

# What Drives the Patterns of Urban Land Use in a Developing Country? The Role of Transport Infrastructures and Natural Amenities<sup>1</sup>

- Ricardo Carvalho de Andrade Lima – Secretaria de Perícia, Pesquisa e Análise do Ministério Público Federal (SPPEA/MPF)
- Raul da Mota Silveira Neto – Programa de Pós-Graduação em Economia da Universidade Federal de Pernambuco (PIMES/UFPE)

## RESUMO:

O trabalho investiga os determinantes do uso do solo urbano para a cidade do Recife, no Brasil. Usando uma base de 98,198 lotes individuais e com um LW multinomial logit, nós evidenciamos que há muito efeito espacial local, o que indica que abordagens paramétricas podem ser inadequadas num contexto intraurbano. Além disso, observa-se alguns padrões gerais de uso do solo: as infraestruturas de transportes (o que inclui avenidas principais, rodovias e estações de metrô) tendem a atrair o uso comercial, enquanto as amenidades naturais (o que inclui os espaços abertos, a praia e o rio) tendem a incentivar o desenvolvimento residencial. Por fim, nota-se que, em média, a probabilidade de não-desenvolvimento do solo é maior em áreas próximas ao centro histórico da cidade e que a verticalização é maior em áreas próximas ao centro e as amenidades naturais.

**Palavras-chaves:** Análise do Uso do Solo Urbano, Estimação Espacial Não-Paramétrica, Transportes e Uso do Solo.

**Área Temática:** Econometria Espacial

## ABSTRACT:

This paper investigates the determinants of urban land use for the city of Recife, Brazil. Using a microdata of 98,198 individual parcels and a locally weighted multinomial logit, we show that there is considerable local spatial effects, which indicates that parametric approaches may be inadequate in an intra-urban context. Additionally, some general patterns of land use are identified: transport infrastructures (which include main avenues, roadways and subway stations) tend to attract commercial activities, while natural amenities (including the public open areas, the beach and the river) tend to encourage residential development. Finally, it is evidenced that, on average, the probability of unimproved lots is higher in areas near the city historical center and the verticalization is greater in areas near the city center and to the natural amenities.

**Keywords:** Urban Land Use Analysis, Spatial Nonparametric Estimation, Transportation and Land Use.

**JEL Classification:** R14, R52, C14.

---

<sup>1</sup> Trabalho concorrente ao Prêmio Paulo Haddad 2018.

## 1. Introduction

Urban land use plays a key role in urban economics since reflects the allocative decisions of firms and households (Duranton and Puga, 2015). Firms define their geographical location following the goal of minimizing costs, while households decide where to live maximizing their utility and considering the tradeoff between commuting costs and housing consumption. In this way, the spatial distribution of urban land use is a consequence of interactions that occur between individuals, businesses, environment and transport systems and the land usually goes to the sector with greater willingness to pay for its use, i.e. the highest bidder (O'Sullivan, 2012). Understanding the patterns of urban land use is of fundamental importance, since several urban problems are associated with land use, such as traffic congestion (Sarzynski et al., 2006, Antipova et al. 2011), violence (Browning et al., 2010, Twinam, 2017, Law et al., 2015) and the environment deterioration (Arnott et al., 2008). In a practical point of view, to know spatial distribution of land use is useful to explain the geography of the social life of the city, support the proper allocation of public services and infrastructure (such as schools, hospitals and roads) and help the formulation of more efficient urban policies (such as master plans, zoning laws and housing programs).

In recent decades, the patterns of urban land use of Brazilian cities experienced significant changes as a direct consequence of the growth of cities and due to intense and quickly urbanization process. The residential capital stock in Brazil practically doubled in the period from 1990 to 2008, an increase of 98.3%, well above the real GDP growth in the period of 80.5% (Institute of Applied Economic Research - IPEA, 2008). The urban space grew horizontally and vertically, although the second type of expansion with a greater intensity. For example, in 2000 about 56.5% of the population living in metropolitan regions resided in the central city, as early as 2010, this proportion fell to 54.8%, indicating a further growth of suburban cities. In relation to verticalization, there was an increase of 11.92% in the proportion of dwellings classified as apartment buildings from 2000 to 2010, contrasting the 2.68% decrease in the proportion of dwellings classified as traditional houses. (Brazilian Demographic Census, 2010). It is necessary to recognize that these phenomena are also observed in other Latin American cities and are related to the rise of gated communities (Borsdorf et al. 2007) and the decentralization of employment and formation of polycentric structures (Fernández-Maldonado et al. 2013). In addition to the recent changes in the spatial configuration of cities, another important characteristic of developing countries is the huge informal housing market. Although the proportion of individuals living in the informal market has declined over time, it is estimated that 21% of the urban population of the Caribbean and Latin America regions still live in informal settlements, which represents a population of approximately 105 million people. (UN-HABITAT, 2016). This strong informality may be a consequence of several factors, such as the intense rural-urban migration, scarcity of housing supply, income inequality and rigid land use regulations (Cavalcanti, Da Mata, and Santos 2017).

Considering the importance of the subject and the recent modifications in the urban configuration of Brazilian cities, the objective of this paper is to investigate the determinants of urban land use analyzing the case of Recife, an important Brazilian urban center and the oldest capital of the country. More specifically, using a novelty micro-data containing information on 98,198 individual lots and employing a semi-parametric spatial discrete choice model, we will associate the categories of land use (residential, commercial, residential multifamily, mixed use and unimproved) with different kinds of urban attributes: accessibility of the lot, the availability of natural amenities, and finally, the local transport facilities. Our study area is particularly interesting for two reasons: firstly, because the city does not have a zoning ordinance that restricts the allocation of land for different categories of use and, secondly, because Recife has natural amenities that generate strong heterogeneities in different spaces. Empirically, these singular characteristics allow us to investigate whether a competitive market allocation of land leads to a clear segregation of uses in the urban space and how natural amenities can shape the city's spatial configuration.

There is a wide body of empirical studies related to ours that can be subdivided into three groups. Firstly, applications include papers that analyzes how works the conversion of forest or agricultural land to urban use (Carrión-Flores et al., 2009; Li et al., 2013; Chakir and Parent 2009; Chakir and Le Gallo, 2013). Second, studies that investigates the changes of land use in an intra-urban context, attempting to

understand the urban attributes that affect land conversions for residential, commercial or industrial use (Verburg et al., 2004; Páez, 2006; Braimoh and Onishi, 2007; Wang et al., 2011; Wrenn and Sam, 2014; Bhat et al., 2015). And finally, the more similar ones to our, that seek to determine the factors that affect the allocation of land to different urban uses, like commercial, residential, industrial and mixed use (McMillen and McDonald, 1999; McMillen and Soppelsa, 2015; Jacob and McMillen, 2015). This last type of research is scarcer, since the most cities have a zoning ordinance that tend to interfere and alter the allocation of land away from what would be determined by a competitive market, making difficult to investigate the natural patterns of land use (McDonald, 2006).

This literature highlights the role of accessibility - commonly measured by proximity to Central Business District (CBD) - and the role of transport infrastructures for explain the patterns of urban land use. For example, Páez (2006) found that the implementation of Union City BART station (California) change the land use in areas close to the new facility. Wang et al. (2011), analyzing the city of Austin (Texas), evidenced that proximity to CBD encourages offices development but discourage commercial and industrial development. Additionally, they showed that areas close to major arterial roads tend to have a greater probability for commercial and industrial conversion. Using a detailed data of 474,190 parcels from Cook County (Illinois), McMillen and Soppelsa (2015) show that residential land use tends to be smaller in regions close to the CBD, to train lines and to major streets. The availability of natural amenities and the biophysical quality of the soil are also recognized important variables to explain the urban land use (Verburg et al., 2004; Wang et al., 2011).

A common concern of these empirical studies is to develop discrete choice models incorporating the spatial dependence hypothesis. It is well recognized that patterns of urban land use is strongly dependent of the neighborhood, i.e., it is undoubtedly a spatial process. The commerce, industry and other business activities tend to develop together to take advantage of agglomeration economies, either through input-output linkages, knowledge spillovers, consumption externalities, or labor pooling (Billings and Johnson 2016). Residential areas also have economies of agglomeration, mainly because they can share the same public infrastructure, like sewage, telecommunications network and electricity. Considering this, the more recent studies incorporated the spatial dependence into discrete choice models in two different ways: through parametric approaches which estimates a global measure of spatial dependence (in line with Anselin, 1988) and through the application of a semi-parametric approach based on extensions of the Locally Weight Regression (LWR) and allowing spatial non-stationary relations among the variables. In the first segment, Chakir and Le Gallo (2013), Wang et al. (2014) and Bhat et al. (2015) found that land use change has a positive autocorrelation, indicating that conversions to a specific use affect the conversions of neighboring parcels for the same use. In the side of semi-parametric approaches, Wang et al. (2011), Wrenn and Sam (2014) and McMillen and Soppelsa (2015) show that although there is a huge spatial variation in the magnitude of factors that affect the allocation of lots to different uses within a particular city, parcels that are geographically close tend to develop very similarly.

