Publication status: This preprint has been published elsewhere.
DOI of the published preprint: https://doi.org/10.1590/0103-6351/6236

Employment sub-centers of a megacity in a developing country: the case of the Municipality of São Paulo, Brazil

Rodger Barros Antunes Campos, André Luis Squarize Chagas

https://doi.org/10.1590/0103-6351/6236

Submitted on: 2021-12-22
Posted on: 2021-12-22 (version 1)
(YYYY-MM-DD)

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Employment sub-centers of a megacity in a developing country: the case of the Municipality of São Paulo, Brazil

Identificação de subcentros de emprego: o caso do Município de São Paulo, Brasil

Rodger Barros Antunes Campos[1]
André Luis Squarize Chagas[2]

[1]Urban and Regional Studies Lab
[2]University of São Paulo, Department of Economics

Abstract: Theoretical models concerned with multiple centers have been brought to the debate on sprawling urban employment. However, empirical methods that identify these centers are not a specific focus of the specialized literature. The purpose of this paper is to apply a new empirical methodology in the identification and characterization of urban employment subcenters (Small Business Districts, SBD) using the case of the Municipality of São Paulo. To this end, we adopted a two-step approach, developing: 1) Exploratory Spatial Data Analysis and 2) Spatial Hedonic Price Model. As a result, we found seven regions that can be considered SBDs.

Keywords: Urban Economics, Center Business District, Subcenter Business District, Spatial Econometrics
JEL Codes: R32, C21

Resumo: Modelos teóricos preocupados com centros múltiplos foram trazidos para o debate sobre o espraiamento do emprego urbano. No entanto, os métodos empíricos que identificam esses centros não são um foco específico da literatura especializada. O objetivo deste artigo é aplicar uma nova metodologia empírica na identificação e caracterização de subcentros de emprego urbano (Pequenos Distritos Comercial, SBD) utilizando o caso do Município de São Paulo. Para tanto, adotamos uma abordagem em duas etapas, desenvolvendo: 1) Análise Exploratória de Dados Espaciais e 2) Modelo de Preço Hedônico Espacial. Como resultado, encontramos sete regiões que podem ser consideradas SBDs.

Palavras-chave: Economia Urbana, Distrito Comercial Central, Distrito Comercial Subcentro, Econometria Espacial

Códigos JEL: R32, C21

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

1. Introduction

Employment decentralization from within the city is one of the changes which foster urban landscape modification, influence wages, housing prices, transportation demands, and worker commuting (Fujita and Ogawa, 1982; Anas and Kim, 1990; Helsley and Sullivan, 1991). The decentralization of employment stems from the suburbanization process (Glaser and Kahn, 2001). Such a process, where economic growth sprawls from the CBD towards the city's edge, utilizes comparative advantages in transportation costs, wages, and land prices (McMillen and Smith, 2003). McMillen (2001b) points out that city growth is a mechanism that promotes a large employment concentration outside of the CBD. This city pattern is typical in big American cities, where suburbs have families, offices, and employment. In some cities, the suburbs aggregate more workers than in the center (McMillen, 2001a). Notably, these new sub-centralities have a meaningful impact on the city structure. Urban Economics literature refers to these new employment clusters as employment subcenters or Small Business Districts (SBD).

SBDs generate competition for land among households and firms (Wheaton, 2004). The low-income family is impacted by this land-use dispute, causing a spatial location change from the new business region towards the city's perimeter due to the upward pressure on rent prices. Then, public transportation demand from the peripheral part of the city to the center increases, disfiguring the concentric pattern of transportation, which features a monocentric town. The commuting patterns become unclear because cross-commuting is one of the household's choices. The worker decision is not constrained to minimize commuting cost, but workers supply their labor force where their skills are required. Therefore, workers have no assurance about commuting distance minimization.

Therefore, identifying sub-centralities improves the knowledge relating microeconomic development issues, such as firm cost structure (congestion, labor force access, land prices, etc.) to workers' welfare and productivity (spatial segregation, commuting, wages, labor supply). Furthermore, SBD can be used as public policy stimulating the creation of new business subcenters, reducing congestions (approximating firms to workers), balancing land prices spatially (less pressure on prices in the CBD region), and reshaping public transport and taxation patterns.

Focusing on Brazil, the fifth-largest country in the world, with an urban population concentration of 84% in cities, 43 million inhabitants live within just 17 cities[a] and produce 34.7% of the Brazilian GDP (IBGE). Regarding the Municipality of São Paulo, 11.446 million inhabitants live within an area of 1,512

[a] Each city has no less than 1 million inhabitants.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

km² and, in 2014, the GDP of the MSP was US$ 237 billion; which was roughly 11.7% of the national GDP, ranking the city as the richest one among all municipalities in the country (SEADE, 2017). Meyer, Grostein, and Biderman (2004) pointed out the employment paths over the districts of São Paulo (MSP). Since the 1960s, service sectors have moved from the old center (Sé district) towards the west and south. Initially, such concentration was observed in the Jardim Paulista and then in the Cerqueira Cezar and Itaim Bibi districts. The latter received many firms due to the construction of the Faria Lima Avenue in the 1970s, which became one of the most important city avenues. From 1970 to 1990, employment sprawling continued toward the Vila Olimpia neighborhood, following Luis Carlos Berrini Avenue. Recently, 2000-2010, the firms are spreading toward the far south of the city and the west of the city zone (see Table 1).

Table 1: Employment Composition Rank

	2000				2012			
Rank	District	Zone Area	Total of Employ.	Relative Particip.	District	Zone Area	Total of Employ.	Relative Particip.
1º	Itaim Bibi	SZ1	139,679	5.96%	Itaim Bibi	SZ1	293,925	7.48%
2º	Jardim Paulista	CZ	108,027	4.61%	Santo Amaro	SZ1	159,349	4.05%
3º	Santo Amaro	SZ1	102,929	4.39%	Jardim Paulista	CZ	156,443	3.98%
4º	Vila Mariana	SZ1	81,897	3.49%	República	CZ	153,793	3.91%
5º	Sé	CZ	80,250	3.42%	Vila Mariana	SZ1	152,247	3.87%
6º	Pinheiros	WZ	80,184	3.42%	Bela Vista	CZ	117,479	2.99%
7º	República	CZ	76,757	3.27%	Pinheiros	WZ	113,005	2.88%
8º	Lapa	WZ	64,122	2.73%	Barra Funda	WZ	103,759	2.64%
9º	Moema	SZ1	64,075	2.73%	Lapa	WZ	102,899	2.62%
10º	Bela Vista	CZ	59,079	2.52%	Moema	SZ1	95,862	2.44%
	Total		2,345,294	-			3,930,013	-

Source: Own elaboration based on data from RAIS/MTE. Center Zone (CZ), South Zone 1 (SZ1), South Zone 2 (SZ2), North Zone (NZ), East Zone 1 (EZ1), East Zone 2 (EZ2), West Zone (WZ)

In this context, this paper aims to identify and characterize sub-centralities in the Municipality of São Paulo, providing the possibility of future assessments about employment decentralization and housing prices, commuting, transportation demands, and wage gradients.

Besides this introduction, the paper is organized into five sections. The next section discusses the monocentric and polycentric city, underlining the employment decentralization process. The third section shows the econometric identification strategy in two stages, the estimated results and SBD

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

characterization. After the main results, we discuss robustness tests. The last section provides final remarks, extension possibilities, and limitations.

2. Monocentric and polycentric cities: theoretical models and empirical strategy

Pioneer economists from the New Urban Economics developed a theoretical model in which job supply was concentrated in the Central Business District (CBD), i.e., a monocentric city. For example, this central assumption is based on the historical context that the city center was likely developed near one transportation node. Hence, results from the monocentric models hypothesize the CBD as a price-anchor, where the distance from the city center increases the negative effect on vector prices. Thus, a distance increase from the CBD can be seen as the transport cost increasing dually. The model's bottom line is that land price must be compensated by distance increases from the central city place to balance firms and households in this economy. In short, the endogenous variables from the model (population density, lot size, land price, wage, etc.) face an inverse spatial cone pattern, and they are estimated as a simple function of the anchor (CBD).

According to Romanos (1979), rent prices are underestimated in areas where sub-centers exist. Also, many cities have more than one business center and/or multiple subcenters. Additionally, as would be expected by monocentric models, property price gradient does not decrease monotonically, relative to the city's CBD, but in polycentric model does.

