Real estate and rapid transit: estimating the market premium of the São Paulo Rail Network

Arthur Bazolli Alvarenga
M.Sc. student in Economics, Universidade Federal de Juiz de Fora (PPGE/UFJF).
E-mail: arthur.bazolli@estudante.ufjf.br
Ari Francisco de Araujo Junior
Department of Economics, Ibmec Minas Gerais (Ibmec-MG).
E-mail: ari.junior@professores.ibmec.edu.br
Admir Antonio Betarelli Junior
Department of Economics, Universidade Federal de Juiz de Fora (UFJF).
E-mail: admir.betarelli@ufjf.br

Abstract

Although rapid transit equipment such as subway lines can benefit the population of an urban center, the value attributed to it cannot be directly estimated, as it is not possible for consumers to acquire such structures in the market. Therefore, this paper uses the hedonic prices model approach to indirectly estimate the consumers' willingness to pay for proximity to subway (Metrô) and commuter rail (CPTM) stations in São Paulo, Brazil. Eight models were estimated, under different underlying hypothesis about the urban structure: monocentric, duocentric, or yet relaxing the centrality assumption, with the inclusion of a cumulative accessibility index. Other aspects of the real estate market were also estimated, such as property's attributes and neighborhood criminality. The results indicate that proximity to subway stations has a positive market premium, whereas commuter rail does not seem to be a relevant feature. In addition, the accessibility index revealed itself significant when considered, showing that its inclusion can possibly make the economic modelling of cities closer to the actual spatial distribution of jobs.
Keywords: real estate, job accessibility, spatial econometrics, urban economics.
JEL codes: C21, R31, R32

1 Introduction

In great urban centers, commuting needs can hardly be supplied by automobiles alone. Bertaud (2018) highlights that the adequate provision of transit infrastructure benefits a region with agglomeration economies by two main reasons: the labor market expands – workers can reach more jobs in less commuting time – and human interactions become more frequent, which leverages innovation and productivity. In contrast, a sub-optimal transit network can lead to efficiency losses to the whole region and reduced social welfare. Labor markets fragment, which reduces the economy's specialization, and transportation costs increase.

Shoup (2003) notes that infrastructure investments can reduce the operational cost of public services and increase the benefitted region's value by more than the construction costs. However, financing difficulties may prevent these investments from happening, which tends to fragment the urban space and to higher infrastructure provision costs. The Access to Opportunities Project, from the *Instituto de Pesquisa Econômica Aplicada* (IPEA)[1], reveals a pattern of spatial inequality in Brazil's urban cores, where central neighborhoods have better infrastructure and more job opportunities than the inner city (Pereira et al., 2019). This patter is the consequence, among other factors, of a history of low investments in mobility in Brazil, out of sync with the rapid growth of the country's urban centers (BNDES Setorial, 2015).

Land value capture (LVC) instruments can finance investments in urban infrastructure. Suzuki *et al.* (2015) examined experiences such as the case of Tokyo: the government promotes an urban

[1] Brazilian Institute of Applied Economic Research.

redevelopment transit-oriented development (TOD)[2], unifying properties and selling a denser, new real estate venture that cover part of the infrastructure costs of public facilities such as rapid transit lines. Brazil is also pioneer in such ventures. After the *Estatuto das Cidades* (Statute of the Cities) bill created a legal framework for urban operations involving the private sector (Brasil, 2001), the city of São Paulo had its first experiences, the Água Espraiada and Faria Lima urban operations (OUs), in which the municipality sells to real estate developers the air rights to densify above the baseline FAR (floor-area ratio)[3] and the revenues are used to urban infrastructure investments. However, these funds were not invested in rapid transit and the experience also had unexpected side effects, resulting from a supply constraint imposed by reducing the baseline FAR reduction in the whole city (Suzuki et al., 2015).

In the academic literature, the valuation of willingness-to-pay (WTP) for accessibility is usually estimated via hedonic price models, which allows the analysis of the implicit demand for different attributes in the housing market (Brueckner, 2011). This methodology led to mixed evidence: Duncan (2008) identified that proximity to light rail stops in San Diego (USA) valued apartments and houses differently; in Phoenix (USA), Atkinson-Palombo (2010) found a similar pattern. In Brazil, Hermann and Haddad (2005) found an ambiguous relationship for São Paulo, while Seabra *et al*. (2016) concluded that Recife's rapid transit has a negative effect in property values. For Nairobi (Kenya), Nakamura and Avner (2021) estimated that job accessibility is highly valued in both formal and informal markets, with a heavier toll on poorer residents.

This work develops a hedonic model for the real estate market of São Paulo, controlling for different property and neighborhood characteristics, with the aim of identifying the locational advantage premiums of proximity to transit stations and public transit accessibility. The study contributes to the field literature in different scales. At the local scale, our econometric strategy allows us to estimate the indirect effects of urban amenities. Also, the level of data concerning properties and the city's characteristics sheds light on the city's housing market, especially concerning recent rail expansions. At the national level, to the best of our knowledge, this paper is pioneer in incorporating into a hedonic model the cumulative accessibility measure developed by IPEA: this has the advantage of directly informing a property's locational advantages derived from access to jobs, whereas using distance to a single point characterized as the core business district (CBD) is not an exact accessibility measure (Higgins and Kanaroglou, 2016). In addition, it allows a finer modelling of urban structure in applied works, relaxing the atomized job premise that underlies the absence of a local accessibility measure.

At a global scale, our evidence shows the major role played by transit stations and job accessibility in a large city in a developing economy. This can further enhance data-driven public policies and stimulate the debate regarding transit financing alternatives in face of budgetary restrictions, common not only in Brazil, but most of Latin America and other developing nations.

The rest of this paper is divided as follows: section 2 gives background information on the city of São Paulo and the theoretical framework. Section 3 presents the data and methodology used, while results are in section 4, and section 5 concludes with final remarks.

2 Background

2.1 The city of São Paulo

Home to 12 million people, São Paulo lies in the heart of the fourth biggest metropolitan region of the world: the São Paulo Metropolitan Region (RMSP), with an additional 10 million inhabitants in the surrounding cities (United Nations, 2019).

---

[2] TOD integrates urban planning and land use with public transit, creating walkable and highly accessible spaces (Suzuki et al., 2015).
[3] FAR is the ratio between computable constructed area ant the lot's size. It determines the building potential of a property. In the case of São Paulo, the municipality sets FAR limits for each zone (Prefeitura do Município de São Paulo, 2022).

