Localization Patterns of Service Activities within Metropolitan Areas: the Case of São Paulo

Edilberto Tiago de Almeida* Raul da Mota Silveira Neto†
Jaime Macedo de Brito Bastos‡ Rubens Lopes Pereira da Silva§

Abstract

Previous evidence shows that productive activity is not distributed randomly in space. In this paper we present a set of evidences about the service location patterns in the largest and most important metropolitan region of Brazil, the São Paulo Metropolitan Region (SPMR). Our results are obtained from a measure based on distances and, therefore, not susceptible to MAUP. We found that 90% of service sectors have significant location patterns. In addition, we also provide evidence on local factors that are associated with observed locational patterns. We found that the greater concentration of human capital is related to higher levels of location.

Keywords: cities, distance-based measure, determinants of geographic concentration, services

Resumo

Evidências anteriores mostram que a atividade produtiva não é distribuída de forma aleatória no espaço. Neste artigo apresentamos um conjunto de evidências sobre os padrões de localização dos serviços na maior e mais importante região metropolitana do Brasil, a Região Metropolitana de São Paulo. Nossos resultados são obtidos a partir de uma medida baseada em distâncias e, portanto, não suscetíveis ao MAUP. Encontramos que 90% dos setores de serviços apresentam padrões de localização significativos. Adicionalmente, são apresentadas evidências sobre os fatores locais que estão associados aos padrões de localização observados. O resultados indicam que a maior concentração de capital humano está relacionada a níveis mais elevados de localização.

Palavras-chave: cidades, medida baseada em distâncias, determinates da concentração geográfica, serviços

JEL classification: R12

Área 10 - Economia Regional e Urbana

*Ph.D candidate, Department of Economics - PIMES, Federal University of Pernambuco - UFPE. E-mail: edilbertotiago@hotmail.com. The authors thank CNPq for their support.

†Associate Professor at the Department of Economics, Federal University of Pernambuco - UFPE.

‡Undergraduate in Economics, Department of Economics, Federal University of Pernambuco - UFPE.

§Ph.D candidate, Department of Economics - PIMES, Federal University of Pernambuco - UFPE.

1 Introduction

After two decades of intense empirical investigations, there is a consensus about the existence of a positive relationship between the spatial agglomeration of activities and productivity (Combes et al., 2012; Behrens et al., 2014; Accetturo et al., 2018; Gaubert, 2018). In fact, given obvious cost of mobility and housing, for example, the own existence of very large urban agglomerations is difficult to conciliate with the presence of rational decision agents without some some benefits from the spatial concentration of activities and people (Puga, 2010; Duranton and Puga, 2014; Thisse, 2018). Two associated branches of the empirical investigation in spatial economics still appear non-consensual, however.

First, although the recent contribution by Duranton and Puga (2004) in formalizing the micro-foundations for urban gains (via sharing, matching, and learning), the debate about the particular empirical relevance of each of these channels is still inconclusive (Krugman, 1991; Andersson et al., 2007; Di, 2011; Wheeler, 2008; Overman and Puga, 2010; Abel and Deitz, 2015; Lychagin et al., 2016). Second, apart of some evidence about particular sectors (Leslie and HUallacháin, 2006; Arzaghi and Henderson, 2008; Klier and McMillen, 2008), not so many advances have been generated in order to understand the spatial distribution of activities within urban centers (Kolko, 2010; Billings and Johnson, 2016; Silva et al., 2019).

Given the more continuity of the urban environments, this lack of rigorous information about the spatial distribution of activities within urban centers was somehow justified, since the use of traditional measures of concentration proposed, for example, by Ellison and Glaeser (1997) are susceptible to Modifiable Area Unit Problem (MAUP) and, thus, can not generate reliable information about the distribution of economic activities. More recently, however, treating the spatial environment as a continuum of points, Duranton and Overman (2005) and Marcon and Puech (2009) proposed measures of spatial location of activities based on distance between agents (firms or workers), the so-called distance-based measures.

Researches that uses continuous measures to document the concentration of productive activity, however, are still scarce, since they demand geocoded information about the agents. But the literature is consistently growing. Duranton and Overman (2005) themselves report levels of concentration of manufacturing activities in the UK, Nakajima et al. (2012) and Inoue et al. (2019) provide for Japan, Vitali et al. (2013) for a set of European countries, Behrens and Bougna (2015) for Canada, and Aleksandrova et al. (2019) for Russian case. Note that most of these works focus on analyzing manufacturing location patterns. Considering service and manufacturing sectors in Germany, the main results of Koh and Riedel (2014)'s work show that 71% of manufacturing and 97% of services have a significant geographic concentration. For France, Barlet et al. (2008) detected that 94% of French service sectors as localized. These findings suggest that services may be more concentrated than manufacturing, especially in large cities (Kolko, 2010).

For Brazil, evidence on the spatial concentration is generally based on measures that are sensitive to the MAUP (Silveira Neto, 2005; Resende and Wyllie, 2005; Lautert and Araújo, 2007; Almeida and Rocha, 2018; Rocha et al., 2019), being the very recent work of Silva et al. (2019) for Recife Metropolitan Region the only exception. The objective of this paper is, thus, contribute to this emergent literature by providing evidence about the patterns of spatial location of services activities in São Paulo Metropolitan Region (SPMR), the largest urban agglomerated of Brazil. In this task, we use geocoded data and the non-parametric method developed by Duranton and Overman (2005) to obtain distance-based measures of location patterns of economic activities. Note that by dealing only with services activities, we produce more reliable counterfactuals, since the statistic significance is obtained by simulations of potential locations. But our research present another advantage compared to the investigation by Silva et al. (2019): the focus on the Brazilian biggest metropolis allows considering a higher number of services activities, since some activities are not present in Recife Metropolitan Region, and thus obtaining a more complete set of information about location patterns of activities.

Our results indicate that 90% of the service sectors studied have significant spatial localization patterns. These results are in line with evidence found for other countries with continuous measures (Koh and Riedel, 2014) and also for Brazil with more limited measures (Domingues et al., 2006; Maciente, 2013). We also find that the levels of location of activities are positively associated with human capital, the degree of product differentiation, and with inter-sector dependence.

The paper is organized as follows. In the next section, we present the characteristics of our area of study and the reasons that make the SPMR representative in Brazil and also in the world. In the third section, the empirical strategy is presented. The last two sections present the results and discussions and the final comments, respectively.

2 São Paulo Metropolitan Region Context

With more than 21 million inhabitants in 2018, according to the Brazilian Institute of Geography and Statistics (IBGE), and constituted by 39 municipalities, the São Paulo Metropolitan Region (SPMR) is the largest urban complex in Brazil and one of the largest in the world, ranking in sixth position among the largest worldwide metropolis, according to a United Nations report of 2014[1]. Located in the most populous and rich state of the country, by agglomerating around 10% of Brazilian population in only 0.1% of the country territory (see Figure 1), the region is an unique Brazilian megalopolis.

Being a genuine reflex of the Brazilian historical pattern of strong spatially concentrated development (Baer, 2002), it is difficult to overstate the importance of the such great urban agglomeration for the Brazilian social and economic realities. Actually, in addition to its very high population density, around 2.7 thousand inhabitants per km^2, the region is recognized largely by its economic relevance for the country. According to data from the Sistema Estadual de Análise de Dados (SEADE), for example, its GDP of 2010 year (around R$ 701.85 billion) was equivalent to about 56% of the state of são Paulo GDP and 20% of Brazilian GDP[2]. In consonance, the SPMR is also responsible for paying a quarter of the taxes in the country and presents the most complex and diversified productive structure of the country. The economic primacy is also accompanied by other distinctive features. At least for Brazilian patterns, the SPMR presents a high Human Development Index HDI (0.794), it is home of important universities (such as the University of São Paulo - USP), of São Paulo-Guarulhos International Airport (the second largest in Latin America), of the most modern medical center in the country, and of the official Brazilian Stock Exchange (Bovespa).