The theoretical literature on polycentric cities should be subdivided into two main strands. The first one deals with the issue of CBD and SBD as an endogenous problem to the optimization process, considering economies of agglomeration and transport cost (Ogawa and Fujita 1980, Fujita and Ogawa 1982, Anas and Kim 1990, Helsley and Sullivan, 1991, Henderson and Slade, 1993, Sasaki and Mun, 1996, Berliant and Konishi, 2000). The main interest is to assess how the city's new spatial pattern impacts the labor and housing market. The factors that lead to a pulverized pattern of workers stem from high transport costs, absence of agglomeration economies, direct sell to families, firms' demand for land, and higher price-elasticity of labor demand. The second group of models is interested in assessing the effect of employment suburbanization on residences' prices and other aspects (Romanos, 1979; Yinger, 1992; Wrede, 2015). These models focus on how workers decide where to live and work and, consequently, how the spatial pattern affects land price, population density, and work-home commuting. The results presented are the decay of land price and wages due to the distancing from the CBD and SBD. Thus, the SBD also constitutes an anchor of real estate prices and wages.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Firms face the tradeoff between localizing in SBD instead of CBD. Firms concentrated in CBD have an agglomeration economy advantage. They do not face friction that discourages firms from going to the suburbs, such as the difficulty of commuting and the low pool of labor supply (McMillen, 2001b).

The emergence of sub-centers is linked to the de-agglomeration process benefits, despite the advantages offered by CBD concentration. Firms in the suburbs can count on some advantages, such as lower cost of land, access to intercity and interstate roads, access to workers over longer distances, and lower wages (because the firm would not have to compensate for commuting). Communication would be viable if firms in the SBD were located close to the CBD and communication innovations were available. Although business diversity may be lower in SBD, the negative effect of non-diversity would be diluted if low transport costs and land offers near CBD were available. Another advantage of the sub-center size is that large SBD offering employment and shopping opportunities reduces household dependence on the center (McMillen, 2001b).

Regarding empirical research, there is no well-established empirical strategy of SBD identification. Initially, the studies used prior regional knowledge as identification, but the results did not fit the data, as McMillen (2001b) pointed out. More objective procedures can be divided into 1) *Cut-off values*: researchers impose some cut-off point on the density of jobs and the total number of jobs (McDonald, 1987; Giuliano and Small, 1991; McDonald and McMillen, 1998; Coffey and Shearmur, 2001; Baumont and Bourdon, 2002; López and Muñiz, 2005); 2) E*conometric models* (Craig and Ng, 2001; McMillen, 2001a, 2001b; McMillen and Smith, 2003); 3) *Spatial statistics* (Plaut and Plaut, 1998; Baumont et al., 2004; Herman and Haddad, 2005; Guillain et al., 2003, and Siqueira, 2014).

McDonald (1987) firstly proposed the cut-off method, applying for the Chicago metropolitan area, and Guiliano and Small (1991) extend the empirical approach, identifying 32 SBD in Los Angeles, United States. Under their method, a sub-center is a set of contiguous areas that have a minimum density of ten employees per acre and hold 10,000 employees. These models received several criticisms due to the discretion of the chosen locations, the sensitivity of the results to the cut-off, and the impossibility of generalization to any other city (Anas et al., 1998; McMillen, 2001b; Baumont et al., 2004).

The following two procedures are transformations of the pioneer method proposed by McDonald (1987). The method considers the residues of a log density specification of employment. The main contribution offered by Craig and Ng (2001) is to eliminate the problem of the discretion of the previous method. The authors use quantile regression focusing on the 95th percentile of the employment density distribution. They consider the rings that present high density and employment, given their knowledge

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

of the city. McMillen's (2001b) criticizes the estimation of symmetric density function around the CBD and the *a priori* researcher's knowledge of employment distribution in the city under analysis.

McMillen (2001a, 2001b) proposed a two-stage model that requires no prior knowledge of the evaluated city. The first stage is close to Craig and Ng (2001) approach, i.e., run employment density logarithm against the distance to the CBD. SBD candidates are those whose residues form clusters statistically significant at 5%. In the second stage, after considering the potential subcenters, a semiparametric model is used. The nonparametric part of the regression adjusted by a flexible Fourier approximation uses the residuals of the first stage as a dependent variable and the distance to the CBD as an explanatory variable. Statistically significant estimated distance coefficients are considered SBD.

The models that use a spatial statistical approach are based on Exploratory Spatial Data Analysis (ESDA). The methodology allows evaluating the spatial pattern of employment level and density. Baumont et al. (2004) identify as sub-centers all high employment levels and statistically significant densities.

All approaches that use gross employment density (employment/area) are sensitive to the area's size. More extensive regions tend to have smaller densities (and vice versa), confounding the identification of subcenters. Besides, Mieszkowski and Smith (1991) point out that the coefficients estimated by gross density tend to overestimate the estimated coefficients.

SBD is characterized by the highest price of land, from which a decrease in these prices is proportional to the distance, and by employment agglomeration, i.e., poles of attraction of firms and workers. The method used in Plaut and Plaut (1998) and Herman and Haddad (2005) may not be the most recommended to identify these centralities. The authors considered the error term of a hedonic price econometric model, and then the kriging approach was applied. However, the latter approach may be sensitive to relevant omitted variables, such as unobservable real estate characteristics or amenities determining the price gradient.

We proposed a two-stage approach, incorporating *a priori* identification of candidates for sub-centrality into the hedonic price model. Thus, in the first stage, it identifies all candidates for SBD using an ESDA-type analysis. In the second stage, we run the hedonic house price model considering explanatory variables distances to SBD candidates (identified in the first stage). The second stage strategy permits controlling the overidentification derived from using employment level concentrated in the polygon centroid. Thus, in both approaches, one can test the null hypothesis that each region is not an SBD.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

3. Exploratory Analysis of Spatial Data and Hedonic Price Model

3.1. The ESDA first stage procedure

The definition of SBD candidate areas is based on the ESDA technique, consisting of the first stage estimation strategy. From this approach, it is possible to obtain measures of global and local spatial autocorrelation. While the Global Moran's I statistic allows us to test the spatial association pattern of jobs (by weighting area) for the whole sample, the Local Moran's I makes it possible to decompose the spatial association regime. The decomposition associate clusters of type HH (high-high), HL (low-low), LH (low-high) and LL (low-low). This is possible because, for each observation of the sample, a specific Moran's I is computed, that is, I_i (where i is the evaluated region and $i = 1, \dots, I$) (Anselin, 1995). Algebraically,

$$I_i = \frac{z_i W_{zi}}{\sigma^2} \qquad (1)$$

Where z_i is the employment deviation by Census tract i; W_{zi} is the average value of employment level deviations in the neighboring tract zones of i and σ^2 is the variance of the value deviation. For inference, the method uses random permutation.

In the second stage, we consider the hedonic price approach. In general terms, the hedonic price model seeks to measure price changes as a function of product quality (Court, 1939, Lancaster, 1966 and Griliches, 1971). Such an approach is widely used in empirical work**.**

Considering a general hedonic model for the house i in years t, the estimated empirical model is as follows:

$$P_{it} = \alpha + \rho \sum w_{ij} P_{it} + \theta CBD_i + \sum_p \tau_p SBD_{p,i} + \sum \gamma_k Z_{kit} + \sum_l \beta_l A_{l,i} + \sum \delta_m T_{m,i} + \sum \kappa_o w_{ij} X_{o,i} + \mu D + \varepsilon_{it}$$
$$\varepsilon_{it} = \lambda \sum w_{ij} \varepsilon_{it} + u_{it} \qquad (1)$$

where, P_{it} is the price of house i in year t, CBD_i is the center-house distance, $SBD_{p,i}$ are the distances from house i to each SBD candidate (previously identified in the first state), Z_{it} are the intrinsic characteristics of house i in t, A_i are the distance vector of house i to each l amenities, T_i is the distances of house i to each m avenues and modes of transportation, D is the time dummy vector, ε_{it} is the spatially autocorrelated error and u_{it} is the random error term. Spatial lags are created by using w_{ij} (neighborhood spatial matrix), under the following conditions: a) $w_{ij} = 0$ for $j = i$, b) $0 < w_{ij} \leq 1$, if j is neighbor of i and c) $\sum_j w_{ij} \leq 1$, in general, it is equal to one. Therefore, $w_{ij} P_{it}$ is house price spatial lag and $w_{ij} X_{m,i}$

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

is covariates spatial lag (Z_{kit}, A_{li}, T_{mi}). The coefficients to be estimated are α, ρ, θ, τ_p, γ_k, β_l, κ_o, μ, λ e σ_u.

Equation 1 specification is not the common understanding neither theoretical nor empirical urban economics literature. According to Hermann and Haddad (2005), studies traditionally adopt the hedonic price function's linear or semi-logarithmic functional form. Thus, the specification of Equation 1 will be estimated considering these two functional forms.