Not so different from other Brazilian cities, the region faces an accessibility challenge. Not only are job opportunities mostly concentrated in the central regions of the capital, but the accessibility is both limited and unequal, as individuals in lower income quintiles usually live farther away. However, public policy in recent decades has been trying to catch up with the transit network deficit: from 2005 to 2018, the subway network (Metrô) has expanded around 80 per cent, from 56 kilometers (CMSP, 2018) to 101 kilometers (CMSP, 2021), and connecting previously underserved regions with the rest of the city. While the subway itself lies completely inside the capital, a commuter rail service, CPTM, reaches other cities of region. Both companies are managed by the state of São Paulo and operate complimentary, with tariff integration. Most of CPTM lines originated from older services dating back to mid-19th century, reorganized in the 1990s.

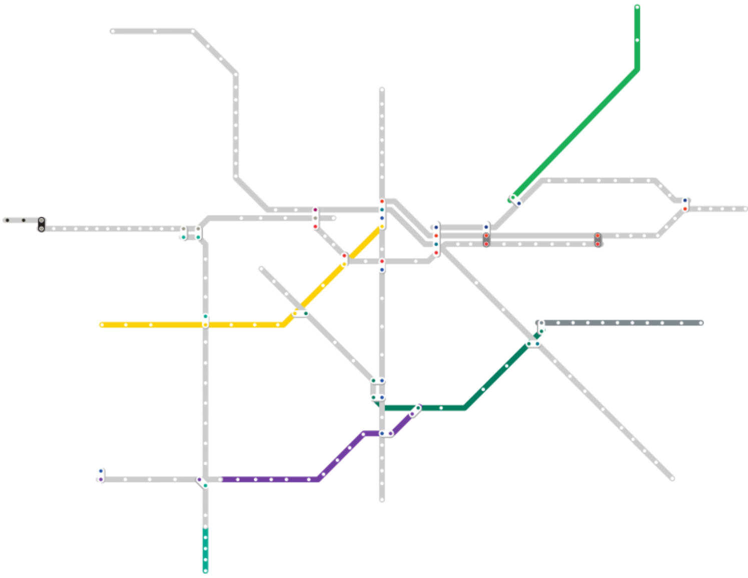

Figure 1: São Paulo's rail network evolution (2005-2018). Source: based on (CMSP, 2021).

2.2 Theoretical Framework

To understand the relationship between accessibility and land value, we depart from Alonso, Muth, and Mills canonical work, the AMM Model (Brueckner, 1987). The model considers a radial city with homogenous households that consume housing (in area) and a composite good and work at the CBD (core business district), a single point in the center of the circle.

In equilibrium, the households' tradeoff is between living closer to the job or consume more housing space, but in a farther region: this leads to a price and density gradient centered on the CBD. Transport costs are exogenous and treated as the opportunity cost of commuting, therefore, investments in transport systems reduce it if they provide accessibility gains (Mills, 1967). Despite the oversimplifying assumptions – such as elastic housing supply and homogenous transport costs – if we relax them the impacts of a commuting cost reduction on land value tend to remain, but unequally: new transit lines lead to an accessibility differential in space, which increases competitive pressure in the local real estate market. At the same time, the predicted density gradient become less steeper: all else equal, part of the households will opt to consume more space in farther regions, as the commuting costs in those places become smaller.

Fujita and Ogawa (1982) developed a model in which households and firms interact in their optimal location choices, relaxing the monocentric assumption. Households choose the work and residence places that maximize consumption of the composite good, while firms employ households and compete with them (and with other firms) for land consumption: the location choice is tied to a local potential, which reflects the agglomeration economies. The bigger this potential is, the more firms will tend to concentrate somewhere, giving rise to centralities.

An equilibrium can happen in multiple forms, from monocentric to acentric, depending on the parameters. The monocentric one happens only if the locational advantages are big enough to compensate the transport costs. Its opposite, the completely mixed configuration, happens when the transport costs are high enough to offset local potentials.

Although they have inherent limitations, both models seem to be coherent with empirical evidence. For instance, Bertaud and Malpezzi (2003) analyzed the spatial distribution in 48 urban centers, using census tract-level populational data and socioeconomic variables. Populational density was calculated for each tract dividing its population by the built-up area (i.e., net from parks, lakes etc.). An exponentially decaying density gradient was identified in most of the cases. This gradient tends to get less intense with an increase in income or a fall in commuting costs, just as predicted by theory. The authors highlight that in cities classified as market-oriented, as Singapore, price and density gradients are aligned. However, places with a strong history of interventions (such as zoning constraints), the density gradient tends to be inverted, but not the price one: in other words, the central locations are the most valued ones, but also the less dense. Moscow would be the extreme case, while Brasília and Johannesburg share similar characteristics.

3 Data and Methodology

The data used in this paper was treated in R (R Foundation, 2021). We used several databases containing residential properties, transit stations, crime data, urban amenities, and city limits, all of them georeferenced. Therefore, we used the R package *sf* (Pebesma, 2018) to calculate topological relations, such as the distance between an apartment and the nearest station. Moran's Index and the spatial weight matrix were calculated via the *spdep* package (Bivand and Wong, 2018), while spatial econometric models were estimated with *spatialreg* (Bivand et al., 2021). Maps and graphs were plotted with the help of *tmap* (Tennekes, 2018) and *ggplot* (Wickham, 2016) packages.

3.1 Data

3.1.1 Real Estate

Housing data was collected in June 2020 from "Quinto Andar", a Brazilian Real Estate company, containing around 5 thousand properties. They were georeferenced in the WGS 84 projection, widely used internationally, but were converted to SIRGAS 2000/UTM Zone 23S projection for compatibility with the other shapefiles.

The database was filtered to consider only apartments for rent, which are in monthly values and liquid from any taxes or fees. In addition, we needed to remove units in the same building, as this was preventing the inversion of the spatial weights matrix, as well as some minor errors, such as an apartment on the 50th floor (an error measure). This resulted in a sample of 1,260 observations, summarized in the **Table 1** below.