Actually, the metropolitan region is part of the Macrometropole Paulista - MMP, a Brazilian expanded metropolitan complex that comprises several metropolitan regions of the state of São Paulo: Campinas, Baixada Santista and Vale do Paraíba and Litoral Norte, three Urban Agglomerations - Jundiaí, Sorocaba and Piracicaba and two Micro-Regions: São Roque and Bragantina. According

[1] According to IBGE, the Metropolitan Regions and Urban Agglomerations are constituted by groups of neighboring municipalities and are established by a complementary state law, according to the determination of article 25, paragraph 3 of the Federal Constitution of 1988, aiming to integrate the organization, planning and execution of public functions of common interest. The list provided by IBGE also includes categories associated with these cuts: metropolitan collars, metropolitan arches, metropolitan expansion areas, metropolitan subdivisions, among others. More details available at: https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial

[2] More details available at: http://www.seade.gov.br/

to data from Companhia Paulista de Planejamento Metropolitano SA[3] (EMPLASA), the SPMR is the main region of the MMP. Most of the cities that form this large urban cluster are intensely integrated and articulated, concentrating a series of diversified activities. Their performance have a direct impact on the Brazilian economy. These characteristics make the SPMR territory an area of strategic importance with potential to leverage state and national competitiveness, also influencing the performance of Brazil in the international context.

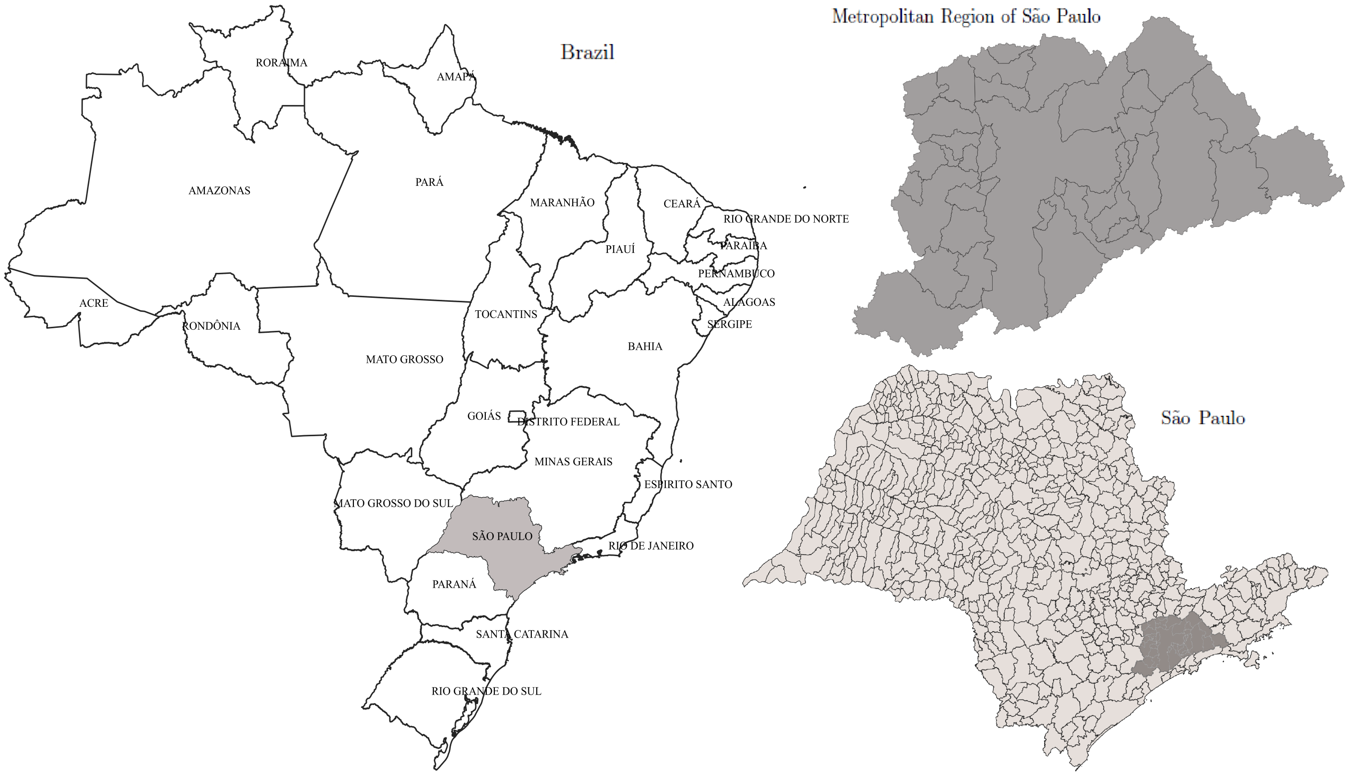


Figure 1. São Paulo Metropolitan Region (SPMR)
Source: Elaborated by the authors.

São Paulo, the main city of the SPMR and capital of the state of são São Paulo, is characterized by being the main political decision-making center of the state. It has a diversified and specialized[4] service center with emphasis on telecommunications, culture, education, health and transportation. The region is a tourism and business center in Latin America, with around 74 thousand events per year, attracting 4.2 million people (Costa, 2013). It is also a management center, hosting headquarters of large transnational companies; an industrial complex, and a financial center.

Table (1) presents descriptive statistics on population, employment, number of firms, CBD distance and diversity of economic activities by municipalities. The set of information allows a brief but useful characterization of important aspects of the economic geography of the SPMR. From the numbers of the refereed table, we notice that, although it is responsible for 56.40% of the population, the city of São Paulo concentrates 71% and 69.52% of the total jobs and firms in the SPMR, respec-

[3]More details available at: https://www.emplasa.sp.gov.br

[4]Previous evidence for Brazil shows that the externalities generated by locally diversified and specialized economic environments are important to explain the spatial configuration of productive activity (see Silva and Silveira Neto (2007, 2009); Fochezatto and Valentini (2010); Rocha et al. (2013) and Almeida et al. (2017)).

tively. In addition, it stands out as the most diversified city (measured by the number of different economic activities). Note also that the further away the municipality is from the center (measured by the CBD of the São Paulo municipality), both the less economic diversified and populous tend to be the municipality. These characteristics are consistent with greater agglomeration gains closer to the municipality of São Paulo.

