FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Deborah Doukas

An analysis of neighborhood and type of location choice

Rio de Janeiro
2021

Deborah Doukas

An analysis of neighborhood and type of location choice

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Grauação em Economia

Área de concentração: Microeconomia Aplicada

Orientador: Pedro Cavalcanti Gomes Ferreira

Rio de Janeiro
2021

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Doukas, Deborah
An analysis of neighborhood and type of location choice / Deborah Doukas. – 2021.
34 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola Brasileira de Economia e Finanças.
Orientador: Pedro Cavalcanti Gomes Ferreira.
Inclui bibliografia.

1. Habitação. 2. Favelas - Rio de Janeiro (Estado). 3. Mercado imobiliário. 4. Mobilidade residencial. I. Ferreira, Pedro Cavalcanti. II. Fundação Getulio Vargas. Escola Brasileira de Economia e Finanças. III. Título.

CDD – 332.320981

Elaborada por Kelly Ayala – CRB-7/7007

DEBORAH DOUKAS

“AN ANALYSIS OF NEIGHBORHOOD AND TYPE OF LOCATION CHOICE’’.

Dissertação apresentado(a) ao Curso de Mestrado em Economia do(a) EPGE Escola Brasileira de Economia e Finanças - FGV EPGE para obtenção do grau de Mestre(a) em Economia.

Data da defesa: 25/03/2021

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Presidente da Comissão Examinadora: Profº/ª **Pedro Cavalcanti Gomes Ferreira**

Pedro Cavalcanti Gomes Ferreira
Orientador

Marcelo Castello Branco Sant'Anna
Membro

Luciene Torres de Mello Pereira
Membro

Em cumprimento ao DECRETO nº 46.970 de 13/03/20 - Poder Executivo do Estado do Rio de Janeiro, DOE nº 047-A em 13/03/20, Art 4ª e Portaria MEC nº 343 de 17/03/20, DOU nº 53 de 18/03/20, que dispõe sobre a suspensão temporária das atividades acadêmicas presenciais e a utilização de recursos tecnológicos (em conformidade à legislação vigente), face ao COVID-19, as apresentações das defesas de Tese e Dissertação, de forma excepcional, serão realizadas de forma remota, inclui-se nessa modalidade membros da banca e discente.

Lucas Jóver Maestri
Diretor

Antonio de Araujo Freitas Junior
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação FGV

Antonio Freitas, PhD
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação
Fundação Getúlio Vargas

Instrução Normativa nº 01/19, de 09/07/19 - Pró-Reitoria FGV
Em caso de participação de Membro(s) da Banca Examinadora de forma não-presencial*, o Presidente da Comissão Examinadora assinará o documento como representante legal, delegado por esta I.N.
*Skype, Videoconferência, Apps de vídeo etc

Abstract

The urban fabric in developing countries is marked by the presence of slums. This paper aims to untangle the driving forces of the housing location choices of dwellers in a framework of multiple neighborhoods whose a fraction of its residences are located inside a slum. We embed a Location Choice Model presented in Bayer, Ferreira, and McMillan (2007) in an urban setting with both slums and multiple neighborhoods. Using the Brazilian 2010 Census, we estimate for heterogeneous residents of *Rio de Janeiro* their willingness-to-pay for several housing characteristics, including its location and access to specific public services, based on their own household features. Our results show that the rental prices in a *Rio de Janeiro*'s neighborhood would decrease by 5.42% if its fraction of households located inside a slum had risen by 10 p.p. and that dwellers who work outside their homes and do not have completed high school are willing-to-pay 47.48% more of their monthly income to live in *Rocinha* instead of living in the city center.

Keywords: Location choice model; Urban Economics; Neighborhoods; Slums

List of Figures

1 Rio de Janeiro's Planning Areas 21
2 Rio de Janeiro's Subdistricts 22

List of Tables

1 Residences in each Planning Area 5
2 Rented residences in each Planning Area 6
3 Percentage of residences in subnormal agglomerates for each Planning Area 6
4 Percentiles of household income in each Planning Area (only for households who rent residences) 7
5 Percentiles of per capita household income in each Planning Area (only for households who rent residences) 8
6 Hedonic Price Regression, First Step 14
7 Hedonic Price Regression, Second Step (% residences) 15
8 Hedonic Price Regression, Second Step (% population) 16
9 Heterogeneity Coefficients - WTP (% Montlhy income) 18
10 Homogeneous coefficients and WTP 19
11 Hedonic Price Regression, First Step - PA and Neighborhood Dummies . . 23

