UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

RUBENS LOPES PEREIRA DA SILVA

AGGLOMERATION WITHIN URBAN AREAS: evidence from Brazil

RECIFE
2017

RUBENS LOPES PEREIRA DA SILVA

AGGLOMERATION WITHIN URBAN AREAS: evidence from Brazil

Dissertação apresentada ao Programa de Pós-Graduação do Departamento de Economia da Universidade Federal de Pernambuco (PIMES) como requisito para obtenção do título de mestre em Ciências Econômicas.

Orientador: Prof. Dr. Raul da Mota Silveira Neto

Coorientadora: Prof.ª Dra. Roberta de Moraes Rocha

RECIFE

2017

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

S586a Silva, Rubens Lopes da Pereira
Agglomeration within urban areas: evidence from Brazil / Rubens Lopes Pereira da Silva. - 2017.
37 folhas: il. 30 cm.

Orientador: Prof. Dr. Raul da Mota Silveira Neto e Coorientadora Prof.ª Dra. Roberta de Moraes Rocha.
Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, CCSA, 2017.
Inclui referências.

1. Aglomeração. 2. Dispersão. 3. Estatística espacial. I. Silveira Neto, Raul da Mota (Orientador). II. Rocha, Roberta de Moraes (Orientadora). III. Título.

331.2 CDD (22. ed.) UFPE (CSA 2018–114)

RUBENS LOPES PEREIRA DA SILVA

AGGLOMERATION WITHIN URBAN AREAS: EVIDENCE FROM BRAZIL

Dissertação apresentada ao Programa de Pós-Graduação do Departamento de Economia da Universidade Federal de Pernambuco (PIMES) como requisito para obtenção do título de mestre em Ciências Econômicas.

Aprovado em: 23 de fevereiro de 2017.

BANCA EXAMINADORA

Prof. Dr. Raul da Mota Silveira Neto
Orientador – UFPE

Prof.ª Dra. Roberta de Moraes Rocha
Coorientadora – UFPE

Prof. Dr. André Matos Magalhães
Examinador – UFPE

Prof. Dr. Eduardo Haddad
Examinador - USP

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter encontrado, me guiado e me sustentado até aqui. Agradeço à minha família por todo o apoio durante essa trajetória e por terem financiando com alegria a minha educação. Sou grato à minha esposa, por todo o seu amor, carinho e companheirismo. Sou grato também aos meus amigos porque nossos encontros sempre aliviaram os momentos difíceis. Agradeço também aos irmãos de fé, que tanto me ajudaram com palavras de conforto e em suas petições a Deus em meu favor.

Agradeço imensamente a dedicação, paciência e apoio do meu orientador, Prof. Dr. Raul da Mota Neto. Sem ele nada disso seria possível. Também agradeço a minha co-orientadora Prof.ª Dra. Roberta Rocha, por suas contribuições sempre valiosas. Sou grato aos meus colegas de pós-graduação pelos comentários e apoio.

Por último, agradeço também à CAPES pelo apoio financeiro que me permitiu realizar esta pesquisa.

RESUMO

O trabalho gera e analisa evidências sobre os níveis de aglomeração das atividades econômicas da Região Metropolitana do Recife nos anos de 2006 e 2011 a partir de medidas de concentração baseadas em distância. Ao analisar setores tanto de indústrias quanto de serviços em um contexto intraurbano, esta pesquisa expande uma literatura que, no Brasil, tem focado nas manufaturas e em diferenças regionais. Os níveis de concentração foram estimados a partir da métrica K_d proposta por Duranton e Overman (2005). Medidas clássicas, como índice de Gini ou E-G, são viesadas para o tamanho das áreas de estudo; porém o K_d, por ser baseado em distâncias, não sofre desse viés e ainda pode ser testado por inferência estatística. É mostrado que a região apresenta um nível notável de aglomeração: 57% e 56% dos setores são concentrados para 2006 e 2011, respectivamente (68% e 71% em amostras estritas). Tais resultados são compatíveis com trabalhos semelhantes para outros países. Dos 20 setores mais aglomerados da RMR, 15 e 16 (2006 e 2011) são serviços. Corroborando com pesquisas anteriores, foi encontrado que atividades relacionadas a computação, advocacia, contabilidade, arquitetura e engenharia mostram tendência para concentração. De forma geral, três conclusões são extraídas do trabalho: 1) em ambos os anos, há mais setores aglomerados do que não, 2) não há muita mudança no percentual de setores aglomerados e dispersos ao longo dos anos e 3) a tendência de aglomeração torna-se mais fraca quando as firmas são ponderadas pelo emprego.

Palavras-chave: Aglomeração. Dispersão. Estatística espacial. Urbano.

ABSTRACT

The research provides an analysis of the agglomeration levels of the economic activities of the Metropolitan Region of Recife in the years of 2006 and 2011 using distance-based measures. In analyzing industries of both manufacturing and services in an intra-urban context, our research expands a literature that, in Brazil, has focused on Manufacturing and regional differences. We estimate the concentrations applying the K_d metric proposed by Duranton and Overman (2005). Classical measures, such as Gini or E-G index, are biased with respect to the size of the study areas. The K_d, being a measure based on distances, does not suffer from this bias and can be tested by statistical inference. We show that the region has a remarkable level of agglomeration: 57% and 56% of the sectors are concentrated for 2006 and 2011, respectively (68% and 71% in strict samples); results that are compatible with similar work for other countries. of the 20 most agglomerated sectors, 15 and 16 (2006 and 2011) are services. Endorsing previous research, we find that activities related to computing, law, accounting, architecture and engineering show a tendency towards localization. Three general conclusions can be extracted from the work 1) in both years, there are many more localized groups than dispersed ones; 2) there is not much difference in terms of share of localized and dispersed industries between years; 3) the tendency towards localization is fainter when firms are weighted by workforce.

Keywords: Agglomeration. Dispersion. Spatial statistics. Urban.

LISTA DE FIGURAS

Figure 1 – Recife Metropolitan Region………...……………………………………..................14
Figure 2 – K_d of Health Care Activities…………………………………………….................19
Figure 3 – Location of Health Care Activities………………………………………………..20
Figure 4 – K_d of Manufacturing and Refining of Sugar……..…………………………….…20
Figure 5 – Location of Manufacturing and Refining of Sugar……..…………………………...21
Figure 6 – Distribution of the EG index…………………………...………………………31

LISTA DE TABELAS

Table 1 - Firms and number of observations... 15

Table 2 - Number of firms: one and three-digit aggregation levels.. 16

Table 3 - Summary statistics for K_d estimations - 2011.. 22

Table 4 - 30 most localized and dispersed groups – Three-digit – Services – 2011……..................... 24

Table 5 - 30 most localized and dispersed groups – Three-digit – Manufacturing – 2011................. 25

Table 6 - Conditionings of within urban location – OLS Linear Regression Estimative………..………..………..………..………..………..………..………..………. 29

Table 7 - Summary statistics for EG estimations – 2011……………………………………………. 32

SUMÁRIO

1 INTRODUCTION…………………………………………………………………...10
2 THEORETICAL ARGUMENTS FOR URBAN LOCATION AND AGGLOMERATION…………………………...…………...…………….…………….....12
3 Data and empirical strategy……………………………………………………...14
3.1 Data……………………………………………………………….........................14
3.2 Estimation methodology...………………………………………….........................16
4 RESULTS………..………………………………………………………………...22
4.1 General Results………....……………………………………….........................22
4.2 Characterizing the urban location…......………………………….........................26
4.3 Evidence from traditional measure and the MAUP…..……………..........................30
5 CONCLUSIONS.……………………………………………………………….....33
REFERENCES...……………………………………………………………......34

1. INTRODUCTION

Recent empirical researches are very convincing in confirming the influence of spatial concentration of urban activities on the local labor productivity or on local total factor (CICCONE AND HALL, 1993; GLEASER AND MARÉ, 2001; ROSENTHAL AND STRANGE, 2004; MORRETI, 2004A, 2004B; GLEASER AND GOTTLIEB, 2009; COMBES ET AL., 2010; HEUERMANN ET AL., 2010; HEUERMANN, 2011; JOFRE-MONSENY ET AL., 2014). Although much needs to be understood about the acting channels of these agglomeration effects, the results appear to justify the renewed value given to modern urban agglomerations.

