Urban Structure of Brazilian Metropolitan Regions: Identification and Characterization of Employment Sub-Centers

Abstract

At the end of the 20th century, Brazil has experienced a strong urbanization process, a consequence of intense rural-urban migration, resulting in approximately 85% of the population living in urban areas. Hence, a set of evidence shows that the populations' quality of life is directly associated with the spatial distribution of the economic activities within the cities. While the factors related to the urban expansion in Brazil are nowadays reasonably well established and exploited in the economic literature, there is a lack of knowledge concerning the structure of Brazilian urban centers, a gap that this article proposes to fill. In this perspective, the main goal of this paper is to identify employment subcenters for some of the most important Brazilian metropolitan regions. Our results suggest that despite a great concentration of employment closer to the CBD, Brazilian metropolitan regions presents a decentralization pattern of economic activities, mainly in the Southeast, evidence that can be partly explained by the characteristics of Brazilian urban centers.

Keywords: Employment Density, Subcenters, Metropolitan Regions, Brazil.

JEL: R10, R12, R30

1 Introduction

At the end of the 20th century, Brazil has experienced a strong urbanization process, as a consequence of intense rural-urban migration, resulting in approximately 85% of the population living in urban areas, which corresponds to only about 1% of the entire national territory, according to the 2010 Demographic Census (IBGE). In this context, Brazilians quality of life is fundamentally linked to the organization and structure of the urban centers. As stated by Da Mata et al. (2007), and more recently by Silva et al. (2017), this process was directly associated with the low productivity of agricultural activity, a rapid industrialization process, the expansion of schooling in urban areas and an improvement in the infrastructure of the cities.

It is important to note that, if, on the one hand, such an urbanization process is related to the factors described above, on the other hand, the employment centers and the spatial distribution of occupations in urban regions is directly linked to benefits associated with higher productivity, as a result of urban agglomerations (Moretti, 2004; Duranton e Puga, 2004; Glaeser, 2010; Baruffi et al.2016). As already well established in the literature, commuting costs and rental prices of housing increases significantly in regions with a higher population concentration. Hence, a necessary condition for the formation of an employment center is that the returns associated with the spatial agglomeration of agents must imply in higher productivity and wages. According to Duranton and Puga (2004), such increasing returns reflect better matching between occupations and workers, the possibility of a greater sharing of services and, ultimately, the knowledge spillover among workers.

In this context, the tradeoff between agglomeration gains and commuting costs is fundamental to analyze the spatial distribution of employment within urban centers. While the benefits associated with better matching, sharing, and learning are favorable forces to the agglomeration of occupations, the negative externalities resulting from greater commuting and the congestion of public goods act in the opposite direction. Hence, the spatial structure of cities and their configuration as monocentric or polycentric results from the performance of these forces. As shown by Fujita and Thisse (2013), for example, when costs arising from negative externalities become sufficiently high, certain activities tend to relocate outside the central business district (CBD). If the agglomeration forces corresponding to these activities are relatively strong, these activities will relocate into secondary centers, called subcenters (Fujita and Ogawa, 1982; Ahlfeldt et al., 2016). Thus, monocentric cities tend to settle in regions that have strong gains from agglomeration and low commuting costs.

There is plenty of evidence in the economic literature about the influence of urban city structures on an important set of factors associated with the quality of life within cities and the social outcomes of individuals in urban centers. Basically, among the influences highlighted, the results indicate the importance of urban structures on the distribution of the population, the possibility of spatial mismatch, the influence on the prices of the urban space (land and urban constructions) and the commuting cost. The evidence available indicates that the spatial distribution of economic activities within cities substantially affects population distribution and density (McDonald and McMillen, 1997, McMillen and McDonald, 1998), real estate prices (Bender e Hang, 1985; Richardson et al. 1990; Edlund et al. 2015), the build profile and land-use intensity (McMillen, 2008; Barr and Cohen, 2010; Ahlfeldt and McMillen, 2015), and the labor market outcomes of its inhabitants, especially through spatial mismatch (Gobilon et al., 2007; Ihlanfeldt, 2008). Moreover, it is possible to emphasize the influence of the urban structure on the commuting. For American cities, Gordon, Kumar and Richardson (1989) show that metropolitan areas with polycentric orientation provide less commuting cost for their residents. Likewise, Giuliano and Small (1991) find evidence that workers located in denser subcenters near the CBD have greater commuting compared to workers located in more distant subcenters. In other words, the urban profile of the city is largely defined by its employment distribution.

Notice that, while the factors associated with fast urban expansion in Brazil are nowadays reasonably well established and exploited in the economic literature, there is a lack of knowledge concerning the structure of Brazilian urban centers, that is, about the characteristics of the spatial distribution of economic activities and people in these urban centers. In this sense, evaluations concerning the monocentric or polycentric structure of Brazilian urban centers or about what kinds of economic activities are clustered (or dispersed) in Brazilian cities remains almost unchartered in the economic literature. Actually, among the few papers that address this theme, we stress out Ingram and Carroll (1981), in which the authors make a comparative analysis of Latin American urban structures. Ohers papers treated specific urban centers; Fernandez-Maldonado et al. (2015) provide an analysis for the case of Metropolitan of Fortaleza, Ramos (2014) considered the case of São Paulo, and, more recently, Belmiro et al. (2018) analysed the case of the city of Recife. Besides using different methodologies for indentifying emplyment subcenters, notice that these few existing studies dealed separetly with specific cases, making it hard to elaborate a more representative profile of the spatial distribution of occupations in the main Brazilian metropolitan regions. Such scarcity of studies in Brazil is, given the experience of the evidence for developed countries, noteworthy when considering the recent costs associated with the Brazilian intra-urban dynamics, whether in terms of rising urban land prices or the commuting time of its residents.

This paper seeks to fill this gap in the literature when analyzing the spatial distribution of economic activities in Brazilian urban centers. Therefore, the main objective of our study is to identify employments subcenters for some of the most important Brazilian metropolitan regions (henceforth MR). To this end, we use georeferenced formal employment information provided by the Institute of Applied Economic Research (IPEA) and apply a two-stage procedure proposed by McMillen (2001) to identify the subcenters. The first stage uses a nonparametric estimation to identify potential subcenters of employment, that is, locations where employment density is exceptionally high even after all spatial trends, are taken into account. The second stage uses a semi-parametric regression to determine if the employment subcenter candidates have significant effects on the employment density function of the study region.

More specifically, the procedure is used to identify employments subcenters for the metropolitan areas of Recife, Salvador, Porto Alegre, Rio de Janeiro, and São Paulo. Recife and Salvador are the two largest and oldest MR in the Northeast of Brazil. Despite having a high GDP, they both have significant heterogeneity in terms of human development, income inequality, and residents living in extreme poverty. Rio de Janeiro and São Paulo are the two largest and most economically important metropolitan regions of Brazil. Both are located in the Southeast, the most developed region of the country, accounting for more than 50% of the national GDP, as well as being the most densely populated. Finally, Porto Alegre represents the largest and most traditional MR in the southern region of the country, characterized by European colonization, strong industrial development and social index above the Brazilian average. Moreover, the Northeast, Southeast, and South of Brazil are the most urbanized and populous regions of the country[1].