Contents

1 Introduction **1**

2 Literature **2**

3 Data **3**

4 Overview of the housing market in *Rio de Janeiro* **5**

5 Two-Stage Hedonic Price Regressions **8**

6 Location Choice Model **9**

6.1 The Environment 9

6.2 Choice of neighborhood and location to live 9

7 Estimation and Identification **10**

7.1 Estimation Procedure: First Step 10

7.2 Second Step and Instruments 12

8 Results **13**

9 Conclusion **20**

1 Introduction

The urban scenario is not always formed by a homogeneous landscape. In fact, most of the cities, especially the metropolitan centers, are formed by really heterogeneous neighborhoods. In developing countries, most of this heterogeneity comes from the differences between the slums and the other areas of the city, which are commonly referred to as the Central Business District (CBD). There are two principal distinguishing characteristics between the slums and the CBD: property rights and median income of the households. However, the spatial and urban economics literature has not yet focused its attention on the diversity among the locations within both the CBD and the slums.

Some questions remain open: why does somebody choose to live in a distant and poor region instead of living in a slum and vice-versa? What are the differences between them? How the particular demographic characteristics of each household clarifies its willingness-to-pay for living in a slum? Also, if someone decides to live in a slum, why choose this specific one instead of others?

These questions are particularly interesting in *Rio de Janeiro*. According to the 2010 census, approximately 22% of *Rio de Janeiro's* population lived in subnormal agglomerates[1]. Each neighborhood in the city belongs to one of the 33 subdistricts of the city[2]. In 32 of them, there is at least one subnormal agglomerate. Moreover, these subnormal agglomerates are way different from each other. For example, *Rocinha* and *Complexo de Manguinhos* are two of the biggest slums in the city. However, *Rocinha's* median income is higher than in poor regions that are not slums, like *Costa Barros*, whereas the same is not true in *Complexo de Manguinhos*.

In this sense, the distribution of income and per capita income of the dwellers who rent their residences are particularly interesting, because their moving costs, if they decide to move to another residence in *Rio de Janeiro*, are smaller than the ones faced by inhabitants who own their homes. Furthermore, the dweller of a slum cannot be the legal owner of the residence that he lives. Hence, the evidence that more than 70% of the households who rent a residence in *Rocinha* could be living in a poor area near to the city center (outside a subnormal agglomerate) is an indication that there are other factors behind the decision of living in a slum, other than just an income constraint.

Therefore, to analyze the driven forces of this location choice, I present a discrete choice model with fixed coefficients, in which each household chooses the residence it is going to live. This model accounts for differences between both the neighborhoods and the regions (i.e., inside and outside a subnormal agglomerate, even in the same neighborhood)

[1] Subnormal agglomerate is a formal definition of what constitutes a slum and will be rigorously explained in the data section.

[2] Also, each subdistrict is placed inside one of the zones regions of the city, namely west zone, south zone, north zone, and city center.

of the residence location. Its estimation embeds the Bayer, Ferreira, and McMillan (2007) structure in this neighborhood and regions framework. Moreover, I use Rio de Janeiro's 2010 Census microdata to estimate the model, as well as the adequacy level data of the residences from the city's *weighting areas*[3] to construct an instrument for the rental prices.

In addition, to investigate the impacts of expanding the fraction of dwellers in a particular neighborhood who live inside a subnormal agglomerate, I develop a two-stage hedonic price regression. The first stage is a regular hedonic price regression with neighborhood dummies. In the second stage regression, we use the neighborhoods' coefficients and the fraction of residences inside a subnormal agglomerate in each neighborhood to find a new coefficient that indicates the size of the penalty in the neighborhoods' rental prices for enlarging their subnormal agglomerates.

The estimated coefficients in the first part of the hedonic prices regressions show that there as significant differences in rental prices among the 33 *Rio de Janeiro*'s subdistricts. Besides, in the second step of the hedonic prices regressions, we find that if the fraction of residences in a neighborhood that are located on subnormal agglomerates increases by ten percentage points, its neighborhood rental prices will decrease, on average, 5.4% in comparison to prices in the city center.