As previously highlighted by Azarghi and Henderson (2008), however, most of the set of evidence were obtained using Manufacturing industries and across urban centers analyses. These circumstances imply that knowledge about the levels and patterns of concentration of economic activities within the urban spaces and associated agglomerations gains is still scarce. Actually, most of the recent empirical investigation about within cities industries spatial distribution considers only with few specific sectors (HUALLACHÁIN AND REID, 1992; SHEARMUR AND ALVERGNE, 2002; GONG AND WHEELER, 2002; HUALLACHÁIN AND LESLIE, 2006; FU, 2007; KLIER AND MCMILLEN, 2008; ARZAGHI AND HENDERSON, 2008; GIULIAN ET AL., 2014). Thus, these investigations neither allow a general comparison across locations of different economic activities, nor a general investigation on possible arguments behind them.

Two important empirical recent researches appear to initiate to fill this gap. Considering the US case, Kolho (2010) showed that service activities are more urbanized than manufacturing, but less concentrated in general, at county level. Billings and Johnson (2016), considering the case of the CMSA of Denver-Boulder-Greelev, applied a co-location indicator based on distance to show higher co-localization rates for transportation, finance and computer services than for the other activities.

The central objective of the research is to use a distance-based metric for evaluating the concentration levels of economic activities in urban Brazil by considering specifically the case of the Recife Metropolitan Region (RMR), the fifth biggest Brazilian metropolitan region. As recognized by Marcon and Puech (2017), by avoiding the MAUP (Modifiable Area Unity Problem), the K_d index suggested by Duranton and Ovreman (2005) is now considered one of the leading functions in spatial economics. Using a unique data set of geocoded firms, we also provide suggestive evidence about economic forces behind the levels of agglomeration within this specific urban space. Besides its monocentric configuration and importance, the RMR is an interesting case because it presents no general zoning for urban land use patterns, which favors market-oriented land allocation.

The investigation brings two contributions to the literature. First, it provides new evidence on location of a large set of economic activities within urban areas using a distance-based measure of location, something still rare even for developed countries[1]. In addition, it considers the patterns of location of firms in an unexplored context, that of a developing country urban area. Usually, urban centers in developing counties are characterized by a lack of functional public transport system and poor urban infrastructure, circumstances that bring new conditionings for location decision of firms within cities. Second, according to best of our known Brazil, the available evidence about agglomeration of activities only considers traditional indexes such as Gini and Ellison-Glaeser index (RESENDE AND WYLLIE, 2005; VIGNARDI, PARRÉ AND GUIMARÃES, 2016).

The set of results indicates that the more concentrated urban activities present higher levels of human capital or higher levels of product differentiation. On the other hand, some of the more dispersed activities belong to manufacturing. These results are consistent with economic arguments for explaining location.

The paper is organized as follows. In the next section, theoretical arguments for urban location are presented and discussed. The data and empirical strategy are presented in section three and, in section four, the results are presented and discussed. Final remarks are presented in the section five.

[1] Most of the available works using DO index consider only the distribution of the manufacturing activities or distribution over the all country; see Duranton and Overmann (2005), Barlet et al.,(2013) amd Koh and Riedel (2014).

2. THEORETICAL ARGUMENTS FOR URBAN LOCATION AND AGGLOMERATION

Traditional locational theory emphasizes the influence of spatial competition for markets, product differentiation, and of consumer externalities for understanding the location of firms (HOTELLING, 1929, EATON AND LIPSEY, 1979; HOLMES AND STEVENS, 2002; CLARK, 2002; KONISHI, 2005; FUJITA AND THISSE, 2013).

At least since Hottelling (1929), it is recognized the role that spatial competition has in generating economic concentration of activities. Note, however, that because of price competition, the Hotelling (1929)'s Nash equilibrium is not necessarily assured when demand is elastic (MULLIGAN AND FIK 1994; ANDERSON AND ENGERS, 1994). In fact, as shown by d'Aspermont et al. (1979), when considering the possibility of price war, an equilibrium with firms' dispersion is quite possible. The situation illustrates the trade-off involved: the possibility of price competition pushes firms away from each other, but competition for markets acts in the opposite direction (FUJITA AND THISSE, 2013). Some product differentiation, thus, appears necessary to restore equilibrium with concentration of firms. As shown by Fujita and Thisse (2013), Nash equilibrium with agglomeration can be obtained in a situation with product differentiation and consumers valuating product variety because these make price competition weaker.

Eaton et al. (1979), Stuart (1979), and Konishi (2005) emphasize another source favoring spatial agglomeration of some urban economic activities, the consumer gains. Some activities, such as retail trade, present spatial concentration because it makes easier for consumers to compare products and it also minimizes transport costs between firms and consumers. In fact, as highlighted by Konishi (2005), spatial concentration of firms in urban centers can result from actions of two opponent forces: more concentrated firms (stores) tend to attract consumers with uncertainty about tastes and expectation of low prices, but, on the other hand, it makes the price lower for the firms. More recently, Billings and Johnson (2016) have argued that the urban spatial distribution of activities is also associated with consumer shop time regularity: because it reduces consumer trips duration, activities demanding more regular and quotidian shops tend to be spreader around residences.

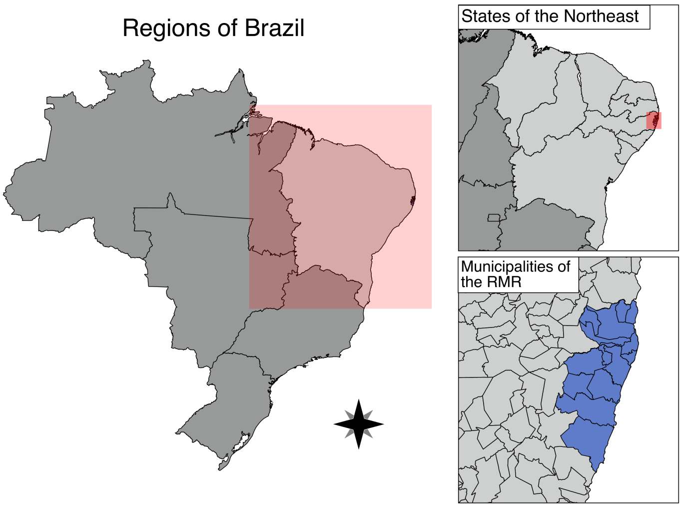
Urban agglomeration gains arguments can also have a role in explaining locations of activities within cities (DURANTON AND PUGA, 2004; BILLINGS AND JOHNSON, 2016). Specifically, if technological spillovers and learning depend on distance, we expect agglomeration of more technological intensive or higher human capital activities within metropolitan areas (KEEBLE AND WILKINSON, 2000; ARBIA ET AL., 2010). Similarly, firm's proximity within cities can also favor accessing informal information and thus promotes a better matching between workers and occupations. Finally, as highlighted by Billings and Johnson (2016), when transport cost is non-negligible, the location of services and input providers can importantly affect the location decision of firms and favor agglomeration within urban areas.

3. DATA AND EMPIRICAL STRATEGY

3.1. Data

With an area of 2,770.452 km^2, the RMR, our study area, is the most important urban agglomeration of Brazilian Northeast region (see Figure 1), the biggest of the North-Northeast regions, and the fifth largest of Brazil. According to 2010 census (IBGE, 2012), the RMR has a population of 3.69 million inhabitants and it is the third most densely populated metropolitan area of the country. The area is located in the state of Pernambuco and is composed by 14 municipalities (Recife, Abreu e Lima, Araçoiaba, Cabo de Santo Agostinho, Camaragibe, Igarassu, Ilha de Itamaracá, Ipojuca, Itapissuma, Jaboatão dos Guararapes, Moreno, Olinda, Paulista, and São Lourenço da Mata).

Figure 1- Recife Metropolitan Region



Since the initial periods of the Portuguese colonization in Brazil in the 16th century, Recife gained economic relevance because of its soil, weather and privileged geographic position, becoming a flourishing place for sugar cane production and maritime trade. Centuries have passed and the southern states developed into the main economic region of Brazil, but Recife remained as one of the most important cities of the country.

Our analysis will make use of three data sets, the CNPJ data (National Registry of Legal Entities) from Minister of Finance, the RAIS/MTE (Annual Report of Social Information) developed by the Brazilian Ministry of Labor and Employment, and the PNAD (a national household survey) from IBGE (Brazilian Institute of Geography and Statistics).

The CNPJ database comprises the registry information of all national companies, from which we have obtained complete firm's addresses. For the RMR, the data is available from 1901 to 2015, totalizing more than 490,000 firms. The RAIS is an annual report that all active formal companies must fill out about their employees. The data is available for the whole

RMR, from 2006 to 2011, comprising more than 7 million observations. From this dataset we know, for each year, if the company was active of not, how many employees it had, its industry code and a set of laborers characteristics, such age and schooling.