Despite having a significant employment concentration around the CBD, the Brazilian metropolitan regions present a decentralization pattern of their economic activities, especially to those in the Southeast. We identify a total of 15 subcenters in the Recife metropolitan region, 13 in Salvador, 55 in Porto Alegre, 52 in Rio de Janeiro, and 84 in São Paulo. We show that these numbers are bigger than correspondent numbers of subcenters obtained using the traditional strategy proposed by Guiliano and Small (1991). Furthermore, we find that subcenters present heterogeneity regarding their degree of influence over the employment density surface in the study regions. While subcenters located closer the CBD presents greater magnitude and global effect, subcenters situated in peripheral regions presents lower employment density and local influence. Based on the results, we found evidence that the non-parametric approach is more suitable for the identification of subcenters in the Brazilian metropolitan regions since it presents the greatest explanatory power when compared to other standard models in the economic literature.

The remainder of the paper is organized as follows: Section 2 describes the procedure used to identify the employment subcenters. Section 3 presents the data and summary statistics. Section 4 presents the main results. Section 5 briefly discusses the results, compare them with the evidence obtained in other studies and discusses how we can understand the results in light of the characteristics of Brazilian urban centers. Finally, Section 6 presents the final considerations and suggestions for future research.

[1]RMR denotes the Recife Metropolitan Region, RMSA denotes the Salvador Metropolitan Region, RMPA denotes the Porto Alegre Metropolitan Region, RMRJ denotes the Rio de Janeiro Metropolitan Region and RMSP denotes the São Paulo Metropolitan Region.

2 Empirical Strategy

Most of studies in theliterature on emplyoment subcenters identification refers to the beginning of the 90s and focuses on American cities. While there is still no established consensus on the most appropriate method for identifying the urban subcenters, we can highlight four identification strategies. The first is based on thresholds for the employment density values and/or total employment of the geographical units (Giuliano and Small, 1991; Song, 1994; Cervero and Wu 1997; McMillen and McDonald 1998; Shearmur and Coffey 2001). Second, spatial data-based methods are used, such as exploratory spatial data analysis (Baumont et al., 2004; Guillain et al., 2004). A third alternative is an identification of peaks in the estimated employment density function or the employment/population relationship (Gordon et al., 1986; McDonald, 1987). Finally, parametric, semiparametric, and noparametric methods can be used (Craig and Ng, 2001; McMillen, 2001; Krehl, 2018).

Due to the simplicity of its application to different regions, the methodology proposed by Giuliano and Small (1991), i.e., cutoff limits for employment density, is commonly used in the literature to identify employment subcenters. Nevertheless, there are some weaknesses with regard to such a procedure. The first is in the fact that the thresholds are defined arbitrarily, being guided by the previous knowledge of the locality. Second, the methodology is sensitive to different geographic units. Finally, strategies based on cutoff limits only identify potential subcenters since they can not infer the statistical significance of each subcenter over the city's employment density function. Furthermore, methods based on exploratory spatial data analysis and peak inspection on the employment density function suffer from the same problem described above, after all, the existence of significant spatial correlations or deviations in the employment density gradient does not guarantee relevance to the locality in the analysis (in terms of influence on the urban structure).

In this research, we use thus a two-stage procedure, as proposed by McMillen (2001), for identifying urban employment subcenters. Among the advantages of using this methodology, we can highlight: i) rule out the need of using thresholds to determine density limits, proposing a more rigorous criterion applicable to different types of urban centers, ii) is less sensible to different geographic units of analysis (e.g. districts, census tracts), iii) explicitly considers the relationship with the CBD (eliminating proximity to this as a factor affecting identification), iv) allows for local variation in the effect of distance from the CBD, which means recognizing variations in urban land patterns, and v) identifies statistically significant local rises in employment density. Moreover, as McMillen (2001) argues, such a procedure allows us to study different urban centers without necessarily having a thorough prior knowledge of the study region under analysis.

In the first stage, we use a non-parametric estimation in order to smooth the natural logarithm of employment density, y, over the distance to CBD. The estimation is performed through locally weighted regressions (LWR), where specific estimates of the employment density with respect to the distance to the employment center are obtained for each geographical unit. This estimation uses a weight matrix of each observation in relation to others. More precisely, we use a geographically weighted regression, where the weights are based on the geographical distance among observations. Let y_i be the natural logarithm of employment density in grid i and $DCBD_i$ the distance of each grid i from the CBD. The estimated regression is

$$y_i = g(DCBD_i) \tag{1}$$

The main idea is to give a higher weight to closer observations when estimating the predicted value of y in a grid i. To obtain a smoother employment density surface, a relatively high window size should be chosen. We choose a window size, which defines the share of geographically closer observations to receive some weight in the estimation, of 50%.

To perform the estimation of equation 1, we must define a kernel function κ_i, which determines the weight given to observation i based on the geographical distance. Different functions can be used. We will follow McMillen (2001) and use a tricube Kernel. Let $d_i(x)$ define the distance between a grid i and a target point x. Ordering the observations such that $d_1(x) < d_2(x) < \ldots < d_n(x)$, we can represent the tricube Kernel in the following equation

$$\kappa_i = \left(1 - \left(\frac{d_i(x)}{d_q(x)}\right)^3\right)^3 I\left(d_i(x) < d_q(x)\right) \tag{2}$$

Where $I(d_i(x) < d_q(x))$ is an indicator function that equals 1 when the condition is satisfied. Hence, all observations beyond the window of the q closest observations are given zero weight in the estimation. In addition, within the window, closer observations are given a higher weight than more distant observations.

The list of subcenter candidates comprises those locations where the residuals are positive and statistically significant at 5% significance level, i.e., $\frac{y_i - \hat{y}_i}{\hat{\sigma}_i} > 1.96$, where $\hat{y}_i$ is the predicted log-employment density estimate at grid i and $\hat{\sigma}_i$ is the estimated standard error for the prediction. To prevent including many near grids with significant residuals as potential subcenters when they cluster together, we restrict the list of subcenter candidates to those grids where the predicted log-employment densities are the greatest among all grids with significant residuals in a 3-miles radius.

The first stage only identifies potential subcenters because, despite detecting local increases in the employment density function trough the residuals, it does not determine whether the location has a statistically significant effect on the overall shape of the employment density function of the region. On the other hand, the second stage is based on a semi-parametric procedure to verify the relevance of each subcenter candidate found in the previous stage. In this step, we verify how the employment density gradient varies when considering the distance to the identified locations as potential subcenters, once also controlled for the distance effect to the CBD.

Let D_{ij} denote the distance between grid i and candidate subcenter j. Define $DCBD_i$ the distance from the grid i to the central business district (CBD). Let S be the number of potential subcenters, where $j = 0, 1, \ldots, S$, thus, we estimate the parameters of the following semi-parametric regression

$$y_i = g(DCBD_i) + \sum_{j=1}^{S}\left(\delta_{1j} D_{ij}^{-1} + \delta_{2j} D_{ij}\right) + u_i \tag{3}$$

Where $g(DCBD_i)$ enters in the equation non-parametrically expressing the relation between the logarithm of the employment density and the distance to the CBD in each grid. The parameters δ_{1j} and δ_{2j} captures the possible influence of the distance to the potential subcenters on the overall employment density function. The variable D_{ij} enter both in level and inverse form. The level form is more desirable when the subcenter influence the entire study region. On the other hand, inverse form is better for capture local effect on the employment density function.