Table 1: Summary of housing data

Variable	Meaning	Min.	Q1	Median	Average	Q3	Max.
rent	Rent, R$	670	1,850	2,975	3,860.40	4,999	19,000
arcond	AC available	0	0	0	0.24	0	1
area	Area (m²)	12	54	80	106.45	138	450
bath	Number of bathrooms	1	1	2	2.16	3	6
floor	Floor number	0	3	6	7.13	10	30
furn	Furnished	0	0	0	0.30	1	1
gas_shower	Gas shower	0	0	0	0.47	1	1
new_ren	New or renovated	0	0	0	0.31	1	1
sauna	Sauna in the Building	0	0	0	0.24	0	1
ward_bed	Wardrobe in bedrooms	0	1	1	0.84	1	1

3.1.2 Neighborhood and Urban Amenities

From the São Paulo's city government "GeoSampa" database ("Geosampa," 2021), we obtained city limits, district boundaries parks, conservational units, and ZEIS (Special Zone of Social Interest) shapefiles. The city has five different ZEIS types, of which two were analyzed: ZEIS 1, 3 and 5. The first of them comprehends informal settlements and social housing and was used as a neighborhood variable as in (Seabra et al., 2016). The other two are areas with vacant buildings (ZEIS 3) or lots (ZEIS 5) in regions well-served by infrastructure[4]: the intuition behind these variables is to analyze if the proximity to abandoned buildings and empty lots is seen as a disamenity. Regarding Parks and Conservational Units (APAs)[5], we calculated the geodetic distance from the property to the nearest amenity (in straight line).

Crime data was collected from the State Public Safety Department website (SSP, 2021) regarding criminal reports registered in 2019 in each precinct (DP) of the city. Data are aggregated at the precinct level, whose limits do not correspond to the city district's boundaries, which can lead to the MAUP – moving areal unit problem (Almeida, 2012). We tested several workarounds in our model, such as crime density (per sq. km of the DP) or absolute numbers, and the best fit was to consider the logarithm of the heinous crimes (BRASIL, 1990) committed in the district that the property is on[6].

3.1.3 Accessibility

Metrô and CPTM stations data was also gathered from GeoSampa and consider only stops operational when the housing data was collected. The geodetic distance in straight line to the closest subway and commuter rail stations were included as separate variables, assuming that the two services, however integrated, may have different willingness to pay rates.

In line with the literature, the locational advantage of a property was estimated through the classical measure, distance to the CBD, but also directly via an accessibility index. We used a measured created by IPEA, which computes the cumulative job opportunities (as a share of the total jobs in the city) that one can reach in 60 minutes using public transport from a given point. In this index, the city is divided in a hexagonal grid comprising a few blocks, and the travel time from the hexagon's centroid and the job opportunities is computed using public transit GTFS data. The accessibility shapefiles were download via the R package *aopdata*, elaborated by the Access to Opportunities Project (Pereira et al., 2019).

Following the suggestion made by Higgins and Kanaroglou (2016), the goal is to measure the locational advantage directly, as the distance to the nearest station, a commonly used measure, is not always able to capture the real accessibility. Another intent is, as mentioned above, to replace the distance to the CBD (or the BDs). Such variable is a simplified version of the urban structure, and, although powerful in providing some insights, it atomizes the complex dynamics of a city's labor market and can lead to imprecisions.

The index is represented in the **Figure 2** below. In one hand, it is visible that the central areas have better accessibility, which reflects not only the greater offer of public transport, but also the high concentration of jobs. On the other hand, we can also note how the rail network plays an important role in the city's accessibility: for instance, the isolated yellow point in the east is Itaquera, a bus, train, and subway terminal. The vertical line in the center of the figure correspond to subway Line 1. the roughly horizontal axis going from the center to the east is a corridor comprising most of lines 3, 11, and 12, and the center-southwest axis is Line 4.

---

[4] For a complete definition, see "Geosampa" (2021).
[5] *Unidades de Conservação Ambiental*.
[6] The crimes considered were murder, assault followed by death, rape, and robbery followed by murder.

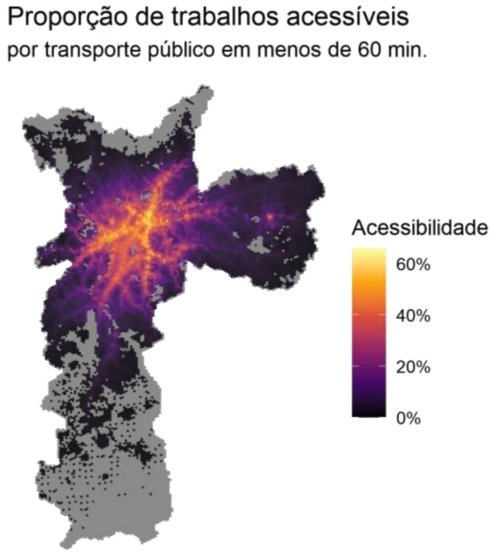

Figure 2: cumulative accessibility index. Source: based on Pereira et al. (2019)

3.2 Methodology

To estimate the hedonic price model, we start with an Ordinary Least Squares (OLS) estimation. In the presence of spatial dependency, however, OLS does not yeld accurate measures, as relevant information remains in the error term. In our model, this can happen if suppliers determine their rent price based on their neighbor's values, which is common in real estate. To incorporate spatial effects, we have also estimated the Spatial Autoregressive Mixed Complete model (SAC), which incorporates spatially lagged variables in the outcome and the error term. This model also has the advantage of rendering direct and indirect impact measures, which further enhances our analyses (Almeida, 2012; Lesage and Pace, 2014; Tyszler, 2006; Wooldridge, 2016).

3.2.1 Hedonic Model

The hedonic model decomposes expenditures with a good in its different attributes, which allow us to address their implicit value (Malpezzi, 2002). In the housing market, the analyzed characteristics are the structural aspects (such as area, bathrooms, and age) and the urban amenities, that influence positively or negatively the property value.

Higgins and Kanaroglou (2016) analyzed over 130 studies made between 1970 and 2010 regarding value capture investments in rapid transit in North America, 106 of them using hedonic models. However, their different functional forms, variables, and modelling choices lead to a wide variety in results, which makes it hard to compare them. Therefore, the authors propose a series of good criteria noted in the best works analyzed, such as: (i) incorporate relative accessibility measures instead of using distance as a proxy, as stated early in this work, (ii) control for factors such as other urban amenities and transit-oriented development (TOD), and (iii) model spatial dependency.

As the real estate market is tied to the local context, the most desirable features in a property and, more so, its surroundings, are not always the same. Therefore, we follow Hermann and Haddad (2005) and Seabra *et al.* (2016) and analyze urban amenities externalities by regressing the rent against the structural variables, letting in the error term the environmental characteristics. According to the spatial distribution of the residuals and their signals, we can compare whether the apparent externalities coincide with the local features, such as affluent neighborhoods, crime, parks, and proximity to employment centers and transit stations.

