel-to-work areas. *SERC Discussion Paper 31*, October 2009.
Duranton, G. Agglomeration and jobs. In: Fischer, M.M.; Nijkamp, P. (eds.) *Handbook of Regional Science*. Spring-Verlag Berlin Heidelberg, pp. 631-648, 2014a.

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

Naritomi, J., Soares, R.R., Assunção, J.J. (2012), Institutional development and colonial heritage within Brazil. *The Journal of Economic History*. Vol. 72, Issue 2, pp. 393-422.

Puga, D. (2010), The magnitude and causes of agglomeration economies. *Journal of Regional Science*, Vol. 50, No. 1, pp. 203-219.

Rosenthal, S.S., Strange, W.C. (2004), 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.

Scitovsky, T. Two concepts of external economies. *Journal of Political Economy*, Vol. 62, No. 2, pp. 143-151, Apr. 1954.

Silva, M.V.B., Silveira-Neto, R.M. (2009), Dinâmica da concentração da atividade industrial no Brasil entre 1994 e 2004: uma análise a partir de economias de aglomeração e da nova geografia econômica. *Economia Aplicada*, Vol. 13, No. 2, pp. 299-331.

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.

Wheeler, C.H. (2006) Cities and the growth of wages among young workers: evidence from the NLSY. *Journal of Urban Economics*. Vol. 60, pp. 162-184.

Yankow, J.J. (2006). Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, Vol. 60, pp. 139-161.