The primary motivation for using a spatial econometric model stems from extensive empirical literature that considers space as an intrinsic feature of the housing market, and consequently, its essentially spatialized database (Arnold, 1987; 1998b). Also, the importance of testing the spatial association is related to two effects overlapping, once that, spatial heterogeneity can generate spatial autocorrelation, and spatial autocorrelation can generate spatial heterogeneity.

To avoid omitted variables, we consider the global effect (direct effect + indirect effect) by estimating spatial econometric models that control these kinds of amenities. According to Can (1992) and Campos and Almeida (2018), two effects can be identified in the decomposition of real estate prices. The first is the effect due to the property's characteristics, and the second is the spatial spillover effect of those composing its neighborhood. Therefore, in terms of econometric analysis, we are interested in the global effect results.

As in Kim, Phipps, and Anselin (2003), Campos (2017) and Campos, and Almeida (2018), the distribution of extrinsic and intrinsic house characteristics are not uniformly distributed in a weighting area, as proposed by Megbolugbe and Hoek-Smit (1996). We seek to overcome this limitation by considering the spatial location of each house and amenity (school, shop, library, etc.) instead of the quantity of or participation amenities in the weighting area. This approach allows units within the weighting area to configure different neighborhoods, depending on the area's size created by the selected distance matrix W cut-off. Another advantage is to avoid the neighborhood discretion imposed by the government in the area composition (district, census tracts, etc.).

As these results are sensitive to different weight matrices, robustness tests on the spatial matrix will be considered below, seeking internal validity for the results found. As for the robustness of the weight matrix, we permit 5% and 10% distance, increasing the initial cut-off, permitting bigger neighborhoods.

Thus, the first stage evaluates the tract areas that form statistically significant clusters by using the ESDA approach. The validation of the tract areas as SBD stems from the hedonic price model. In this

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

second stage, the identification strategy considers real estate in space for each year. The amenities (schools, parks, hospitals, etc.) and transportation (subway, train, access to the avenues and roads) were also geocoded. The geocoded data allows a) to identify the location of the properties, amenities, and transportation in the city, b) to calculate the Euclidean distance of each property to the public amenities and transportation within the municipality.

After controlling property prices for intrinsic characteristics, amenities, and access to transportation, we assumed that the distance of each property to each SBD suggested by the first stage should be statistically significant and with negative signs to be considered sub-centers. This identification strategy allows making inferences about the SBD and evaluating if the econometric model of hedonic prices of the real estate market behaves as predicted by the theoretical models.

3.1. The Econometric second stage estimation

As a starting point for the estimation of the Hedonic Price Model, it is considered that the OLS estimation, used in Equation 1, contains the following restrictions for the spatial parameters $\rho = \lambda = \kappa_o = \tau_p = 0$. This estimation allows for evaluating if control variables can solve the possible spatial dependence pointed out in the literature, permitting to assess if the price gradient decreases due to the distance from the CBD.

The Moran’s I test is used to verify the existence of spatial dependence in OLS model residuals. It follows the empirical approach Halleck Vega and Elhorst (2015) proposed, focusing on the models that consider the spatial lag of the explanatory variables (SLX). According to Gibbons and Overman (2012) and Halleck Vega and Elhorst (2015), the SLX model should start when the assessment requires a spatial econometric approach. The SLX model is the most flexible of all spatial specifications, although it faces some disadvantages, such as multicollinearity and causality interpretation difficulties. It is easily interpretable in terms of signal, magnitude, and level of significance. Besides, it also avoids all the criticisms that fall into the models of global spillover when estimating SDM, SDEM, or GNS type models (Halleck Vega and Elhorst, 2015).

4. Database

We used a collection of databases. In the first stage estimation, the employment data of Annual Social Information Report (RAIS – *Relação Anual de Informações Sociais*) produced by the Ministry of Labor is used to identify the SBD. It covers all formally established (incorporated) organizations (public and private) and workers with a labor card and is considered a census of formal workers. RAIS is restricted

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

to the formal labor market, and it is a limitation of our analysis since we are not considering the informal labor market. However, in 2013, formal employment represented 66% of the labor force localized in the Metropolitan Area of São Paulo according to the National Household Sample Survey (PNAD – *Pesquisa Nacional por Amostra de Domicílios*). The geocoding of firms and aggregation of workers by Census tracts were conducted by the Municipal Secretariat for Urban Development (SMDU – *Secretaria Municipal de Desenvolvimento Urbano)* of the Municipality of São Paulo (MSP). The MSP has 310 Census tracts, according to the Brazilian Institute of Geography and Statistics (IBGE).

In the second stage, real estate formal market data are used. The database was obtained from the Brazilian Company for Heritage Studies (Embraesp) with 4,758 thousand vertical and horizontal newly constructed housing between January 2006 and March 2013 for the municipality of São Paulo (MSP). The base covers 271.861 apartments and houses in the period. According to the 2010 Census, the number of domiciles (formal and informal) is 3,573,509 units in the MSP. Even though we are considering about 6% of total domiciles in the MSP, each property launch price captures the average market price behavior of each neighborhood composed of many existing apartments and houses. Furthermore, the Embraesp database permits the understanding that property launching growth rate and housing typologies are spatially heterogeneous, reflecting the non-isotropic pattern of land use policy, consumer behavior, and financial viability of constructor companies. This database has already been widely used in applied work (Biderman, 2001; Maciel, 2010; Nadalin, 2010; Campos, 2017; Campos and Almeida, 2018).

Embraesp database provides information on the size of the property (floor area and total area), the number of bedrooms, bathrooms, parking lot spaces, elevators, apartment units per floor, building floors, and a dummy distinguishing between an apartment or house typologies). Other information is not considered, such as the composition of the condominium (swimming pools, playground, courts, etc.). To overcome such a lack of information, we used the difference between the total area and the useful floor area as a proxy for the condominium size per apartment.

An important feature of the database is to deal with new residences, which excludes the need to consider property depreciation. Regarding monetary restatement, all prices were adjusted by the IGP-DI index and updated to December 2013. The database is geocoded using the ZIP code, street, and number information, and the geographic coordinate system used is SIRGAS 2000 with UTM Zone 23S projection. The same geographical projection is used for the maps.

To control for urban infrastructure, we calculate the minimum Euclidian distance from each property to transportation sites (train and subway stations, main avenues, bus corridor, and highway roads) and

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

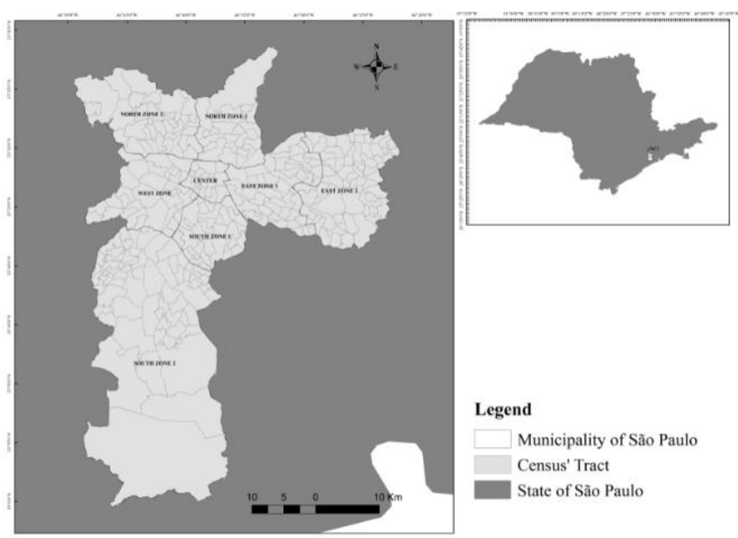
amenities variables (urban parks, libraries, cultural centers, art galleries, theaters, concert halls, theater movies, malls, race stadium, resorts, sports centers, clubs, courts, soccer stadiums, elementary school, and high school, hospitals, health centers, and *favelas*).

5. Results

5.1. First Stage: Exploratory Analysis of Spatial Data

For the specific case of the city of São Paulo, Figure 1 shows the spatial configuration of the municipality by macro zones and Census tracts.

Figure 1 – Municipality of São Paulo – SP



Source: Own elaboration

Due to the *ad hoc* characteristic of the neighborhood composition, different weighting matrices are considered. In this specific evaluation, binary contiguity (queen and rook), distance, and k nearest neighbor weight matrices are used. In contiguity matrices, w_{ij} is equal to 1 if i and j are contiguous, and zero otherwise. The used distance weight matrix considers the centroid distances of the tracts, considering the Euclidean distance. Mathematically, $w_{ij} \leq d_{ij}$. In this specification, the maximum distance for at least one neighbor is 8 Km, avoiding the island issue in the econometric estimation. Concerning the k neighbor weight matrix, we chose six closest neighbors because it is the average number of neighbors in the database.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

First, consider the results of Global Moran's I test. Table 2 shows the results for the employment variables using the weighting matrices discussed above. All results are statistically significant at 1%. Note that all tests show a positive spatial correlation. This result indicates a similarity's relationship between the level of employment and the spatial location of the employment level of the neighbors, suggesting that the MSP employment spatial pattern is concentrated in the neighborhoods where neighbors are similar to each other.