By merging firms present in both CNPJ and RAIS data sets and using geocoding techniques, we have obtained a unique data set containing firm's location[2] and employees' information. We could identify the location of 92,3% of the firms of the RMR region in the years of 2006-2011 from RAIS. This comprehends more than 68,000 firms that, over 6 years, totalize more than 249,000 lines of information at the firm level.

Note that some constraints were also imposed on this dataset before starting the analysis. For space reasons, the results and analysis are focused mainly on the last year of the database (2011), although we also use some evidence for 2006 as illustration. To guarantee that the analysis would reflect the reality of the firms' location behavior of an economic activity, we considered only industries with at least 50% of the jobs in the formal market in both years according to data from PNAD. Finally, in the same way as Billings and Johnson (2016), we focused our analysis on industries that had at least 10 firms in both years of study, but, for robustness purposes, we also reported results without this last restriction. Our final database corresponds to 61.027 observations comprising 44.007 plants.[3] Information about the data processing and adjustment is reported in Table 1.

Table 1 – Firms and number of observations

	Observations	**Plants**	**% of plants**
RAIS firms (2006-2011)	275,958	76,036	100.0%
RAIS firms identified in CNPJ database	256,720	70,148	92.3%
RAIS geocoded firms (precisely located)	249,430	68,046	89.5%
RAIS geocoded firms (precisely located) for 2006 and 2011	85,589	61,597	81.0%
RAIS geocoded firms (precisely located) for 2006 and 2011, with at least 50% of formality	61,509	44,328	58.3%
RAIS geocoded firms (precisely located) for 2006 and 2011, with at least 50% of formality and from industries with at least 10 plants	61,027	44,007	57.9%

Source: the information was obtained by merging to data sets: RAIS/MTE information about firms and the information of firm address from CNPJ database.

[2] We use the software QGIS for geo-processing.

[3] 10,288 firms for 2006 only; 16,699 firms for 2011 only; 17,020 firms present in both years.

In Brazil, the CNAE (*Classificação Nacional de Atividade Econômicas*) is the official classification system of industries. Although there have been some revisions of the CNAE, here we use the original 1995 version. The posterior revisions have not changed it much and, by using this variant, we avoided losing information when associating data from different datasets (CNPJ, RAIS and PNAD). We use two levels of aggregation: sections (a one-digit code), the most aggregated (17 sections or industries), and groups (a three-digit code), with intermediary level of aggregation (218 groups or industries). For our study, due database constrains (as described above), our working dataset contains 8 sections and 85 groups, with 55 belonging to Services sector. Table 2 presents descriptive statistics.

Table 2 – Number of firms: one and three-digit aggregation levels

	2006	2011
Number of Firms	26,987	34,040
Number of CNAE groups (three-digit)	85	85
Number of CNAE sections (one-digit)	8	8
Average of employees per plant	25.8	28.8

Source: the information was obtained by merging RAIS/MTE information about firms with the information of firm address from CNPJ database

3.2 Estimation methodology

Duranton and Overmann (2005) described five ideal proprieties for a spatial agglomeration index: 1) it must be comparable across different industries; 2) the general pattern of firms' agglomeration is controlled; 3) the industrial concentration is controlled; 4) the empirical results are not biased in respect to scale and aggregation; and 5) the results' significance is testable. The first two requirements are essential to any concentration measure. Traditional metrics, such as the Gini index, already satisfy them. The third requirement is also satisfied by the index proposed by Ellison and Gleaser (1997), which assumes a null hypothesis of spatial randomness that accounts for industrial concentration. Neither Gini nor EG indexes, however, can satisfy the fourth propriety, as they require points to be first aggregated in delimited areas before executing the analysis, which introduces the Modifiable Areal Unit Problem (MARCON AND PUECH, 2009).

The issue with MAUP is twofold. First, as it is discussed in section 5, the results of the current investigation are sensitive to different spatial aggregations. Second, because firms are aggregated into spatial units, all firms within a region are treated equally – points near each other have no more weight than points in opposed extremes. Furthermore, estimations are biased for industries that agglomerate near the limits of areas. But a metric that treats space continuously, such as the K_d, can avoid these problems (ARBIA AND ESPA, 1996; LANG ET AL., 2016). As it is distance-based, the K_d treats each point individually, without aggregating them into spatial units.

The fifth property is important because, if an index assumes a null hypothesis of random distribution, any evaluation can only be probabilistic. Hence, a correct assessment of concentration requires a hypothesis test. The K_d can test the significance of its results by generating confidence bands via Monte-Carlo simulations. As the K_d satisfies all five properties, it can be considered an adequate concentration measure.

Recently, new distribution-based agglomeration measures have been proposed. Marcon and Puech (2009) designed the M function, an extension to Ripley's functions that respects all five proprieties. Lang et al. (2016) also proposed the *m* function, a density function of M. Despite these developments, the K_d has been the leading function in economic studies to measure spatial concentration (MARCON AND PUECH, 2017) and we follow, therefore, the DO's approach. They use a density function K_d, which counts the average number of pairs of points for each distance, using a kernel function for weighting the pairs of observations according their distance deviation from the referred distance. Then, K_d values are, thus, compared to confidence bands obtained under the null hypothesis of random distribution of firms. Shortly, this method consists of four steps:

1. Obtain the value of the index using kernel density estimation. The first step is to obtain the bilateral distances between firms. For an industry with *n* plants, the Euclidian distance is calculated for all pairs of points resulting in $\frac{n\,(n-1)}{2}$ bilateral distances. Define $r_{i,j}$ as the distance between establishments i and j. Given n points, the K-density estimator for each point i of an industry at a distance r is obtained as:

$$\widehat{K_d}(r) = \frac{1}{n(n-1)h} \sum_{j=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{r-r_{i,j}}{h}\right) \quad , \qquad (1)$$

whereas h is the bandwidth of the Gaussian distribution[4] and f is the Gaussian Kernel-function:

$$f = K\left(\|x_i - x_j\|, r\right) = \frac{1}{h\sqrt{2\pi}} exp\left[-\frac{\left(\|x_i - x_j\| - r\right)^2}{2h^2}\right] \quad , \qquad (2)$$

[4] The ideal h value is the one which minimizes the mean integrated square error. As the density function is unknown, the h is estimated. D-O (2005) follow Silverman (1986), who defines $h = 0.9An^{-1/5}$ and $A = \min(\sigma, interquantile\ range/1.34)$ arguing that it "will do very well for a wide range of densities".

in which x_i is the reference point and x_j is a neighbor. The maximum value of the Kernel function is reached when the distance between points i and j is equal to r and decreases according to a Gauss distribution with SD h as the distance deviates from r. There is also a variation of the $\widehat{K_d}$ in which firms are pondered by employees or other information, the $\widehat{K}^{emp}$:

$$\widehat{K}^{emp}(r) = \frac{1}{\sum_i \sum_{j=i+1} w(x_i)\, w(x_j)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w(x_i)\, w(x_j)\, f\left(\frac{r - r_{i,j}}{h}\right) \quad (3)$$

2. Construct counterfactuals. Counterfactuals are generated by sampling (without replacement) the number of firms in the industry that is being analyzed from the overall population of sites. This controls for the activities' overall agglomeration. Given an industry *m*, each sample is a pseudo-*m* industry, for which a kernel-density $\widetilde{K}_d$ is estimated. For each *m*, 1000 $\widetilde{K}_d$ counterfactuals are generated. These are null hypothesis simulations, which form the confidence interval.

3. Global confidence bands. Following DO, we consider distances from 0 up to the median of all bilateral distances of the data[5]. This extension is divided in 512 equal parts and a $\widetilde{K}_d$ is measured for each one. Then, for every industry and distance, the iterations are ranked. The lower bound $\underline{\widetilde{K_d}(r)}$ and upper bound $\overline{\widetilde{K_d}(r)}$ will be the values such that no less than 95% of all randomly generated $\widetilde{K}_d$, across the whole distance spectrum, lies, respectively, below and above these bands.

4. Identify localized and dispersed industries. Once we have our confidence bands, we can verify if the estimated values for the $\widehat{K_d}$ indicate concentration, dispersion or randomness. If $\widehat{K}_{d_m}(r) > \overline{\widetilde{K}_{d_m}(r)}$ for at least one *r*, we say the industry is localized. If $\widehat{K}_{d_m}(r) < \underline{\widetilde{K}_{d_m}(r)}$ for at least one *r* and the industry is not localized[6], we say the industry is dispersed. For each industry *m*, the localization and dispersion indices are defined, respectively, as below:

$$\Gamma_m(r) \equiv \max\left(\widehat{K}_{d_m}(r) - \overline{\widetilde{K}_{d_m}(r)}, 0\right)$$

(4)

$$\psi_m(r) \equiv \begin{cases} \max\left(\underline{\widetilde{K}_{d_m}(r)} - \widehat{K}_{d_m}(r), 0\right) & if \ \sum_{r=0}^{512} \Gamma_m(r) = 0 \\ 0 & otherwise \end{cases}$$

(5)

Finally, we construct condensed measures by summing all values across the spatial extension:

[5] The range is from 0km to 35km in our case.