Table 1. Characteristics of the São Paulo Metropolitan Region in 2015

Cities	Population		Employment		Firms		Distance	Employ. per	Number of
	Total	%	Total	%	Total	%	CBD (Km)	Firms	sectors
São Paulo	12,176,866	56.40	3,611,638	71	256,655	69.52	0	14.07	82
Guarulhos	1,365,899	6.30	215,408	4.20	16,737	4.53	16	12.87	77
São Bernardo do Campo	833,240	3.90	162,659	3.20	13,197	3.57	19	12.32	75
Santo André	716,109	3.30	163,678	3.22	12,677	3.43	24	12.91	74
Osasco	696,850	3.20	112,747	2.20	8,929	2.42	22	12.63	73
Barueri	271,306	1.30	202,019	4.00	7,177	1.94	30	28.14	79
Mogi das Cruzes	440,769	2.00	70,544	1.40	6,285	1,70	57	11.22	74
São Caetano do Sul	160,275	0.70	80,735	1.60	4,889	1.32	14	16.51	71
Diadema	420,924	2.00	43,814	0.86	4,348	1.18	21	10.07	69
Cotia	244,694	1.10	50,139	1.00	3,889	1.05	31	12.89	70
Mauá	468,148	2.20	37,064	0.73	3,680	1.00	27	10.07	70
Suzano	294,638	1.40	30,693	0.60	3,554	0.96	44	8.63	67
Carapicuíba	398,611	1.80	26,035	0.51	2,889	0.78	26	9.01	68
Taboão da Serra	285,570	1.30	41,087	0.80	2,849	0.77	30	14.42	71
Santana do Parnaíba	136,517	0.60	37,745	0.74	2,373	0.64	40	15.9	70
Itaquaquecetuba	366,519	1.70	22,123	0.43	2,180	0.59	36	10.15	63
Embu das Artes	270,843	1.30	27,284	0.54	2,061	0.56	27	13.23	65
Ribeirão Pires	122,607	0.60	11,350	0.22	1,503	0.41	55	7.55	63
Itapevi	234,352	1.10	21,466	0.42	1,480	0.40	40	14.5	59
Itapecerica da Serra	173,672	0.80	15,213	0.30	1,370	0.37	34	11.1	62
Poá	116,530	0.50	19,455	0.38	1,341	0.36	42	14.5	65
Ferraz de Vasconcelos	191,993	0.90	7,287	0.14	1,091	0.30	45	6.68	58
Jandira	123,481	0.60	9,517	0.19	1,077	0.29	34	8.83	60
Franco da Rocha	152,443	0.70	9,055	0.18	959	0.26	47	9.44	49
Francisco Morato	174,008	0.80	6,727	0.13	808	0.22	48	8.32	48
Arujá	88,455	0.40	5,533	0.10	794	0.22	45	6.96	60
Cajamar	75,638	0.40	12,353	0.24	729	0.20	41	16.94	54
Mairiporã	98,374	0.50	4,618	0.09	698	0.19	37	6.61	59
Caieiras	100,129	0.50	8,889	0.18	537	0.15	38	16.55	50
Santa Isabel	56,792	0.30	3,971	0.08	447	0.12	61	8.88	45
Vargem Grande Paulista	51,702	0.20	3,034	0.05	417	0.11	44	7.27	49
Embu-Guaçu	68,856	0.30	2,282	0.04	380	0.10	49	6	41
Guararema	29,451	0.10	2,237	0.04	313	0.08	79	7.15	45
Rio Grande da Serra	50,241	0.20	1,245	0.02	234	0.06	50	5.32	26
Juquitiba	31,235	0.10	1,614	0.03	207	0.06	72	7.8	35
Biritiba Mirim	32,251	0.10	819	0.01	135	0.04	79	6.06	27
São Lourenço da Serra	15,667	0.10	3,298	0.06	133	0.04	54	24.8	32
Salesópolis	17,022	0.10	626	0.01	100	0.03	101	6.26	20
Pirapora do Bom Jesus	18,604	0.10	179	0.00	48	0.01	55	3.73	17

Source: Elaborated by the authors based on RAIS dataset.

According to SEADE data for 2017, the economically active population of SPMR comprises more than 11 million individuals (the largest in Brazil), 15% of which are employed in Manufacturing, 7% in construction and 18% in Commerce. By contrast, 60% of the population is employed in the service sector, which shows how important this activity is for the functioning of the region. Actually, in the SPMR are located most of the Brazilian headquarters of large industrial, commercial and financial complexes. To meet the demands of firms and inhabitants, an intricate network of increasingly sophisticated services was built, which is characterized by its deep relationship between several sectors of activities, such as information technology, finance, consulting, and marketing. This multiple network of relationships between people is characteristic of modern urban centers and vital for the existence of cities.

In this research we use geo-referenced information (complete address identified spatial unities or geocoded) about formal firms using official information from the RAIS data (Annual Report of Social Information) of the Minister of Labor and Employment. More specifically, from a total of 359,248 formal firms belonging to the service sectors from RAIS database, identified through the National Registry of Legal Entities (CNPJ), we worked on a sample of 352,625 geocoded firms. Additionally, in order to work on a reliable location information of economic activities, we restrict our sample to firms with at least 50% of formality. Our final restricted sample, therefore, is composed by 332,705 geo-referenced representing the location of 5,086,035 jobs and 92.61% of the total RAIS firms. Figure (2) presents the spatial distribution of these firms among the municipalities of the SPMR and Table (2) sum up the procedures for adjusting the data set.

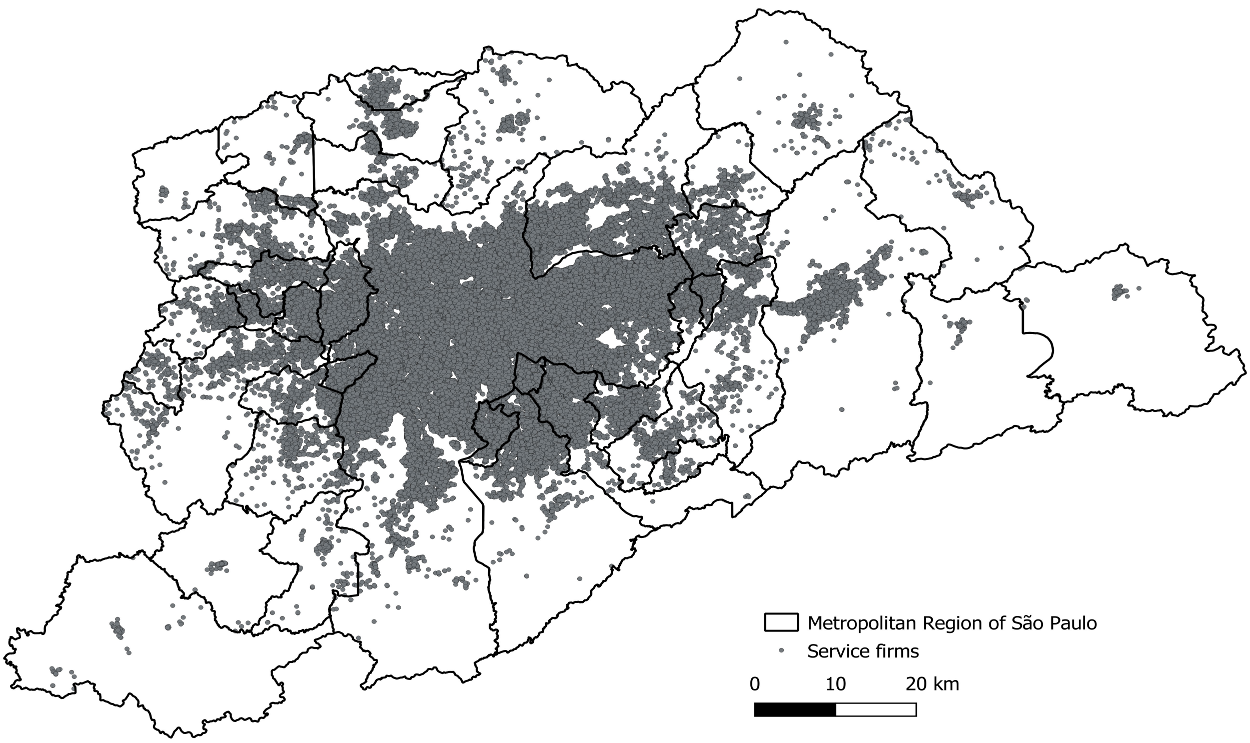


Figure 2. Localization of firms of services in Metropolitan Region of São Paulo
Source: Elaborated by the authors based on a unique dataset.

Note that the set of evidence about the patterns of location of firms of different activities of Services in SPMR is obtained by considering a 3 digits level of sector desegregation using the official CNAE 1.0 (Classificação Nacional de Atividade Econômicas), a worldwide compatible classification generated by IBGE. This is the common level of sector desegregation used in similar studies (Duranton and Overman, 2005; Behrens and Bougna, 2015; Aleksandrova et al., 2019) and, in the case of

Table 2. Summary of data set

	Employment	# plants	% plants
RAIS firms	5,449,044	359,248	100%
RAIS geocoded firms	5,325,769	352,625	98.16%
RAIS firms with at least 50% of formality	5,205,670	338,942	94.35%
RAIS geocoded firms with at least 50% of formality	5,086,035	332,705	92.61%

Source: Elaborated by the authors based on RAIS dataset.

CNAE, comprises 218 groups of different economic activities (including Agriculture, Manufacturing, and Services). After the above described constraints, we finally worked with 81 activities of services (sectors) out of a total of 92 services activities of CNAE.

3 Methodology

As pointed out earlier, more traditional measures of concentration of productive activity are susceptible to MAUP. In this research we use a measure based on distances proposed by Duranton and Overman (2005) that overcomes this problem. The procedure for constructing this metric consists of four steps.