Table 2 - Moran's I Statistic

Weight Matrix	Moran's I	P-value
Distance	0.365373	0.000
Queen	0.592239	0.000
Rook	0.595834	0.000
K-neighbors (K6)	0.469755	0.000

Source: Own elaboration based on data from RAIS/MTE.

To identify the SBD candidates, we consider the Local Moran's I, which composes the LISA (Local Indicator of Spatial Association) indicators. Table 3 shows the total of tracts area statistically significant at least 5%. The results are much more sensitive to the distance-formatted weighting matrix.

Although the weighting matrices (queen, rook, and k neighbors) point to different results, they have smaller deviations from the aggregate result than the distance weight matrix. To consider the SBD candidates from the city of São Paulo, we only take those that are statistically significant for all specifications and those areas characterized as HH and HL clusters. In this case, with at least 5% significance, 37 tract areas are found as candidates for SBD, and all of them are labeled as HH only, as displayed in Table 3. For the three spatial weighting matrix specifications (queen, rook, and k neighbors), only one weighting area is not significant for the rook and queen specification simultaneously. As a result, 36 weighting areas are considered as SBD candidates.

Table 3 - Total of Tract Areas by Cluster

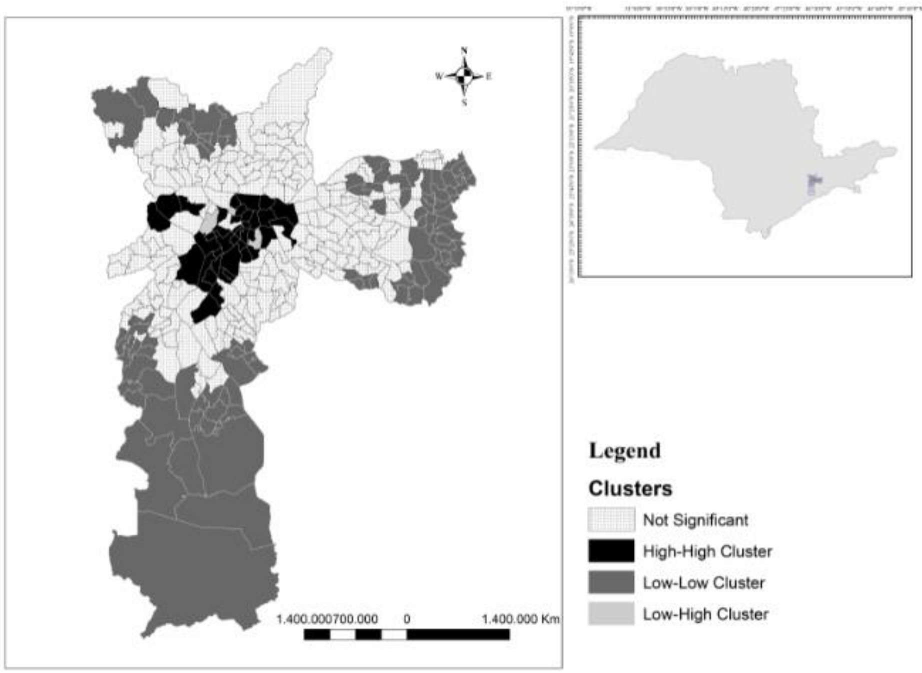
Cluster	Rook	Queen	6 neighbors	Distance
HH	47	52	38	74
LL	74	75	91	116
LH	4	5	4	36
HL	0	0	0	3
Not significant	185	178	177	81

Source: Own elaboration based on Local Moran's I results. All results are statistically significant by at least 5%.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Regarding the localization of these tract areas (Figure 2), the above-average spatial concentration (HH) of the employment level is located in almost all of the Center Zone (CZ), part of the South Zone 1 (SZ1) and South Zone 2 (SZ2), part of the West Zone (WZ) and small participation in the East Zone 1 (EZ1).

Figure 2 – Employment Level Clusters



Source: own elaboration, using results of Local Moran's I test (6 nearest neighbors)

5.2. Second Stage: Hedonic Prices Model

The econometric approach seeks to validate the SBD candidates identified in the previous subsection. However, in the first approximation, the models will be estimated without SBD controls. This exercise seeks to shed light on the behavior of the coefficient that measures the distancing effect from CBD by considering the Municipality of São Paulo as a monocentric city.

It is noted that the CBD coefficient is not statistically significant for any of the functional forms (Table 4, columns 1-2). By controlling the model by the quadratic effect of CBD distancing, the coefficient signals diverge from the theoretical models and are not statistically significant (Table 4, columns 3-4). Two possibilities may justify these results: a) spatial dependence may be distorting the signals and/or tests and/or b) omission of (a) relevant variable(s) (SBD), in which case the city of São Paulo would be polycentric.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Table 4: OLS Model

	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
CBD	-2.9E^{-02}	-1.5E^{-06}	3.5E^{-02}	6.5E^{-06}	-4.2E^{-02}	-5.8E^{-06}	-4.8E^{-03}	-2.2E^{-06}
CBD²	-	-	-3.1E^{-06}	-3.9E^{-10}	-	-	-1.9E^{-06}	-1.8E^{-10}
CONST	4.0E^{+03}	8.1E^{+00}	3.9E^{+03}	8.1E^{+00}	3.9E^{+03}	8.1E^{+00}	3.8E^{+03}	8.0E^{+00}
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SBD	No	No	No	No	Yes	Yes	Yes	Yes
Time Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R²	0.73	0.78	0.73	0.78	0.74	0.79	0.77	0.79
F test	278***	361***	273***	354***	170***	220***	170***	220***
Moran's I test	3.97E^{-02}	3.87E^{-02}	4.00E^{-02}	3.93E^{-02}	1.90E^{-02}	1.58E^{-02}	1.88E^{-02}	1.60E^{-02}

Note: $p \leq$ *** 1%; $p \leq$ ** 5% e $p \leq$ * 10%.

Tests on residues (Moran's I tests) point to non-spatial randomness (Table 4, columns 1-4). When using the SLX[b] estimator to correct spatial dependence (Table 5, columns 9-11), the estimated coefficients are theoretically counterintuitive (overall effect) and statistically significant. These models estimated without control for sub-centralities indicate that the price of land is higher in the fringes of the city - a result that is not observed. Moreover, the null hypothesis of spatial randomness is not rejected. The exception is the log-level functional form model with control for the quadratic form of the distancing from CBD, which global effect results agree with the theoretical models (Table 5, column 12). However, these counterintuitive results open space to evaluate the second hypothesis: do the non-intuitive and statistically non-significant results of the CBD coefficients stem from the omission of SBD controls?

Controlling for all SBD candidates in the econometrics estimation, we are assuming for cross-commuting. That is, workers have the possibility of choosing distant business centers even though there are business centers closer to their homes (Plaut and Plaut, 1988; Wrede, 2015). This specification is more flexible when compared to the approach of choosing the nearest center of the worker's residence as proposed by Fujita and Ogawa (1982) and Baumont, Ertur, and Gallo (2004).

The results derived from the estimation of Equation 1 (controlling for SBD) by OLS (Table 4, columns 5-8) are supported by the urban theoretical models (decreasing gradient price) - for both functional forms and specifications of the CBD variable. However, tests on spatial randomness reject the

[b] All results are discussed under the total effects (direct + indirect effects).

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

null hypothesis. Due to spatial dependence, the signals and tests may be inaccurate. Therefore, to overcome the OLS estimator limitation, we estimate the econometric model by the SLX estimator.

As the specification that considers the quadratic form of distancing is not very explanatory and generally does not present statistical significance for the coefficients that capture such effect, the models estimated by SLX will not be demonstrated in this specification. The results derived from the SLX estimator show the signs of the overall effects following the theory and are statically significant. Moran's I tests do not reject the null hypothesis of spatial randomness (Table 6, columns 17-22).