A variety of alternatives can be used to estimate $g(DCBD_i)$ such as locally weighted regressions, flexible Fourier forms or cubic splines. McMillen (2001) asserts that the choice of $g(\cdot)$ makes little difference. The central idea is to allow flexibility and variability in the employment density gradient in relation to the distance among different locations in the urban region.

In this sense, a very attractive form is using cubic splines, as proposed by Anderson (1982). Traditional methods of estimating urban density functions, such as negative exponential functions, consider density patterns to be monotonic, i.e., with density decreasing as distance to CBD increases. However, the employment density patterns in several cities does not follow this pattern. Thus, the estimation of the employment density-distance relationship through cubic splines allows a greater flexibility to the gradient. Furthermore, as McMillen (2008) argues, other alternatives like non-parametric estimator are more difficult to apply and have few advantages when the nonlinearity is restricted to a single variable.

Let x denote the distance variable, which is separated in equal intervals and a cubic function is applied to each region. Define as knots the boundaries between the intervals. Let x_0 be the minimum value, x_1,x_2 and x_3 the knots, and x_4 the maximum value. Therefore, the distance between each knot is defined as $(x_4 - x_0)/4$. We can represent the splines cubic function as

$$
\begin{aligned}
g(DCBD_i) \approx \alpha + \beta_1(x_i - x_0) + \beta_2(x_i - x_0)^2 + \beta_3(x_i - x_0)^3 + \\
\gamma_1 D_1(x_i - x_1)^3 + \gamma_2 D_2(x_i - x_2)^3 + \gamma_3 D_3(x_i - x_3)^3 + \epsilon_i
\end{aligned} \tag{4}
$$

Where D_k are dummy variables that equals one when $x_i > x_k$ for $k = 1,2,3$. After obtaining $g(DCBD_i)$ from equation 4, the equation 3 is estimated by ordinary least squares (OLS).

To avoid multicollinearity problems due to many subcenter candidates, the final list of subcenters is obtained from a reverse stepwise regression procedure. Initially, the equation 3 is estimated with all potential subcenters and the variable whose coefficient has the smallest t value is then eliminated. The reduced equation is estimated again, and this routine is repeated until all distance variables are positive and significant at the 20% level. Therefore, the final list of subcenters includes the locations with positive coefficients on either δ_{1j} or δ_{2j} at the end of the procedure.

Notice that, for comparison purposes, we also estimate two more benchmark models in the literature on urban economics (McMillen, 2008; Krehl, 2018). The idea is to evaluate whether the two-stage procedure improves outcomes and the ability to explain the spatial distribution of activities. The first model consists of a linear relationship between the employment density and the distance to the CBD, as described in the following equation

$$
y_i = \beta_0 + \beta_1 DCBD_i + u_i \tag{5}
$$

The second model is a linearized version of the negative exponential model, which is given by

$$
\ln(y_i) = \beta_0 + \beta_1 DCBD_i + u_i \tag{6}
$$

Both equations 5 and 6 are estimated by OLS. Furthermore, we also consider the cubic splines model described by equation 4 and the second stage without the nonparametric part.

Finally, we also compare our results with those obtained when applying the simpler and traditional strategy proposed by Guiliano and Small (1991). Briefly, these authors proposed identification of subcenters through thresholds limits for total emplyoment and density. Specifically, after buiding spatial grids of 1 squared km in the metropolitan phisical areas, we adopt their values of a total of minimum of 10,000 emplyoments and of 2,500 emplyoments/squared, respectively, for total emplyoment of a subcenter and for the grid density belonging to a subcenter.

3 Data

Our main data source for identifying employment subcenters is the *Relação Anual de Informações Sociais* (RAIS) for the year of 2015. This is an administrative dataset maintained by the Brazilian Ministry of Labor. The RAIS consists of a high-quality panel with information about the characteristics of contracts between firms and employee, for all individuals formally employed, disaggregated at the municipality level. It is determined by law that firms and workers to fill in the report annually, with some penalties in case of non-compliance, such as loss of benefits granted to the firms. Thus, agents have incentive to provide accurate information.

More precisely, we use the georeferenced microdata from RAIS, which contains information about the address of each establishment registered. The georeferenced dataset was obtained from the Institute of Applied Economic Research (IPEA). Using the location of each firm, we obtain information regarding the employment location. It is worth emphasizing that one limitation of RAIS is the lack of information on workers who are not formally employed. This is an important feature in the Brazilian context, where informality rates exceed 40% of all workers. However, there are two factors that reduce such adversity in our study. First, and more importantly, the rate of formal jobs in metropolitan regions is higher than the national average. According to the *Pesquisa Nacional de Amostra por Domicílio* (PNAD) from 2015, the percentage of formal employees in the metropolitan region of Recife is 62,78%, 62,71% in Salvador, 68,25% in Rio de Janeiro, 73,50% in Porto Alegre, and 72,84% in São Paulo, while the national average is 57%. Second, the spatial distribution of the informal workers is like the formal workers, therefore, there is not considerable loss of information when we identify the employment subcenters.

We choose the metropolitan regions of Recife, Salvador, Porto Alegre, Rio de Janeiro, and São Paulo in order to obtain a comprehensive profile of the Brazilian Metropolitan Regions. Opreationally, we divide the metropolitan regions into grid cells of $1km^2$ to avoid possible endogeneity problems between choosing different geographic units (for example, political administrative disctricts). Consequently, the number of employees in each grid equals employment density. Table 1 provides some summary data on the study regions.

Table 1: Descriptive Statistics of Study Regions

	Recife	Salvador	P. Alegre	R. de Janeiro	S. Paulo
Area (km^2)	2770	4375	10346	6744	7946
Number of Cities	14	13	34	21	39
Employees	809,414	771,647	1,039,716	2,778,799	6,280,832
Residents (millions)	4.044	3.899	4.317	12.699	21.571
Number of grids with employment	684	555	1640	2406	3610
Average employment density	1183	1390	633	1165	1739
Median employment density	197.5	250	106	153	306.6
Share of core city employment in total employment (%)	60.45	75.24	53.30	74.37	67.23

Source: Elaborated by the authors.

Metropolitan regions area ranges from 2,770 square kilometers in Recife to 10,346 square kilometers in Porto Alegre. The metropolitan region of São Paulo has both the largest number of formal workers with approximately 6.3 million, as well as the largest population, with 21.5 million inhabitants. Moreover, the metropolitan regions of Recife and Salvador have similar patterns of both employment and population. Finally, one characteristic of the metropolitan regions is that formal employment is highly concentrated in the core cities of each region. This aspect is also evidenced by Fernández-Maldonado et al. (2014) and Ingram and Carroll (1981), who showed that Latin American cities have a higher concentration of employment around the CBD.