First, the pairwise distances of all plants in a service sector are calculated and we estimate the density of their distribution. Second, we construct a counterfactual distribution assuming that the plants of a particular service sector are located randomly among all possible locations. We use all service sectors in the SPMR as counterfactual. Third, we define the confidence bands. The second step is replicated 1,000 times – following Duranton and Overman (2005), for some specific sectors tests of robustness with 2,000 and 10,000 simulations are performed – to build the confidence intervals of 1,000 counterfactual K densities. Finally, we test whether a particular service sector has patterns of location or dispersion or is randomly distributed, comparing the observed distribution of bilateral distances with the confidence bands obtained in the sampling process. Below we detail each step.

Firts step: Obtain the value of the index using kernel density estimation. Using our previously described geocoded database, the distance between a pair of plants (i, j) is initially calculated. For n plants, we have $n(n\text{-}1)$ pairs of distances. Using a kernel density function, the density of the bilateral distances at any target distance can be calculated according to:

$$\hat{K}d(r) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{r - r_{i,j}}{h}\right) \tag{1}$$

where $_{i,j}$ is distance between plants i and j; h is bandwidth[5] and; f is the the standard Gaussian kernel function:

$$f = k(\|x_i - x_j\|, r) = \frac{1}{h\sqrt{2\pi}} exp\left(-\frac{(\|x_i - x_j\| - r)^2}{2h^2}\right) \tag{2}$$

where x_i is the reference point and x_j is the neighbor.

[5]Following Silverman (2018), the ideal bandwidth for the Gaussian kernel function is $1.06sn^{-0.2}$ where s is the standard deviation of $n(n\text{-}1)$ bilateral distances.

The maximum of the kernel function is reached when the distance between points i and j are equal to r and decreases according to the Gaussian function when the distance deviates from r.

Second step: Counterfactual densities. Using our data set for all service establishments in the SPMR, that is, we consider the location of all service establishments as possible locations for a predetermined sector m. For each of these locations, we randomly draw as many as the number of firms in the sector m and we assign a firm of m randomly. We calculate the bilateral distances considering the random draw and obtain the K-density. This procedure ensures that we are controlling for the geographical location pattern of the entire service sector within the metropolitan area and for the differences between the size of the sectors, thus allowing comparisons of location patterns between sectors.

Third step: For each sector m, we repeat the second step 1,000 times. This implies that for each distance r, we have an estimated 1,000 K-density set. Following Duranton and Overman (2005), we consider the distances from 0 to the median, $\bar{r}$, of all bilateral distances in the sample. Additional discussion and alternative ways to define the range can be found in the Behrens and Bougna (2015) and Aleksandrova et al. (2019) works for Canada and Russia, respectively, in a regional context considering the distribution of manufacturing across the country[6]. Our analysis is then restricted to $r \in [0, \bar{r}]$. For each subdivision of the sample and each kilometers in this interval a $\hat{K}d(r)$ is estimated. The lower $\hat{K}d_{lo}(r)$ and upper $\hat{K}d_{hi}(r)$ limits are defined as less than 95% of the $\hat{K}d(r)$ estimated between $\hat{K}d_{lo}(r)$ and $\hat{K}d_{hi}(r)$.

Fourth step: Identifying which service sectors are localized and which ones are dispersed. Any deviation from the confidence band constructed in step 3 indicates the location or dispersion of the firms belonging to the service sector in the SPMR. A sector m can be classified as follows:

i) If $\hat{K}d_m(r) > \hat{K}d_{m,hi}(r)$ for at least one r, the sector is localized.
ii) If $\hat{K}d_m(r) < \hat{K}d_{m,lo}(r)$ for at least one r and $\hat{K}d_m(r) < \hat{K}d_{m,hi}(r)$ for all r[7], the sector is dispersed.

We can then define measures of localization (Γ_m) and dispersion (Ψ_m) for each r, as below:

$$\Gamma_m(r) \equiv \max\left(\hat{K}d_m(r) - \hat{K}d_{m,hi}(r), 0\right)$$

$$\psi_m(r) \equiv \begin{cases} \max\left(\hat{K}d_{m,lo}(r) - \hat{K}d_m(r), 0\right) & \text{if } \sum_{r=0}^{\bar{r}} \Gamma_m(r) = 0 \\ 0 & \text{otherwise} \end{cases}$$

From these definitions, we generate condensed indices that summarizes the information across all $r \in [0, \bar{r}]$:

$$\Gamma_m = \sum_{r=0}^{\bar{r}} \Gamma_m(r) \qquad \text{and} \qquad \psi_m = \sum_{r=0}^{\bar{r}} \psi_m(r)$$

[6] Differently from Duranton and Overman (2005), Behrens and Bougna (2015) and Aleksandrova et al. (2019) analyze large countries in territorial extension and, therefore, it is important to define an adequate range. For Canada, Behrens and Bougna (2015) use 800 kilometers, and to Russia, Aleksandrova et al. (2019) use 1,000 kilometers. For more details on the computational implementation see Aleksandrova et al. (2019).

[7] When a sector shows peaks of concentration, it is possible that other points of the curve will fall below the lower confidence bound as a form of compensation. This happens because the values are normalized to sum 1; it does not imply dispersion.

These last indices allow us to measure the amount of localization or dispersion of an industry with a single value. A higher Γ_m (Ψ_m) indicates a more localized (dispersed) sector m. These measures allow comparisons between sectors, which in turn admit that sectors are ranked according to their level of localization or dispersion.

Finally, since K-density is a distribution function for sector m, we can calculate its cumulative (CDF) for $r \in [0, \bar{r}]$:

$$\mathsf{CDF}_m = \sum_{r=0}^{\bar{r}} \hat{K}d_{obs}(r) \tag{3}$$

We can interpret the larger values of CDF_m as indicating that sector m, for a given distance, presents a pattern of greater geographic concentration. And, alternatively as highlighted by Aleksandrova et al. (2019), we can interpret this as the probability that two randomly drawn firms in a given sector m will be distant from each other up to a maximum of $\bar{r}$ kilometers.

4 Results

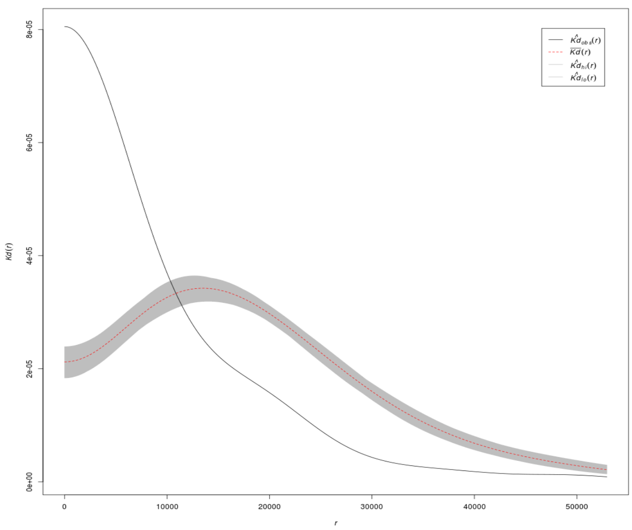
In this section, we present Duranton and Overman (2005) K-densities for 81 sectors (3-digits) of services based on the CNAE for all SPMR and, then, discuss economic arguments associated with the their patterns of location. Notice that, in order to obtain a reliable set of evidence, based on the information from the Pesquisa Nacional por Amostra de Domicílios (PNAD) of 2015, We only consider the sectors with more than 50% of the total formal employment. For each sector of services, we also compute local and global confidence bands based on 1,000 random permutations, as explained previously.