[6] As the values are normalized to sum 1, it is expected that peaks of concentration will be compensated by points below the lower confidence bound. This, however, does not imply dispersion.

$\Gamma_m = \sum_{r=0}^{512} \Gamma_m(r)$ $\psi_m = \sum_{r=0}^{512} \psi_m(r)$

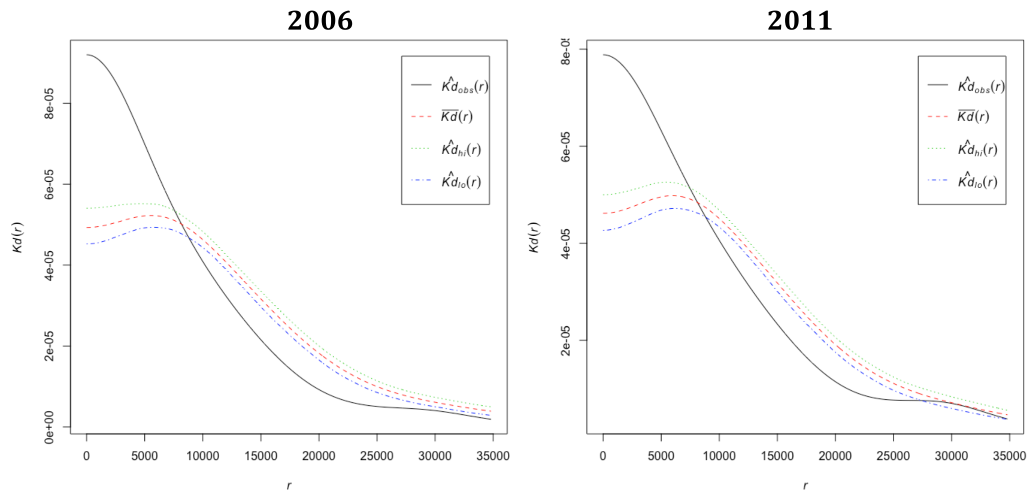
(6)

The higher the index, the higher is the deviation from randomness of the industry's spatial distribution. These measures allow us to rank sectors by degree of localization or dispersion.

We highlight some cases of sectorial agglomeration that illustrate the information generated by K_d measures. The graphics presented next depict the K_d value (vertical axis) for each distance in the horizontal axis. In each graphic, the black line is the estimated value; the green and blue dashed line are the upper and lower confidence band, respectively; while the red dashed line is an average of both bands.

The sector Health Care Activities, for instance, presents a consistent localization pattern: it has the 5th highest Γ_m for 2006 and the 9th highest for 2011 (Table 5). This can easily be seen in Figure 1, as the estimated K_d line stays well above the confidence interval from 0 to approximately 7km for both years, revealing that the localization phenomena of this sector happens at relatively short distances. From its map (Figure 2), one can also notice how localized this sector is and where the clusters occur. It's also noticeable that its firms agglomerate near the coastal area and mainly in the CBD of the capital of the state, Recife.

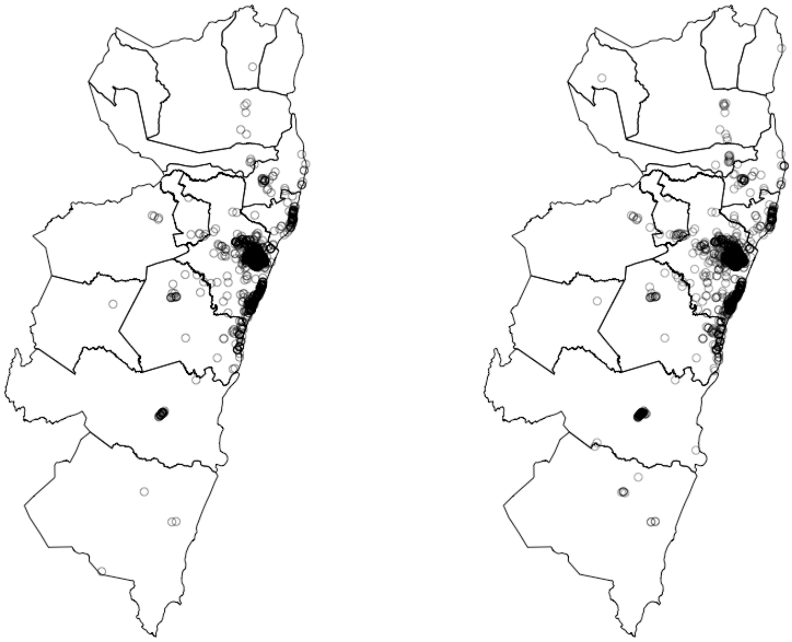
Figure 2 – K_d of Health Care Activities



Source: Elaborated by the author.

Figure 3 - Location of Health Care Activities

2006 **2011**

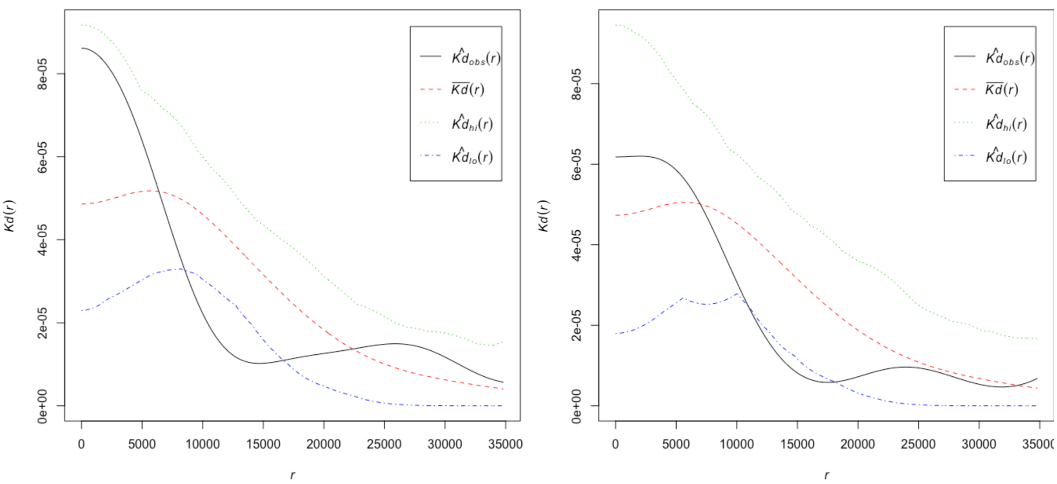


Source: Elaborated by the author.

A benefit of using distance-based measures of location is that it can equally identify cases of agglomeration and dispersion. The industry of Manufacture and Refining of Sugar, for instance, shows a persistent pattern of dispersion in both years of 2006 and 2011 (Figures 3 and 4).

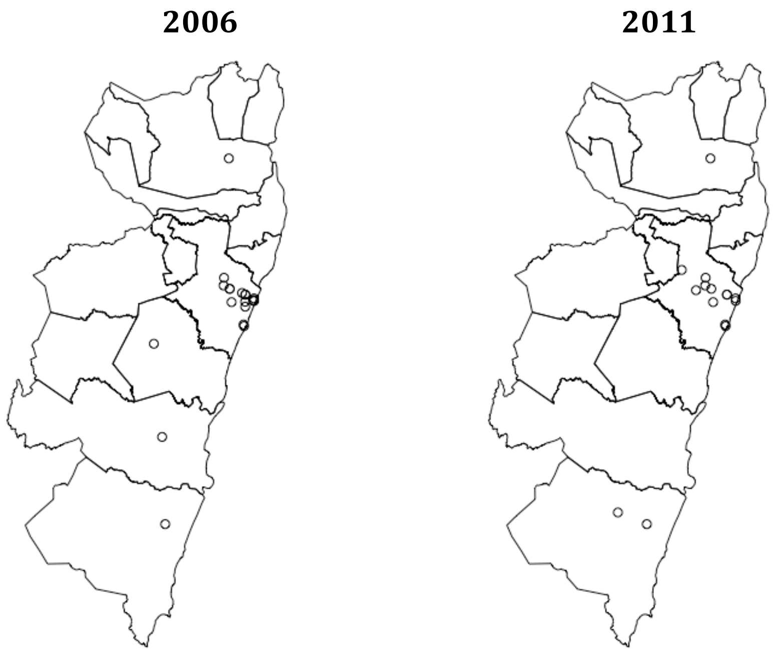
Figure 4 – K_d of Manufacture and Refining of Sugar

2006 **2011**



Source: Elaborated by the author.