Table 5: SLX Model (I)

	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level	Level-Level	Log-Level
	(9)	(10)	(11)	(12)	(13)	(15)	(15)	(16)	(17)	(18)	(19)	(20)
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
CBD	1.7E-01*	5.3E-05*	2.4E-01*	5.4E-05**	1.9E-01*	5.5E-05*	2.5E-01*	5.8E-05*	2.3E-01*	6.5E-05*	2.9E-01*	6.9E-05***
WCBD	9.3E-02	7.1E-06	-1.6E-01	-1.1E-04**	-4.5E-03	8.2E-07	-2.3E-01	-1.1E-04**	-2.1E-02	-4.4E-06	3.0E-02	-5.6E-05
CBD²	-	-	-2.8E-06	4.3E-10	-	-	-3.5E-06	2.0E-10	-	-	-4.8E-06	-1.9E-10
WCBD²	-	-	1.2E-05	5.4E-09***	-	-	1.0E-05	4.9E-09**	-	-	-2.1E-06	2.3E-09
CONST	2.2E+3*	7.2***	2.3E+3*	7.3***	2.3E+2*	7.6***	2.8E+3*	7.5***	1.9E+3	7.6***	3.2e+3*	7.5***
Controle	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SBD	No	No	No	No	No	No	No	No	No	No	No	No
Time Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R²	0.76	0.81	0.76	0.81	0.76	0.80	0.76	0.81	0.76	0.80	0.76	0.8
F statistic	180***	230***	178***	226***	179***	228***	174***	223***	179***	228***	175***	223***
Moran's I test	-2.8E-03	-6.1E-03	-3.1E-03	-5.7E-03	-2.9E-03	-4.8E-03	-3.2E-03	-5.1E-03	-3.3E-03	-5.2E-03	-3.5E-03	-4.9E-03
Weight Matrix	Distance (threshold)				5% *times* threshold				10% *times* threshold			

Note: $p \leq$ *** 1%; $p \leq$ ** 5% e $p \leq$ * 10%.

From this econometric exercise, it is possible to find evidence in favor of the existence of at least one other central region in the city of São Paulo. Therefore, models that disregard this sub-centrality estimate biased coefficients and heteroskedastic residues.

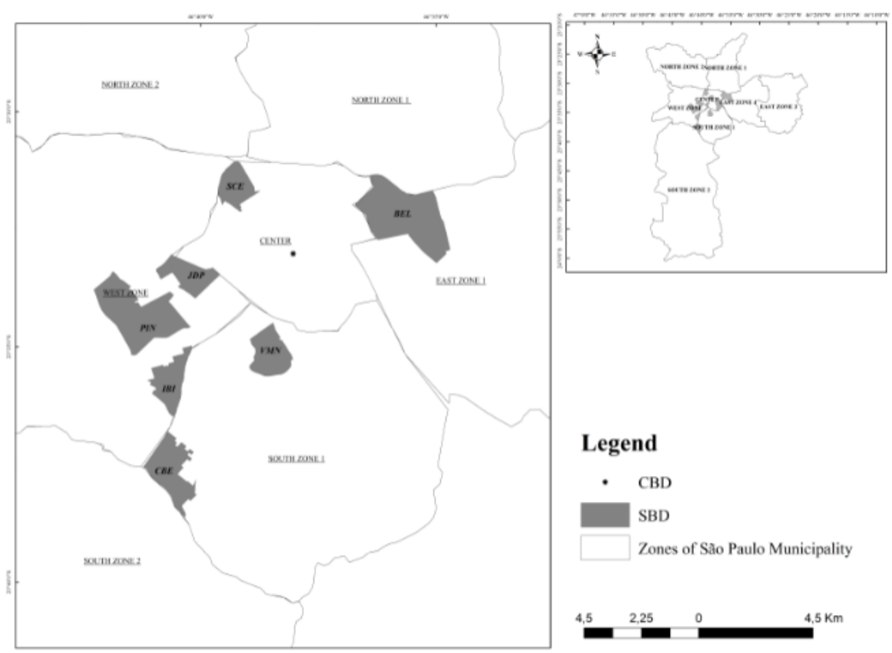
The SLX model controlling for SBD presents the best theoretical and econometric adjustment. The CBD coefficients show the signals theoretically highlighted. Moreover, the discussion about SBD considers the log-level functional form since it presents higher R²-adjusted when compared to the level-level specification.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Duranton and Puga (2015) point out no agreement regarding a consistent definition of SBD. Many authors have considered the contiguity areas as an empirical strategy to overcome the challenge of quantifying SBD (Guiliano and Small, 1991; McMillen, 2003; Baumont, Ertur, and Gallo, 2004). Following the theoretical models, areas whose estimated coefficients are statistically significant and negative are taken as SBD. Under this criterion, from the estimated model, the city of São Paulo presents seven SBDs, which give a negative and statistically significant global effect (direct and indirect).

Figure 3 shows the geographical localization of each identified SBD (Table 6, column 1) and the city's main avenues. The SBDs are in the respective districts of the city: Santa Cecilia (SCE), Belém (BEL), Jardim Paulista (JDP), Pinheiros (PIN), Itaim Bibi (IBI), Campo Belo (CBE), and Vila Mariana (VLM)[c]. Geographically, SBDs are identified in Central Zone (CZ), South Zone 1 (SZ1), West Zone (WZ) and East Zone 1(EZ1).

Figure 3 – The SBD of the Municipality of São Paulo



Source: Own elaboration based in data from RAIS and IBGE

Our results are in line with the effects observed in the municipality. Studies for the metropolitan area of São Paulo report findings close to ours, even using a different methodological approach, different sizes of polygons, and different databases. Siqueira (2014)'s findings report seven SBD for the metropolitan

[c] Respectively, the variables A15, A24, A17, A10, A6, A7, A21 in the econometric model.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

area of São Paulo[d] (SPMA), being three of them in the municipality of São Paulo localized in the historic center, Paulista Avenue – Faria Lima Avenue corridor and Santo Amaro. Campos (2018) reports three SBD for the SPMA, but only one in the MSP, aggregating all the seven SBD found in this paper. Siqueira (2014) and Campos (2018) identification strategies are based on the relative concentration of employment for SBD identification, justifying why they found less SBD than us. In this way, their findings may be labeled as global SBD (relative to the SPMA), while ours is local SBD (relative to the MSP).

Table 6: SLX Model (II)

	LOG-LEVEL			LOG-LEVEL	LOG-LEVEL	LINEAR	LINEAR	LINEAR
		(17)		(18)	(19)	(20)	(21)	(22)
	Direct Effect	Indirect Effect	Global Effect	Global Effect	Global Effect	Global Effect	Global Effect	Global Effect
CBD	$5.43E^{-05***}$	$-1.95E^{-04***}$	$-1.41E^{-04}$	$-2.08E^{-04}$	$-1.28E^{-04}$	$-2.38E^{-01}$	$-4.55E^{-01}$	$-1.36E^{-01}$
A1	$-2.92E^{-05***}$	$7.30E^{-04***}$	$7.01E^{-04}$	$5.94E^{-04}$	$-3.22E^{-05}$	$2.09E^{+00}$	$1.52E^{+00}$	$-1.2E^{-01}$
A2	$-3.59E^{-06}$	$9.54E^{-05}$	-	-	-	-	-	-
A3	$-4.98E^{-06}$	$3.92E^{-04}$	-	-	-	-	-	-
A4	$3.37E^{-05}$	$-1.87E^{-04}$	-	$-1.04E^{-03}$	$-8.73E^{-04}$	$-2.49E^{+00}$	$-5.15E^{+00}$	$-2.73E^{+00}$
A5	$4.45E^{-05}$	$8.81E^{-04}$	-	-	-	-	-	-
A6	$-3.87E^{-05}$	$-3.40E^{-03***}$	$-3.40E^{-03}$	$-3.50E^{-03}$	$-1.83E^{-03}$	$-1.05E^{+01}$	$-8.90E^{+00}$	-
A7	$-7.04E^{-06}$	$-2.58E^{-03***}$	$-2.58E^{-03}$	$-2.27E^{-03}$	-	$-6.61E^{+00}$	$-4.60E^{+00}$	-
A8	$6.79E^{-05}$	$4.11E^{-0}3^{***}$	$4.11E^{-03}$	$4.44E^{-03}$	$2.37E^{-03}$	$4.36E^{-01}$	$4.09E^{-01}$	$3.86E^{-01}$
A9	$9.06E^{-0}7$	$-4.22E^{-04}$	-	$-5.95E^{-04}$	$-7.13E^{-04}$	-	$-2.59E^{+00}$	$-2.22E^{+00}$
A10	$-3.91E^{-05}$	$-4.00E^{-03***}$	$-4.00E^{-03}$	-	-	-	-	-
A11	$7.79E^{-05}$	$4.44E^{-03***}$	$4.44E^{-03}$	$3.16E^{-03}$	-	$1.01E^{+01}$	-	-
A12	$-7.29E^{-05}$	$2.71E^{-03*}$	$2.71E^{-03}$	-	-	-	-	-
A13	$-4.03E^{-05}$	$3.87E^{-04}$	-	$3.11E^{-03}$	$2.69E^{-03}$	-	$1.06E^{+01}$	-
A14	$7.25E^{-06}$	$-1.35E^{-03}$	-	-	-	-	-	-
A15	$1.62E^{-05}$	$-4.85E^{-03***}$	$-4.85E^{-03}$	$-4.29E^{-03}$	$-2.50E^{-03}$	$-1.14E^{+01}$	$-8.95E^{+00}$	-
A16	$5.82E^{-06}$	$1.36E^{-03**}$	$1.36E^{-03}$	$1.41E^{-03}$	$1.69E^{-03}$	$6.59E^{+00}$	$6.61E^{+00}$	$5.38E^{+00}$
A17	$-1.71E^{-04*}$	$-5.18E^{-03***}$	$-5.35E^{-03}$	$-3.65E^{-03}$	$-4.74E^{-03}$	$-2.46E^{+01}$	-	$-1.36E^{+01}$
A18	$2.58E^{-04*}$	$2.12E^{-03}$	$2.58E^{-04}$	$2.32E^{-04}$	-	$1.03E^{+00}$	-	-
A19	$-1.08E^{-04}$	$4.13E^{-03}$	-	$8.48E^{-03}$	-	-	$3.10E^{+01}$	-
A20	$9.83E^{-05}$	$2.35E^{-03}$	-	$1.39E^{-04}$	$1.39E^{-04}$	$4.93E^{-01}$	$6.02E^{-01}$	$6.48E^{-01}$
A21	$-1.47E^{-04**}$	$-5.77E^{-03***}$	$-5.92E^{-03}$	$-6.18E^{-03}$	$-8.01E^{-03}$	$-2.30E^{+01}$	$-1.90E^{+01}$	$-5.74E^{-01}$