4 Results

4.1 General results

Table 2 presents the main results obtained from the estimation of the two-stage procedure proposed by McMillen (2001). The first row presents the results of the first stage, where a locally weighted regression is estimated. We identify 37 potential subcenters in Recife metropolitan region, 38 in Salvador, 81 in Porto Alegre, 96 in Rio de Janeiro, and 125 in São Paulo. The second of row of Table 2 reveals a total of 15 subcenters in Recife, 13 in Salvador, 55 in Porto Alegre, 52 in Rio de Janeiro, and 84 in São Paulo metropolitan region remain statistically significant at the end of the second stage.

Furthermore, the descriptive evidence shown in Table1 suggests a high employment concentration in the core cities of the metropolitan regions. Indeed, rows 4-5 of Table 2 suggest that subcenters located in the core cities, and thus close to the CBD, have much higher employment density than subcenters located in peripheral municipalities. These findings indicate that subcenters located farthest from the CBD have only local influence on the employment density surface, while subcenters near the CBD have global influence. Note, also, that numbers of subcenters identified in this research are much higher than the ones obtained by apllying the strategy of Guiliano and Small (1991) (presented in the lat line of Table 2), This result suggests that this last approach is not able to identify emplyoment agglomerations that importantly affect the urban centers structures. Actually, we note that this common approach appear able to identify only emplyoment subcenters closeer to the core city.

Table 2: Identifying subcenters - Two-stage procedure estimation results

	Recife	Salvador	P. Alegre	R. de Janeiro	S. Paulo
Potential subcenters (first stage)	37	38	81	96	125
Subcenters (second stage)	15	13	55	52	84
Number of subcenters in the core city	5	6	13	18	25
Average employment density in the core city	3678.2	3081.2	2069.8	5064.1	2441.2
Average employment density out the core city	971.5	934.4	619.7	1557.2	1980.4
Subcenters - Guiliano and Small (1991)	6	7	3	16	21

Source: Elaborated by the authors.

Moreover, Table 3 provides an overview of the estimated models and their respective goodness-of-fit measures. Rows 1-2 consists of R^2 adjusted obtained from estimating equations 5 and 6, respectively, which compare the two benchmark models on literature. Row 3 shows the adjusted R^2 from estimating the cubic spline model, as described by equation 4. Finally, rows 4-5 provides the R^2 adjusted obtained from estimating the two-stage model, without $g(DCBD_i)$ and complete, respectively.

Table 3: Comparing goodness-of-fit among different models - R^2 adjusted

	Recife	Salvador	P. Alegre	R. de Janeiro	S. Paulo
OLS	0.008	0.088	0.052	0.076	0.144
Negative exponential model	0.030	0.194	0.067	0.161	0.401
Cubic splines	0.215	0.331	0.140	0.194	0.464
Second stage without $g(DCBD_i)$	0.413	0.444	0.405	0.386	0.637
Second stage complete	0.426	0.505	0.460	0.417	0.663
Subcenters - Guiliano and Small (1991)	0.214	0.407	0.314	0.326	0.214

Source: Elaborated by the authors.

The first evidence that emerges when analyzing Table 3 is that the fit of the model is much smaller when considering the urban structure as purely monocentric, for all metropolitan regions, i.e., the linear relationship between the density and the distance to CBD is insufficient to explain the pattern of employment density of urban centers. When considering a non-linear relationship, such as cubic splines, the explanatory power of the model increases compared to linear models. However, when considering the distance to the employment subcenters identified in this study, the adjusted R^2 increases considerably, indicating the importance of considering these secondary centers when modeling the employment density surface. For all study regions, the results obtained using the two-stage procedure proposed by McMillen (2001) presented the highest goodness-of-fit measure among all models evaluated. Note that this comparasion includes the regressions using the emplyoment subcenters identified by apllying Guiliano and Small (1991) aproach (last line of Table 3).

Based on these results, we find evidence that the nonparametric approach is more suitable for subcenter identification in the Brazilian metropolitan regions. We identify statistically significant subcenters for all metropolitan regions under study (row 2 in Table 2). Considering the model fit of the second stage without the CBD variable $g(\cdot)$ we obtain considerable explanatory power regarding the employment density surface in each study region (row 4 in Table 3). In addition, we also highlight the heterogeneity of the influence of theses subcenters. While subcenters located near the CBD presents higher employment density and global influence, subcenters situated in peripheral regions presents lower employment density and local influence. These differences can be associated as peculiarities of the Brazilian metropolitan regions, that we briefly discussed later.

Finally, despite the importance of the employment subcenters, we can not neglect the CBD relevance, since it concentrates a significant fraction of the total employment in the metropolitan regions. The metropolitan areas of Recife and Salvador, both in the Northeast, have about 26% of total employment concentrated in the CBD. In turn, the metropolitan region of Porto Alegre has approximately 20%, while the regions of Rio de Janeiro and São Paulo have 16.09% and 13.47%, respectively. Besides suggesting some regional differences (that we discuss in the next section), these numbers are quite expressive if we consider that the CBD corresponds to less than 1% of the total area of the metropolitan regions.

4.2 Subcenters in Brazilian Metropolitan Regions

The set of Figures 1-5 presents the locations of the subcenters (together with the CBD) of the five Brazilian metropolitan regions analysed in this research. We describe each of the results in the following paragraphs.

Figure 1 presents the 15 subcenters identified for the Recife Metropolitan Region. Notice that most secondary employment centers are located in the contiguous region of Olinda (two subcenters), Recife (five subcenters) and Jaboatão dos Guararapes (one subcenter), the most economically important municipalities of the RMR. We also identify smaller sub-centers in the cities of Ipojuca and Cabo de Santo Agostinho (in the south), which are associated with the presence of the Suape harbor complex. Furthermore, the five subcenters located near to the CBD of Recife (e.g. close the neighborhoods of Casa Forte and Boa Viagem, in Recife) have higher employment density, while subcenters located in the more peripheral regions have lower employment density, in accordance with the results found in our semiparametric estimation.

Figure 2 shows the locations of the 13 subcenters of Salvador Metropolitan Region. Like the RMR, most of the identified subcenters are located in the main economic areas, being these the municipalities of Salvador (6 subcenters), Lauro de Freitas (2 subcenters), and Camaçari (one subcenter). Due to the large employment concentration in the city of Salvador,

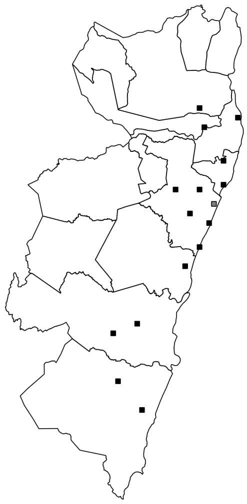

Figure 1: Subcenters in Recife Metropolitan Region

10 of the 13 subcenters are located near the CBD in the Southeast of RMSA. We also identify smaller subcenters in the municipalities of Candeias, Simões Filho, and Dias D'Avila.