At this point, it should be noted that most empirical studies are conducted for North American cities and therefore very little is known about the determinants of urban land use in the developing world. The existing evidence indicates that the patterns of land use are sensitive to distinct institutional contexts, perhaps because of the different urbanization levels between countries. For example, for Lagos (Nigeria), Braimoh and Onishi (2007) found that residential development tends to be larger in areas close to the Central Business District (CBD), while McMillen and Soppelsa (2015) show just the opposite. In addition, much of the previous literature has prioritized the theoretical development of spatial discrete choice models, neglecting the empirical investigation of land use determinants. This created a large gap between methodological advances and empirical land use analyzes. Therefore, given the recent changes in the spatial configuration of cities in developing countries - mainly the increased urban sprawl, the verticalization and the huge informal housing market - and the lack of empirical studies analyzing urban land use in these countries, it is extremely important to develop new research in the area.

We intend to contribute in this sense, analyzing the determinants of land use for the case of Recife, a large Brazilian metropolis. As Recife has similar urban trends to other Latin American cities and the spatial distribution of land use is market oriented, we believed that, with some care, the results can be expanded to other cities of the developing world. Our paper also innovates when analyzing the factors

that affect the allocation of land for tall buildings and for unimproved lots, contributing to the literature on determinants of verticalization (Barr, 2010, Ahlfeldt and McMillen, 2017) and to studies that analyze the vacancy rates (Morandé et al., 2010; Nadalin and Iglori, 2016). Despite the strong increase in verticalization in Brazil, the empirical evidence about its determinants is practically non-existent.

The rest of the paper is organized as follows: in section 2, we discuss the main theoretical predictions regarding urban land use, in section 3 we briefly present the urban trends and characteristics of Recife and describe the database, section 4 shows details of the empirical strategy, section 5 presents the results and the discussion, section 6 presents some aspects related to the predictive capacity of the empirical models, and finally, section 7 presents the final remarks.

## 2. Theoretical Background: the landowner decision and the bid-rent functions

The patterns of urban land use are determined by economic interactions between the landowner and the different sectors of the urban economy (housing, manufacturing, commerce, offices and others). From the landowner perspective, he is confronted with the decision to allocate parcels of land - with constant quality and size - for different categories of use. In addition, it is assumed that the landowner is risk-neutral and has the objective of maximize the present discounted value of all rental flows of land parcels. Considering these assumptions, the decision rule that stems from a dynamic optimization problem (described in detail, for example, by Lubowski, 2002) is to choose, in each period of time  $t$ , the land use that generates the highest net return. The net return on land in use  $k$  is equal to the return generated by the parcel of land allocated in use  $k$  minus the potential opportunity cost of conversion.  $k$  is an element of the total set of different land uses, denoted by  $K = (1, \dots, k)$ . A landowner with land in use  $j$  will choose the use  $k$  at time  $t$  that satisfies:

$$ArgMax_k (R_{kt} - rC_{jkt}) \geq R_{jt} \quad (1)$$

Where  $R_{kt}$  represents the expected return of the land in use  $k$  at time  $t$ ,  $r$  is the discount rate and  $C_{jkt}$  is the expected marginal cost of convert land in use  $j$  to use  $k$  at time  $t$ . Therefore, the landowner's optimal decision is to convert the land in use  $j$  to the use  $k$  if the return of the second category exceeds the return of the first category (after deducting the conversion costs), not convert if the return in use  $j$  exceeds the return in use  $k$ , and finally, the landowner can be indifferent between the different categories of land use. The description of the landowner's behavior is useful to illustrate one of the main results of the urban economy: the land goes to the urban sector with greater willingness to pay for its use, or, in other words, the highest bidder (Fujita, 1989; O'Sullivan, 2012). Usually, the amount the sectors are willing to pay for land is called bid rent.

The bid-rent is a function of several factors, ranging from the biophysical quality of the soil, to the availability of transport infrastructure and other public goods. Considering the close link between bid-rent and land use, in the present paper, we will associate the categories of land use with three different types of urban attributes: the degree of accessibility of the lot (measured as the distance to the CBD), the proximity to transport facilities and, finally, the availability of natural amenities.

About the first aspect, the traditional Alonso-Muth-Mills model (Alonso, 1964; Muth, 1969; Mills, 1972) is useful to predict how the land will be allocated in different regions of the same urban area. One of the main conclusions of the model is to show that in equilibrium, the price of land decreases according to the distance from the CBD (Brueckner, 1987). As a consequence, when deciding where to live, households face a trade-off between consuming more housing (living farther from the CBD) or getting closer to the workplace and reduce the commuting costs (living closer from the CBD). That is, residential development in areas close to the CBD is encouraged by accessibility and, on the other hand, discouraged by higher land prices. The willingness to sacrifice accessibility for space is heterogeneous among different households and varies according to income (Duranton and Puga, 2015). In addition, because land is more expensive in areas close to the CBD, house builders choose to replace land by more capital and thus produce taller buildings (Fujita and Thisse, 2013; Ahlfeldt and McMillen, 2017). In relation to

business - such as commerce, industrial activities and offices – these tend to benefit from proximity to the CBD, due to potential agglomeration gains.

Transport infrastructures also play a key role in determining the patterns of urban land use. As described by McDonald (2006), a new transportation project changes the bid-rent of agents to match the savings in transport costs, and, consequently, affect the allocation of land to different uses. Firstly, the business sector is strongly benefited by the proximity of transport infrastructures: it facilitates the goods flow, reduces the freight costs, and improves access to the consumer market (O'Sullivan, 2012). Thus, the bid-rent for commercial or industrial uses is expected to be greater in places closer to the transport facilities, since these areas tend to be more profitable. In a recent paper, Billings and Johnson (2016) show that proximity to transport infrastructures explain a great part of the industrial agglomeration in Denver-Boulder-Greeley CMSA. In relation to the effects on residential land use, the transport proximity can increase or decrease the bid-rent of households. Easier access to transport infrastructures has the benefit of reducing costs related to commuting, improving access to amenities and public services and reducing the shop trips distance (McDonald, 2006). However, in areas very near to infrastructures, a reduction in the bid-rent may occur, due to the predominance of negative externalities such as congestion, pollution and noise. The bid-rent function for mixed land use behaves in a very particular way. Mixed land use is a response to situations where commuting costs are relatively high and agglomeration forces are low (Wheaton, 2004, Fujita and Thisse, 2013). Under these circumstances, the firms tend to locate more disperse and households are incentivized to work near of their residences, encouraging mixed use conversions. Thus, is expected that mixed use be associated with locations more distant from transport facilities and with the peripheral area of the city (in relation to CBD), where the capacity for agglomeration gains is limited.

It's also necessary to consider the role of environmental quality in household and business locational decisions. In this sense, Cho (2001) and Brueckner et al. (1999) expanded the Alonso-Muth-Mills model by incorporating the availability of natural amenities into the utility function of households. In the residential sector, is expected that bid-rent will be greater in places close to natural amenities. In such places, families benefit from having access to a more beautiful view, a cleaner air or more leisure and recreational options. However, with the only exception of specific sectors such as touristic and gastronomic, firms are not directly benefited by having a localization closer to natural amenities, and, therefore do not tend to have a greater bid-rent in areas with good environmental quality.

In summary, the theoretical results points that the degree of lot centrality and the proximity to transport infrastructures tends to increase the bid-rent of the business sector and generate an inconclusive effect on the households bid-rent (which will depend on the preferences of each of them). Areas close to the CBD also tend to have taller buildings and a smaller proportion of mixed land use. On the other hand, the household's willingness to pay for land use tends to increase in the vicinity of natural amenities, while the bid-rent of the business sector is only affected by this attribute in very particular cases.

### **3. Context and Data**

#### **3.1 Urban characteristics of the study area**

Our study area comprises the municipality of Recife, the main city of the Metropolitan Region of Recife (RMR) and the capital of Pernambuco state. The RMR is one of the main urban agglomerations in Brazil, being the richest metropolis of the North and Northeast regions (with a GDP of R\$ 75.8 billion, according to Brazilian Institute of Geography and Statistics, 2015) and the eighth richest in the country. The metropolitan region has a population of about 3.94 million of residents and Recife concentrates 41.2% of this contingent.

In recent decades, the city has undergone strong modifications in its urban structure, and, like other Latin American cities, it experienced processes of intense verticalization, urban sprawl and economic decentralization in relation to its historical center. As shown by Silveira-Neto (2016), in the period from 1991 to 2000, the districts closer to the Recife center grew by an average of 5.5% while the peripheral districts grew by 19.9%. This tendency of sprawl was attenuated in the most recent period (2000 to 2010), where both areas grew by about 8%. However, despite being an old city (founded in

1537) and having experienced some urban sprawl and employment decentralization, Recife remains with an eminently monocentric structure. In a recent paper, Belmiro et al. (2017) identified the centers and sub-centers of Recife and showed that the peaks in employment density are concentrated in a single region that is located in the intermediations of the historic center of the city, indicating the presence of a monocentric pattern. The authors also show that there is a negative and convex relationship between density and distance to the CBD, which reinforces the monocentric structure of Recife and confirm the Alonso-Muth-Mills model predictions.

Although the urban sprawl has been limited, the city grew to the sky: from 2000 to 2010, there was a 14.5% increase in the proportion of dwellings located in apartment buildings (Brazilian Demographic Census, 2010). The more vertical growth of the city is a consequence of two factors: the lack of urban infrastructure in peripheral regions (such as sewage), which inhibits the land use in suburban areas and, in second place, the higher income and the massive housing subsidies, which have led to increases in demand for areas with good infrastructure and closer to the city center (Silveira-Neto, 2016). Both factors valorized the land located in the qualified areas, encouraging the exchange of land by capital and, thus, the construction of taller buildings. In addition to these factors, the zoning code may also have facilitated the recent verticalization of Recife.