4.1 Measuring Location Patterns

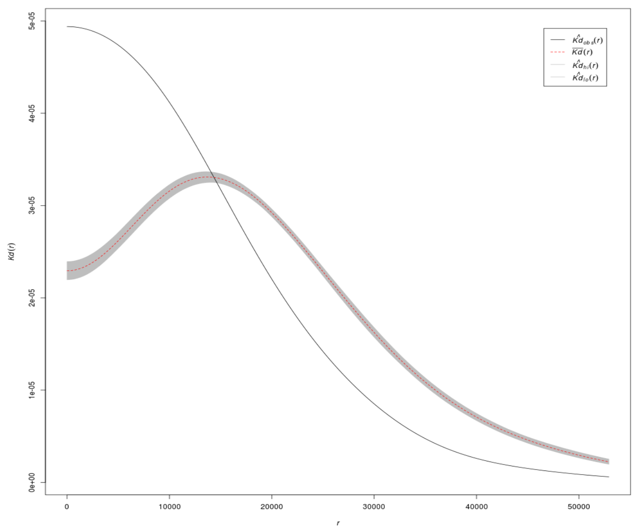
Our baseline results indicate that 90% of the service sectors in the SPMR presents statistically significant location patterns. In order to illustrate the generated set of evidence, the following Figure (3) presents the estimated baseline results with examples of location ((a) and (b) top two graphics), randomness ((c) left bottom graphic) and dispersion ((d) right bottom graphic).

The observed value of the index in the service sector is represented by the solid line, $\hat{K}d_{obs}(r)$. Upper and lower confidence bands are represented, respectively, by $\hat{K}d_{hi}(r)$ and $\hat{K}d_{lo}(r)$; these bands contain 95% of the counterfactual distributions, so when the solid line $\hat{K}d_{obs}(r)$ is between them we can not reject (at the 5% level) the null hypothesis that the localization pattern of the particular service is random. Remember that if the line $\hat{K}d_{obs}(r)$ is above the upper limit of the confidence band, the distances between firms are over represented in comparison to randomness, and this is interpreted as a situation of defined pattern of location. On the other hand, when line $\hat{K}d_{obs}(r)$ is below the lower limit of the confidence band, the distances between firms are underrepresented in comparison to spatial randomness, and this is interpreted as a situation of dispersion. In the horizontal axis, the distances are represented in meters and its limit of 52.9 km (half of the biggest distance between firms) was established following the suggestion of Duranton and Overman (2005) .

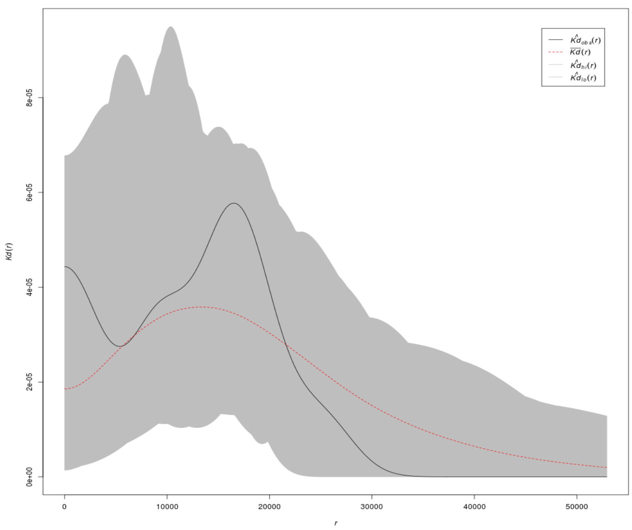
Consistent with the importance of the principal CBD in the capital, Figure 3 (a) indicates a location pattern of the *Auxiliary activities of financial intermediation* - CNAE 3 digits 671 - up to approximately 10 km. This finding corresponds to the location pattern where firms are disproportionately located over short distances. Figure 3 (b) shows a similar location pattern for *Wholesale of personal and household goods* - CNAE 3 digits 514 - located at a distance of up to 14 km. On the other hand, Figure 3 (c) shows the location pattern of the *Interurban rail transport* - CNAE 3 digits 601 - where we do not reject the null hypothesis of location randomness. And, in line with expectations, once it tend to follow the residential distribution, Figure 3 (d) shows the *Ductwork* - CNAE 3 digits 603 - with dispersion at distances close of 10 km.



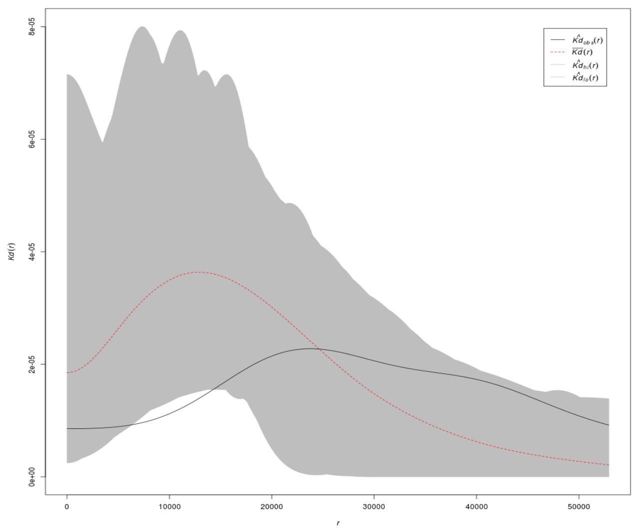
(a) Auxiliary activities of financial intermediation



(b) Wholesale of personal and household goods



(c) Interurban rail transport



(d) Ductwork

Figure 3. K-density estimates for selected CNAE (3-digit) of services
Source: Elaborated by the authors based on the RAIS data.

Table (3) outlines the set of evidence and presents a summary of the results about the location patterns for all the service sectors. We notice that only one sector was classified as a dispersed sector and seven sectors as randomly distributed. On the other hand, the percentage of activities with a clear location pattern (around 90%) is higher compared to previous evidence using the same location measuring strategy. For developed countries, this set of evidence indicates that most of manufacturing activities have patterns of geographical location (Barlet et al., 2008; Nakajima et al., 2012; Koh and Riedel, 2014; Behrens and Bougna, 2015; Aleksandrova et al., 2019), but this percentage is never achieved. For the service sectors, the evidence is scarcer. Our results, however, are in line with those obtained by Koh and Riedel (2014), for Germany, and Silva et al. (2019), for the Metropolitan Region of Recife. These last authors particularly found a higher share of activities with defined patterns of location for Services (71.7%) than for Manufacturing activities (64.7%). The even higher share of Services activities presenting a defined patterns of location in the current study may both reflect difference between the two metropolitan regions and the fact that these authors have treated Services and Manufacturing activities simultaneously (what matters in contractual simulations). Anyhow, notice that, unlike most of manufacturing activities, most of Service activities strongly depend on interactions with customers (obvious examples are the financial intermediation and consulting services that serve customers in different industries (Koh and Riedel, 2014)). As highlighted by Kolko (2010), this characteristic not only tends to make these activities more urbanized than Manufacturing ones, but also to restrict the spatial scope of its location determinants. These characteristics apparently favor identifying location patterns for Services in a given urban area.

Table 3. Summary of geographic concentration patterns for SPMR services

	N = 332,705 plants	
Status	Number	Percentage
Localized services	73	90.12%
Random	7	8.64%
Dispersed services	1	1.23%
Localization (average) $\bar{\Gamma}\|_{\Gamma_m>0}$	0.000818	
Dispersion (average) $\bar{\Psi}\|_{\Psi_m>0}$	0.000114	
Total	81	100.00%

Source: Elaborated by the authors based on estimates. Notes: The values of $\bar{\Gamma}|_{\Gamma_m>0}$ and $\bar{\Psi}|_{\Psi_m>0}$ report average for all significantly localized services and for all significantly dispersed, respectively.

Table (4) presents the 30 service sectors (CNAE 3-digits) with the highest value of the index Γ_m, i.e, the most localised activities, together with the value of Ψ_m for the unique dispersed activity. We initially highlight the fact that most of these activities belongs to the areas of finance (for example, *Auxiliary activities of financial intermediation* - CNAE 3 digit 671 and *Other financial intermediation activities* - CNAE 3 digit 659) and commerce (for example, *Commercial lease* - CNAE 3 digit 654 and *Wholesale pf personal and household goods* - CNAE 3 digit 514) or involve using a higher level of human capital (for example, *Software consulting* - CNAE 3 digit 722 and *Legal, accounting and business advisory activities* - CNAE 3 digit 741). On the other hand, among the sectors in which it was not possible to reject the hypothesis of randomness are *retailing of used articles* - CNAE 3 digit 525, *car rental* - CNAE 3 digit 589 and *intercity rail transport* - CNAE 3 601 digit present, for example.