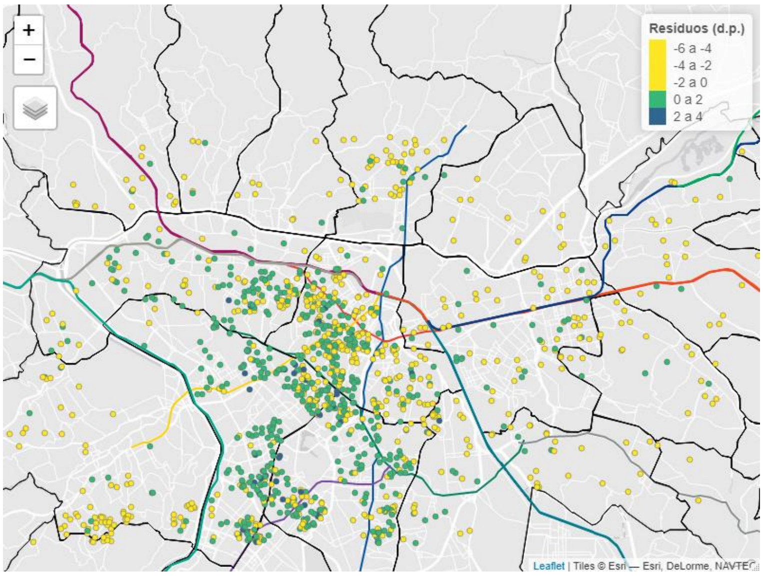

Figure 3: spatial distribution of the residuals from the structural regression.

The **Figure 3** above represents the spatial distribution of the residuals from the structural equation superposed on the subway and commuter rail lines. The color indicates the structural equation residual, in standard deviations, with positive values indicating a high expected impact from the urban amenities on the rent price. At first sight, we can see that the observations are concentrated in an approximately rectangular area, centered at Praça da Sé – the city's historical center. The apartments become sparse towards the edges of the region and are mostly concentrated near the central regions.

The map shows a patter similar to the one found in Hermann and Haddad's (2005) previous study for São Paulo: the affluent regions show mostly positive residuals, while the negative ones can generally be found in the inner city. Two examples are the negative residuals cluster near Paraisópolis, a ZEIS 1 settlement in the bottom left of the figure, and a positive cluster in the Pinheiros and Vila Mariana regions, southwest of the center. The Sé region, however, has a diverse pattern, encompassing from upscale neighborhoods like Higienópolis and Pacaembu to the now degraded Campos Elísios, as well as the historical downtown.

The visual analysis indicates negative residuals close to ZEIS zones of the three kinds. This relationship is not so clear with the parks, which are scarce, and the conservational units, which are mostly in the edges of the city and far from residential areas. The exception is the Ibirapuera Park, visible in **Figure 3**, as many properties around it have positive residuals. Regarding the accessibility variables, we can note that residuals are also higher for properties closer to Metrô and CPTM lines. The residuals also suggest a land value gradient – conditional on structural characteristics – irradiating not from the traditional Praça da Sé, but rather from Pinheiros and Vila Mariana regions, to where the financial district relocated in the last decades.

As for the accessibility variables, it is useful to look at the results found in the literature. Hermann and Haddad (2005) found no statistical significance for the São Paulo subway, while commuter rail was identified as an amenity. In the case of Recife, Seabra *et al.* (2016) found no relevant effect for the city's rail network. As the authors of both works highlight, the discouraging results might come from the ambiguous nature of these services (as they also generate noise pollution, for instance), or a low accessibility potential: specifically, Recife's Metrorec does not reach the most important and valued regions of the city, while São Paulo's Metrô, in 2005, was significantly smaller. That way, an updated study may identify a greater relevance of the subway network in rent prices, as it became a more important accessibility instrument.

As stated in the previous subsection, Higgins and Kanaroglou (2016) debate accessibility measures and land use. Defining accessibility as the potential to reach opportunities distributed on space (Páez et al., 2012), they suggest that using proximity to transit stations as a proxy for accessibility can be inaccurate, for not considering directly the actual accessibility gains provided by that equipment. This can be the case if a subway line does not serve a city's main job centers, or if access by car or bike is more competitive.

3.2.2 Econometric strategy

The regressions follow a general form,

$$P = f(S, E, A), \tag{1}$$

where P is the rent price, S is the structural variables vector, E comprises the environmental aspects, and A are the accessibility measures – distance to the CBD, distance to the transit stations, and the cumulative accessibility index.

The structural variables are the ones described in . The chosen environmental variables were the geodetic distances to the ZEIS 1, ZEIS 5, and conservational units, in kilometers, and the natural log of heinous crimes reported in the property's police precinct. Distance to parks and ZEIS 3 were not significant, so they were removed from the sample.

$$E = dist\_conserva + dist\_zeis1 + dist\_zeis5 + ln(crime). \tag{2}$$

Finally, the accessibility measures were tested in eight different combinations summarized in the **Table 2** below. This allows us to check if the accessibility index is a better measure of the locational advantages than the atomic structure assumed by the standard models and the distance to transit stations. Regression (1) assumes a duocentric model. In (2), we consider the monocentric model with the city's CBD at Praça da Sé, while (3) is also monocentric, but assuming that the Faria Lima Avenue is now a more accurate measure of the city's main job center. Sceneries (4) to (8) differ by incorporating the cumulative job opportunities that can be reached by public transit in 60 minutes, in different configurations. From (4) to (6), we still incorporate the distance measures to the business districts, as in (1) to (3). In scenery (7), however, we remove the BDs/CBD variables; finally, in (8), we also drop the distance to transit station variables and let the accessibility index alone represent the locational advantages.

Table 2: accessibility measures.

Variable	Description	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
dist_bd	Distance to the closer BD (duocentric)	•			•				
dist_se	CBD = Sé		•			•			
dist_farialima	CBD = Faria Lima			•			•		
dist_metro	Distance to the closest Metrô station	•	•	•	•	•	•	•	
dist_cptm	Distance to the closest CPTM station	•	•	•	•	•	•	•	
cmatt60	Cumulative job accessibility by transit, x10				•	•	•	•	•

We tested for spatial dependency with the Moran's Index. A positive and significant value indicates that properties tend to cluster in a similar fashion; that is, units with high rents are surrounded by units with high rent and the cheaper ones are neighbor to affordable units. In contrast, a negative Moran's I suggest that expensive units are surrounded by cheaper ones, and so on. The spatial weights matrix (W) was specified according to Baumont's (2004) procedure: first, we estimate an OLS model; then, we test for spatial autocorrelation in the residuals via Moran's I with several W matrices. The matrix that yields the most significant statistic is the ideal one, as it best captures the spatial dependency.