Unlike major cities such as São Paulo and Brasília, the land-use legislation of Recife<sup>2</sup> does not impose any type of direct constraint on the allocation of land to particular uses. For example, there are no legal divisions between residential and commercial areas. The zoning code of the city is characterized by restricting only the intensity of use (through maximum floor-area ratios and height limits) and the land-use in areas of environmental protection, of social interest and near of historical and cultural buildings. For this reason, the urban planning legislation is criticized as being very permissible for any kind of prejudice and adverse land use (Medina, 1997). In this way, the market interactions, rather than a government planner, determine the great part of the spatial distribution of land use in the city.

### 3.2 Data

Considering the purpose of explaining the factors that are associated with a particular urban land-use, we will adopt a novelty and official database that contains information of 98,198 individual lots in the city of Recife. This includes all parcels with property rights, leaving aside the informal ones. The database contains the following variables: area, geographical coordinates, land status (active or unimproved), number of floors, year of building construction and the form of land use. The database was updated in 2013, and is built by the local government through the Geographic Information System of Recife (RESIG). Initially, the parcels are divided into the following categories of land-use: Commercial<sup>3</sup> (7.78%), Industrial (0.24%), Mixed Use (0.98%), Residential (70.52%), Residential - Multifamily (6.69%), Unimproved (11.27%) and Special Classes (2.49%). Mixed use refers to lots that cover simultaneously residences and business activities. The last type of land use refers to schools, hospitals, religious temples, public buildings and other unrecorded institutions. For descriptive purposes, Figure 1 shows the spatial distribution of lots by categories of land use.

Although a simple visual inspection does not reveal specific patterns, the following general characteristics can be observed in the Figure 1: the predominance of multifamily residential lots on the seashore, the existence of residential agglomerations in the suburban areas, and the concentration of commercial activities in the vicinity of the CBD. Additionally, it is noteworthy that there are few lots with industrial and mixed use. The first fact is that Recife is a predominantly service-oriented city, with around 85% of GDP originated from this activity (Brazilian Institute of Geography and Statistics, 2015). The industrial plants of the Metropolitan Region are located in other municipalities, excluded from our study area. Low mixed use can be a consequence of the high agglomeration forces that exist in the city -

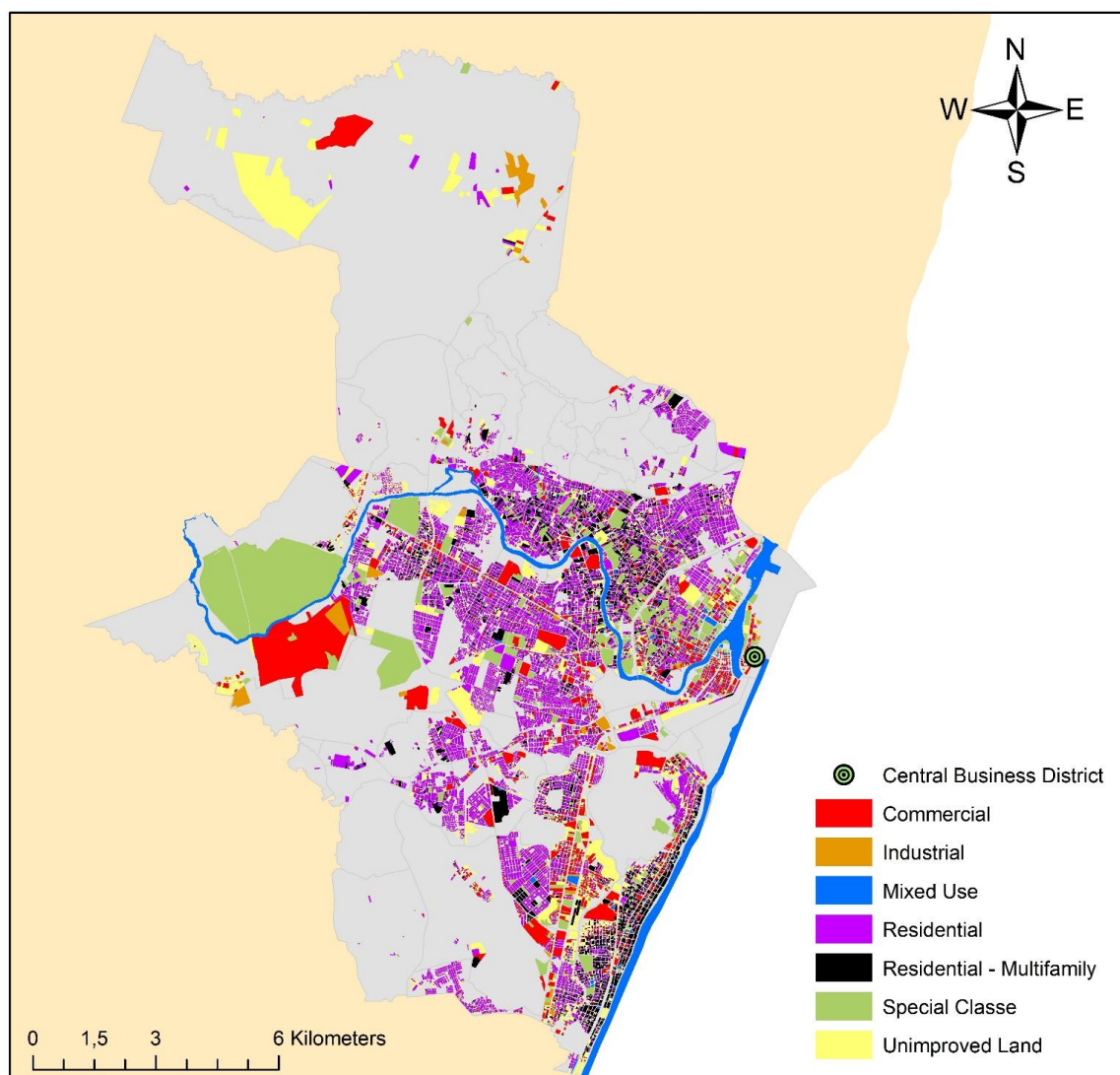
---

<sup>2</sup> The land-use legislation of Recife comprises the following ordinances: Law 16,176/1996, Law 16,719/2001 and Law 17,511/2008.

<sup>3</sup> This category includes a wide range of private business activities: retail stores, offices, parking's, gas stations, financial institutions, hotels and others.

since Recife remains with a strong monocentric structure - and by the lack of urban planning tradition that values mixed land use.

**Figure 1 - Spatial distribution of lots by categories of land-use.**



Source: Constructed based on Geographic Information System of Recife, 2013.

The explanatory variables were constructed with a geographic information system (GIS) program that allows calculating the distance in relation to each particular lot. Initially, we had taken as CBD the traditional historical center of Recife, known as *Marco Zero* or *Rio Branco* square. According to official records, this place is considered the starting point of Recife. As shown by Belmiro et al. (2017), the city's employment density peak are localized in the vicinity of the historic center and, in addition, about 59.75% of all commercial parcels of the city are located within a radius of 5km. In relation to transport variables, we included in our analysis the distance (in meters) in relation to the three major infrastructures of the city: the subway stations, the main avenues<sup>4</sup> and the roadways. Finally, the natural amenities include the *Capibaribe* River, the beach and the main public open spaces<sup>5</sup>. Table 1 describes the summary statistics of explanatory variables.

<sup>4</sup> We chose the main avenues as being those classified as arterial routes by the city's transit agency. In appendix (available upon request to authors), we list these main avenues.

<sup>5</sup> We use only the thirty largest open spaces in the city, which includes parks, squares and green areas. In appendix, we list these areas.

**Table 1 – Summary Statistics of Lots in Recife**

	Mean	Std. Deviation	Maximum	Minimum
Dist. CBD	6,131.2	2,335.3	18,297	26.20
Area (m <sup>2</sup> )	579.6	17,845.8	3,968,246	0.020
Dist. Roadways	3,090.9	1999.8	7912	14.624
Dist. Main Avenues	506.1	542.7	8,739	1.198
Dist. Subway Station	2,575.3	1,890.7	14,844	14.01
Dist. Public Open Areas	1,039.0	763.2	11,998	1.20
Dist. Beach	4,830.5	2,457.0	18,544	49.925
Dist. <i>Capibaribe</i> River	2,465	2,143.0	9,765	6.282

Note: The number of observations is 98,198 and the distances are measured in meters.

#### 4. Empirical Strategy: the Locally Weighted Regression

As a way to associate the categories of land use with different urban attributes, we will apply a non-parametric strategy of spatial modeling: the locally weighted regression (LWR). This approach is useful in an intra-urban context, to understand non-linear relationships and to allow spatial heterogeneity between the estimated coefficients. In the study of urban land use, it is very common that the relationships between the variables are different in each particular geographical area. For example, as discussed in subsection 2.2, residential development may be both attracted and driven away by proximity to transportation infrastructures. Locations very close to transport facilities tend to be more crowded, polluted and noisy, which discourages residential land use. However, locations relatively more distant (but still close) tend to attract families who wish to benefit from greater accessibility provided by transport facilities. When recognized that there is a significantly spatial variation in a particular phenomenon, a common practice in applied econometrics is to estimate different functions for different geographical areas. The LWR simply formalize this heuristic approach (McMillen and McDonald, 2004).