Furthermore, the heterogeneity coefficients estimated in the location choice model reveal that the dwellers who work outside their homes and do not have completed high school are willing-to-pay almost 50% more of their monthly income to live in *Rocinha* instead of the city center. On the other hand, the ones who also work outside their home but have completed high school have a higher willingness to pay to live in the city center than in *Rocinha*.

Additionally, we found that the dwellers who work outside their houses have a higher willingness-to-pay for living in poor neighborhoods (that are not entirely composed by subnormal agglomerates) if they have completed their high school and that bigger and poorer households are more willing-to-pay to live in *Rocinha* than the other ones.

2 Literature

The recent studies in spatial economics have been able to incorporate in their models the heterogeneity of different locations and the interactions between them (Redding and Rossi-Hansberg (2017)). In Diamond (2016), a structural spatial equilibrium model is estimated to analyze the impact in terms of welfare and the causes of workers' skill geographic sorting among United States cities between 1980 and 2000. The author finds

[3]This is the second smaller geographic region in Brazil. The only smaller region is called *setor censitário*, which resembles the census tract of the US Census Bureau.

that the wage gap between college and high school graduates was not the only driver of inequality between these two populational groups. The differences in the level of rents and amenities among the cities that attract more a group than the other also plays a major role in the rise of inequality. Similarly to her approach, the amenities and level of rents among the neighborhoods will play a significant role in attracting individuals with different characteristics to distinct locations in the model presented in my project.

There are also studies that address location choices within cities, as in Cavalcanti et al. (2019). They present a housing choice model with heterogeneous agents. In their model, the choice is between formal and informal housing. The two main reasons for someone to choose to live in an informal location (slum) are the high costs attributed to formal housing and the several building constraints that apply only to formal residences. The major difference between their approach and the one used in this project is that while they are focused on finding the income thresholds that make an agent choose between formal or informal housing, I am interested in explaining what drives a different housing decision among agents with the same income level.

Ferreira et al. (2016) use a dynamic model to analyze the workers' choice of where to work and live. There are three living locations: rural zones, slums and city propers, the last two are part of the urban zone. Additionally, there are two labor markets: one in the rural zone and one in the urban zone. Living in the urban zone (both in and outside slums) gives the worker access to the urban labor market, which is more productive than the rural one. In contrast, the costs of living in the urban zone are higher than in the rural zone, even in the slums (in this case, it is a utility cost). Also, the impact of moving to the city on the educational attainment of the children is very expressive. They find that individuals moving from rural zones to city slums raise the country's productivity and provide upward mobility to their children. Even though I do not have the data about the work location in terms of neighborhoods, I will somehow incorporate the effect of the work location in the model used in this project.

3 Data

I will first clarify some definitions. The definitions of slums and subnormal agglomerates can vary depending on the reference being used. The definition of subnormal agglomerate stated here is the same given by the *Instituto Brasileiro de Geografia e Estatistica* (IBGE) in the Brazillian 2010 Census. Therefore, a subnormal agglomerate is a set formed by at least 51 households that do not have formal property rights and either have urbanization outside current standards or do not have the adequate provision of essential public ser-

vices[4]. The essential public services available in the data are: electricity, sewage system, water supply system, and garbage collection. What I define as a slum in this project is the sum of all subnormal agglomerates. This distinction is important because one neighborhood might have more than one subnormal agglomerate.

I define as a poor area the collection of households that:
i) Are not localized in a subnormal agglomerate, and
ii) Are in a neighborhood which belongs to one of the two *planning areas* of the city with the lowest median income.

The city of Rio de Janeiro has five *planning areas* (PA)[5]. **PA 1**, which composes the city center, includes the following neighborhoods: *Portuaria, Centro, Rio Comprido, Sao Cristovao, Ilha de Paqueta, Santa Teresa.* **PA 2**, which has a high median household income and is near to the center, includes *Zona Sul, Tijuca,* and *Vila Isabel.* **PA 3**, which has a low median household income and is near to the city center, is formed by *Zona Norte* without *Tijuca* and *Vila Isabel.* **PA 4**, which has a high median household income and is far from the city center consists of *Barra da Tijuca, Jacarepagua*, and *Cidade de Deus.* Finally, **PA 5**, which has a low median household income and is far from the city center, is formed by *Realengo, Bangu, Campo Grande, Santa Cruz, Guaratiba*, i.e., *Zona Oeste* without PA 4.