While the somewhat discretionary choice of the W matrix can lead to different coefficients, LeSage and Pace (2014) shows that most of this problem derives not from changing W itself, but due to model misspecification. Therefore, we now turn our attention to determining the model that best represents the data generating process. Tyszler (2006) suggests to begin with the SAC model: if only one of ρ (dependent variable lag) or λ (error lag) are significative, we should analyze respectively the SAR (spatial autoregressive) or the SEM (spatial error) models, while if both are relevant, we should proceed with the complete specification, which was the case in our work. The SAC is specified as follows (Almeida, 2012):

$$y = \rho W_1 y + X\beta + \xi, \tag{3}$$
$$\xi = \lambda W_2 \xi + e$$

where W_1y is the $n \times 1$ vector of spatial lags in the dependent variable, $\rho \in (-1, 1)$ is the spatial autoregression coefficient, $X\beta$ are the conventional regressors from the classical model and their coefficients, $W_2\xi$ is the $n \times 1$ vector of spatial lags in the error term, $\lambda \in (-1, 1)$ is the spatial autoregressive parameter of the error term, and e is a stochastic disturbance. The subscripts in the W matrices denote that they can be different; however, the same matrix was used as we have no *a priori* reason for changing them. The autoregressive parameters are restricted to the (-1,1) interval to prevent an explosive behavior.

The SAC model can be estimated via Maximum Likelihood (ML), but with two drawbacks: it is not consistent to heteroskedasticity, and depending on the number of observations, it can be computationally unfeasible due to the two jacobians required by the spatial weights matrices. Kelejian e Prucha (1998) suggested a generalized spatial two-stage least squares estimation (GSTSLS). Using W_1X and W_1^2X as instruments, it is possible to obtain heteroskedasticity-consistent estimators; besides, it is computationally simpler and faster than the MV method. Here, we conducted our main analysis using the GSTSLS estimators, while MV rendered similar results.

While in an OLS regression the beta coefficient of a variable can be interpreted as its partial derivative, $\widehat{\beta_r} = \partial y/\partial x_r$, this is not the case in the spatial models. According to LeSage and Pace (2009), this interpretation relies on the assumption – valid in the classical model – that variables are independent. However, as the regressors are spatially dependent, a change in x in a region can affect y's value somewhere else. This, in turn, can impact again the value of the other region and so forth, in a feedback effect.

Therefore, the correct way to assess the variables' influences on the rent is estimating its direct and indirect impacts and computing the average value across observations. The average **direct** impact of a variable is similar to the classical model's coefficient, as it indicates how the change in the value of x_k in region i, x_{ki}, affects the outcome of the own region, y_i:

$$\bar{M}(k)_{direct,k} = n^{-1} \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_{ki}}. \tag{4}$$

The **indirect** measure shows the impact in y_i due to a change in x_{kj}, with $i \neq j$, and is harder to estimate. However, as the **total** impact is the average of the partials of all y_i with respect to all x_{kj}, including $i = j$, we can calculate the indirect impact by subtracting the direct from the total, $\bar{M}(k)_{indirect} = \bar{M}(k)_{total} - \bar{M}(k)_{direct}$.

$$\bar{M}(k)_{total,k} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_{kj}} \tag{5}$$

4 Results

The eight models were first estimated through OLS. The Moran's I was best estimated using a 3 nearest neighbors matrix that resulted in a positive and significative statistic in al sceneries, as shown in **Table 3** below.

Table 3: Moran's I statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Moran's I	0.0536***	0.0549***	0.0482***	0.0387***	0.0318***	0.0099***	0.0365***	0.0443***

Note: *** = significant at 0,1%.

Following Tyszler's (2006) procedure, our estimation of the SAC model via ML rendered significant ρ and λ coefficients, with the exception of ρ for model 6. The SAR model also resulted in a relevant ρ, which gives more reason to use SAC over SEM. The statistics for these parameters are in the and the below.

Table 4: parameters of the SAC models estimated via ML.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ	0.2761	0.3473	0.0474	0.2945	0.3541	-0.0106	0.3477	0.3475
p-value	0.0008	0.0000	0.6878	0.0002	0.0000	0.9317	0.0000	0.0000
λ	0.6497	0.5877	0.7360	0.5863	0.4654	0.7898	0.5432	0.5385

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p-value	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000

Table 5: parameters of the SAR models estimated via ML.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ	0.4229	0.4416	0.3800	0.3977	0.4190	0.3666	0.4121	0.3820
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

As the results had heteroskedastic residuals, the SAC models were also estimated via GSTSLS, with similar impact measures to those obtained with ML, and can be found in **Table 6** below. There has been higher volatility between the sceneries than in the OLS estimations, however, the statistical significance from most variables remained, as well as their direction. The biggest differences were in the accessibility measures.

The structural variables' results are in line with the literature. However, we emphasize that some dummy variables such as AC and gas shower availability might represent high-end characteristics, and not just these specific features. Therefore, their indirect impacts can reflect not exactly an appreciation in a unit's rent when its neighbor installs an AC unit, but rather a spillover effect that happens on neighborhoods when higher quality constructions increase land values in the whole region.

Among the environmental variables there was more heterogeneity. The coefficient of $dist\_conserva$ points that each kilometer of distance from a conservational unit decreases the rent between 2.23 and 5.12 per cent in total effects. However, it is necessary to be cautious when interpreting this variable: as these features lies mostly in the edges of the city, far from almost all the observations, it is possible that the variable does not measure exactly the amenity that they provide in the housing market. As for the ZEIS zones, getting one kilometer away from a ZEIS 5 lead to an increase in the rents between 6.85 and 11.06 per cent. However, while ZEIS 1 shows similar results to the work of Seabra *et al.* (2016) for the Brazilian city of Recife, in most cases the results were not statistically relevant. The variable log $(crimes)$ indicates that, on average, an increase in one percentual point in the number of victims in heinous crimes is correlated with 0.20 per cent cheaper rents, varying from 0.16 per cent in scenery (7) to 0.22 per cent in model (7). The estimates were considerably smaller in the OLS models, that do not consider spatial dependency, reinforcing the importance of modelling it correctly when analyzing georeferenced variables.