Through the LWR is possible to get different coefficients for each particular observation. Thus, unlike the traditional parametric techniques, the method allows spatial non-stationary. The usual procedure for estimation is the repetition of weighted least squares for each observation. However, in the case of discrete dependent variables, the estimation method involves the maximization of likelihood functions (Tibshirani and Hastie, 1987). In this paper, as we are interested in understanding the determinants of urban land use, the dependent variable will be discrete with  $M$  categories. In this way, like McMillen and McDonald (1999) and Wang et al. (2011), we apply a non-parametric multinomial logit. Considering there are  $M$  categories of land use, the local estimates are obtained by maximizing the following pseudo log-likelihood function for each observation:

$$\ln L_i = \sum_{j=1}^N w_{ij} [I_{0j} \ln(P_{0j}) + \dots + I_{Mj} \ln(P_{Mj})] \quad (2)$$

Where  $I_{Mj}$  is an indicator variable that takes one when the category  $M$  is chosen and takes zero, otherwise. The term  $P_{Mj}$  is the probability of choosing the alternative  $M$  and, is defined as follows (normalizing the base alternative to zero):

$$P_{Mj} = \frac{\exp(\beta'_{Mi} x_j)}{1 + \sum_{s=1}^M \exp(\beta'_{si} x_j)} \quad (3)$$

In equation (3), the term  $x_j$  is a vector of  $k$  explanatory variables that have potential to affect the urban land use, and  $\beta_{Mi}$  is the associated vector of coefficients. The maximization of the equation (2) produces  $n$  distinct estimates of the vectors  $(\beta_1, \dots, \beta_k)$  for each category of land use  $M$ , which we denote by  $\beta_{Mik}$ .



Before maximize the equation (2), it is necessary to establish a way of associating observations that are spatially close in order to obtain the spatial weights,  $w_{ij}$ . This is done through a Kernel weighting function, which having as input the geographical coordinates of  $i$  and  $j$ , generates the spatial weight  $w_{ij}$ . This type of non-parametric process follows the most fundamental law of spatial econometrics: the shorter the distance between  $i$  and  $j$ , the greater will be the weight given to these points. A range of kernel functions can be arbitrarily used at this stage. For example, a function that is widely used is the tri-cube:

$$w_{ij} = [1 - (d_{ij}/d_{iQ})^3]^3 I(d_{ij} < d_{iQ}) \quad (4)$$

Where  $d_{ij}$  is the Euclidean distance between point  $i$  and point  $j$  (based on geographical coordinates),  $d_{iQ}$  is the distance between  $i$  and the  $Q$ -th nearest neighbor and  $I(d_{ij} < d_{iQ})$  is an indicator function that equals one only when the condition is satisfied. Thus, this function put zero weight in the points that have a distance to the target point  $i$  higher than the  $Q$ -th neighbor and put a non-zero weight that decreases with the distance to those points that are below the threshold. The number  $Q$  is called the window size of the kernel function and is conceptually similar to the bandwidth parameter found in other kinds of kernel functions, these parameters determine how fast the weights decrease with the distance (McMillen and McDonald, 2004). Although the choice of kernel function is not sensible, the choice of the window size or bandwidth is critical to the estimation process (Wheeler and Paéz, 2010). Usually, the choice of these parameters can be made based on an interactive method of Cross-Validation (CV), through the maximum value of the log likelihood function or based on the Akaike Information Criterion (AIC).

Finally, it is important to note that the estimated vector of  $\beta_{Mik}$  does not correspond to the marginal effects of an increase of an explanatory variable  $k$  on the probability of selecting alternative  $M$ . In a multinomial logit approach, the marginal effects are given by (Croissant, 2011):

$$\frac{\partial P_{Mi}}{\partial x_{ik}} = P_{Mi} \left( \beta_{Mik} - \sum_s P_{Si} \beta_{sik} \right) \quad (5)$$

The second term of the expression in brackets represents the weighted average of the coefficients for all alternatives. The weights are the probabilities of choosing these alternatives. From the above expression, it is clear that the signal of the estimated coefficients may differ from the signal of the corresponding marginal effects.

## 5. Results and Discussion

In this section, we will present the results of the estimation of our land use model and discuss the key findings, comparing them to the theoretical predictions and to other empirical studies. In this way, subsection 5.1 briefly presents the results of the parametric multinomial logit and subsection 5.2 shows the results of the locally weighted estimation.

### 5.1 Standard Multinomial Logit Estimates

Initially, we chose to use five categories of land use in the empirical model: unimproved, residential, commercial, multifamily residential and mixed use. These categories correspond to 97.51% of the total number of observations. We opt to merge the industrial category with commercial ones, since the number of industrial parcels is not very representative (only 0.24% of total) and most of the manufactures in Recife is of light aspect, resembling commercial establishments. The parcels classified as “Special Classes” were not employed in the estimation, since they are characterized by generic uses and, in most cases, are not subject to market rules (such as public buildings, hospitals and churches). Thus, Table 2 presents the marginal effects of the multinomial logit model with five categories of urban land use. Additionally, the results of the estimated coefficients (with unimproved use normalized to zero, as a base category) are available in table 3A of the Appendix. The model was estimated by the maximum

likelihood method and all the explanatory variables are in logarithmic format. Each column of Table 2 represents a specific category.

The figures in Table 2 indicate that the greater the distance to the Central Business District (CBD), the greater the probability of unimproved parcels, and the lower the likelihood of commercial, mixed use and residential land use. That is, both households and businesses are attracted by the degree of lot centrality. More specifically, a 1% increase in lot distance to CBD reduces the probability of residential use by 5.76% and reduces the probability of commercial use by 8.59%. However, the use of lots for apartment buildings is positively affected by distance to the CBD, a counterintuitive result in face to the urban economy theory, as discussed in subsection 2.2. In relation to transport infrastructures (main avenues, highways and subway stations), we note that mixed use and commercial activity is attracted by these facilities, while residential is discouraged. The residential multifamily use is more common near main avenues and to roadways, which may be a consequence of the high land value in these places that incentive capital-intensive buildings.

The numbers of column (5) indicates that mixed use is more frequent in the proximity of the CBD and transport facilities, like the commercial parcels. Although this result is counterintuitive in relation to the theoretical motivation for mixed use (low agglomeration forces and high commuting costs), this suggests that the commercial side of mixed use is more preponderant than the residential side in determining the geographical location of the lot. Regarding natural amenities, a variety of patterns are observed: open areas (which include parks, green areas and squares) attract both types of residential use. However, the beach and the river proximity encourages commercial use, multifamily residential, mixed use and the unimproved of lots.

**Table 2 – The determinants of urban land use: standard Multinomial Logit marginal effects**

	(1)	(2)	(3)	(4)	(5)
	Unimproved	Residential	Commercial	R. Multifamily	Mixed Use
Dist. CBD	0.1454*** (0.0034)	-0.0576*** (0.0045)	-0.0859*** (0.0021)	0.0064*** (0.0020)	-0.0082*** (0.0006)
Area	0.0388*** (0.0013)	-0.1408*** (0.0019)	0.0440*** (0.0009)	0.0533*** (0.0008)	0.0047*** (0.0003)
Dist. Roadways	-0.0117*** (0.0014)	0.0443*** (0.0021)	-0.0225*** (0.0011)	-0.0081*** (0.0010)	-0.0019*** (0.0004)
Dist. Main Avenues	0.0195*** (0.0010)	0.0019* (0.0012)	-0.0141*** (0.0005)	-0.0056*** (0.0005)	-0.0017*** (0.0002)
Dist. Subway	-0.0059*** (0.0013)	0.0116*** (0.0017)	-0.0145*** (0.0007)	0.0101*** (0.0008)	-0.0018*** (0.0002)
Dist. Open Areas	0.0424*** (0.0013)	-0.0355*** (0.0017)	0.0024*** (0.0008)	-0.0089*** (0.0007)	-0.0005* (0.0002)
Dist. Beach	-0.0602*** (0.0018)	0.1377*** (0.0027)	-0.0341*** (0.0011)	-0.0384*** (0.0010)	-0.0051*** (0.0004)
Dist. <i>Capibaribe</i> River	-0.0529*** (0.0012)	0.0682*** (0.0017)	-0.0063*** (0.0009)	-0.0075*** (0.0008)	-0.0015*** (0.0003)
N. Observations	95708				
Log-Likelihood	-70884,156				
McFadden R <sup>2</sup>	0.1942				

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The standard errors are in parentheses. The outcome variables are the categories of urban land use: Unimproved, Residential, Commercial, Residential – Multifamily or Mixed Use. The explanatory variables are in logarithm format and the model were estimated by maximum likelihood.

Like the traditional regression framework, a shortcoming of the parametric multinomial logit is that the estimated parameters represent only the mean values. Thus, the estimation process implicitly assumes that all parcels of the city are affected in the same manner by the different urban attributes, regardless of their geographical location. That is, it imposes a single parametric and linear specification to

explain all local relations. This approach is too restrictive in an intra-urban context, where it is feasible that there is strong spatial variation (McMillen and McDonald, 2004; Wang et al., 2011). In the next subsection, we will present the locally weighted multinomial logit estimation, a less arbitrary approach that allows parameters to vary smoothly in the space.