Previous studies[8] have also shown that the financial activities have significant location patterns in France (Barlet et al., 2008), Germany (Koh and Riedel, 2014) and Brazil in the study of Silva et al. (2019) for the Recife Metropolitan Region. The evidence obtained here, thus, confirms once more

[8]Using the index of Ellison and Glaeser (1997) for the US, the Kolko (2010) study finds that the service sectors are also heavily concentrated.

a clear pattern of location for these activities within the Brazilian megalopolis. As for the group of activities involving higher levels of human capital, similar results were also obtained by Billings and Johnson (2016), for the Denver-Boulder-Greeley metropolitan area, and Silva et al. (2019). We notice that this group includes very different kinds of services (for example, *Architectural and engineering services and specialized technical advice* - CNAE 3 digit 742 and *Health care* - CNAE 3 digit 851), a fact that suggests the role of human capital for their location patterns. Interestingly, both groups of activities tend to present defined location patterns at short distances: while for the financial activities the average (among activities of the group) maximum distance for a observed location pattern is 10.2 km, for human capital based activities the same average distance is 13.0km.

The activities of commerce presenting a clear location pattern include both retail and wholesale activities (for example, *Wholesale pf personal and household goods* - CNAE 3 digit 514 and *Non-specialized retail* - CNAE 3 digit 521). In both situations, note that they involve a great variety of differentiated products and, thus, they location patterns may act reducing research and transport costs for consumers and bringing demand externalities for firms of differentiated products (Konishi, 2005). Similar results were also obtained by Silva et al. (2019) for the metropolitan region of Recife. Notice also that for this group of activities the average maximum distance for a observed location pattern is longer (about 29.2km) than for the aforementioned activities, suggesting less importance for geographic proximity.

Note that the location patterns observed in Table (4) through the index Γ_m inform about the existence and relative relevance of a location pattern of firms in different economic activities, but not directly about their geographic concentration levels. The results in Table (5) which presents the ranking of the sectors with the 30 activities with the largest CDF_m bring this information by registering the probability that two randomly drawn firms of a given sector would be distant form each other up to 52.9 km. Interestingly, these new results indicate that, in general, the most strongly localized service economic activities are also the most agglomerated (Aleksandrova et al., 2019). In addition, note, particularly, that half of all activities belongs to finance or services involving higher human capital.

4.2 Characterizing services location patterns in SPMR

In this subsection, we present evidence about economics forces associated with the patterns of location of service sectors in the SPMR using simple multivariate regressions. Far from obtaining causal relationships, something not allowed by our dataset, we simply look for theoretical based associations between the identified patterns of location of firms and economic arguments. Briefly, these economic arguments for understanding firms location patterns derives both from traditional location theories and sources of urban agglomeration gains.

The patterns of location of firms, according to locational theory arguments, are associated both with the conditions of spatial competition and product differentiation and consumer externalities (Hotelling, 1929; Clark, 2002; Konishi, 2005; Fujita and Thisse, 2013), product differentiation turns price competition weaker allowing spatial proximity for firms that compete with each other. As argued by Konishi (2005), consumer externalities turn stronger this possibility. More recently, Billings and Johnson (2016) highlighted that consumer shop time regularity also matter for the firms patterns of location, once product and services that impose more frequent and regular shops tend to locate nearer residences.

But firms location patterns in an urban context can also be influenced by the possibility of urban agglomeration gains (Ciccone and Hall, 1996; Acemoglu and Angrist, 2000; Glaeser and Mare, 2001; Moretti, 2004; Fu, 2007; Greenstone et al., 2010; Duranton, 2016; Chauvin et al., 2017; Dingel

Table 4. Most localized (30 sectors), most dispersed and random groups (CNAE 3 - digit) services

CNAE 3 digits	Service name	# of plants	
	Most localized		Γ_m
671	Auxiliary activities of financial intermediation	841	0.003275
654	Commercial lease	12	0.002589
704	Building condominiums	30,657	0.002495
514	Wholesale of personal and household goods	7,008	0.002038
724	Database activities and online distribution of electronic content	466	0.001959
702	Rental of real estate	1,558	0.001909
659	Other financial intermediation activities	560	0.001884
722	Software consulting	2,352	0.001819
632	Auxiliary activities for transportation	6,449	0.001788
633	Activities of travel agencies and travel organizers	2,574	0.00177
744	Advertising	2,649	0.001715
661	Life and non-life insurance	357	0.001712
662	Supplementary pension	62	0.001614
519	Wholesale trade of goods in general or not included in the previous groups	2,259	0.001557
721	Hardware Consulting	957	0.001546
701	Incorporation and purchase and sale of real estate	3,155	0.001523
741	Legal, accounting and business advisory activities	16,512	0.001459
602	Other land transport	18,176	0.001386
653	Nonmonetary intermediation - other types of deposits	97	0.001251
803	Higher education	530	0.001227
521	Non-specialized retailing	12,097	0.001224
742	Architectural and engineering services and specialized technical advice	3,593	0.001222
655	Other credit granting activities	616	0.001159
455	Finishing works	5,881	0.001148
851	Health care activities	25,482	0.00101
621	Air transport, scheduled	136	0.000967
631	Moving and storing loads	769	0.000934
723	Data processing	1218	0.000916
900	Urban cleaning and sewage and related activities	624	0.000880
410	Water collection, treatment and distribution	194	0.000830
	Most dispersed		Ψ_m
603	Ductwork	16	0.000114
	Random		–
525	Used-vehicle retailing	413	–
743	Material and product testing; quality analysis	293	–
611	Cabotage and long-distance sea transport	8	–
802	High School	400	–
711	Car rental	689	–
612	Other aquatic transports	18	–
601	Interurban rail transport	10	–

Source: Elaborated by the authors based on estimates.
Note: Γ_m and Ψ_m are computed at 52.9 kilometers distance.

Table 5. Most geographically concentrated service sectors in SPMR

CNAE 3 digits	Service name	CDF_m
654	Commercial lease	0.00975
611	Cabotage and long-distance sea transport	0.00970
662	Supplementary pension	0.00969
601	Interurban rail transport	0.00969
621	Air transport, scheduled	0.00969
612	Other aquatic transports	0.00968
671	Auxiliary activities of financial intermediation	0.00968
704	Building condominiums	0.00966
514	Wholesale of personal and household goods	0.00965
721	Hardware Consulting	0.00965
653	Nonmonetary intermediation - other types of deposits	0.00964
519	Wholesale trade of goods in general or not included in the previous groups	0.00964
722	Software consulting	0.00963
632	Auxiliary activities for transportation	0.00962
655	Other credit granting activities	0.00962
622	Air transport, non-scheduled	0.00962
634	Activities related to the organization of cargo transportation	0.00961
744	Advertising	0.00961
659	Other financial intermediation activities not specified above	0.00961
702	Rental of real estate	0.00959
661	Life and non-life insurance	0.00959
729	Other computer activities, not specified above	0.00959
741	Legal, accounting and business advisory activities	0.00958
633	Activities of travel agencies and travel organizers	0.00958
803	Higher education	0.00957
663	Health plans	0.00957
672	Activities auxiliary to insurance and occupational pensions	0.00957
516	Wholesale of machinery, apparatus and equipment for agricultural, commercial, office, industrial, technical and professional uses	0.00956
724	Database activities and online distribution of electronic content	0.00956
725	Maintenance and repair of office and computer machinery	0.00956

Source: Elaborated by the authors based on estimates.
Note: CDF_m is computed for all $r \in [0, 52.9\mathsf{km}]$.

et al., 2019). These gains can arise both from spatial proximity to services and inputs offers (better sharing) and from spatial proximity to similar activities involving technological spillovers and learning (typically, activities involving higher levels of human capital).