Figure 2: Subcenters in Salvador Metropolitan Region

Figure 3 displays the 55 subcenters identified for the Porto Alegre Metropolitan Region. Most subcenters are located in a center-north region relative to the CBD. As we argued before, the spatial distribution of emplyoment of here is quite different from those of the two northern regions, being the municpaility of Porto Alegre (the capital) relatively less important. Actually, it is possible to highlight the formation of two groups of subcenters. The first are those subcenters situated in the municipalities of Porto Alegre, Canoas and Gravataí, again, the economic center of the RMPA. These cities have a total of 20 subcenters in the center-south region. The second group is located in the northern region of the RMPA, composed by the municipalities of Novo Hamburgo, Ivoti, Campo Bom, Dois Irmãos, and

Estância Velha, with a total of 10 subcenters. Other subcenters near the CBD are located in Viamão and Guaíba.

Figure 3: Subcenters in Porto Alegre Metropolitan Region

Figure 4 presents the 52 subcenters identified for the Rio de Janeiro Metropolitan Region. Like the other Metropolitan Regions, a considerable number of subcenters are located in the core city of the metropolitan region, i.e., in Rio de Janeiro, which has a total of 21 subcenters. Furthermore, our results also show a high number of subcenters situated in the eastern region, composed by the municipalities of Niterói, São Gonçalo, and Itaboraí, with a total of 10 subcenters. Finally, we also have an agglomeration of subcenters in the north of Rio de Janeiro, with emphasis on the cities of Nova Iguacu, Duque de Caxias and São João de Meriti, which together have 13 subcenters.

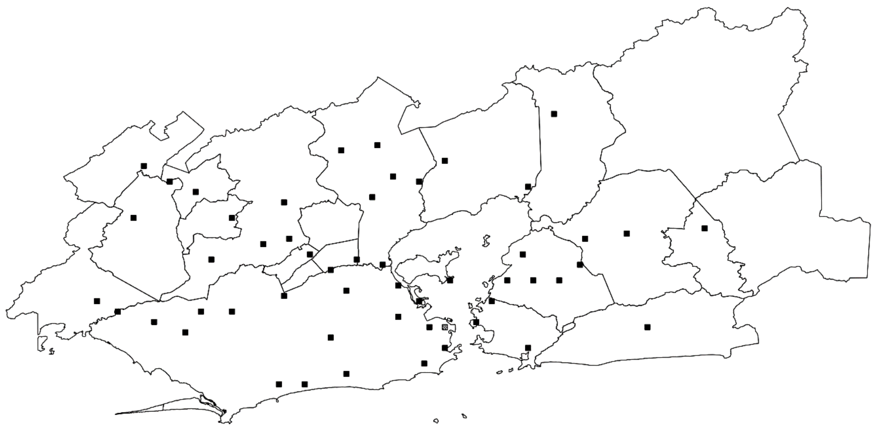

Figure 4: Subcenters in Rio de Janeiro Metropolitan Region

Finally, for the São Paulo Metropolitan Region, we identify a total of 84 subcenters, which are distributed heterogeneously among the five sub-regions, as displayed in Figure 5. Again, a large number of subcenters are located in the central region of São Paulo, close to the Consolação and Sé districts. However, unlike the other study regions, the RMSP does not present a large discrepancy between the employment density of the subcenters inside and outside the core city. This fact is due to the existence of large industrial centers located in the municipalities neighboring São Paulo. We can highlight, for example, the industrial complex of Paulista ABC, located in the southeast and composed by the municipalities of Santo André, São Bernardo do Campo, and São Caetano do Sul. The RMSP has also important industrial complexes in the municipality of Osasco, in the west, and in Guarulhos, in the east, which are regions that have a high number of subcenters with high employment density.

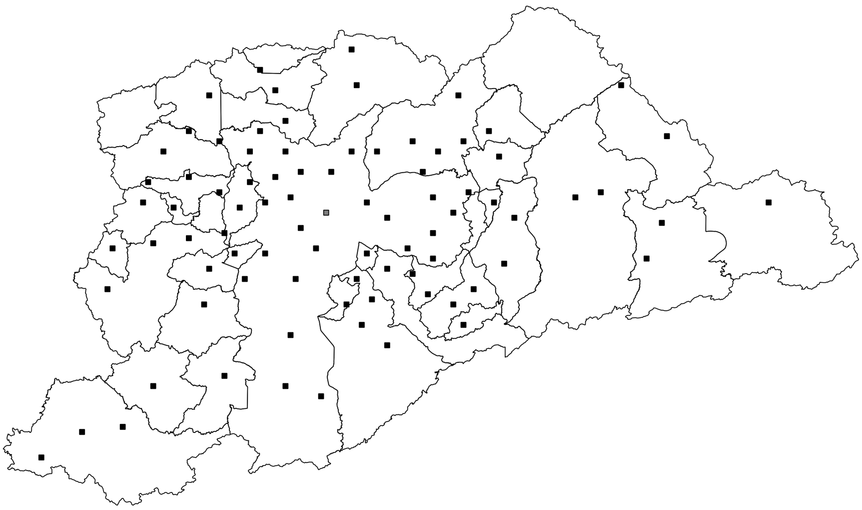

Figure 5: Subcenters in São Paulo Metropolitan Region

5 Discussion

As previously described, Brazilian cities present a set of peculiarities that resembles those of Latin America urban centers. These characteristics distinguish them from American and European cities, which are the focus in most of the literature dealing with the identification of employment subcenters (Guiliano and Small, 1991; McMillen, 2001; Krel, 2018). Hence, it is natural to ask whether and to what extent the results that we found for the pattern of the employment distribution in Brazilian metropolitan regions differ from those found for developed countries and correspond to the evidence available for other Latin American cities.

Considering the number of subcenters recently identified for large cities in Germany by Krehl (2018) apllying the same methodology, in general, we found a larger number of subcenters in Brazilian urban agglomerations. Thus, our results is more similar to those obtained by MacMillen (2001) for US metropolitan regions. More speficially, while Krehl (2018) identifies a total of 15 subcenters in Munich, 10 in Cologne, 8 in Stuttgart and 16 in Frankfurt, McMillen (2001) finds a total of 33 subcenters in Chicago, 28 in Dallas, 25 in Houston, 19 in Los Angeles, 2 in New Orleans and 22 in San Francisco. Note also that, despite using different approaches, we found a number of subcenters for the Recife and Salvador

Metropolitan Region that is similar to the number of subcenters identified by Fernandez-Maldonado et al. (2015) for Fortaleza (where 11 subcenters were identified by the authors), another important metropolitan region of the Northeast of Brazil.

These Brazilian difference relative to Germanian evidence and similarity with respect to US pattern are consistent with the relation bewteen employment and population sprawls in the respective enviroments. As argued by Krehl (2018), different from US urbanization, in German, the more restrict urban land regulation implied a less heterogeneous population distribution within the cities and contributed for more emplyoment sprawl. This has generated a smaller difference bewteen spatial emplyoment and population distributions within cities and, thus, lower number of emplyoment subcenters. Following more the US pattern of land regulation, Brazilian metropolis present more salient diference between population and emplyoment distributions and, thus, a higher number of emplyoment subcenters.

Actually, Figures 6-10, which show the number of employees per number of residents in each grid (the ratio increases as the grid colors changes from orange to red), make clear the above mentioned difference between population and emplyoment distributions for Brazilian metroplitan regions. These figures show the presence of regions with the greater relative concentration of employment, mainly in the central regions. The Brazilian situation is thus quite different from the strong employment sprawl found by Krehl (2018), wherein only in a few areas the number of workers exceeds the resident population.