## 5.2 LW Multinomial Logit Estimates

In order to obtain a set of coefficients for each particular observation, we estimate a Locally Weighted Multinomial Logit model through the log likelihood function given by (2). We used a tri-cube kernel weight function and, based on the Cross-Validation method, we chose a 40 percent Window Size. In addition, like McMillen and Soppelsa (2015), we employ the adaptive decision tree approach (Loader, 1999) to estimate the model in a smaller number of target points, which significantly reduces the estimation time. Thus, Figure 2 shows the Kernel density functions of the distribution of estimated coefficients, considering the unimproved use as the base category.

Initially, Figure 2 shows a strong heterogeneity in the estimated coefficients, so that in most distributions, there are both positive and negative values. For example, it is noted that proximity to the CBD may attract or displace any category of land use, indicating that urban areas are much more complex than the relationships established in the monocentric model. In this sense, there are few situations where each specific Kernel density contains only coefficients with a common signal. One such case is the distribution of the coefficients related to the main avenues and to commercial land use (density “J”), where all coefficients assume negative values. This reveals that proximity to major avenues encourages commercial use in all parcels of the city, without exception. Furthermore, almost all the coefficients that measure the relationship between public open areas and residential land use (both single and multifamily, density “E” and “S”) have a negative sign, revealing that this type of amenity strongly support the residential development.

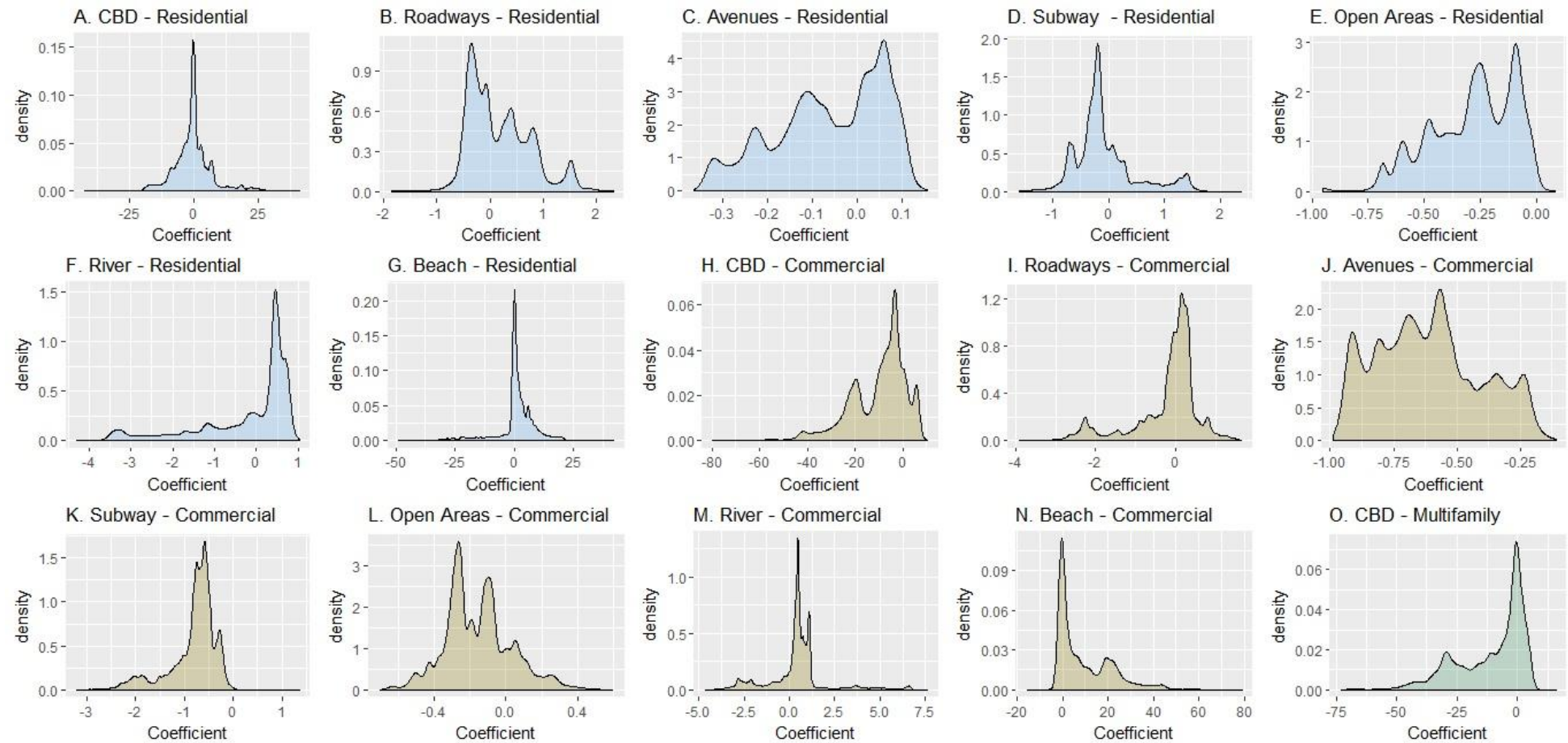
Another interesting aspect of the Kernel densities of Figure 2 is that in each of them there is a limited set of parcels that develop similarly among each other but diverges from the complete set, which generates the shape with peaks and valleys. For example, we observe that the distribution of coefficients related to proximity to the beach (density “U”) indicates that this natural amenity strongly encourages some parcels for multifamily use (the magnitude of the coefficients reaches up to -24.5) but exerts little or no influence in the large part of the lots used for this purpose.

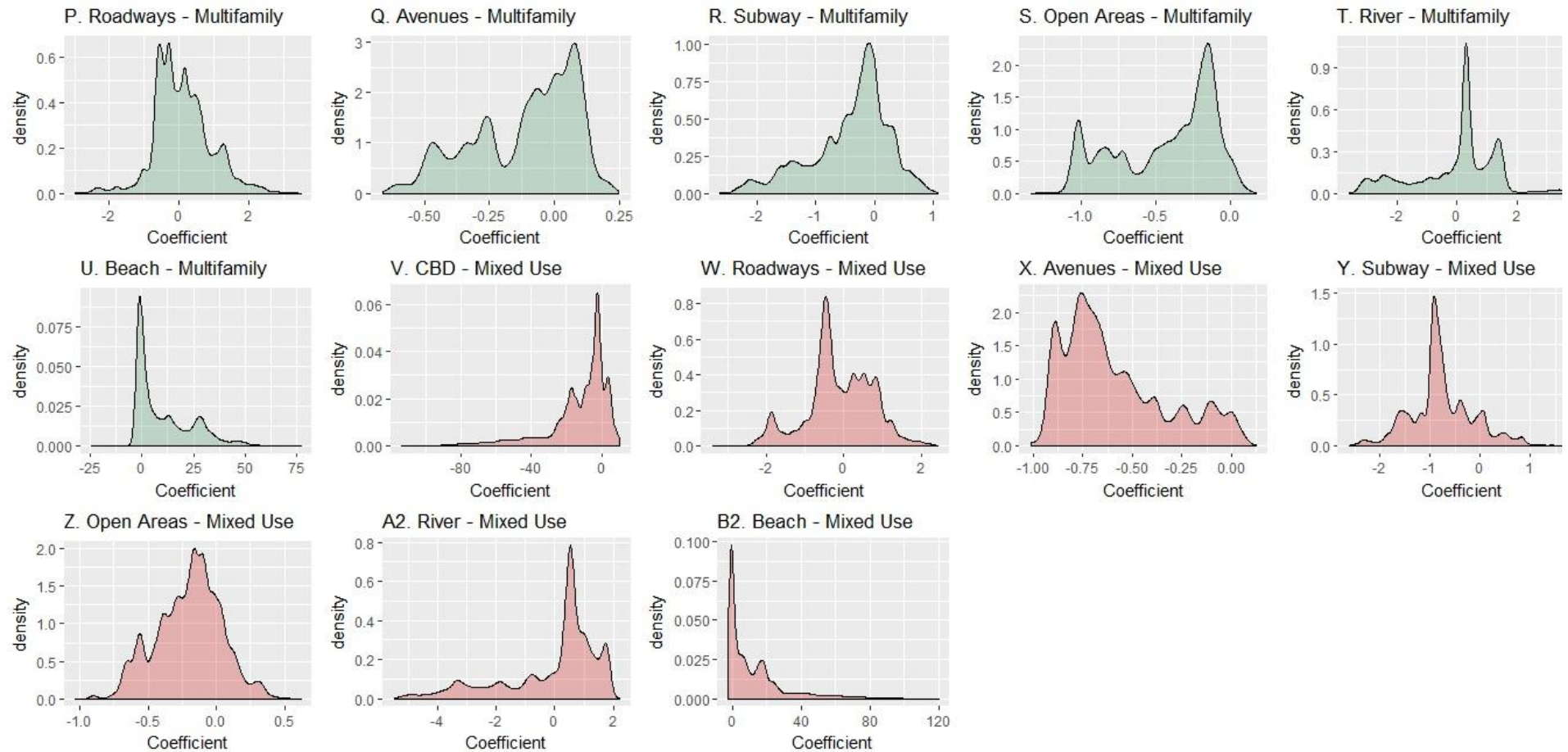
In order to facilitate interpretation of the results and to identify a general pattern of urban land use, Table 3 presents the mean, standard deviation, maximum and minimum of the marginal effects associated with each coefficients<sup>6</sup>.