The second point of the previous definition is essential so that we can both distinguish the poor area of the CBD and keep it closer, in terms of income, to the slum. Lastly, I define the Central Business District (CBD) as the households that belong neither to the slum nor to the poor area. Note that the same neighborhood can have households in the slum and in the CDB or in the slum and in the poor area, but it cannot have households in both the poor area and the CDB. These three definitions will guide the estimation of the model and the evaluation of our results.

Henceforth, I will present the 2010 Brazilian Census data set. It contains information about the individuals and their households at the *weighting areas* level. More specifically, it consists of information about income, rent, educational attainment, physical characteristics of the household, access to essential public services, and the status of the household, that is, if it is rented, owned or something else. Since the dweller of a slum cannot be the owner of the residence that he lives, I will restrict our sample only to individuals that rent the household where they live.

The location choice in this project will be made in the city of *Rio de Janeiro.* Since it has 33 subdistricts, I am using them as neighborhoods. However, it is important to notice that some cities have only one subdistrict, despite having several neighborhoods. In order

[4]For more information about it, go to https://biblioteca.ibge.gov.br/visualizacao/periodicos/92/cd_2010_aglomerados_subnormais.pdf

[5]In the Appendix section, there are maps containing *Rio de Janeiro*'s *planning areas* and subdistricts.

to analyze this type of cities, one would have to make its neighborhoods compatible with the Municipal Mesh of census tracks of the 2010 Census.

Finally, I will also use dwellers data, such as income, household size, as well as the level of education and the workplace of the chief of the family to estimate the heterogeneous coefficients that will be specified in section 7.

4 Overview of the housing market in *Rio de Janeiro*

This section aims to elucidate some aspects of the housing market in *Rio de Janeiro*. In particular, I will present evidence that many of the slum dwellers in the city could be living in residences outside a subnormal agglomerate in any region of the city, but they decide not to.

Table 1 shows the distribution of residences among the city's planning areas (PA)[6]. There are more than 2 million residences in the city and more than one-third of them is located on PA 3, which has a median household income lower than the city center (also known as PA 1) and is also near to it. Besides, we separated one of the biggest subnormal agglomerates in the city (*Rocinha*) from its PA. That is because *Rocinha* is the only neighborhood in *Rio* that is solely composed of subnormal agglomerates.

Table 1: Residences in each Planning Area

Planning Areas	All Residences	All Residences (%)
PA 1	104721	4.88
PA 2 ex-Rocinha	380856	17.76
Rocinha	23399	1.09
PA 3	791926	36.93
PA 4	309265	14.42
PA 5	534278	24.91
Rio de Janeiro	2144445	100

Table 2 also exhibits the number of dwellers in each PA and its share in comparison to the rest of the city. However, its information is restricted to the residences that are rented. This distinction to the previous table is important because we will focus our analysis only on the rented residences. The reason why we are doing this is that a residence located inside a subnormal agglomerate cannot be formally owned. Hence, if a household wants to buy a residence, it has to buy it in a place outside the slum.

[6]The planning areas are defined in the previous section.

Table 2: Rented residences in each Planning Area

Planning Areas	Rented Residences	Rented Residences (%)
PA 1	35020	7.32
PA 2 ex-Rocinha	101480	21.20
Rocinha	8556	1.79
PA 3	176936	36.97
PA 4	79549	16.62
PA 5	77050	16.10
Rio de Janeiro	478591	100

Besides, we can determine which fraction of the residences in each planning area are located inside a subnormal agglomerate. In the table below, it is stated that less than 10% of the residents in the rich area near the city center (PA 2 ex-Rocinha) live inside a subnormal agglomerate, while more than 30% of the rented residences in the rich area far from the city center are located in a subnormal agglomerate.

Table 3: Percentage of residences in subnormal agglomerates for each Planning Area

Planning Areas	All Residences (%)	Rented Residences (%)
PA 1	29.47	16.84
PA 2 ex-Rocinha	8.29	6.02
Rocinha	100	100
PA 3	23.60	18.24
PA 4	23.97	35.58
PA 5	14.90	9.86
Rio de Janeiro	19.89	18.54

Using the information revealed in Table 3, we are going to define income thresholds for each PA that makes a household able to afford to live there. In table 4, we have the percentiles of household income in each PA. The exercise here is to assume that, for a particular PA, its slum dwellers are the ones with the lowest incomes. Therefore, for each PA, the income percentiles of the dwellers living in rented residences that are higher than the fraction of the residents who live inside a subnormal agglomerate indicate how much a household must earn to afford to live outside a slum in that PA.