More specifically, Figures 6 and 7 reveals that, in the Recife and Salvador metropolitan regions, respectively, there is a very large concentration of employment in relation to the number of residents in the regions closer to the CBD. Figures 8 and 9 show the employment rate per resident population for the metropolitan areas of Porto Alegre and Rio de Janeiro, respectively. Besides the large concentration of employment in the central areas, the RMRJ has a great share of employment per residents in the south, while the RMPA has a high concentration of employment in the north. Finally, Figure 10 indicates that São Paulo metropolitan region presents several centers with high employment density, being these situated in the central area, and in the regions of ABC Paulista (southeast), Osasco and Barueri (west), and in Guarulhos (east). Therefore, we can conclude that, in Brazil, when considering the population distribution, employment is clearly concentrated. This structure is much closer to that found for the United States cities and quite different from the more decentralized pattern in the German cities.

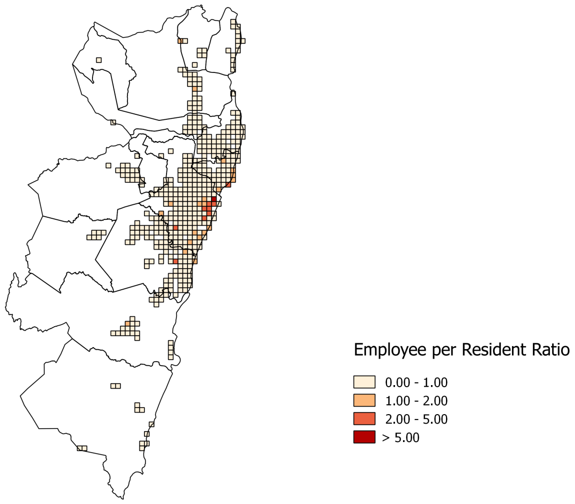

Figure 6: Employees per number of residents - RMR

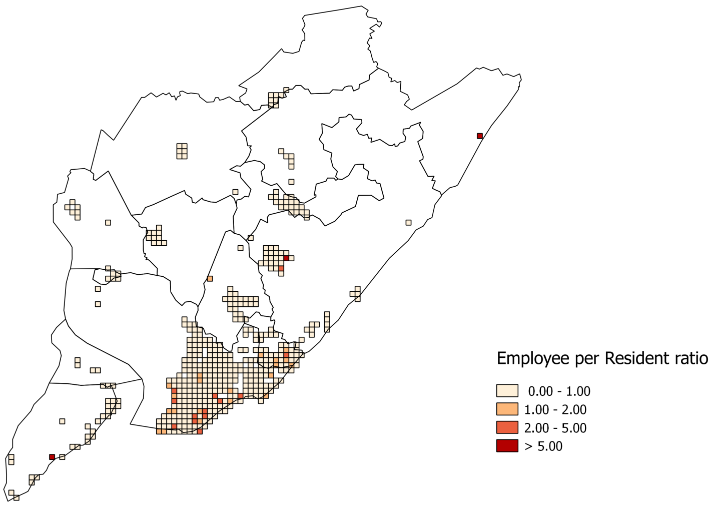

Figure 7: Employees per number of residents - RMSA

Although some similarities of emplyoment distribution when compared to US experience, there are also important particular characteristics of the Brazilian urbanization that certaintly contribute for the identified patterns. These characteristics, in essence, are associated with Brazilian lower economc development and higher level regional disparities. Regarding to last this point, note, for example, that the above set of evidence also indicates that there are also important regional differences among the observed patterns of emplyomet distribution across Brazilian metropolitan regions. First, notice that there is great similarity in such standards for the MR of the Northeast of Brazil (Recife and Salvador). Such patterns of emplyoment distribution clearly differs, on the other hand, from the situation found for Porto Alegre, despite their similar population sizes. In addition, it easy to observe that the largest and economically strongest Brazilian MR (São Paulo and Rio de Janeiro) also have the highest

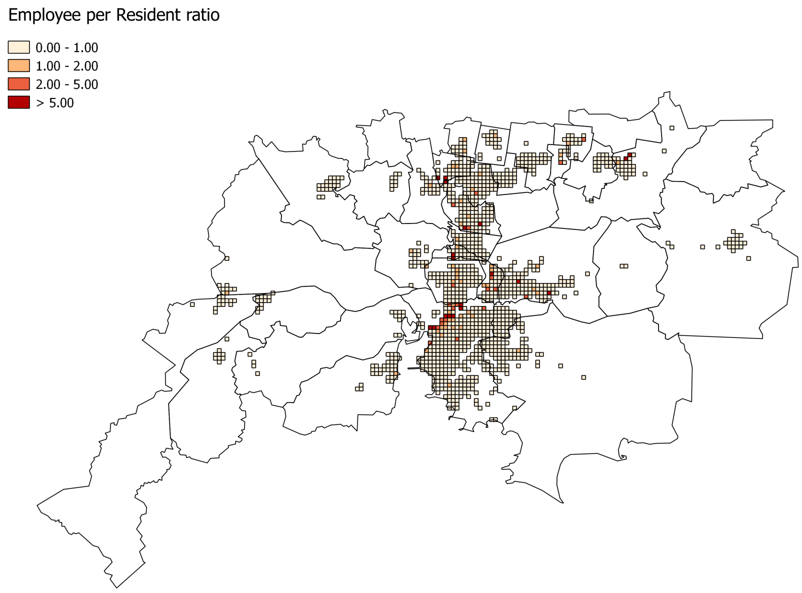

Figure 8: Employees per number of residents - RMPA

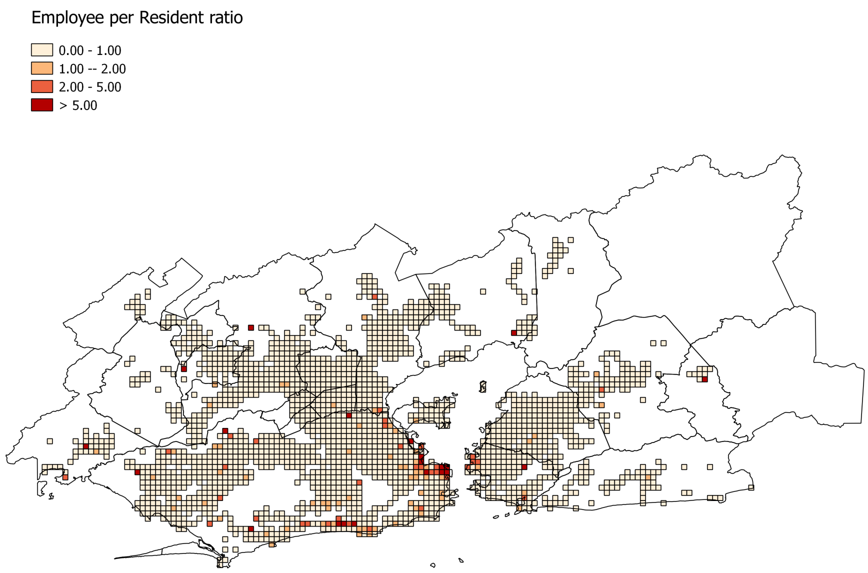

Figure 9: Employees per number of residents - RMRJ

number of subcenters, as somehow expected. Even though not corroboratory (a task beyond the purposes of this research), we argue that these results are also in clear agreement with the general and regional specificities of Brazilian urbanization.