---

<sup>6</sup> Before interpreting the results of Table 3, noted that, unlike the parametric models, it is not clear how to get an aggregate measure of statistical tests for the average marginal effects. Issues regarding statistical inference still need to be improved by the theoretical LWR literature.

**Figure 2 – Kernel Densities of Coefficient Distribution, LW Multinomial Logit.**





Note: The model is estimated by the locally weighted maximum likelihood using a 422 target points and a tri-cube kernel function with a 40 percent window size. The number of observations is 95,708. The outcome variables are the categories of urban land use: Unimproved, Residential, Commercial, Residential – Multifamily and Mixed Use. Where the Unimproved use is defined as the base category. All the explanatory variables are in logarithm format. The blue distribution represents the coefficients associated to residential land use, the yellow represents the coefficient associated to commercial land use, the green represents the coefficients associated to residential multifamily use, and the red distribution, those related to the mixed use.

**Table 3 – The determinants of urban land use: LWR Multinomial Logit marginal effects**

	(1)	(2)	(3)	(4)	(5)
	Unimproved	Residential	Commercial	R. Multifamily	Mixed Use
Dist. CBD	-0.0707 (1.0066) [-6.70, 16.22]	0.9549 (1.3943) [-4.42, 9.89]	-0.3685 (0.4731) [-9.99, 2.21]	-0.4557 (0.9758) [-9.32, 1.21]	-0.0599 (0.1171) [-2.41, 0.25]
Area	0.0319 (0.0425) [-0.21, 0.19]	-0.1446 (0.0956) [-0.47, 0.02]	0.0450 (0.0467) [-0.12, 0.32]	0.0631 (0.0692) [-0.01, 0.42]	0.0046 (0.0059) [-0.032, 0.07]
Dist. Roadways	0.0052 (0.0494) [-0.58, 0.49]	0.0395 (0.1229) [-0.50, 0.55]	-0.0460 (0.1051) [-0.53, 0.19]	0.0145 (0.0622) [-0.41, 0.75]	-0.0027 (0.0119) [-0.15, 0.24]
Dist. Main Avenues	0.0134 (0.0181) [-0.01, 0.17]	0.0147 (0.0279) [-0.06, 0.16]	-0.0262 (0.0262) [-0.22, 0]	0.0010 (0.0116) [-0.06, 0.21]	-0.0029 (0.0041) [-0.06, 0.01]
Dist. Subway	0.0022 (0.0728) [-0.48, 0.25]	0.0477 (0.0982) [-0.15, 0.49]	-0.0330 (0.0398) [-0.50, 0.13]	-0.0114 (0.0454) [-0.42, 0.31]	-0.0055 (0.0089) [-0.10, 0.03]
Dist. Open Areas	0.0311 (0.0368) [-0.02, 0.20]	-0.0265 (0.0359) [-0.18, 0.12]	0.0073 (0.0144) [-0.08, 0.17]	-0.0111 (0.0273) [-0.21, 0.07]	-0.0009 (0.0055) [-0.06, 0.02]
Dist. Beach	0.0580 (1.1330) [-15.97, 8.38]	-0.8201 (1.6179) [-10.04, 4.28]	0.2428 (0.4771) [-1.78, 9.91]	0.4725 (1.0036) [-1.90, 9.71]	0.0524 (0.1216) [-0.13, 2.55]
Dist. Capibaribe River	0.0196 (0.1361) [-0.55, 0.79]	-0.0525 (0.1472) [-0.92, 0.25]	0.0328 (0.0969) [-0.27, 1.05]	-0.0003 (0.0600) [-0.66, 0.54]	-0.0002 (0.0070) [-0.17, 0.04]

Note: The model is estimated by the locally weighted maximum likelihood using a 422 target points and a tri-cube kernel function with a 40 percent window size. The number of observations is 95,708. The outcome variables are the categories of urban land use: Unimproved, Residential, Commercial, Residential – Multifamily and Mixed Use. All the explanatory variables are in logarithm format. The standard deviations are in parentheses and the minimum and maximum are in brackets.

From Table 3, it can be observed that the greater the distance to the CBD, the smaller - on average - the probability of the unimproved land. This indicates that there is, in some way, a concentration of empty or vacant spaces near the city historical center. Nadalin and Iglioni (2016) found a very similar result for the case of São Paulo. According to the authors, a feasible explanation for the proliferation of empty uses in the historic center is that the place has lost some of its attractiveness as the city became more polycentric and thus began to receive a smaller share of public investments. In addition, in these places, there is a greater incentive for landowners to wait for higher prices rather than sell or rent their properties, since any investment in renewing the historic center tends to prompt rapid local gentrification (Brueckner and Rosenthal, 2009). In relation to column (2) and (3), it can be identified that although CBD can both attract and displace commercial and residential land use, commercial development tends to be higher in areas close to the center (negative mean), while the residential use tends to be lower in these areas (positive mean). The strong variability of the marginal effects associated with residential use (ranging from -4.42 to 9.87) indicates that in fact, households have heterogeneous preferences regarding the accessibility-space choice (Duranton and Puga, 2015). Column (4) of Table 3 shows that, on average, the probability of land use for apartment buildings is negatively correlated with distance to CBD, a result predicted by urban economics theory. Locations closer to the city center tend to have more expensive lands, which encourages capital intensive (rather than land intensive) constructions (Fujita, 1989; Ahlfeldt and McMillen, 2017). Mixed land use also tends to occur near the CBD (column 5), although there are

parcels that are positively influenced. This result may be explained by the fact that a great part of the buildings of the historical center of Recife were constructed in the eighteenth and nineteenth centuries, a period characterized by high commuting costs and by a public administration that employed urban planning traditions that prioritized the mixed land use. The historical characteristic of the city ends up offsetting the agglomeration forces that exist in the center that could incentives the conversion of mixed uses to purely commercial uses. This evidence is similar to that of McMillen and McDonald (1999), who analyzed the city of Chicago in 1920 and showed that mixed use was more common near the CBD and near the commuter train stations.

In relation to transport infrastructures, Table 3 presents a series of interesting results. Firstly, the proximity to the roadways and main avenues affects very similarly the different categories of land use. On average, these facilities attract the business sector and deviate households (both standard and multifamily residences). Similarity to Figure 2, there are no cases where the proximity of the main avenues discourages commercial use, since the marginal effects only assume negative or zero values (ranging from -0.22 to 0). As discussed earlier, roadways and avenues are key variables for the locational choice of firms, since these infrastructures facilitate the products flow, the commuting of employees and allow easier access to the consumer market (O'Sullivan, 2012, Billings and Johnson, 2016). That is, these infrastructures attract large population flows and generate agglomerative gains. The empirical literature of urban land use also points out that the greater proximity to roadways, freeways, highways and main avenues encourage the commercial and industrial development (Braumoh and Onishi, 2007; Wang et al., 2011; Jacob and McMillen, 2015). From column (1), we observe that parcels distant from these transport infrastructures tend to have a greater probability of unimproved. From column (5), we note that mixed use is more common near these facilities, reinforcing the idea that the commercial side of mixed use is more decisive than the residential side for determining the geographical location of the lot. In relation to residential sector, the results of column (2) and column (4) indicate that, on average, households choose to live in areas further away from roadways and main avenues, an indication that negative externalities (noise, pollution and crowding) overlap the benefits related to greater accessibility.

To understand better this effect, Figure 3 plot the correlation between the distance to the main avenues and the corresponding marginal effects of residential use (each point corresponds to a particular lot). Note that there is a strong non-linearity: as the distance from the avenues increases, the marginal effects decrease and reach negative values, indicating that the households begin to appreciate these transportation facilities. This behavior is consistent with the idea that the negative externalities generated by the main avenues are only concentrated in parcels very close to this infrastructure and support the empirical studies showing that local residents are commonly opposed to the installation of transport infrastructures in their immediate proximity (Ahlfeldt and Maennig, 2015).