Figure 5 – Location of Manufacture and Refining of Sugar



Source: Elaborated by the author.

4. RESULTS

4.1 General results

General results for the K_d estimations for year 2011 are presented in Table 3, both for non-weighted (Panel A) and weighted cases (Panel B)[7]. The first notable evidence is the localization amount: in the extended sample, 56% of the firms were globally localized (Panel A). Although the scales of analysis are different, this evidence is similar to what was obtained in previous studies. Duranton and Overman (2005) obtained a value of 52% for UK's manufactures, while Nakajima et al. (2012) obtained 50% for Japanese firms. On the other hand, we found that only 5% (2% for activities with 10 or more firms) of the activities is classified as dispersed ones.

When the sample is restricted to groups with at least 10 firms, the share of localization increases to 71%, suggesting that most groups with few firms are randomly distributed or dispersed. These numbers resemble those provided by Barlet et al. (2013) for the location analysis of France industries (63%), and by Koh and Riedel (2014) for the location patterns of manufactures in Germany (71%). Note that the results obtained when firms are weighted by workforce indicate higher share of dispersed and randomly localized firms, and smaller share of localized ones, indicating higher level of localization for firms than for employment.

To sum up, three main conclusions can be drawn from these results: 1) there are many more localized groups than dispersed ones in the RMR and this confirms the trend for location of firms in an urban environment; 2) industrial sectors with fewer firms appear more dispersed or less located; and 3) there isn't much difference in terms of share of localized and dispersed industries between years.

Table 3 – Summary statistics for K_d estimations – 2011

	Non-weighted			
Panel A	10 or more Firms		All firms	
	Number	**Percentage**	**Number**	**Percentage**
Localized	60	71%	68	56%
Random	23	27%	48	39%
Dispersed	2	2%	6	5%
Average Γ_m	0.00079		0.00074	
Average Ψ_m	0.00019		0.00012	

Continua na próxima página

[7] Due space restriction, we only present the estimative for the year of 2011. The values for the year of 2006 are similar and can be made available by the authors upon request.

Continuação da página anterior

Panel B	Weighted			
	10 or more Firms		All firms	
	Number	**Percentage**	**Number**	**Percentage**
Localized	34	40%	45	37%
Random	39	46%	63	52%
Dispersed	12	14%	14	11%
Average Γ_m	0.00088		0.00073	
Average Ψ_m	0.00005		0.00005	

Source: Elaborated by the author.

In Tables 4 and 5, we list the value of Γ_m and ψ_m indexes for the 30 industries (with at least 10 firms) that present highest location or dispersion levels for 2011 (62 of 85 sectors), together with number of firms (N), respectively, for Services and Manufacturing activities[8]. Two general evidences must be highlighted from the levels of location of industries. First, while representing 64.7% of all industries considered, the Services activities comprise 71.7% (43 out of 60) of industries presenting location. Thus, the activities present a stronger tendency to location. This is consistent with Koh and Riedel (2014), who also highlight a notable tendency of service industries towards localization in Germany, as 61% of them are concentrated. Furthermore, the levels of location in Services also appear higher than for Manufacturing. From the specific numbers of Tables 4 and 5, we perceive that around 62.8% of located Services industries presents higher values for Γ_m than the value of the index for the most located Manufacturing activity (*Other Food Products Manufactures*).

Industries' specific levels of location are also very informative. Among the more concentrated industries are activities related to informatics systems, development of computer programs, business advisory and support, credit grant, legal, and accounting – industries that recently Billings and Johnson (2016) also found to be co-agglomerated in Denver-Boulder-Greeley metropolitan area. We notice that, according to information from RAIS, these are activities with a highly-qualified workforce (the data show that 34% of the workers of this group have at least a college degree). Thus, the higher level of concentration of these services is consistent with benefits of technological spillovers and labor market pooling (ROSENTHAL AND STRANGE, 2001; HOLMES AND STEVENS, 2002; KOLKO, 2010).

Second, we also perceive a second distinct group of concentrated industries that correspond to retail industries of differentiated goods (group codes 521, 525, 524). Although only four 3-digit industries were highlighted, these comprehend a wide range of sub-activities.

[8] The list of the other activities together with their indexes is presented in the appendix.

Retail trade of other products in specialized shops (group 524), for example, contains pharmacy, perfumery and cosmetics, home and personal use devices, informatics, musical instruments, construction material, books and magazines, and others. The higher level of concentration of these activities appears also consistent with economic arguments suggested for the location of activities in urban areas. The products of these groups are similar to each other, but not exactly equal; thus, they can maintain some market power by differentiating product characteristics, quality and servicing arrangements[9].

Third, a few manufacturing industries also show concentration: Forging, stamping, powder metallurgy and metal treatment services (group 283) and Manufacture of miscellaneous chemical products and preparations (group 249). These sectors show an average peak of agglomeration at 30km, while the peak of agglomeration for industries in general happens at 11km. The evidence is consistent with idea that factories usually take great space, much more than service industries and stores. Tables 4 and 5 also show cases of dispersion. The dispersion of *Manufacture and Refining of Sugar* sector (group 156) is compatible with previous literature for Brazil (using the Ellison-Glaeser index). Lautert and Araújo (2007), for example, reported low levels of concentration for traditional non-durable consumer goods industries. In the case of RMR, the dispersion of this sector is also consistent with the historical pattern of production of sugar cane, present in most of the 14 municipalities. In the same way, the dispersion of *Trade, Maintenance and Repair of Motorcycles, Parts, Spare Parts and Accessories* (group 504) in 2011 is not entirely unexpected considering the studies of Camargo (2006), who found that auto industry activities in Brazil went through a deep process of dispersion from 1996 to 2001.

Table 4 – 30 most localized and dispersed groups – Three-digit – Services - 2011

Group	Γ_m	Localized	N
721	0.0033223	Consultancy in informatics Systems	64
911	0.0029977	Activities of Business, Patronal and Professional organizations	96
912	0.0029100	Activities of Union organizations	197
655	0.0028787	Other Credit Grant Activities	72
741	0.0024960	Legal, Accounting and Business Advisory Activities	1191
722	0.0022520	Development of Computer Programs	130
501	0.0022456	Trade in Retail and Wholesale of Automotive Vehicles	356
602	0.0021879	Other Land Transportation	666
851	0.0018081	Health Care Activities	1376

Continua na próxima página

[9] Fashion apparel and perfumery appear to be notable examples of this argument.

Continuação da página anterior

621	0.0017397	Air Transport, Regular	19
514	0.0016545	Wholesale Trade in Personal and Domestic Uses	741
704	0.0013920	Condominiums	4045
222	0.0012133	Printing and Related Service for Third Parties	252
659	0.0011541	Other Activities of Financial intermediation	42
521	0.0010391	Non-Specialized Retail Trade	1686
221	0.0010180	Edition; Editing and Printing	146
652	0.0009769	Monetary intermediation - Demand Deposits	286
525	0.0009470	Retail Trade of Used Articles, in Shops	38
742	0.0009332	Architectural and Engineering Services and Specialized Technical Advice	278
703	0.0008931	Real Estate Activities on Account of Third Parties	194
751	0.0008756	State Administration and Economic and Social Policy	131
723	0.0008580	Data processing	94
930	0.0008104	Personal services	884
801	0.0006568	Pre-primary and Elementary Education	917
702	0.0006528	Real estate management	65
701	0.0006123	Real Estate Lawyers	212
516	0.0005275	Wholesale Machines, Devices and Equipment for Agricultural, Commercial, office, industrial, Technical and Professional	263
283	0.0004816	Forging, Stamping, Powder Metallurgy and Metal Treatment Services	50
803	0.0003993	College Education	62
712	0.0003754	Rental of Other Transport Means	28
	Ψ_m	**Dispersed**	
504	0.0001580	Trade, Maintenance and Repair of Motorcycles, Parts, Spare Parts and Accessories	236

Source: RAIS/MTE with estimations by the author.