We initially underline the precarious conditions of urban mobility that afflicts the great cities of Brazil, which is a consequence of the low investment in roads and collective transportation and the still relatively low-income levels that prevents the widespread use of individual vehicles (Pero and Stefaneli, 2015). The commuting problems in the Brazilian urban centers seem extremely relevant for the understanding of the spatial distribution of occupations.

The last two decades presents a fast increase of the automotive fleet, when the number of vehicles increases from 19.9 million in 2000 to 40 million in 2012, according to Martine et al

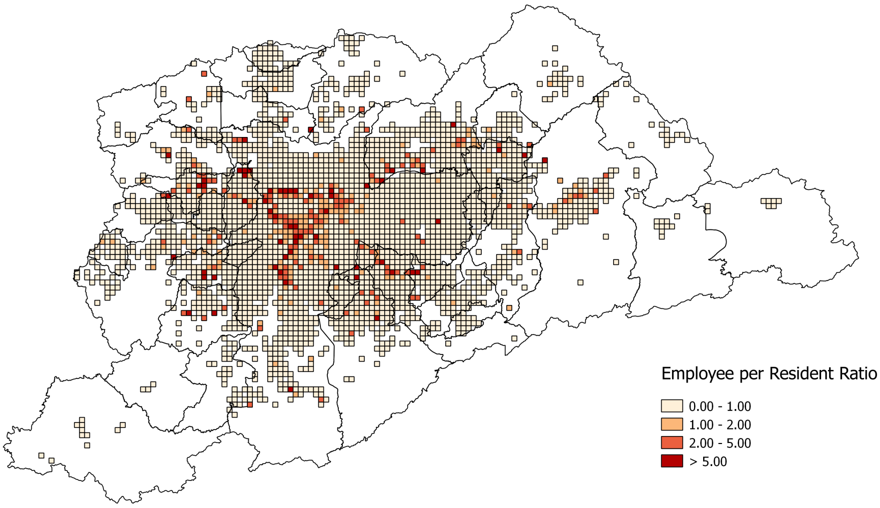

Figure 10: Employees per number of residents - RMSP

(2012). This scenario, encouraged by a greater credit supply, tax reductions on industrialized products, and fuel price freezes, generated a significant increase in the commuting time to work. Pero and Stefanelli (2015) showed that workers situated in São Paulo and Rio de Janeiro metropolitan region have the highest commuting time in Brazil. Apart from that, Recife and Salvador are among the metropolitan regions with the greatest growth rate of the commuting time, being 17.8% and 27.1%, respectively. On the other hand, the Porto Alegre metropolitan region presents the shortest commuting time among Brazilian MR. According to Pereira and Schwanen (2013), a possible explanation is the better distribution of economic activities, controlled urban expansion, and more efficient public transport systems.

Notice that, consistent with Fujita and Oghawa (1982) and Fujita and Thisse (2013), in such a scenario of increasing difficulties of urban mobility, the rise in the daily commuting cost for workers and expensive interaction among firms lead to a trend towards the decentralization of economic activities and weakening of the initial monocentric configuration in the urban centers. This implication is entirely consistent with the presence of a large number of subcenters in the Brazilian metropolitan regions. In other words, the deficiencies regarding the transportation infrastructure present in the Brazilian urban centers, which reflects the low investment in public transport, potentially exert a pressure on the urban roads as the income expansion tends to be accompanied by a fast expansion of the fleet, resulting in higher mobility costs. Such movements, therefore, hinder the employment concentration in few localities.

But notice that this potential push for employment decentralization is also attenuated by the action of urban centripetal forces. In the last half of the 20th century, Brazilian urban centers experienced an intense migrant flow from rural areas. The fast growth of the metropolitan population in a few decades prevented the cities from providing the necessary infrastructure to accommodate the new residents. As a consequence, migrants with fewer resources settled in the periphery or in empty central areas, such as hills and steep slopes. These subnormal clusters, later called favelas (Nadalin, 2018), are inadequate for formal urban occupation because they do not follow urban planning guidelines. Hence, they end up lacking the provision of essential services such as sanitation, education, and security. The expansion of these regions reflects the main socio-economic problems of the country, such as high levels of income inequality and high rates of urban violence.

Recently, Nadalin (2018) showed the existence of a high heterogeneity concerning the favelas spatial distribution among Brazilian metropolitan regions. In Recife and Salvador, favelas are situated in the central regions, with 23.9 % and 26.9 % of the population of these MR, respectively, living in subnormal clusters. The metropolitan regions of Rio de Janeiro and São Paulo have spatially dispersed favelas, with 14.4 % and 10.8 % of the residents located in these areas, respectively. Finally, the Porto Alegre metropolitan region, characterized by above-average social indices, presents only 7.4 % of the population living in informal housing.

Actually, it is worth mentioning a chronic problem of many Brazilian urban centers: the lack of infrastructure and the precarious provision of essential basic services. Table 4 presents the percentage of households with access to basic sanitation or garbage collected daily for the metropolitan areas under study. The results indicate important regional differences in the endowments for these services.

Table 4: Urban enviroment quality - Share of households with access to basic sanitation, daily garbage collected, and living in slums (%)

	Basic Sanitation	Daily Garbage Collected	Presence of Slums
Recife	57.1	85.8	23.9
Salvador	91.6	54.3	26.9
Porto Alegre	89.2	95.5	7.4
Rio de Janeiro	93.6	90.0	14.4
São Paulo	95.1	93.4	10.8

Source: Elaborated by the authors based on the PNAD 2015 and Demographic Census of 2010

As can be seen in Table 4, the metropolitan areas of Recife and Salvador, both in the Northeast, present a serious problem of home service infrastructure deficit, with only 57.1% of RMR households with access to basic sanitation and 54.3% of RMSA households with garbage collected daily. Such deficiencies certainly represent resistance to the decentralization of economic activities and employment since they limit the qualified urban space with the necessary infrastructure to productive activities.

To sum up, the lack of adequated infrastructure put pressure on urban roads which, in turn, can not accommodate the large fleet of vehicles. Aligned to the low investment in public transport, this generates a high commuting cost. These facts appear to encourage the decentralization of economic activities and, consequently, the formation of new employment subcenters. In this perspective, the high number of subcenters found in the São Paulo and Rio de Janeiro metropolitan regions is consistent with the costly commuting. Concerning the Northeast metropolitan regions, note that both Recife and Salvador metropolitan area were primarily monocentric, since their economies were established around the port of their respective cities. In addition, as evidenced by Table 4, both MR present deficiencies in the provision of home service infrastructure, which acts as a counterforce to the employment decentralization. Thus, in spite of the growth of commuting in these regions, which stimulates the decentralization and the emergence of subcenters, the presence of restraining forces ends up slowing this process. Such movement can be evidenced by the large share of employments situated in the CBD.