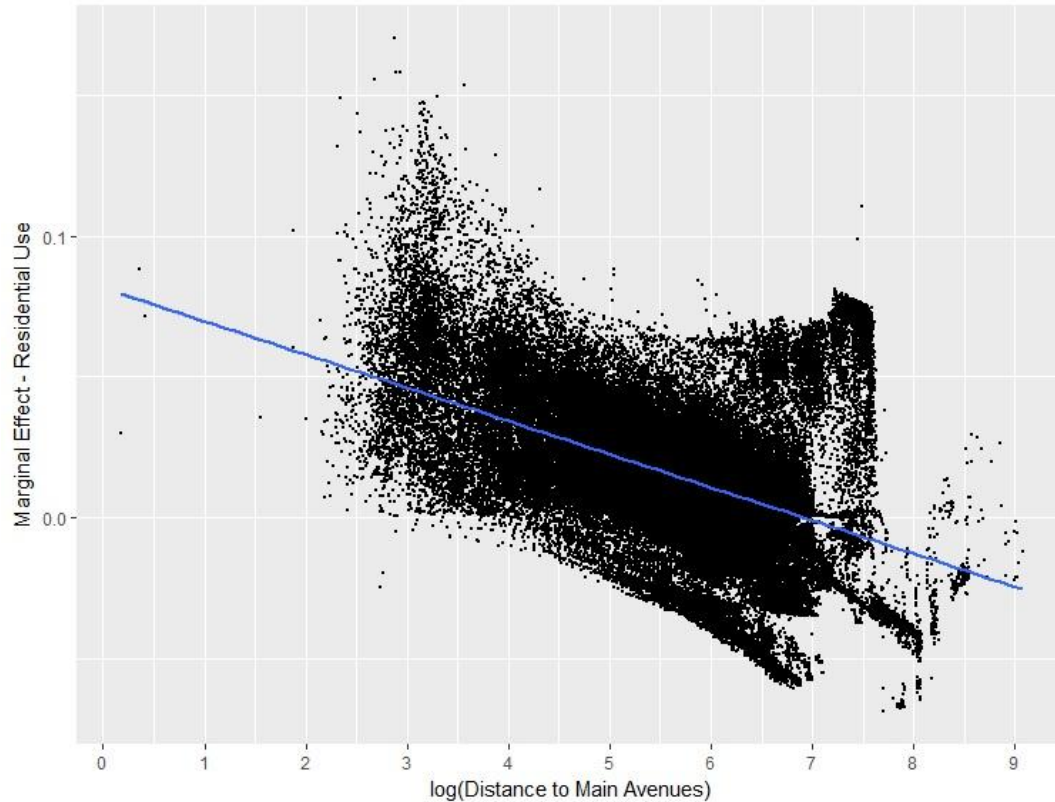
The subway stations also tends to discourage land use for residential purposes, since the average marginal effect assumes a positive value. Like above, this can be a consequence of negative externalities generated by the subway, especially noise, which is intensified because the train track of the city is not underground. Additionally, this explanation is in accordance with the study of Seabra et al. (2016), which shows that proximity to subway reduces the housing prices in Recife. Column (3) and (5) reveals that subway closeness increases the probability of commercial activities and mixed uses, a result well expected, since the subway station attracts a large influx of potential consumers.

The last three lines of table 3 shows the relations between natural amenities (public open areas, *Capibaribe* River and the Beach) and the five land use categories. Looking at the average marginal effect, a general result is easily identified: all three amenities tend to attract the residential sector and discourage commercial activities and the unimproved of lots. The only counterintuitive result is that the beach reduces – on the average - the probability of residential multifamily use. A simple visual inspection of the city's seashore reveals that it is replete of apartment's buildings (Figure 1). The large variability of the marginal effects (ranging from -1.9 to 9.71) and the fact that the seashore concentrates only a small proportion of the total number of apartments in the city can explains this result. Anyway, these evidences ratifies the urban economic theory that point to natural amenities as key factor for the households locational choices, influencing the spatial allocation of land use in cities (Brueckner et al., 1999). In addition, the empirical literature of urban land use finds similar results to ours. For example, Braimoh and



Onishi (2007) show that residential development tends to be closer to water bodies, and McMillen and Soppelsa (2015) finds that proximity to parks and to Lake Michigan increases the probability of residential land use. The hedonic prices applications also evidenced the importance of natural amenities: Seabra et al. (2016) shows that the proximity to Capibaribe River and to the Beach appreciate the price of housings in Recife.

**Figure 3 - Correlation Between Marginal Effect of Residential Use and Distance to Main Avenues.**



Note: The Marginal Effects were extracted from the specification described in Table 3, the distance to main avenues are in logarithm format and each point corresponds to a particular lot. The blue line corresponds to the regression line.

To conclude, the figures in Table 3 show a general pattern of urban land-use: while transport infrastructures are critical to business activity, the natural amenities play a key role in the attraction of residential sector. Note that these results diverge from the estimations of Table 2, showing that is extremely important to consider the wide local coefficient variation and the respective non-linearities. In the next section, we will analyze the predictive capacity of the empirical models.

## 6. Model Predictive Performance

In order to evaluate and compare the predictive performances of the parametric and the locally weighted Multinomial Logit, Table 4 shows the land use predictions of both approaches, comparing them with the actual land use. As a starting point, the first number of each cell in the table shows the predictions constructed under the assumption of randomly land use, using the sample proportions. For example, note that 11074 of the lots are unimproved, representing 11.57% of the total, so if we randomly predict the land uses, we would predict that  $0.1157 \times 69250 = 8012.6$  of the truly residential parcels are unimproved. The second number of each cell presents the prediction of the standard multinomial logit and the third number shows the prediction obtained with the LW multinomial logit. Thus, from the first cell, it can be observed that the parametric model accurately predict 851 unimproved parcels, while the non-parametric model correctly predict 1947 of these. The diagonal elements are of greater interest, since they represent the number of correct predictions of each particular empirical model.



**Table 4 – Predictive Performance.**

Actual Use	Predicted Land Use					Total
	Unimproved	Residential	Commercial	R. Multifamily	Mixed Use	
Unimproved	1281.3	8012.6	907.6	760.2	112.2	11074
	851	9318	306	599	0	
	1947	8232	330	565	0	
Residential	8012.6	50106.2	5675.6	4753.8	701.8	69250
	610	67413	695	532	0	
	859	67080	733	578	0	
Commercial	907.6	5675.6	642.9	538.5	79.5	7844
	88	5250	1822	684	0	
	138	4694	2410	602	0	
R. Multifamily	760.2	4753.8	538.5	451.0	66.6	6570
	56	4696	232	1586	0	
	91	4401	254	1824	0	
Mixed Use	112.23	701.84	79.49	66.58	9.83	970
	2	642	228	98	0	
	3	616	252	99	0	
Total	11074	69250	7844	6570	970	95708
	1607	87319	3283	3499	0	
	3038	85023	3979	3668	0	

Note: The first number of each cell presents the random prediction (based in sample proportions), followed by the parametric prediction and LWR predictions.

Thus, the diagonal numbers of Table 4 reveals that the LW Multinomial Logit has a better predictive capacity compared to the other two approaches, since it can predict correctly a greater amount of parcels, with the exception of residential land use. The best predictive performance of the locally weighted models is also evidenced in other studies (For example, in McMillen and McDonald, 2004). The numbers of Table 4 reveals that it is extremely difficult to predict the allocation of land for mixed use with the empirical models, which may be a consequence of the low proportion of lots in this category (only 1% of the total).

## 7. Conclusions

In last decades, the Latin American cities experienced intense modifications in their spatial configurations, including the increase of verticalization, urban sprawl, slum formation and the formation of polycentric structures. Faced with these new tendencies and in response to urban problems, public policies towards cities started to gain greater priority from central governments. In the Brazilian case, several policies were adopted, including the creation of a federal law that establishes general guidelines for subdivision of lots and zoning (Federal Law 6,766/79), the creation of the Ministry of Cities in 2003 and the implementation of a housing subsidies policy of approximately R\$ 34 billion.

In this context of cities growth and greater awareness of urban challenges, it becomes important to understand how work the interactions between households, business and the urban space. Thus, the objective of this paper was to explain the patterns of urban land use for the city of Recife, a large Brazilian metropolis. More specifically, with a novelty microdata of 98,198 individual lots, we investigated the role of transport infrastructures, lot centrality and the availability of natural amenities on the explanation of different categories of urban land use. Note that like other cities in developing countries, Recife has a large informal housing market. This makes our exercise particularly interesting since it allows us to understand to what extent the competitive market is able to explain the allocation of land use in a heavily informal city.

Through a locally weighted Multinomial Logit, we detected that there are considerable local spatial effects, indicating that parametric approaches are inadequate to explain land use in an intra-urban context. Additionally, we also noted that the LWR approach has a better predictive performance compared to parametric counterpart. Although there are significant local spatial variations, we identified some general patterns of land use: transport infrastructures (which include main avenues, roadways and subway stations) tend to attract commercial use, while natural amenities (including the public open spaces, the beach and the river) tend to encourage residential development. In addition, our results indicate that the probability of unimproved is greater in areas further away from natural amenities and transport facilities and in areas close to the CBD. It has also been shown that, on average, verticalization is greater in areas close to the CBD and to natural amenities. Although it is completely expected that the intensity of use be higher in more valued areas, the proliferation of uses that generate negative externalities (such as multifamily residences) near rivers and the seashore indicates that the zoning ordinances of the city was ineffective in encouraging the private internalization of these externalities. Finally, another interesting result is that mixed use occurs in areas near transportation facilities and the CBD, indicating that the commercial side of mixed use is more decisive than the residential side for determining the geographical location of the lot.

In general, our results indicate that the patterns of urban land use in Recife do not differ significantly from those of North American cities (McMillen and Soppelsa, 2015; Wang et al., 2011), and that a great part of the theoretical conclusions of urban economics are empirically corroborated. In practical terms, our results can be useful to improve the city urban planning and help to formulate public policies for urban land. Firstly, the choice of locations for the implementation of social housing programs should prioritize areas that have a natural vocation for residential development. The definition of urban zoning must also respect the equilibrium obtained through the competitive market - which generates a clear segregation of uses - in order to avoid conflicts and the increase of unimproved parcels. In addition, the decision to build transport infrastructures should also consider the potential changes in land use of the vicinity areas. Finally, the high number of vacant lots in the city (especially in the historic center) indicates that there is still a gap for urban growth, without the need to increase verticalization or sprawl.