In order to capture part of these arguments, based on available information, we built indicators of product differentiation, human capital and vertical integration (linkages) for each economic activity. For the product differentiation, we built the proxy proposed by Silva et al. (2019). The index is based on the sector differentiation of the structure of occupation, being defined by:

$$\mathsf{D}_m = \frac{1}{\mathsf{N}_m \mathsf{H}_m} \sum_{i=1}^{\mathsf{N}_m} \sum_{k=i}^{\mathsf{H}} (\mathsf{y}_{mik} - \bar{\mathsf{y}}_{mk})^2 \tag{4}$$

where N_m is the number of firms of service sector m; H_m is the number of kinds of occupations of service sector m; y_{mik} corresponds to the share of occupation k in the total kinds of occupation of firm i of service sector m; and $\bar{\mathsf{y}}_{mk}$ is the average shares of occupations k of service sector m. The index capture the idea that the more differentiated the service within the sector, the more diverse is its occupational structure. More specifically, for example, both here for SPMR and in Silva et al. (2019), we get the fuel retail trading activity (an activity known by its homogeneous product) with very low indicator.

Notice that the influence of product differentiation on location patterns of firms may depend of the level of differentiation itself (while homogeneous or weakly differentiated products may brings price competition, extremely differentiated products has no impact on firms' location patterns). Thus, similarly to Silva et al. (2019), in the following econometric specification, we allow for a non-linear association between product differentiation and the location patterns of firms.

We also use he percentage of workers with at least one college in the degree in the sector, $\mathsf{college}_m$, to measure its intensity of human capital using. This represents the potential of technological spillovers and learning of the sectors and, thus, may influence the location choices of firms of the sectors (Ciccone and Hall, 1996; Acemoglu and Angrist, 2000; Glaeser and Mare, 2001; Moretti, 2004; Fu, 2007; Greenstone et al., 2010; Duranton, 2016; Chauvin et al., 2017; Dingel et al., 2019).

In line with the arguments that the sharing of suppliers and inputs can favor the agglomeration of firms, we use the ratio of the number of different occupations per firm of the sector as an indicator of its vertical integration level, Share_m. Because this indicator can also vary according to the size of the firms, we also use an indicator for the firms size (number of workers per firm),$\mathsf{Workers/firm}_m$.

Finally, following Koh and Riedel (2014) and Silva et al. (2019), we also use the mean age of the workers, Age_m as a control variable to capture potential influences arising from different conditions of workers' mobility or differentiated demographic patterns across economic activities.

Formally, we estimate the influence of these factors through the following equation:

$$\Gamma_m = \theta + \beta_1 \mathsf{college}_m + \beta_2 \mathsf{Share}_m + \beta_3 \mathsf{D}_m + \beta_4 \mathsf{D}_m^2 + \beta_5 \mathsf{Age}_m + \beta_6 \mathsf{Workers/firm}_m + \epsilon_m \tag{5}$$

where Γ_m is the location index for service sector and ϵ_m is an error term.

Table (6) presents the results of the estimates of equation (5) by ordinary least squares (OLS). Except for the differentiation index (for which we explore non-linear effects), the variable are used in log. forms since it facilities interpretation and reduce the number of decimal cases of the estimated coefficients. In the columns of (I) to (III) we present simple association specifications, that is, considering each of the explanatory variables previously described separately. In columns (IV) and (V) we consider the three indicators together, respectively, without and with the controls variables.

We note that the estimated coefficient for the variable $\mathsf{college}_m$ (column (I)) is positive and statistically significant, indicating that the patterns of location of service sectors are associated with the share of skilled workers. Note that this result is robust across different multivariate specifications. Thus, the evidence, also obtained also by Koh and Riedel (2014) and Silva et al. (2019), suggests that location patterns of firms are positively influenced by the human capital involved, a result consistent with the importance of technological spillovers and learning within them.

Note also, that, although the results from columns (II) and (III) for individual associations, respectively, for vertical integration and differentiation indicators do not present statistical significance, they present expected signals. Furthermore, the effects appear statistical significant when more variables are included (columns (IV) and (V)). More specifically, from both columns (IV) and (V), we observe a negative association between the degree of vertical integration and the firms location pattern, suggesting that more interdependence positively affect it. Again, this result is line with the evidence provided by Silva et al. (2019) and Behrens and Bougna (2015) and is consistent with the importance of proximity to services and inputs offers.

The estimated parameters associated with our proxy for differentiated products (D_m) was positive for the linear variable and negative for the quadratic one. These results suggest that product differentiation positively affect the spatial pattern of location of firms for lower levels of differentiation and has none or negative effects for higher levels of differentiation. Again, the results are similar to those obtained by Silva et al. (2019) and perfectly consistent with the the trade-off highlighted by the location theory (Fujita and Thisse, 2013): product differentiation may turn weaker price competition effect, but this attenuation matters more for less differentiated products.

Table 6. Conditionings of within urban localization (linear regression estimation).

Variable	(I)	(II)	(III)	(IV)	(V)
$\ln(\mathsf{college}_m)$	0.00019**			0.00028***	0.00029***
	(0.00008)			(0.00007)	(0.00007)
$\ln(\mathsf{Share}_m)$		-0.00007		-0.0002***	-0.00025**
		(0.00007)		(0.00007)	(0.00009)
Differentiation (D_m)			0.15534	0.40003*	0.43616
			(0.23898)	(0.23901)	(0.26435)
Differentiation^2 (D_m^2)			-48.54928	-77.06726**	-83.77263**
			(38.95908)	(38.06548)	(39.81494)
$\ln(\mathsf{Age}_m)$					0.00078
					(0.00117)
$\ln(\mathsf{Workers/firm}_m)$					0.00006
					(0.00008)
Constant	0.00107***	0.000668***	0.00067***	0.00076***	-0.00223
	(0.00017)	(0.00011)	(0.00016)	(0.0002)	(0.00409)
R^2	0.08432	0.01465	0.01605	0.18025	0.19321
Adjusted R^2	0.07273	0.00218	-0.00917	0.13711	0.12779
F-statistics	6.11533**	1.01621	8.07425***	5.99979***	3.93324***
Observations	81	81	81	81	81

Source: Elaborated by the authors based on estimates.
Notes: The dependent variable is $\ln(\Gamma_m + 1)$. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Heteroskedastic robust errors are given in parentheses.

5 Conclusion

We present a detailed analysis of the firms location patterns of the service sectors in the SPMR, the largest Brazilian metropolitan region and one of the biggest worldwide urban agglomeration. The set of evidence was obtained using geocoded information and a distance-based measure that allowed us to overcome one of the main problems of traditional concentration measures, MAUP. In addition, in an exploratory way, the research presented evidence about the economic arguments that may be associated to the identified localization patterns.

We highlight three general aspects about this set of evidence about the patterns of distribution of service activities in SPMR:

i) Most of the service sectors analyzed (around 90%) present a deviation from the random pattern, indicating a defined location pattern. This result is in line with the models of formation of cities that indicate that more specialized sectors are more productive when concentrated spatially close (Combes et al., 2012; Behrens et al., 2014; Accetturo et al., 2018; Gaubert, 2018).

ii) Financial and human capital based activities present location patterns at short distances, indicating their spatial agglomeration tendencies. These results confirm previous evidence in different contexts (Koh and Riedel, 2014; Silva et al., 2019). On the other hand, the activities of commerce presenting tend present location patterns at average distances, what is consistent with the relevance of proximity to residences.

iii) Our evidence about the economic arguments associated with the observed location patterns suggests that these defined patterns are positively associated with human capital, the degree of product differentiation, and with inter-sector dependence localization of firms over shorter distances. Thus, this exploratory set of evidence is consistent with both the relevance of learning and sharing sources of agglomeration gains and with spatial competition arguments of traditional location theory.

However, much is still necessary for deeper understanding of location patterns in Brazilian urban centers. First, no evidence is available about about the the patterns of joint location between pairs of different economic activities. Second, obviously, much work and energy are necessary in order to effectively explain the identified location patterns (maybe using policy historical experiments). Both directions, actually, are in our research agenda.