[d] São Paulo Metropolitan Area is composed of 39 municipalities spread over an area of 7,946 km².

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

A22	5.32E^{-05}	3.74E^{-03***}	3.74E^{-03}	3.97E^{-03}	4.92E^{-03}	1.22E^{+01}	7.48E^{+00}	9.27E^{+00}
A23	4.55E^{-05**}	1.50E^{-03***}	1.55E^{-03}	1.59E^{-03}	4.80E^{-05}	7.92E^{+00}	7.68E^{+00}	2.16E^{-01}
A24	-3.46E^{-05}	-9.03E^{-04*}	-9.03E^{-04}	-	-	-7.22E^{+00}	-6.84E^{+00}	-
A25	4.05E^{-05}	-3.58E^{-04}	-	-	-	-	-	-
A26	-1.65E^{-04}	8.27E^{-03}		-1.99E^{-04}	-2.53E^{-04}	3.21E^{+01}	3.25E^{+01}	-7.31E^{-01}
A27	2.62E^{-04**}	-1.80E^{-03}	2.62E^{-04}	2.80E^{-04}	2.57E^{-04}	1.28E^{+00}	-2.59E^{+01}	1.17E^{+00}
A28	8.70E^{-05}	-3.27E^{-02}	-	-2.93E^{-02}	-3.43E^{-02}	-	-	-1.55E^{+02}
A29	-2.93E^{-05}	1.13E^{-02}	-	-	-	-	-	-
A30	-2.71E^{-04}	1.85E^{-02}	-	-	1.81E^{-02}	5.57E^{+01}	6.08E^{+01}	8.45E^{+01}
A31	3.92E^{-05}	-4.97E^{-04}	-	-	-	-	-	-
A32	-3.90E^{-05}	-2.08E^{-03}	-	-2.81E^{-03}	-6.08E^{-05}	-9.53E^{+00}	-1.04E^{+01}	-2.24E^{-01}
A33	-3.58E^{-05}	1.61E^{-04}	-	-	-	-	-	-
A34	4.06E^{-05}	2.44E^{-03}	-	-	3.38E^{-03}	-	-	-
A35	-1.10E^{-04}	-8.16E^{-03}	-	-8.70E^{-03}	-	-	-2.47E^{+01}	-
A36	1.68E^{-04*}	4.66E^{-03}	1.68E^{-04}	6.96E^{-03}	5.74E^{-03}	1.20E^{+00}	1.11E^{+00}	1.03E^{+00}
CONST	7.61***	-	7.61***	7.95***	7.497.95***	1.79E^{03}	3.11 E^{03*}	1.99
Weight Matrix		K1		K2	K3	K1	K2	K3
Control		Yes		Yes	Yes	Yes	Yes	Yes
Time Dummy		Yes		Yes	Yes	Yes	Yes	Yes
Adjusted R^2		0.8155		0.8139	0.8115	0.7793	0.7762	0.7731
F statistic		134***		132***	130***	107***	105***	103***
Moran's I test		-5.43E-03		-4.79E-03	5.16E-03	-4.40E-03	-1.99E-03	-3.13E-03

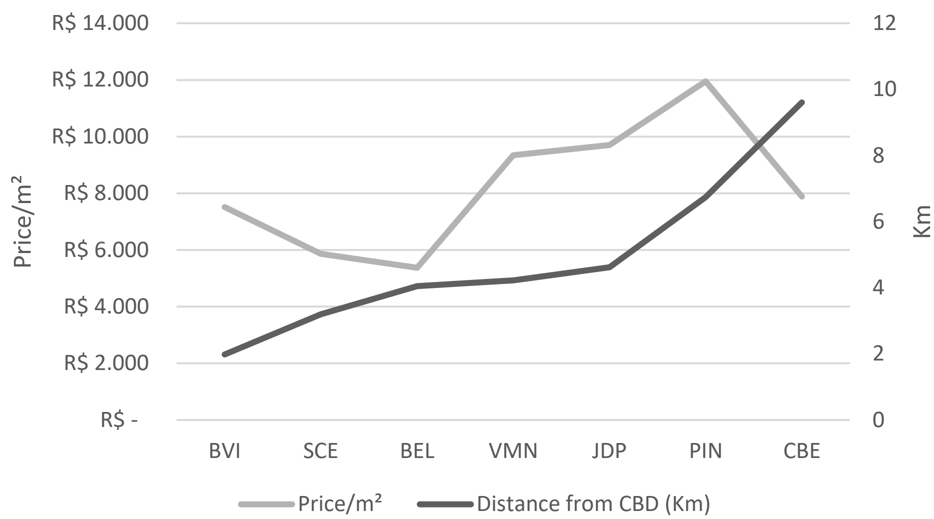
Note: $p \leq$ *** 1%; $p \leq$ ** 5% e $p \leq$ * 10%. Direct effect is the effect of the distancing of CBD and SBD on the price of real estate *i*. Indirect effect is the effect of the distancing of the neighbors *j* of CBD and SBD on the price of real estate *i*. The Total Effect is the sum of the statistically significant coefficients. The coefficients whose statistical significance is not observed were not reported. The spatial weight matrices are K1 (threshold), K2 (5% increase in threshold), and K3 (10% increase in threshold).

5.3. Sub-centralities Characterization in São Paulo Municipality: employment specialization, wage, and educational level

A descriptive evaluation of housing price, employment specialization, wage, and education provides the best understanding of each SBD, underlining the heterogeneity amongst sub-centers. Considering all identified SBD, the average prices/m² is R$ 8.230,00, and the standard deviation is R$ 2.295,00. Housing prices are higher in Pinheiros than elsewhere. Figure 4 displays a positive correlation between average prices/m² and distance from CBD, while monocentric models point out a negative correlation. Such behavior supports our findings in terms of the existence of sub-centers.

Figure 4 – Housing Price/m² by SBD

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.



Considering the activities sectors provided by IBGE, we focus on activities concentration in each SBD. In general, SBDs are concentrated on service activities. Table 7 shows the highest concentrations of employment are in the construction (Campo Belo), credit institutions, insurance and capitalization (Itaim Bibi), management and administration, real estate, securities, technical service (Vila Mariana, Santa Cecilia, Pinheiros), and medical, dental, and veterinary services (Jardim Paulista). Belém is the only one to have the highest concentration of employees in the manufacturing activities.