Table 5 – 30 most localized and dispersed groups – Three-digit - Manufacturing - 2011

Group	Γ_m	**Localized**	**N**
158	00005389	Other Food Products Manufactures	779
251	0.0004811	Rubber Products Manufacturing	30
247	0.0004763	Manufacture of Soaps, Detergents, Cleaning Products and Perfumery Products	73
281	0.0004363	Manufacture of Metal Structures and Heavy Boiler Works	135
252	0.0004335	Manufacture of Plastic Products	169
249	0.0004069	Manufacture of Chemical Products	28
296	0.0003177	Manufacture of Other Machinery and Equipment for Specific Use	38
263	0.0002247	Concrete, Cement, Fiber cement, Gypsum and Stock Manufacture	100
289	0.0000875	Manufacturing of Miscellaneous Metal Products	116
181	0.0000565	Clothing Manufacturing	437

Continua na próxima página

Continuação da página anterior

274	0.0000550	Metallurgy of Non-ferrous Metals	17
248	0.0000476	Manufacture of Paints, Varnishes, Enamels, Lacquers and Related Products	22
214	0.0000420	Manufacturing Crafts Miscellaneous Paper, cardboard and paperboard	45
401	0.0000414	Production and distribution of electricity	47
264	0.0000343	Manufacture of Ceramic Products	27
155	0.0000063	Grinding, Manufacture of Amylaceous and Balanced Rations for Animals	43
284	0.0000001	Manufacture of Cutlery, Locksmiths' and Hand Tools	72
	$\boldsymbol{\Psi_m}$	**Dispersed**	
156	0.0002313	Manufacture and Refining of Sugar	15

Source: RAIS/MTE with estimations by the author.

4.2 Characterizing the urban location

Based on section 2 discussions, we provide here exploratory and suggestive evidence about possible arguments explaining industries' location in RMR. Basically, using linear regressions, we present evidence about the association between locations of industries and the arguments we previously highlighted.

First, we expect agglomeration of more technological intensive or higher human capital activities within metropolitan areas (KEEBLE AND WILKINSON, 2000; ARBIA ET AL., 2009; ARBIA ET AL., 2010). According to Kolko (2010), for example, education may be related to concentration as workers and firms benefit from market labor pooling and technological spillovers. Thus, we classify the industries by the percentage of workers with at least a college degree and use this variable as a conditioning of urban location ($College$). Data limitation makes impossible to clearly to know the channels through which the variable affects location, but we use average employees age and percentage of workers involved with research and development (R&D) as controls for this effect. As argued by Koh and Riedel (2014), older or traditional activities may have different incentives to agglomerate than present-day more modern ones, that employ younger and more-educated people. In the same way, it is also possible that dispersed universities, which employ higher educated researchers and professors, affect the results.

Second, product differentiation makes price competition weaker and thus favors firms' localization. In order to capture this effect, we consider a measure associated with the productive structure differentiation of the firms within sectors. The measure corresponds to an indicator of variability of different kinds of occupation among the firms of an industry. Considering the industry *j,* we take the following value for it:

$$D_j = \frac{\sum_{i=1}^{N}\left[\left(x_{ji1}-\bar{x}_{j1}\right)^2+\left(x_{ji2}-\bar{x}_{j2}\right)^2+\cdots+\left(x_{jiM}-\bar{x}_{jM}\right)^2\right]/M_j}{N_j} \qquad (7)$$

where N_j is the number of firms of *j,* M_j is the number of kinds of occupation of the industry j, x_{ji1} corresponds to the share of occupation 1 in the total kinds occupation of firm i of industry j, x_{jiM} is the correspondent share for occupation *M*, and $\bar{x}_{j1}$ and $\bar{x}_{jM}$ are, respectively, average shares of occupations 1 and *M* of industry *j*. In Table A1 of the appendix, we list the industries ranked by the D_j index. Coherent with our expectations, the five activities presenting the lowest value of D_j include the activities *State Administration and Economic and Social Policy*, *Retail Trade of Combustive, Hotels and Other types of Accommodation,* and *Health Care Activities.*

We expect that greater industrial product differentiation is associated with bigger occupations variability or, in other words, that product homogeneity inside an industry is negatively associated with occupational variability among the firms. In spite of these expectations, it is interesting to recognize that the effect may also depend on the level of the differentiation itself. In fact, the economic argument for agglomeration relies on some homogeneity of firms' products, but, in this context, product differentiation relaxes price competition and thus favors firms' localization. The influence of increasing differentiation, however, may even be null (or a negative one) on the location of firms of activities presenting higher product differentiation because there was no initial price competition or gains for the consumers from getting information about products. Thus, there would be a non-linear relationship between localization and product differentiation and we test for this possibility by considering also a quadratic term in the specifications of the regressions.

Third, the location of services and input providers can importantly affect the location decision of firms and favor agglomeration (BILLINGS AND JOHNSON, 2016). As the proximity of input services can favor location, more vertically integrated firms tend to show weaker interest in agglomeration. We could not obtain information about input-output linkages, so we use a proxy of vertical integration of the industry by measuring the ratio of number of different occupations per firm ($Oc./firm$). Arguably, because they perform more tasks, more vertically integrated firms tend to present higher number of different occupations. Because bigger firms also tend to present higher number of different kinds of occupations, we also use the ratio number of workers per firm as control.

Finally, following Koh and Riedel (2014), we also consider the percentage of workers in an industry who are engaged in manual activities ($Man.$) A higher percentage is expected to be present in industries with high transportation costs, which is a decisive factor for the location choice of a firm. While this factor should be stronger in regional location analysis, we also investigate its presence in our urban context.

To sum up and formally, we consider the following specification:

$$\Gamma_j = \alpha + \beta_1 College_j + \beta_2 D_j + \beta_3 D_j^2 + \beta_4 (Oc./firm)_j + \beta_5 Man_j + X_j\gamma + \varepsilon_j \qquad (8)$$

Where Γ_j is the location index for industry j, X_j is the set of controls variables above discussed, Greek letters are parameters, and ε_j is an error term.

Results are presented in Table 6, both for all activities (columns (I) and (II)) and for Services activities only (Columns (III) and (IV))[10]. Besides, columns (I) and (III) present results without controls variables and Columns (II) and (IV) consider additional variables; note, however, that including additional variables scarcely changes the estimative.

Independently of the particular specification, the results indicate positive association between the human capital variable (percentage of workers with a college degree) and the localization index and a negative association between the number of different kinds of occupation per firm and localization. On the other hand, we get no evidence of association of the percentage of workers engaged in manual activities and agglomeration of firms. The first result is similar to one obtained by Koh and Riedel (2014) for the distribution of activities in German using the same index and is also consistent with the importance of localization for activities using more human capital. Similarly, the negative association between number of different kinds of occupations per firm and the localization index appears to be consistent with the weaker relevance of localization in sectors with more vertically integrate firms (and less dependent on input services) as suggested by the results of Billings and Johnson (2016). As for control variables, both columns (II) and (IV) indicate a negative association between the percentage of workers involved with research and development occupations and firms' localization, suggesting that these activities are spread across the urban center.

We also note that, while for the sample of all activities we got no evidence of association between product differentiation and location of the firms, when considering only the Services activities (columns (III) and (IV) of Table 5), we obtained a non-linear relationship. This difference is not unexpected, once the need of face-to-face contact implies

---

[10] Note that, because the value of the dependent variable involves four decimal cases, we present coefficients estimative using five decimal cases or even scientific notation.

stronger importance of consumer mobility costs for Services and tends to make inner-city locational competition more important (ARZAGHI AND HENDERSON, 2008; JOFRE-MONSENY ET AL., 2014). Note that the positive (Differentiation) and negative (Differentiation^2) estimated coefficients suggest that the degree of concentration of firms first increases and then decreases with the level of product differentiation of the industry. More specifically, from the values of the coefficients of column (IV) of Table, we note that the increasing of the industrial localization happens until a value of 0.01133 (=0.11222/9.90216) of the differentiation index, which, in our sample of Services activities, includes 94.4% of them.

Although the set of evidence is far from guaranteeing any causal relationship, we believe that our exploratory results are consistent with economic arguments for understanding firms' location within Recife Metropolitan Region.

Table 6 – Conditionings of within urban location – OLS Linear Regression estimative. Dependent variable is the location index Γ_m.