These general characteristics of Brazilian urban centers do not seem to easily explain, however, the formation of subcenters in the Porto Alegre metropolitan region. The stability of commuting time aligned to a lower presence of informal dwellings suggest a more balanced expansion of the urban center. Therefore, the high number of subcenters found for the RMPA may reflect other more particular issues, such as the historical process of colonization of the region characterized by stronger presence of more autonomous foreign communities in different municipalities.

6 Concluding Remarks

Result of a strong rural-urban migration process, Brazil is currently a remarkably urban country, with a great share of the population living in the cities. This fact implies that the quality of life of his residents is directly related to the urban infrastructure and its services. In this sense, there is a robust set of evidence that exposes the influence of the employment centers and subcenters on the factors associated with the quality of life in the cities. Despite this, there are few studies that analyze the spatial distribution of occupations in Brazilian urban centers.

The main goal of this paper was to identify urban employment subcenters for some of the most important metropolitan regions in Brazil. Based on the georeferenced employment information provided by the Institute of Applied Economic Research (IPEA), we use a two-stage procedure proposed by McMillen (2001) and we identify a total of 15 subcenters in the Recife Metropolitan Region, 13 in Salvador, 55 in Porto Alegre, 52 in Rio de Janeiro and 84 in São Paulo. Moreover, our results suggest that the semiparametric approach is the most adequate to identify the employment subcenters in the Brazilian metropolitan regions since it generates the highest explanatory power when compared to other standard models.

In spite of the great employment concentration near the CBD, the Brazilian metropolitan regions present a decentralization trend of the economic activities. Significant differences in the R2 adjusted when including distance variables to subcenters in the model indicate the presence of multiple employment subcenters and their importance to characterize the pattern of employment density in the study regions. Different from the developed country cities previously studied in the literature, Brazilian urban centers have a strong presence of informal housing, disparities in urban infrastructure among cities and high commuting costs, mainly in the Northeast region of Brazil. Thus, these peculiarities generate an incentive to the decentralization of occupations and it is consistent with the high number of subcenters identified and their relative importance. However, the higher infrastructure deficit of the domiciliary services in the metropolitan areas of Recife and Salvador appears to act as deterrent force avoiding greater employment decentralization.

Future research related to this topic involves the expansion of the identification of employment subcenters for the other Brazilian metropolitan regions, such as Belo Horizonte, Curitiba, and Fortaleza. In addition, the investigations may characterize subcenters regarding the type of sectoral activity predominant in each locality and verify how these subcenters may influence commuting costs, verticalization, land use patterns, spatial mismatch, and population distribution in Brazilian cities.

References

Ahlfeldt, G. M., & McMillen, D. P. (2015). The vertical city: the price of land and the height of buildings in Chicago, 1870-2010.

Anderson, J. E. (1982). Cubic-spline urban-density functions. Journal of Urban Economics, 12(2), 155-167.

Baumont, C., Ertur, C., & Le Gallo, J. (2004). Spatial analysis of employment and population density: the case of the agglomeration of Dijon 1999. Geographical analysis, 36(2), 146-176.

Barr, J., & Cohen, J. P. (2014). The floor area ratio gradient: New York City, 1890–2009. Regional Science and Urban Economics, 48, 110-119.

Cervero, R., & Wu, K. L. (1997). Polycentrism, commuting, and residential location in the San Francisco Bay area. Environment and Planning A, 29(5), 865-886.

Craig, S. G., & Ng, P. T. (2001). Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. Journal of Urban Economics, 49(1), 100-120.

Da Mata, D., Deichmann, U., Henderson, V. J., Lall, S. V., & Wang, H. G. (2005). Determinants of city growth in Brazil. The World Bank.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Handbook of regional and urban economics (Vol. 4, pp. 2063-2117). Elsevier.

Edlund, L., Machado, C., & Sviatschi, M. M. (2015). Bright minds, big rent: gentrification and the rising returns to skill (No. w21729). National Bureau of Economic Research.

Fernández-Maldonado, A. M., Romein, A., Verkoren, O., & Parente Paula Pessoa, R. (2014). Polycentric structures in Latin American metropolitan areas: Identifying employment subcentres. Regional Studies, 48(12), 1954-1971.

Fujita, M., & Thisse, J., Economics of agglomeration: cities, industrial location, and globalization. Cambridge university press, 2013.

Fujita, M., & Ogawa, H. (1982). Multiple equilibria and structural transition of nonmonocentric urban configurations. Regional science and urban economics, 12(2), 161-196.

Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. Regional science and urban economics, 21(2), 163-182.

Glaeser, E. & Maré, D. (2001). Cities and Skills. Journal of Labor Economics, 19(2), 316-342.

Gobillon, L., Selod, H., & Zenou, Y. (2007). The mechanisms of spatial mismatch. Urban studies, 44(12), 2401-2427.

Ingram, G. K., & Carroll, A. (1981). The spatial structure of Latin American cities. Journal of urban Economics, 9(2), 257-273.

Krehl, A. (2018). Urban subcenters in German city regions: Identification, understanding, comparison. Papers in Regional Science, 97, S79-S104.

Lopes Pereira da Silva, R. & Silveira-Neto, R. & Rocha, R. (2019). Localization patterns within urban areas: evidence from Brazil. Area Development and Policy. 4. 1-20.

McDonald, J. F. (1987). The identification of urban employment subcenters, Journal of Urban Economics 21,242-258.

McMillen, D. P., & McDonald, J. F. (1997). A nonparametric analysis of employment density in a polycentric city. Journal of Regional Science, 37(4), 591-612.

McMillen, D. P., & McDonald, J. F. (1998). Suburban subcenters and employment density in metropolitan Chicago. Journal of Urban Economics, 43(2), 157-180.

McMillen, D. P. (2001). Nonparametric employment subcenter identification. Journal of Urban economics, 50(3), 448-473.

McMillen, D. P. (2006). Testing for monocentricity. A Companion to Urban Economics. Oxford: Blackwell, 128-140.

Moretti, E. (2004). Workers' education, spillovers, and productivity: evidence from plant-level production functions. American Economic Review, 94(3), 656-690.

Moretti, E. (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of econometrics, 121(1-2), 175-212.

Nadalin, V. G., & Mation, L. F. (2018). Localização intraurbana das favelas brasileiras: o papel dos fatores geográficos. IPEA.

Pero, V., & Stefanelli, V. (2015). A questão da mobilidade urbana nas metrópoles brasileiras. Revista de Economia Contemporânea, 19(3).

Ramos, F. R. (2014). Três Ensaios sobre a Estrutura Espacial Urbanas em Cidades do Brasil Contemporâneo: economia urbana e geoinformação na construção de novos olhares. Tese de Doutorado, São Paulo, Fundação Getúlio Vargas.

Firmino Costa da Silva, D., Elhorst, J. P., & Silveira Neto, R. D. M. (2017). Urban and rural population growth in a spatial panel of municipalities. Regional Studies, 51(6), 894-908.

Song, S. (1994). Modelling worker residence distribution in the Los Angeles region. Urban studies, 31(9), 1533-1544.