Table 7: Sectoral Composition by SBD

Sector Activities	Subcenters						
	Belém	Vila Mariana	Jd. Paulista	Santa Cecília	Pinheiros	Campo Belo	Itaim Bibi
Agriculture, forestry, breeding, plant extractivism	0,0%	0,0%	0,0%	0,0%	0,4%	0,0%	1,2%
Manufacturing	24,1%	3,7%	2,1%	9,2%	4,7%	4,5%	2,2%
Food, beverages, and Industrial utility services ethyl alcohol industry	4,2%	0,1%	0,1%	0,7%	1,5%	0,2%	2,0%
Industrial utility services	0,0%	0,0%	0,1%	0,0%	0,2%	0,0%	3,0%
Construction	4,0%	6,6%	2,8%	7,6%	16,7%	23,7%	5,6%
Retail trade	10,5%	9,6%	10,2%	14,3%	13,9%	19,0%	8,7%
Wholesale trade	6,2%	3,7%	2,5%	4,4%	2,9%	3,0%	3,8%
Credit institutions, insurance, and capitalization	1,0%	1,4%	2,5%	0,9%	5,6%	2,4%	26,0%
Management and administration, real estate, securities, technical service	23,3%	34,0%	12,4%	38,4%	28,6%	21,6%	17,8%
Transport and communications	3,3%	3,0%	3,6%	4,7%	4,6%	5,7%	3,2%
Accommodation, food, repair, maintenance,	16,4%	20,8%	10,4%	15,1%	17,4%	13,6%	13,9%

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

Medical, dental, and veterinary services	4,8%	9,6%	49,2%	2,0%	1,3%	2,4%	9,0%
Teaching	2,2%	7,5%	4,0%	2,7%	2,1%	4,0%	3,7%

Source: Elaboration based on data from RAIS 2013.

Itaim Bibi and Pinheiros are districts with the highest concentration of workers receiving more than 5 minimum wage monthly, as shown in the Table 8. On the other hand, Belém and Campo Belo are the subcenters with the largest share of workers receiving less than 2 minimum wages monthly (60% and 54%, respectively). While the first two SBD demand highly skilled people for jobs supplied, the last two are concentrated in manufacturing and civil construction activities, which demand workers with less schooling.

Table 8: Wage Composition by SBD

Minimum Wage	Subcenters						
	Belém	Vila Mariana	Jd. Paulista	Santa Cecília	Pinheiros	Campo Belo	Itaim Bibi
More than 20	1%	1%	2%	1%	6%	1%	6%
20 – 10	2%	3%	6%	2%	10%	3%	12%
10 – 5	8%	9%	15%	8%	17%	9%	18%
4 – 2	31%	39%	44%	46%	36%	32%	34%
1,51 – 2	26%	25%	18%	23%	16%	29%	13%
1,5 – 1,1	30%	17%	12%	17%	12%	23%	13%
Less/iqual than 1	4%	4%	2%	4%	3%	2%	4%

Source: Own elaboration based on data from RAIS 2013.

One possibility of ascertaining the low productivity is considering the years of study of workers in each SBD. Table 9 allows an evaluation of the schooling composition of workers in each SBD and a correlation with this wage behavior, although with little rigor. In Pinheiros, 42% of the employees have at least finished high school (complete high school + incomplete higher education), and 34% have at least a completed college degree. Itaim Bibi (higher wage composition of workers amongst all identified SBDs) presents 40% of the workers with at least complete high school and 40.4% with at least a bachelor’s degree.

Table 9: Workers Educational Composition by SBD

Education	Subceters						
	Belém	Vila Mariana	Jd. Paulista	Santa Cecília	Pinheiros	Campo Belo	Itaim Bibi
Incomplete High School	30%	32%	16%	34%	25%	42%	20%
Complete High School	54%	43%	45%	50%	34%	36%	31%
Incomplete Higher Education	5%	5%	4%	4%	8%	5%	9%
Complete Higher Education	11%	19%	34%	11%	33%	17%	39%
Master´s	0,2%	1%	0,4%	1%	1%	0,3%	1%
PhD	0,0%	0,3%	1%	0,3%	0,1%	0,1%	0,4%

Source: Own elaboration based on data from RAIS 2013.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

On the other hand, Belém concentrates 54% of its workforce with complete high school and 30% with incomplete high school. Campo Belo also presents its workforce concentrated in the ranges with few years of schooling.

6. Robustness Tests

The main concern in this section is the validity of the identification strategy, i.e., whether the identification strategy correctly identifies subcenters. We carry out robustness tests of the SBD identification, trying different spatial matrix specifications, and changing the functional form.

We initially considered different spatial weight matrices for the model without SBD controls and the functional forms level-level and log-level. Then, we tested distance weight matrices increasing the initial threshold by 5% and 10% (8.4 Km and 8.8 Km). This procedure allows us to test the omission of a relevant variable when estimating models disregarding SBD.

From Table 5 (columns 13-20), one can conclude that the signs and significance of the estimated coefficients remain statistically significant, and the signals keep going against the theoretical approach for both spatial weight matrices and functional forms. These results support the internal validity of the estimates and underline the hypothesis that the model is biased when omitting SBD controls. For the log-level functional form and quadratic specification of the CBD variable, the estimated coefficient is not sensitive to the changes of the spatial weight matrices. This result provides additional evidence against models that use the quadratic form of CBD as a control.

As described in the previous section, the coefficients of the CBD variable estimated by OLS are robust to the different functional forms and specifications, presenting expected signals in all models (Table 4, columns 5-8); however, such models have spatial dependence. When estimating the model with the control by SBD using the SLX estimator, the total effect of the CBD presents the expected signals for both matrices of weight considered (Table 6, columns 18 and 19). The results validate the econometric estimation and support the polycentric hypothesis of São Paulo.

The Pinheiros region is the only one losing statistical significance when the weight matrices are changed, but the signal stays as expected. In this model, the Morumbi district (A9), an adjacent neighbor of Pinheiros, presents a statistically significant effect with the expected signal. It is impossible to test if the subcenter is on the census tract border of both regions, and we do not consider that this slight difference in just one model is that important. Then, we continue advocating to characterize Pinheiros as a business subcentrality.

The Morumbi district starts to present the statistically significant coefficient with the sign expected from the theoretical models after increasing the distance cut-off by 10%. However, the coefficient of region A7 ceases to be statistically significant, although the signals remain as expected. It should be noted that the coefficients that

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

estimate the commuting effect concerning the areas A4, A28, and A32 are statistically significant for both specifications of the spatial weight matrices.

When considering the functional level-level form (Table 6, column 20-22), it is noted that the same areas identified as SBD are identified compared to the econometric model estimated with the log-level functional form. The exception is the Pinheiros area that is not significant, but Morumbi does.

Additionally, when Equation 1 is estimated using level-level function form, the econometric estimation identifies more SBD than log-level form; however, the R^2-adjusted of these models is lower than the semi-logarithmic models, as shown before.

7. Final Remarks

Theoretically, SBDs concentrate "forces" capable of segregating due to the increase of the land price, yet are capable, to some extent, of decongesting urban roads when it can reduce the commuting time within the municipality. The SBD cannot be seen simply as the consequence of the firm’s decision in the space, searching by agglomeration gain, in a way autonomous and independent. Contrarywise this, it should be considered an important public urban policy mechanism capable of facilitating the matching between workers and firms in the labor market. For instance, by reducing the land taxes, the policymaker can incentivize firms to move outside the center, increasing the city's land value and rebalancing the municipality's land price.

Due to the SBD impact on the economy, the sub-centers identification has long been evaluated in the international literature. However, in Brazil, few studies on sub-center identification in large cities exist. Therefore, in this article, we proposed a two-stage identification approach, seeking to eliminate the discretion of the suggested empirical methods. The first stage aims to identify candidates for SBD using an ESDA approach. In the second stage, using a Hedonic Spatial Price Model is an innovation to the proposed methods since it allows the testing of the alternative hypothesis of the existence of SBD.

In this article, seven sub-centers were identified for the municipality of São Paulo. All sub-centers identified are in areas of great relative importance in the job offer within the city. All of them are located in regions with great road mobility or railway and subway services.

Besides identifying the subcenters, it was possible to characterize the activities in which each SBD is specialized, as well as the wage and educational decomposition among SBD workers. Notably, SBD with a higher concentration of employees receiving low wages presents a higher concentration of workers with lower schooling levels and concentrated in retail trade activities. SBDs concentrated in financial activities have higher remuneration for workers and a higher concentration of workers with a higher educational level than the former. Thus, it was possible to notice that the identified sites are centralities in the MSP and are heterogeneous.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

The identification of SBDs is sensitive to geographical scales, as already pointed out by Anas et al. (1998). Thus, by changing the weighting areas by districts of the municipality or gridding the municipality's total area into cells of 1 km², the results can be changed. One way to overcome this limitation would be to consider the spatial location of each firm and stress the modifiable areal unit problem. In this way, the identification of the SBD would not be restricted to the geographical polygon boundary proposed by the IBGE. Then it would be possible to identify a corridor-shaped SBD, for example. If all identified SBD connect to each other, instead of 7 SBDs we would identify fewer SBD.