As for the accessibility variables, proximity to subway stations is appreciated in the real estate market. The impacts are higher in regressions (1) to (3), which do not incorporate the accessibility index. When incorporated, the effect of $dist\_metro$ is around a quarter smaller when compared to the previous situations. The variable has its higher values in the duocentric structure represented in scenery (1): the impacts were 6.26 per cent directy, representing the rent's fall for each kilometer further from a station, and 4.22 per cent in indirect impact: this is the spillover effect of an appreciation (depreciation) of the neighborhood caused by the proximity (distance) from a station. The results for all SAC models are generally higher than the a-spatial ones and, more so, with less variability between sceneries. Moreover, the introduction of the accessibility index rendered $dist\_metro$ insignificant in model (5) via OLS, which did not happen when modelling the spatial effects.

The distance to CPTM's commuter rail stations showed an inverse relationship (indicating a disamenity), but no relevant impact in the rent. We highlight, however, that our data set covers only the capital, which houses about half the population of the whole metropolitan region. In São Paulo, there is more variety of transport options than in the rest of the metropolitan region, with the CPTM lines being superposed to Metrô lines and bus corridors in some regions. Thus, it is possible that the commuter rail lines do not represent a significant accessibility gain in the capital, while in the rest of the region, CPTM probably offers higher advantage – although we cannot demonstrate that in this work. This is not a coincidence, as Metrô's subway lines are design to cover the core of São Paulo with shorter distances between stations, while CPTM brings people from the surrounding cities to the capital and has more spaced stations. Even though the statistic was rejected in all sceneries, the reasoning behind the signal is similar to what Seabra *et al.* (2016) found for Recife (whose system is more similar to CPTM than to São Paulo's subway): physical characteristics, such as it being mostly at-grade or elevated, and stations less integrated to the urban tissue than the subway ones, may be perceived as negative externalities such as noise pollution.

The three variables of distance to BD/CBD pointed to the occurrence of the prices gradient preconized by the theory when $cmatt60$ is not included. The duocentric setting (1) yielded the highest impact measures: each kilometer farther away from a business district decreases rents by 4.07 per cent in total effects. Decomposing them, the direct impact is 2.30 per cent, while being in a farther neighborhood has an indirect effect – due to the neighbors' cheaper rents – of 1.77 per cent. The monocentric settings (2) and (3) had similar values.

When introducing the cumulative job accessibility measure ($cmatt60$) from the model (4) onwards, the distance to BD/CBD measures become irrelevant at the 5 per cent significance level, while the distance to the nearest subway station remains important. It is possible that the variables are so correlated that it is hard to disentangle their effects, which would explain to the loss of significance; however, the variance inflation factors (VIF) analysis made for the OLS regressions did not indicate any important multicollinearity problem: the worst case was observed in scenario (4) with a 6.5 FIV for $dist\_cbd$, still not problematic, and range from 1 to 3.5 for most variables.

The difference between incorporating or not other accessibility metrics is very high for $cmatt60$ estimates, as the total impact doubles from scenario (4) to (8), when the other variables are dropped. In the fourth setting (duocentric with accessibility index), each 10 additional percentual points of jobs accessible by public transit increases rents in 3.09 per cent directly and 2.36 per cent indirectly. In model (7), when $dist\_metro$ and $dist\_cptm$ are included but $dist\_cbd$ is not, the same 10 per cent variation in accessibility leads to 7.13 per cent higher rents: 3.83 per cent of them through direct effects and 3.30 per cent indirectly, for being in an accessible neighborhood. Finally, when distance to transit stations variables are also removed, $cmatt60$ represents a 10.80 percent premium on the rent, 5.97 per cent directly and 4.83 per cent indirectly.

Table 6: impact measures for the models estimated using GSTSLS.

(continues)

	(1)			(2)			(3)			(4)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log(area)	0.5154***	0.3687***	0.8841***	0.5172***	0.4503***	0.9675***	0.5091***	0.2514*	0.7605***	0.5136***	0.3811***	0.8947***
bath	0.0736***	0.0526**	0.1262***	0.0714***	0.0621***	0.1335***	0.0742***	0.0367*	0.1109***	0.0752***	0.0558***	0.131***
floor	0.0061***	0.0043*	0.0104**	0.006***	0.0052**	0.0112**	0.0061***	0.003*	0.0091**	0.006***	0.0044*	0.0104**
arcond	0.1499***	0.1072**	0.2571***	0.1513***	0.1317***	0.2831***	0.1486***	0.0734*	0.222***	0.1487***	0.1103***	0.2591***
gas_shower	0.0921***	0.0659**	0.1579***	0.0917***	0.0798**	0.1715***	0.0901***	0.0445*	0.1346***	0.0921***	0.0683**	0.1604***
furn	0.1141***	0.0816**	0.1958***	0.1147***	0.0998***	0.2145***	0.1132***	0.0559*	0.1691***	0.1148***	0.0852***	0.2000***
new_ren	0.0899***	0.0643**	0.1541***	0.0908***	0.079**	0.1698***	0.0895***	0.0442*	0.1336***	0.0883***	0.0655**	0.1538***
gym	0.0843***	0.0603**	0.1447***	0.0854***	0.0744**	0.1598***	0.0827***	0.0409*	0.1236***	0.0833***	0.0618**	0.1451***
sauna	0.0784**	0.0561*	0.1345**	0.0795**	0.0692*	0.1486**	0.0783**	0.0387'	0.1170*	0.078**	0.0579*	0.136**
dist_conserva	-0.0307***	-0.0219**	-0.0526***	-0.0233***	-0.0203**	-0.0437***	-0.0247***	-0.0122*	-0.0369***	-0.0268***	-0.0199**	-0.0467***
dist_zeis1	0.0222	0.0159	0.0381	0.025	0.0217	0.0467	0.0304'	0.015	0.0453'	0.0263	0.0195	0.0458
dist_zeis5	0.0492***	0.0352***	0.0843***	0.0561***	0.0488***	0.1049***	0.0444**	0.0219*	0.0663**	0.0469***	0.0348***	0.0816***
log(crime)	-0.1136***	-0.0813**	-0.1949***	-0.1113***	-0.0969**	-0.2083***	-0.1105***	-0.0546*	-0.1651***	-0.1167***	-0.0866**	-0.2032***
dist_metro	-0.0646***	-0.0462**	-0.1107***	-0.0541***	-0.0471**	-0.1013***	-0.0736***	-0.0363*	-0.1099***	-0.0474***	-0.0352**	-0.0826***
dist_cptm	0.0141	0.0101	0.0243	0.0057	0.005	0.0107	0.0031	0.0015	0.0047	0.0124	0.0092	0.0216
dist_cbd	-0.0233*	-0.0166*	-0.0399*							-0.0125	-0.0093	-0.0218
dist_se				-0.016*	-0.0139*	-0.0299*						
dist_farialima							-0.0185*	-0.0091*	-0.0277*			
cmatt60										0.0304*	0.0226*	0.053*

Note: *** significant at 0,1%, ** significant at 1%, * significant at 5%, ' significant at 10%.