References

Abel, J. R. and Deitz, R. (2015). Agglomeration and job matching among college graduates. *Regional Science and Urban Economics*, 51:14–24.

Accetturo, A., Di Giacinto, V., Micucci, G., and Pagnini, M. (2018). Geography, productivity, and trade: Does selection explain why some locations are more productive than others? *Journal of Regional Science*, 58(5):949–979.

Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9–59.

Aleksandrova, E., Behrens, K., and Kuznetsova, M. (2019). Manufacturing (co) agglomeration in a transition country: Evidence from russia. *Journal of Regional Science*.

Almeida, E. T. and Rocha, R. d. M. (2018). Labor pooling as an agglomeration factor: Evidence from the brazilian northeast in the 2002–2014 period. *EconomiA*.

Almeida, E. T. d., de Rocha, R. d. M., and Gomes, S. M. F. P. O. (2017). Economias de aglomeração e o crescimento das indústrias intensivas em tecnologia: evidências para o nordeste no período 2002-2014. *Revista Brasileira de Estudos Regionais e Urbanos*, 11(4):467–494.

Andersson, F., Burgess, S., and Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128.

Arzaghi, M. and Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, 75(4):1011–1038.

Baer, W. (2002). *Economia Brasileira*. NBL Editora.

Barlet, M., Briant, A., and Crusson, L. (2008). Concentration géographique dans l'industrie manufacturière et dans les services en france: une approche par un indicateur en continu. *Documents de Travail de la DESE-Working Papers of the DESE.*

Behrens, K. and Bougna, T. (2015). An anatomy of the geographical concentration of canadian manufacturing industries. *Regional Science and Urban Economics*, 51:47–69. https://doi.org/10.1016/j.regsciurbeco.2015.01.002.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553.

Billings, S. B. and Johnson, E. B. (2016). Agglomeration within an urban area. *Journal of Urban Economics*, 91:13–25. https://doi.org/10.1016/j.jue.2015.11.002.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? cities in brazil, china, india and the united states. *Journal of Urban Economics*, 98:17–49.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, 86(1):54–70. https://doi.org/10.3386/w4313.

Clark, G. L. (2002). London in the european financial services industry: locational advantage and product complementarities. *Journal of Economic Geography*, 2(4):433–453. https://doi.org/10.1093/jeg/2.4.433.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594. https://doi.org/10.3982/ECTA8442.

Costa, M. A. C. (2013). Caracterização e quadros de análise comparativa da governança metropolitana no brasil: arranjos institucionais de gestão metropolitana (componente 1): região metropolitana de são paulo rmsp.

Di, Addario, S. (2011). Job search in thick markets. *Journal of Urban Economics*, 69(3):303–318.

Dingel, J. I., Miscio, A., and Davis, D. R. (2019). Cities, lights, and skills in developing economies. Technical report, National Bureau of Economic Research. http://www.nber.org/papers/w25678.

Domingues, E. P., Ruiz, R. M., Moro, S., and Lemos, M. B. (2006). Organização territorial dos serviços no brasil: polarização com frágil dispersão. *Estrutura e dinâmica do setor de serviços no Brasil. Brasília: IPEA.*

Duranton, G. (2016). Determinants of city growth in colombia. *Papers in Regional Science*, 95(1):101–131. https://doi.org/10.1111/pirs.12225.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106. https://doi.org/10.1111/0034-6527.00362.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. Elsevier.

Duranton, G. and Puga, D. (2014). The growth of cities. In *Handbook of Economic Growth*, volume 2, pages 781–853. Elsevier. https://doi.org/10.1016/B978-0-444-53540-5.00005-7.

Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105(5):889–927.

Fochezatto, A. and Valentini, P. J. (2010). Economias de aglomeração e crescimento econômico regional: um estudo aplicado ao rio grande do sul usando um modelo econométrico com dados de painel. *Revista EconomiA*, 11(4):243–266.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the boston metropolitan area. *Journal of Urban Economics*, 61(1):86–111. https://doi.org/10.1016/j.jue.2006.06.002.

Fujita, M. and Thisse, J.-F. (2013). *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge university press.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–53.

Glaeser, E. L. and Mare, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, 39(153):41–57.

Inoue, H., Nakajima, K., and Saito, Y. U. (2019). Localization of collaborations in knowledge creation. *The Annals of Regional Science*, 62(1):119–140.

Klier, T. and McMillen, D. P. (2008). Evolving agglomeration in the us auto supplier industry. *Journal of Regional Science*, 48(1):245–267.

Koh, H.-J. and Riedel, N. (2014). Assessing the localization pattern of german manufacturing and service industries: a distance-based approach. *Regional Studies*, 48(5):823–843.

Kolko, J. (2010). Urbanization, agglomeration, and coagglomeration of service industries. In *Agglomeration Economics*, pages 151–180. University of Chicago Press.

Konishi, H. (2005). Concentration of competing retail stores. *Journal of Urban economics*, 58(3):488–512. https://doi.org/10.1016/j.jue.2005.08.005.

Krugman, P. R. (1991). *Geography and trade*. MIT press.

Lautert, V. and Araújo, N. C. M. d. (2007). Concentração industrial no brasil no período 1996-2001: uma análise por meio do índice de ellison e glaeser (1994). *Economia Aplicada*, 11(3):347–368.

Leslie, T. F. and HUallacháin, B. Ó. (2006). Polycentric phoenix. *Economic Geography*, 82(2):167–192.

Lychagin, S., Pinkse, J., Slade, M. E., and Reenen, J. V. (2016). Spillovers in space: Does geography matter? *The Journal of Industrial Economics*, 64(2):295–335.

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

Marcon, E. and Puech, F. (2009). Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, 10(5):745–762.

Moretti, E. (2004). Estimating the external return to higher education: Evidence from cross-sectional and longitudinal data. *Journal of Econometrics*, 120(1-2):175–212.

Nakajima, K., Saito, Y. U., and Uesugi, I. (2012). Measuring economic localization: Evidence from japanese firm-level data. *Journal of the Japanese and International Economies*, 26(2):201–220.

Overman, H. G. and Puga, D. (2010). Labor pooling as a source of agglomeration: An empirical investigation. In *Agglomeration Economics*, pages 133–150. University of Chicago Press.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219. https://doi.org/10.1111/j.1467-9787.2009.00657.x.

Resende, M. and Wyllie, R. (2005). Aglomeração industrial no brasil: um estudo empírico. *Estudos Econômicos (São Paulo)*, 35(3):433–460.

Rocha, R. d. M., Araújo, J. E. S., and de Almeida, E. T. d. (2019). As indústrias da tranformação são concentradas geograficamente? um teste empírico para o brasil (2002-2014). *Nova Economia*.

Rocha, R. d. M., Bezerra, F. M., and de Mesquita, C. S. (2013). Uma análise dos fatores de aglomeração da indústria de transformação brasileira. *Revista EconomiA*.

Silva, M. V. B. and Silveira Neto, R. d. M. (2007). Crescimento do emprego industrial no brasil e geografia econômica: Evidências para o periodo pós-real. *Revista EconomiA*.

Silva, M. V. B. d. and Silveira Neto, R. d. 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*, 13(2):299–331.

Silva, R. L. P. d., Silveira, Neto, R. d. M., and Rocha, R. (2019). Localization patterns within urban areas: evidence from brazil. *Area Development and Policy*, pages 1–20.

Silveira Neto, R. d. M. (2005). Concentração industrial regional, especialização geográfica e geografia econômica: evidências para o brasil no período 1950-2000. *Revista Econômica do Nordeste, Fortaleza*, 36(2):189–208. https://ren.emnuvens.com.br/ren/article/view/732.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Thisse, J.-F. (2018). Human capital and agglomeration economies in urban development. *The Developing Economies*, 56(2):117–139. https://doi.org/10.1111/deve.12167.

Vitali, S., Napoletano, M., and Fagiolo, G. (2013). Spatial localization in manufacturing: a cross-country analysis. *Regional Studies*, 47(9):1534–1554.

Wheeler, C. H. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38(2):101–118. https://doi.org/10.1016/j.regsciurbeco.2008.01.011.