8. References

ANAS, A.; ARNOTT, R.; SMALL, K. Urban spatial structure. Journal of Economic Literature, Vol.36, No.3, p.1426-1464, 1998.

ANAS, A.; KIM, I. General equilibrium models of polycentric land use with endogenous congestion and job agglomeration. Journal of Urban Economics, v. 28, p. 318-325, 1990.

ANSELIN, L. Local indicators of spatial association – (LISA). Geographical analysis, v. 27, n. 2, p. 93-115, 1995.

BERLIANT, M.; KONISHI, H. The endogenous formation of a city: Population agglomeration and marketplaces in a location-specific production economy. Regional Science and Urban Economics, v. 30, p. 289.324, 2000.

BAUMONT, C.; BOURDON, F. Centres secondaires et recomposition économique des espace urbain, le casda communauté de l'aglomération dijonnaise (1990; 1999). LATEC Working Paper 4, Universidade de Bougogne, Dijon, 2002.

BAUMONT, C.; ERTUR, C.; GALLO, J. L. Spatial analysis of employment and population density: the case of the agglomeration of Dijon, 1999. Geographical analysis, v. 36, n. 2, 2004.

CAMPOS, R. B. A. O mercado imobiliário residencial no município de São Paulo: uma abordagem de preços hedônicos para determinação de oferta e demanda de imóveis. Nova Economia, vol. 27, p. 303-337, 2017.

CAMPOS, R. B. A.; ALMEIDA, E. S. Efeito vizinhança e efeito adjacência nos preços de imóveis residenciais no município de São Paulo. Estudos Econômicos, v. 48, n. 1, p. 5-38, 2018.

CAN, A. Specification and estimation of hedonic housing price models. Regional Science Urban Economics, v. 22, p. 53–474, 1992.

COFFEY, W. J.; SHEARMUR, R. G. The identification of employment centers in Canadian metropolitan areas: the example of Montreal, 1996. The Canadian Geographer, v. 45, n. 3, p. 371-386, 2001.

COURT, A. T. Hedonic price indexes with automotive examples. In: The dynamics of automobile demand, Nova York, The General Motors Corporation, p. 99-117, 1939.

DIXIT, A.; STIGLITZ, J. Monopolistic competition and optimum product diversity. American Economic Review, v. 67, p. 297-308, 1977.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

FUJITA, M.; OGAWA, H. Multiple Equilibria and Structural Transition of Non- Monocentric Urban Configurations. Regional Science and Urban Economics v.12, n. 2, 161–191, 1982.

GIULIANO, G.; SMALL, K. A. Subcenters in the Los Angeles region. Regional Science and Urban Economics, vol. 21, p. 163-182, 1991.

GLASER, E. L.; KAHN, K. A. Decentralized employment and the transformation of the American city. Working Paper 8117, NBER, 2001.

GREENE, D. Recent trends in urban spatial structure. Growth and Change. v. 10, p. 29-40, 1980.

GRILICHES, Z. Hedonic Price Indexes and the Measurement of Capital and Productivity: Some Historical Reflections. In Triplett, J.(org.), Fifty Years of Economic Measurements, University of Chicago Press, 1990.

GUILLAIN, R.; GALLO, J. L.; BOITEUX-ORAIN, C. The evolution of the spatial and sectoral patterns in Ile-de-France over 1978-1997, in '50th North American Meetings of the Regional Science Association International (RSAI), 2003.

HALLECK VEGA, S.; ELHORST, J. P. The SLX model. Journal of Regional Science, v. 55, p. 339–363, 2015.

Helsley, R. W., and Sullivan, A. M. Urban subcenter formation. Regional Science and Urban Economics, v. 21, 255–275, 1991.

HENDERSON, J.V.; SLADE, E. Development games in non-monotonic cities. Journal of Urban Economics, v. 34, p. 207-229, 1993.

HERMANN, B.; HADDAD, E. A. Mercado imobiliário e amenidades urbanas: a view through the window. Estudos Econômicos, v. 35, n. 2, p. 237-269, 2005.

IBGE (2013). Produto Interno Bruto a preços correntes e Produto Interno Bruto per capita segundo as Grandes Regiões, as Unidades da Federação e os Municípios - 2010-2013. Retrieved from http://www.ibge.gov.br/home/estatistica/economia/pibmunicipios/2010_2013/default_xls.shtm [acessed July 2016].

KIM, W. C; PHIPPS, T. T; ANSELIN, L. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management, v. 43, p. 24-39, 2003.

LANCASTER, K. J. A New Approach to Consumer's Theory. Journal of Political Economy, v. 74, p. 132-157, 1966.

McDONALD, J. F.; McMILLEN, D. P. Suburban subcenters and employment density in metropolitan Chicago. Journal of Urban Economics, v. 43, n. 3, p. 157-180, 1998.

McDONALD, J. F. The identification of urban employment subcenters. Journal of Urban Economics, v. 21, p. 242-258, 1987.

McMILLEN, D. P.; SMITH, S. C. The number of subcenters in large urban areas. Journal of Urban Economics, vol. 53, nº 3, p. 321-338, 2003.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

McMILLEN, D. P. Nonparametric employment subcenter identification. Journal of Urbana Economics, v. 25, n. 2, p. 15-27, 2001a.

McMILLEN, D. P. Polycentric urban structure: the case of Milwaukee. Economic Perspective, v. 25, n. 2, p. 15-27, 2001b.

MEGBOLUGBE, I. F.; HOEK-SMIT. Understanding neighborhood dynamics: a review of the contributions of William G. Grigsby. Urban Studies, v. 33, n.10, p.1779-1795, 1996.

MIESZKOWSKI, P.; SMITH, B. Analyzing urban decentralization: the case of Houston. Regional Science and Urban Economics, Esevier, v. 21, n. 2, p. 183-199, 1991.

MEYER, R; GROSTEIN, M. D.; BIDERMAN, C. São Paulo Metrópole. Editora da Universidade de São Paulo: Imprensa Oficial do Estado de São Paulo, 2004.

OGAWA, H., e FUJITA, M. Equilibrium Land Use Patterns in a Non-Monocentric City. Regional Science and Urban Economics, v. 20, n. 4, p. 455–475, 1980.

PLAUT, P. O. e PLAUT, S. E. Endogenous Identification of Multiple Housing Price Centers in Metropolitan Areas. Journal of Urban Economics, v. 7, 193-217, 1998.

ROMANOS, M. C. Household Location in a Linear Multi-Center Metropolitan Area. Regional Science and Urban Economics, v. 7, p. 233–250, 1979.

SASAKI, K.; MUN, Sei-Il. A Dynamic Analysis of a Multiple-Center Formation in a City. Journal of Urban Economics, v. 40, n. 3, p. 257–278, 1996.

SEADE. Fundação Sistema Estadual de Analise de Dados. Retrieved from http://www.seade.gov.br/ [accessed November 2016]

SMDU. Secretaria Municipal de Desenvolvimento Urbano. Retrieved from http://ces.ibge.gov.br/base-de-dados/metadados/mte/relacao-anual-de-informacoes-sociais-rais [accessed November 2016]

WHEATON. W. Commuting, congestion, and employment dispersal in cities with mixed land use. Journal of Urban Economics, v.55, n. 3, p.417-438, 2004.

WREDE, M. A continuous spatial choice logit model of polycentric city. Regional Science and Urban Economics, v. 53, p. 68-73, 2015.

YINGER, J. Urban Models with More than One Employment Center. Journal of Urban Economics, v. 31, n. 2, p. 181–205, 1992.

Article approved for publication in volume 31, number 3, 2021 of the Revista Nova Economia. Article in editing final stage. This text is a preliminary version of the article accepted for publication, made available in the SciELO *Preprints* database in December 2021. DOI: http://dx.doi.org/10.1590/0103-6351/6236.

About the authors

Rodger Barros Antunes Campos – rodgerantunes@gmail.com
Associate Researcher in the Regional and Urban Economics (NEREUS), University of São Paulo, São Paulo, SP, Brazil.
ORCID: https://orcid.org/0000-0001-5478-6510

André Luis Squarize Chagas – achagas@usp.br
Departments of Economics, University of São Paulo, São Paulo, SP, Brazil.
ORCID: https://orcid.org/0000-0001-7870-8736

About the article

Submission received on May 21, 2020. Approved for publication on January 21, 2021.

Author’s contributions

Rodger Barros Antunes Campos: Data collection, modeling and results analysis, writing and proofreading.
André Luís Squarize Chagas: Results analysis, writing and proofreading.

Conclict of interest statement

The authors declare that there are no conflicts of interest of personal, commercial, academic or financial nature between the authors and the publication of this article.