## References

- Ahlfeldt, G.M., Maennig, W., 2015. Homevoters vs. leasevoters: A spatial analysis of airport effects. *J. Urban Econ.* 87, 85–99.
- Ahlfeldt, G.M., McMillen, D.P., 2015. The Vertical City: The Price of Land and the Height of Buildings in Chicago 1870-2010 (No. 180), SERC Discussion Paper.
- Alonso, W., 1964. *Location and Land Use*. Harvard University Press, Cambridge, MA.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*.
- Antipova, A., Wang, F., Wilmot, C., 2011. Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Appl. Geogr.* 31, 1010–1018.
- Arnott, R., Hochman, O., Rausser, G.C., 2008. Pollution and land use: Optimum and decentralization. *J. Urban Econ.* 64, 390–407.
- Barr, J., 2010. Skyscrapers and the skyline: Manhattan, 1895-2004. *Real Estate Econ.* 38, 567–597.
- Belmiro, C., Rodrigues, F., Silveira-Neto, R.M., 2017. De Centro Exportador a Polo de Serviços- a Cidade do Recife ainda é Monocêntrica Uma Investigação Empírica sobre sua Atual Configuração, in: VI Encontro Pernambucano de Economia. pp. 1–19.
- Bhat, C.R., Dubey, S.K., Alam, M.J. Bin, Khushefati, W.H., 2015. A new spatial multiple discrete-continuous modeling approach to land use change analysis. *J. Reg. Sci.* 55, 801–841.
- Billings, S.B., Johnson, E.B., 2016. Agglomeration within an urban area. *J. Urban Econ.* 91, 13–25.
- Borsdorf, A., Hidalgo, R., Sánchez, R., 2007. A new model of urban development in Latin America: The gated communities and fenced cities in the metropolitan areas of Santiago de Chile and Valparaíso. *Cities* 24, 365–378.
- Braimoh, A.K., Onishi, T., 2007. Spatial determinants of urban land use change in Lagos, Nigeria. *Land use policy* 24, 502–515.

- Browning, C.R., Byron, R. a., Calder, C. a., Krivo, L.J., Kwan, M.-P., Lee, J.-Y., Peterson, R.D., 2010. Commercial Density, Residential Concentration, and Crime: Land Use Patterns and Violence in Neighborhood Context. *J. Res. Crime Delinq.* 47, 329–357.
- Brueckner, J.K., 1987. The structure of urban equilibria: A unified treatment of the muth-mills model, in: Mills, E. (Ed.), *Handbook of Regional and Urban Economics*. Elsevier, pp. 821–845.
- Brueckner, J.K., Rosenthal, S.S., 2009. Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich? *Rev. Econ. Stat.* 91, 725–743.
- Brueckner, J.K., Thisse, J.F., Zenou, Y., 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *Eur. Econ. Rev.* 43, 91–107.
- Carrión-Flores, C., Flores-Lagunes, A., Guci, L., 2009. Land Use Change : A Spatial Multinomial Choice Analysis, in: *Agricultural & Applied Economics Association*. pp. 1–26.
- Cavalcanti, T., Da Mata, D., Santos, M., 2017. On the Determinants of Slum Formation. *Econ. J.*, Unpublished manuscript Forthcomin.
- Chakir, R., Le Gallo, J., 2013. Predicting land use allocation in France: A spatial panel data analysis. *Ecol. Econ.* 92, 114–125.
- Chakir, R., Parent, O., 2009. Determinants of land use changes: A spatial multinomial probit approach. *Pap. Reg. Sci.* 88, 327–344.
- Cho, C.-J., 2001. Amenities and Urban Residential Structure: An Amenity-embedded Model of Residential Choice. *Pap. Reg. Sci.* 80, 483–498.
- Croissant, Y., 2011. Estimation of multinomial logit models in R: The mlogit Packages An introductory example, CRAN-R.
- Duranton, G., Puga, D., 2015. Urban Land Use, in: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), *Handbook of Regional and Urban Economics*. Elsevier Inc., pp. 467–560.
- Fernández-Maldonado, A.M., Romein, A., Verkoren, O., Parente Paula Pessoa, R., 2013. Polycentric Structures in Latin American Metropolitan Areas: Identifying Employment Sub-centres. *Reg. Stud.* 0, 1–18.
- Fujita, M., 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge university press.
- Fujita, M., Thisse, J.-F., 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge university press.
- Jacob, B., McMillen, D., 2015. Border Effects in Suburban Land Use. *Natl. Tax J.* 68, 855–874.
- Law, J., Quick, M., Chan, P., 2015. Open area and road density as land use indicators of young offender residential locations at the small-area level: A case study in Ontario, Canada. *Urban Stud.*
- Li, M., Wu, J., Deng, X., 2013. Identifying Drivers of Land Use Change in China: A Spatial Multinomial Logit Model Analysis. *Land Econ.* 89, 632–654.
- Loader, C., 1999. *Local Regression and Likelihood*. Springer, New York.
- Lubowski, R.N., 2002. Determinants of land-use transitions in the United States: econometric estimation of a Markov model, in: *Forestry and Agriculture Greenhouse Gas Modeling Forum*. pp. 8–11.
- McDonald, J.F., 2006. Urban Transportation and Land Use, in: Arnott, R.J., McMillen, D. (Eds.), *A Companion to Urban Economics*. Wiley-Blackwel, pp. 281–291.
- McMillen, D., McDonald, J.F., 2004. Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and Application, in: Anselin, L., Florax, R., Rey, S. (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. New York: Springer-Verlag, pp. 225–229.
- McMillen, D., McDonald, J.F., 1999. Land use before zoning : The case of 1920 ' s Chicago. *Reg. Sci. Urban Econ.* 29, 473–489.
- McMillen, D., Soppelsa, M.E., 2015. a Conditionally Parametric Probit Model of Microdata Land Use in Chicago. *J. Reg. Sci.* 55, 391–415.
- Medina, L.L., 1997. A Legislação de Uso e Ocupação do Solo do Recife, in: *7o Encontro Nacional Da Associação de Pós-Graduação e Pesquisa Em Planejamento Urbano e Regional - ANPUR*. pp. 540–566.
- Mills, E., 1972. *Studies in the Structure of the Urban Economy*. Johns Hopkins University Press, Baltimore, MD.
- Morandé, F., Petermann, A., Vargas, M., 2010. Determinants of urban vacant land: Evidence from

- Santiago, Chile. *J. Real Estate Financ. Econ.* 40, 188–202.
- Muth, R., 1969. *Cities and Housing*. University of Chicago Press, Chicago, IL.
- Nadalin, V., Iglori, D., 2016. Empty spaces in the crowd. Residential vacancy in Sao Paulos city centre. *Urban Stud.*
- O'Sullivan, A., 2012. *Urban Economics*, 8th ed. McGraw-Hill Irwin.
- Páez, A., 2006. Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. *J. Transp. Geogr.* 14, 167–176.
- Sarzynski, A., Wolman, H.L., Galster, G., Hanson, R., Gordon, P., Richardson, H., 2006. Testing the Conventional Wisdom about Land Use and Traffic Congestion : The More We Sprawl , the Less We Move ? *Urban Stud.* 43, 601–626.
- Seabra, D., Silveira Neto, R.D.M., Menezes, T., 2016. Amenidades Urbanas e Valor das Residências: Uma Análise Empírica para a Cidade do Recife. *Econ. Apl.* 20, 143–169.
- Silveira Neto, R.D.M., 2016. Mais Carros, Menos Espraçamento e Menos Mobilidade? A (I)Mobilidade Urbana Recente na Região Metropolitana do Recife, in: *V Encontro Pernambucano de Economia*. pp. 1–22.
- Tibshirani, R., Hastie, T., 1987. Local likelihood estimation. *J. Am. Stat. Assoc.* 82, 559–567.
- Twinam, T., 2017. Danger zone: Land use and the geography of neighborhood crime. *J. Urban Econ.* 100, UN-HABITAT, 2016. *Urbanization and Development: Emerging Futures*. Nairobi, Kenya.
- Verburg, P.H., Eck, J.R.R. Van, Nijs, T.C.M. De, Dijst, M.J., Schot, P., 2004. Determinants of land-use change patterns in the Netherlands. *Environ. Plan. B Plan. Des.* 31, 125–150.
- Wang, Y., Kockelman, K.M., Damien, P., 2013. A spatial autoregressive multinomial probit model for anticipating land-use change in Austin, Texas. *Ann. Reg. Sci.* 52, 251–278.
- Wang, Y., Kockelman, K.M., Xiaokun, W., 2011. Anticipation of Land Use Change through Use of Geographically Weighted Regression Models for Discrete Response. *Transp. Res. Rec.* 111–123.
- Wheaton, W.C., 2004. Commuting, congestion, and employment dispersal in cities with mixed land use. *J. Urban Econ.* 55, 417–438.
- Wheeler, D., Paéz, A., 2010. Geographically Weighted Regression, in: Fischer, M., Getis, A. (Eds.), *Handbook of Applied Spatial Analysis*. Springer-Verlag Berlin Heidelberg, pp. 461–507.
- Wrenn, D.H., Sam, A.G., 2014. Geographically and temporally weighted likelihood regression: Exploring the spatiotemporal determinants of land use change. *Reg. Sci. Urban Econ.* 44, 60–74.