	All Activities		Services	
	(I)	(II)	(III)	(IV)
College	0.00001**	0.00002**	0.00002**	0.00002**
	(6.83e-06)	(8.10e-06)	(8.32e-06)	(9.90e-06)
Differentiation	0.06251	0.05647	0.107341*	0.11222*
	(0.05307)	(0.05389)	(0.06057)	(0.06438)
Differentiation^2	-2.97453	-3.00885	-4.70982**	-4.95108**
	(1.79590)	(1.80704)	(2.09011)	(2.21941)
Occupations/firm	-0.00012***	-0.00012***	-0.00024***	-0.00029***
	(0.00004)	(0.00004)	(0.00006)	(0.00008)
Manual tasks	7.54e-07	8.79e-07	7.83e-06	8.72e-06
	(3.82e-06)	(3.81e-06)	(7.06e-06)	(7.05e-06)
Age		0.00039		0.00027
		(0.00043)		(0.00051)
Eng. and Scient.		-0.00006**		-0.00006**
		(0.00003)		(0.00003)
Workers/firm		-6.22e-08		1.91e-07
		(2.40e-07)		(2.69e-07)
Constant	0.00047**	-0.00142	0.00037	-0.00096
	(0.00023)	(0.00162)	(0.00027)	(0.00241)
F statistics	3.67***	3.10***	3.75***	2.26**
R^2	0.2020	0.2368	0.2159	0.2548
R^2adjus.	0.11405	0.1564	0.1343	0.1223
Observations	85	85	55	55

Obs.: Heteroscedastic robust error in parenthesis. "***", "**", and "*" indicate statistical significance, respectively, at 1%, 5%, and 10% levels. "College" indicates percentage of workers of the industry with a college degree; "Differentiation" corresponds to the index of equation (7); "Occupations/firm" is the industry occupations/firm ratio; "Manual tasks" indicates the percentage of workers in an industry who are engaged in manual activities; "Age" corresponds to the industry workers average age; "Eng. and Scient." indicates percentage of workers of the industry in engineering and science occupations; "Workers/firm" corresponds to the industry workers/firms ratio.

4.3 Evidence from traditional measure and the MAUP

In order to compare our results with those obtained using a more traditional concentration index, we generate new evidence of industries' localization using the EG index (ELLISON AND GLEASER, 1997). Note that a significant advantage of a distance-based measure is to deal with the traditional MAUP. Thus, we now show how severe this problem can be in an urban context.

The Ellison-Glaeser index is a composition of two metrics of concentration: G, that measures raw geographic concentration, and the Herfindahl index, that captures the plant size distribution (HERFINDAHL, 1950). The combination of both forms an index capable of measuring spatial concentration while considering each industry's economics of scale. The E-G index is defined as follows:

$$\gamma_{EG} = \frac{G-(1-\sum_i x_i^2)H}{(1-\sum_i x_i^2)(1-H)} \qquad (9)$$

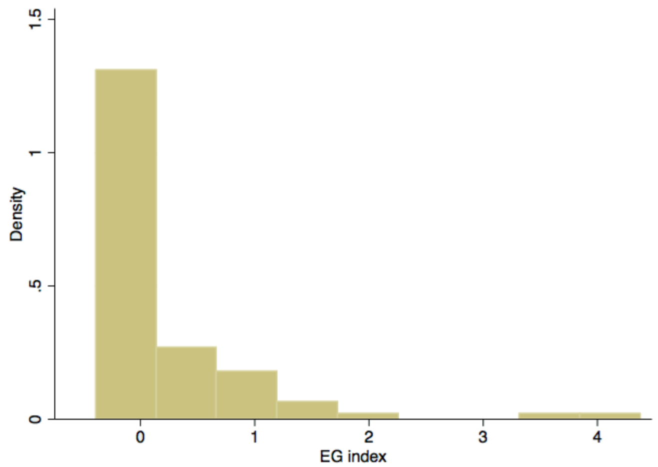
in which x_i is the share of location i in the overall employment. G and H are defined as:

$$G = \sum_i (s_i - x_i)^2 \qquad H = \sum_j z_j^2 \qquad (10)$$

whereas s_i is location i's share of employment within its industry, and z_j is firm j's share of employment within its industry. The industries can be classified into three levels of agglomeration based on how much the values depart from zero. $\gamma_{EG} < 0.02$ indicates weakly concentration; $0.02 < \gamma_{EG} < 0.05$ indicates intermediate and $\gamma_{EG} > 0.05$ expresses strong concentration.

The EG index was calculated for all 85 groups (3-digit sectors) initially using the 14 municipalities of the RMR as spatial units. The γ_{EG} values were averaged so that we can assess the general concentration. Figure 5 shows the distributions of the EG index for the set of industries. Notice that the distributions are skewed; there is a gap in higher levels with no industries and then a few sectors appear with very high values (around value 4). The mean value of γ_{EG} was 0.283 in 2011, which is classified as 'strong concentration'.

Figure 6 – Distribution of the EG index



The value of EG index, together with its components H (Herfindahl index), and G (the mean raw concentration), are presented in Table 6 for municipalities. According to the numbers, in 2011, 69% of the industries appear as concentrated when the EG methodology is applied. In the range of very concentrated sectors (γ_{EG}>0.05), we obtained 44% of the industries. In contrast, estimations of the K_{emp}[11] ('weighted' values in Table 1) show concentration levels varying from 33% to 40%, which confirms the idea that "the EG approach is less precise in identifying agglomeration patterns (as it is not based on a statistical test for deviations from randomness)", as Koh and Riedel (2014) pointed out.

Similar to Alecke et al. (2006) and Koh and Riedel (2014), we calculate the index for three different spatial aggregation levels to observe how the EG is affected when spatial unit varies. Besides municipalities (14 units), we ran the analysis for sub-districts (37 units) and micro-regions (3 units). These additional evidences are also presented in Table 6, along with the percentages of industries that lies in different classification intervals.

The results confirmed that the index is sensitive to the spatial unit choice. The mean value of γ_{EG} is positive for 84% of the industries in 2011 when we consider sub-districts. On the other hand, when a more aggregated spatial unit is chosen (micro-regions), the concentration drops to 51% (in the same year). Note that the percentage of strongly concentrated industries also drops when the aggregation level increases. This unusual pattern probably is associated with firms' dimension and the intangible nature of their outcome. In an intra-urban context, firms are generally small and a lot of them are service-related needing face-to-face contact. Their clusters, thus, tend to occur inside small spatial units (ARZAGHI

[11]As the E-G weights firms by number of employees, it is only comparable to the K_{emp} (which permits weighting) and not with the K_d.

AND HENDERSON, 2008). As you move to more disaggregated units, the comparison changes to a few small areas with very high concentration and many small areas with weak or zero presence of firms. Consequently, the γ_{EG} will be likely higher towards smaller geographic units. The index clear sensitivity to geographic units makes the EG methodology imprecise for concentration analysis also in an urban context.

Table 7 - Summary statistics for EG estimations – 2011

	H	**G**	**EG**
Sub districts	0,128	0,190	0,111
Municipalities	0,128	0,171	0,283
Micro regions	0,128	0,088	0,743
	EG > 0	**0 ≤ EG ≤ 0,05**	**EG > 0,05**
Sub districts	84%	26%	58%
Municipalities	69%	26%	44%
Micro regions	51%	25%	26%

Source: Elaborated by the author.

5. CONCLUSIONS

This paper analyzed the location of the economic activities of a Brazilian metropolitan region, the RMR, using the distance-based measure of localization proposed by Duranton and Overman (2005). Using a distance-based measure for studying firm location is still scarce and practically unexplored in intra-urban developing country context. Thus, two general contributions are provided. First, it applies the measure of localization for studying the location of firms in an important developing country and, second, it examines the localization pattern for both manufacturing and services industries in an intra-urban context.

We found that, of the 20 most agglomerated sectors, 15 and 16 (in 2006 and 2011, respectively) are services. We highlight three general evidences: 1) there are many more localized groups than dispersed ones in the RMR and this confirms the trend for the location of firms in an urban environment; 2) industrial sectors with fewer firms appear more disperse or less localized; and 3) there isn't much difference in terms of share of localized and dispersed industries between years. Endorsing previous research, we find that computing, law, accounting, architecture and engineering industries show a tendency towards higher levels of localization.

We also found that a non-distance-based measure of localization such as the traditional E-G (ELLISON AND GLEASER, 1997) overestimates localization levels and clearly suffers from the MAUP. As the K_d is distance-based and can be tested statistically, it does not require spatial aggregation and provide more accurate results.

Finally, we provide exploratory and suggestive evidence about the association between different economic arguments traditionally used for explaining localization and the value of the localization index across industries. Consistent with economic interpretation, this additional evidence indicates that localization levels are positively associated with higher human capital and negatively associated with a proxy of vertical integration of firms. Furthermore, specifically for Services activities, our results are consistent with non-linear effects of product differentiation on firm localization (localization positively associated with product differentiation for lower levels of differentiation and negatively for higher ones).

The research could be expanded in at least two directions. Once available, the results could incorporate informal sector firms, generally present in developing counties. Even being a minor problem for big metropolitan regions, such as Recife, this incorporation would add generality to the analysis. Second, the identification of any causal relationship between economic arguments and firm localization imposes the use of appropriated identification strategies not yet applied in this research.

REFERENCES

ALECKE, Björn et al. Are there really high-tech clusters? The geographic concentration of German manufacturing industries and its determinants. **The Annals of Regional Science**, v. 40, n. 1, p. 19, 2006.