2
(end)

	(5)			(6)			(7)			(8)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log(area)	0.5158***	0.4079***	0.9237***	0.5103***	0.3314**	0.8417***	0.5131***	0.4273***	0.9404***	0.5143***	0.3959***	0.9102***
bath	0.0738***	0.0583***	0.1321***	0.0757***	0.0491**	0.1248***	0.0751***	0.0625***	0.1376***	0.0768***	0.0591***	0.1359***
floor	0.0059***	0.0047**	0.0106**	0.006***	0.0039*	0.0098**	0.0059***	0.0049**	0.0108**	0.0058**	0.0045*	0.0103**
arcond	0.1496***	0.1183***	0.2679***	0.1482***	0.0962**	0.2445***	0.149***	0.1241***	0.2731***	0.1477***	0.1137***	0.2614***
gas_shower	0.0922***	0.0729***	0.165***	0.091***	0.0591*	0.1501***	0.0908***	0.0756***	0.1664***	0.0907***	0.0698**	0.1606***
furn	0.1151***	0.0911***	0.2062***	0.1145***	0.0743**	0.1888***	0.1153***	0.096***	0.2114***	0.1186***	0.0913***	0.2099***
new_ren	0.0891***	0.0704**	0.1595***	0.088***	0.0572*	0.1452***	0.0887***	0.0738**	0.1625***	0.087***	0.0669**	0.1539***
gym	0.085***	0.0672**	0.1523***	0.0823***	0.0534*	0.1357***	0.0828***	0.0689**	0.1517***	0.0788***	0.0607**	0.1395***
sauna	0.0795**	0.0629**	0.1424**	0.0779**	0.0506*	0.1284**	0.0779**	0.0649*	0.1429**	0.0807**	0.0621**	0.1428**
dist_conserva	-0.0235***	-0.0186**	-0.0422***	-0.0232***	-0.0151*	-0.0382***	-0.0207***	-0.0173**	-0.038***	-0.0128**	-0.0098*	-0.0226*
dist_zeis1	0.0239	0.0189	0.0429	0.0321*	0.0208'	0.0529*	0.0347*	0.0289*	0.0636*	0.0343*	0.0264*	0.0607*
dist_zeis5	0.055***	0.0435***	0.0985***	0.0438***	0.0284*	0.0722***	0.0432***	0.036***	0.0791***	0.0448***	0.0345***	0.0793***
log(crime)	-0.113***	-0.0894***	-0.2024***	-0.1174***	-0.0762*	-0.1937***	-0.1205***	-0.1004***	-0.2209***	-0.121***	-0.0931***	-0.2141***
dist_metro	-0.0374**	-0.0296*	-0.0670**	-0.0507***	-0.0329*	-0.0837***	-0.0462***	-0.0385**	-0.0847***			
dist_cptm	0.0088	0.0070	0.0157	0.0068	0.0044	0.0112	0.0058	0.0049	0.0107			
dist_cbd												
dist_se	-0.0120'	-0.0095'	-0.0216'									
dist_farialima				-0.0077	-0.005	-0.0127						
cmatt60	0.0314**	0.0248*	0.0562*	0.0335**	0.0218*	0.0553**	0.0376***	0.0313**	0.0689***	0.058***	0.0446***	0.1026***

Note: *** significant at 0,1%, ** significant at 1%, * significant at 5%, ' significant at 10%.

5 Final Remarks

The hedonic models analyzed allowed us to identify that the main attributes in São Paulo's real estate market follow the economical intuition, preserving the effects' signals although varying the values due to the different theoretical assumptions tested. While the distinction between neighborhood aspects and those exclusively related to accessibility is subtle, our results strongly suggest that rapid transit infrastructure implies a market premium. The relationship between rents and distance to subway stations was significant and positive in all scenarios: this provides even more evidence of rapid transit's positive externalities, as it points out to a relevant willingness-to-pay for transit proximity.

By including the accessibility index as an exploratory variable and the loss of significance of the classical distance to CBD measures, we show that the new measure can replace the previous ones. Therefore, we can bring urban economics' empirical analysis closer to the spatial structure observed in real cities, relaxing the atomized jobs premise. Also, as the distance to subway stations remains a relevant feature even when including the accessibility index, not only the variables can be used complimentary, but indicates that the subway network might bring more amenities than accessibility alone.

Some factors that contributed to the specification of the models were the abundancy of information on residential properties for rent and the high level of georeferenced features publicized by the municipality: controlling for these variables makes the estimated relation from accessibility to land value clearer. By modelling the spatial dependency, it was possible to eliminate the classical model's estimation bias and, moreover, distinguish direct effects from indirect effects (which suggest spillover and feedbacks) giving more evidence regarding the spatial-economical dynamics in course. Finally, the two-stages estimation (GSTSLS) provided results robust to heteroskedasticity.

The choice of using an on-line advertisements database had some implications: the scarcity of properties in the inner city (which limits our sample spatially), the absence of homes in informal settlements, and rent values that reflect the supply-side instead of an exact market equilibrium, as would be the case of a transactions database. Other limiting factor is the cross-section nature of the database, which limits our ability to inquire causal relationships.

A future extension that would further enrich the analysis are obtaining data for different periods, enabling a panel setting to estimate, for example, the impact of the recent subway station openings. Finally, the ongoing Covid-19 pandemics catalyzed an ongoing process of change in the way individuals and companies interact with space, for instance, with the widespread adoption of e-commerce and working from home.

6 References

Almeida, E., 2012. Econometria Espacial Aplicada, 1st ed. Alínea, Campinas.