ARBIA, Giuseppe; ESPA, Giuseppe. **Statistica economica territoriale**. Cedam, 1996.

ARBIA, Giuseppe et al. Detecting the existence of space–time clustering of firms. **Regional Science and Urban Economics**, v. 40, n. 5, p. 311-323, 2010.

BARLET, Muriel; BRIANT, Anthony; CRUSSON, Laure. Location patterns of service industries in France: A distance-based approach. **Regional Science and Urban Economics**, v. 43, n. 2, p. 338-351, 2013.

BEHRENS, Kristian; BOUGNA, Théophile. An anatomy of the geographical concentration of Canadian manufacturing industries. **Regional Science and Urban Economics**, v. 51, p. 47-69, 2015.

BILLINGS, Stephen B.; JOHNSON, Erik B. Agglomeration within an urban area. **Journal of Urban Economics**, v. 91, p. 13-25, 2016.

CAMARGO, Otávio Silva. As mudanças na organização e localização da indústria automobilística brasileira (1996-2001). 2006.

CENSO, I. B. G. E. Instituto Brasileiro de Geografia e Estatística. Fornecido em meio eletrônico: [www.ibge.gov.br/home/estatistica/populacao/censo2010/], v. 20, n. 03, 2010.
CICCONE, A. HALL, R. E. (1993). Productivity and the density of economic activity. National Bureau of Economic Research.

CLARK, Gordon L. London in the European financial services industry: locational advantage and product complementarities. **Journal of Economic Geography**, v. 2, n. 4, p. 433-453, 2002.

COMBES, Pierre-Philippe et al. Estimating agglomeration economies with history, geology, and worker effects. In: **Agglomeration economics**. University of Chicago Press, 2010. p. 15-66.

D'ASPREMONT, Claude; GABSZEWICZ, J. Jaskold; THISSE, J.-F. On Hotelling's" Stability in competition". **Econometrica: Journal of the Econometric Society**, p. 1145-1150, 1979.

DURANTON, Gilles; OVERMAN, Henry G. Testing for localization using micro-geographic data. **The Review of Economic Studies**, v. 72, n. 4, p. 1077-1106, 2005.

DURANTON, Gilles; PUGA, Diego. Micro-foundations of urban agglomeration economies. In: **Handbook of regional and urban economics**. Elsevier, 2004. p. 2063-2117.

EATON, B. Curtis; LIPSEY, Richard G. Comparison shopping and the clustering of homogeneous firms. **Journal of Regional Science**, v. 19, n. 4, p. 421-435, 1979.

ELLISON, Glenn; GLAESER, Edward L. Geographic concentration in US manufacturing industries: a dartboard approach. **Journal of political economy**, v. 105, n. 5, p. 889-927, 1997.

FERNÁNDEZ-MALDONADO, Ana María et al. Polycentric structures in Latin American metropolitan areas: Identifying employment sub-centres. **Regional Studies**, v. 48, n. 12, p. 1954-1971, 2014.

FU, Shihe et al. Smart café cities: Testing human capital externalities in the Boston metropolitan area. **Journal of Urban Economics**, v. 61, n. 1, p. 86-111, 2007.

FUJITA, Masahisa; THISSE, Jacques-François. **Economics of agglomeration: cities, industrial location, and globalization**. Cambridge university press, 2013.

GLAESER, Edward L.; MARE, David C. Cities and skills. **Journal of labor economics**, v. 19, n. 2, p. 316-342, 2001.

GLAESER, Edward L.; GOTTLIEB, Joshua D. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. **Journal of Economic Literature**, v. 47, n. 4, p. 983-1028, 2009.

GONG, Hongmian; WHEELER, James O. The location and suburbanization of business and professional services in the Atlanta area. **Growth and Change**, v. 33, n. 3, p. 341-369, 2002.

HERFINDAHL, Orris C. **Concentration in the steel industry**. 1950. Columbia University.

HEUERMANN, Daniel; HALFDANARSON, Benedikt; SUEDEKUM, Jens. Human capital externalities and the urban wage premium: Two literatures and their interrelations. **Urban Studies**, v. 47, n. 4, p. 749-767, 2010.

HEUERMANN, Daniel. Human capital externalities in Western Germany. **Spatial Economic Analysis**, v. 6, n. 2, p. 139-165, 2011.

HOLMES, Thomas J.; STEVENS, John J. Geographic concentration and establishment scale. **Review of Economics and Statistics**, v. 84, n. 4, p. 682-690, 2002.

HOTELLING, Harold. Stability in competition. In: **The Collected Economics Articles of Harold Hotelling**. Springer, New York, NY, 1990. p. 50-63.

IPHAN. História - Recife (PE). Available at <http://portal.iphan.gov.br/pagina/detalhes/1449/>. Accessed on May of 2017.

LECAILLON, Jacques. Marshall (Alfred)-Principles of economics. **Revue économique**, v. 14, n. 14, p. 318-318, 1963.

KEEBLE, David; WILKINSON, Frank. **High-technology clusters, networking and collective learning in Europe**. Routledge, 2017.

HYUN-JU, K.; RIEDEL, Nadine. Assessing the Localization Pattern of German Manufacturing and Service Industries. **Regional Studies**, v. 48, n. 5, 2010.

KOLKO, Jed. Urbanization, agglomeration, and coagglomeration of service industries. In: **Agglomeration economics**. University of Chicago Press, 2010. p. 151-180.

LANG, Gabriel; MARCON, Eric; PUECH, Florence. Distance-based measures of spatial concentration: introducing a relative density function. 2016.

LAUTERT, Vladimir; ARAÚJO, Nilton Clóvis Machado de. Concentração industrial no Brasil no período 1996-2001: uma análise por meio do índice de Ellison e Glaeser (1994). **Economia Aplicada**, v. 11, n. 3, p. 347-368, 2007.

KLIER, Thomas; MCMILLEN, Daniel P. Evolving agglomeration in the US auto supplier industry. **Journal of Regional Science**, v. 48, n. 1, p. 245-267, 2008.

KONISHI, Hideo. Concentration of competing retail stores. **Journal of Urban Economics**, v. 58, n. 3, p. 488-512, 2005.

MARCON, Eric; PUECH, Florence. A typology of distance-based measures of spatial concentration. **Regional Science and Urban Economics**, v. 62, p. 56-67, 2017.

MARCON, Eric; PUECH, Florence. Measures of the geographic concentration of industries: improving distance-based methods. **Journal of Economic Geography**, v. 10, n. 5, p. 745-762, 2009.

MORETTI, Enrico. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. **Journal of econometrics**, v. 121, n. 1-2, p. 175-212, 2004.

MULLIGAN, Gordon F.; FIK, Timothy J. Price and Location Conjectures in Medium-and Long-Run Spatial Competition Models. **Journal of Regional Science**, v. 34, n. 2, p. 179-198, 1994.

NAKAJIMA, Kentaro; SAITO, Yukiko Umeno; UESUGI, Iichiro. Measuring economic localization: Evidence from Japanese firm-level data. **Journal of the Japanese and International Economies**, v. 26, n. 2, p. 201-220, 2012.

HUALLACHÁIN, Breandán Ó.; REID, Neil. The intrametropolitan location of services in the United States. **Urban Geography**, v. 13, n. 4, p. 334-354, 1992.

RESENDE, Marcelo; WYLLIE, Ricardo. Aglomeração industrial no Brasil: um estudo empírico. **Estudos Econômicos (São Paulo)**, v. 35, n. 3, p. 433-460, 2005.

ROSENTHAL, Stuart S.; STRANGE, William C. The determinants of agglomeration. **Journal of urban economics**, v. 50, n. 2, p. 191-229, 2001.

ROSENTHAL, Stuart S.; STRANGE, William C. Evidence on the nature and sources of agglomeration economies. In: **Handbook of regional and urban economics**. Elsevier, 2004. p. 2119-2171.

SHEARMUR, Richard; ALVERGNE, Christel. Intrametropolitan patterns of high-order business service location: A comparative study of seventeen sectors in Ile-de-France. **Urban**

Studies, v. 39, n. 7, p. 1143-1163, 2002.

SILVERMAN, Bernard W. Monographs on Statistics and Applied Probability. In: **Density estimation for statistics and data analysis**. Chapman & Hall, 1986.

STUART, Charles. Search and the spatial organization of trading. **Studies in the Economics of Search**, p. 17-33, 1979.

VIGNANDI, Rafaella Stradiotto; PARRÉ, José Luiz; DE FREITAS GUIMARÃES, Paulo. Measures of Industry Agglomeration in Brazil: a Study Addressing Neighboring Effects. **Análise Econômica**, v. 34, n. 65.