Atkinson-Palombo, C., 2010. Comparing the Capitalisation Benefits of Light-rail Transit and Overlay Zoning for Single-family Houses and Condos by Neighbourhood Type in Metropolitan Phoenix, Arizona. Urban Studies 47, 2409–2426. https://doi.org/10.1177/0042098009357963

Baumont, C., 2004. Spatial effects in housing price models : do housing prices capitalize urban development policies in the agglomeration of Dijon (1999) ? (Research Report). Laboratoire d'économie et de gestion (LEG).

Bertaud, A., 2018. Order without design: how markets shape cities. MIT Press, Cambridge, Massachusetts.

Bertaud, A., Malpezzi, S., 2003. The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition 103.
Bivand, R., Millo, G., Piras, G., 2021. A review of software for spatial econometrics in R. Mathematics 9. https://doi.org/10.3390/math9111276
Bivand, R., Wong, D.W.S., 2018. Comparing implementations of global and local indicators of spatial association. TEST 27, 716–748.
BNDES Setorial, 2015. BNDES Setorial, n. 41, mar. 2015. BNDES Setorial 396.
Brasil, 2001. LEI Nº 10.257, DE 10 DE JULHO DE 2001.
BRASIL, 1990. LEI Nº 8.072, DE 25 DE JULHO DE 1990.
Brueckner, J.K., 2011. Lectures on urban economics. MIT Press, Cambridge, Mass.
Brueckner, J.K., 1987. Chapter 20 The structure of urban equilibria: A unified treatment of the muth-mills model, in: Handbook of Regional and Urban Economics. Elsevier, pp. 821–845. https://doi.org/10.1016/S1574-0080(87)80006-8
CMSP, 2021. Metrô São Paulo | Quem somos - saiba sobre transporte metropolitano [WWW Document]. Metrô de São Paulo. URL http://www.metro.sp.gov.br/metro/institucional/quem-somos/index.aspx (accessed 4.11.21).
CMSP, 2018. Linha do tempo – Metrô 50 anos [WWW Document]. Metrô de São Paulo. URL http://50anos.metrosp.com.br/index.php/linha-do-tempo/ (accessed 4.11.21).
Duncan, M., 2008. Comparing Rail Transit Capitalization Benefits for Single-Family and Condominium Units in San Diego, California. Transportation Research Record 2067, 120–130. https://doi.org/10.3141/2067-14
Fujita, M., Ogawa, H., 1982. Multiple equilibrium and structural transition of non-monocentric urban configurations 36.
Geosampa [WWW Document], 2021. . Sistema de Consulta do Mapa Digital da Cidade de São Paulo. URL http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx (accessed 5.17.21).
Hermann, B.M., Haddad, E.A., 2005. Mercado imobiliário e amenidades urbanas: a view through the window. Estud. Econ. 35, 237–269. https://doi.org/10.1590/S0101-41612005000200001
Higgins, C.D., Kanaroglou, P.S., 2016. Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg. Transport Reviews 36, 610–634. https://doi.org/10.1080/01441647.2016.1174748
Kelejian, H., Prucha, I., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating A Spatial Autoregressive Model with Autoregressive Disturbances. The Journal of Real Estate Finance and Economics 17, 99–121. https://doi.org/10.1023/A:1007707430416
Lesage, J., Pace, R., 2014. The Biggest Myth in Spatial Econometrics. Econometrics 2. https://doi.org/10.3390/econometrics2040217
LeSage, J.P., Pace, R.K., 2009. Introduction to spatial econometrics, Statistics, textbooks and monographs. CRC Press, Boca Raton.
Malpezzi, S., 2002. Hedonic Pricing Models: A Selective and Applied Review, in: O'Sullivan, T., Gibb, K. (Eds.), Housing Economics and Public Policy. Blackwell Science Ltd, Oxford, UK, pp. 67–89. https://doi.org/10.1002/9780470690680.ch5
Mills, E.S., 1967. An Aggregative Model of Resource Allocation in a Metropolitan Area. The American Economic Review 57, 197–210.
Nakamura, S., Avner, P., 2021. Spatial distributions of job accessibility, housing rents, and poverty: The case of Nairobi. Journal of Housing Economics 51, 101743. https://doi.org/10.1016/j.jhe.2020.101743

Páez, A., Scott, D.M., Morency, C., 2012. Measuring accessibility: positive and normative implementations of various accessibility indicators. Journal of Transport Geography 25, 141–153. https://doi.org/10.1016/j.jtrangeo.2012.03.016

Pebesma, E., 2018. Simple features for r: Standardized support for spatial vector data. The R Journal 10, 439–446. https://doi.org/10.32614/RJ-2018-009

Pereira, R.H.M., Braga, C.K.V., Serra, B., Nadalin, V.G., 2019. Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras – 2019. Texto para Discussão (TD) 2535 : Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras – 2019.

Prefeitura do Município de São Paulo, 2022. Coeficiente de Aproveitamento [WWW Document]. Gestão Urbana SP. URL https://gestaourbana.prefeitura.sp.gov.br/coeficiente-de-aproveitamento-ca/ (accessed 7.6.22).

R Foundation, 2021. R: The R Project for Statistical Computing [WWW Document]. URL https://www.r-project.org/ (accessed 6.13.21).

Seabra, D.M. da S., Silveira Neto, R.D.M., Menezes, T.A. de, 2016. Urban amenities and Real State Value: An Empirical Analysis for Recife City. Econ. Aplic. 20, 143. https://doi.org/10.11606/1413-8050/ea147474

Shoup, D.C., 2003. Is under-investment in public infrastructure an anomaly?, in: Methodology for Land and Housing Market Analysis. Taylor & Francis e-Library, London.

SSP, 2021. Dados Estatísticos do Estado de São Paulo [WWW Document]. Secretaria de Segurança Pública do Estado de São Paulo. URL http://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx (accessed 10.31.21).

Suzuki, H., Murakami, J., Hong, Y.-H., Tamayose, B., 2015. Financing Transit-Oriented Development with Land Values: Adapting Land Value Capture in Developing Countries, Urban Development. The World Bank. https://doi.org/10.1596/978-1-4648-0149-5

Tennekes, M., 2018. tmap: Thematic maps in R. Journal of Statistical Software 84, 1–39. https://doi.org/10.18637/jss.v084.i06

Tyszler, M., 2006. Econometria espacial: discutindo medidas para a matriz de ponderação espacial (Dissertação). Escola de Administração de Empresas de São Paulo, São Paulo.

United Nations, 2019. World Urbanization Prospects 2018: Highlights. New York.

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Wooldridge, J., 2016. Introdução à econometria: uma abordagem moderna, 6th ed. Cengage Learning, São Paulo.