For instance, less than 20% of the residents of PA 1 lived in subnormal agglomerates, so the households there with an income higher than one thousand *reais* were living outside the slum. Here we can see that more than 60% of *Rocinha*'s households have a higher income than this, so we can say that at least 60% of Rocinha's households could be living

in PA 1 (that area around the center). We can use this same line of reasoning for other planning areas as well. Hence, for example, in PA 2 ex-*Rocinha* (rich and near to the city center), less than 10% of people live in subnormal agglomerations. So whoever earns more than 1250 *reais* in *Rocinha* could also be living there. This value is between the .50 and .60 percentiles, so between 50% or 40% of *Rocinha*'s households could be living there.

Table 4: Percentiles of household income in each Planning Area (only for households who rent residences)

Planning Areas	.10	.20	.30	.40	.50	.60	.70	.80	.90
PA 1	650	1000	1200	1510	1810	2200	2706	3510	5513
PA 2*	1250	2000	2970	4000	5000	6310	8100	11010	16000
Rocinha	510	700	910	1020	1210	1448	1700	2040	2500
PA 3	600	960	1200	1500	1800	2200	2800	3600	5400
PA 4	700	1020	1350	1700	2050	2800	4000	6100	11080
PA 5	510	736	1000	1200	1500	1800	2240	2994	4440
Rio de Janeiro	660	1000	1300	1620	2022	2700	3600	5200	9000

*Ex-Rocinha

Finally, the households located inside subnormal agglomerates are known for having a larger number of dwellers than the households from the other parts of the city. Taking that into consideration, we analyze the percentiles of *per capita* household income in each PA. However, the results remain very similar. Table 5 determines that 30% of the households that rent residences in *Rocinha* could be living outside a subnormal agglomerate in any PA of the city. Moreover, more than 70% of the households in *Rocinha* could be living in a poor area near to the city center (PA 3) outside a subnormal agglomerate and more than 90% of *Rocinha*'s households could be living outside a subnormal agglomerate in a poor area far from the city center (PA 5).

Table 5: Percentiles of per capita household income in each Planning Area (only for households who rent residences)

Planning Areas	.10	.20	.30	.40	.50	.60	.70	.80	.90
PA 1	271.67	403.33	523.33	655.00	803.33	1000.00	1300.00	1780.00	2651.67
PA 2*	577.50	1000.00	1475.00	2000.00	2500.00	3250.00	4250.00	5833.33	8350.00
Rocinha	233.33	333.33	400.00	466.67	510.00	600.00	700.00	850.00	1067.50
PA 3	250.00	355.00	500.00	575.00	700.00	870.00	1100.00	1500.00	2240.00
PA 4	277.50	430.00	550.00	700.00	930.00	1182.00	1666.67	2766.67	5000.00
PA 5	200.00	287.50	386.67	500.00	585.00	706.67	900.00	1207.00	1900.00
Rio de Janeiro	256.25	400.00	510.00	677.50	875.00	1150.00	1571.00	2452.50	4300.00

*Ex-Rocinha

5 Two-Stage Hedonic Price Regressions

This section aims to evaluate how the process of enlarging the fraction of the residences in a specific neighborhood that is located inside a subnormal agglomerate affects, on average, the neighborhood rental prices. The construction of this two-stage framework allows us to incorporate the aggregate characteristics of the neighborhoods in terms of their fraction of slum dwellers, even though we cannot observe the specific residences which are inside a subnormal agglomerate. We start with a well-known hedonic price regression, described below:

$$ln(P_h) = p_h = \beta X_h + \sum_{j=1}^{J} \theta_j D_{j,h} + \xi_h \tag{1}$$

Where P_h is the rent of a residence h, X_h are the physical characteristics of h, and D_j are the neighborhood dummies for j = 1, ..., J.

For the second stage regression, we have two assumptions: (i) the effects of the physical characteristics of a residence is the same, whether or not it is located inside a subnormal agglomerate, (ii) for any neighborhood, the penalty for amplifying the fraction of residences inside the slums is the same. Hence, the second stage regression is:

$$\theta_j = \phi S_j + u_j \tag{2}$$

Where S_j is fraction of residences inside a subnormal agglomerate in neighborhood j. The coefficient ϕ indicates the extention penalty. Futhermore, in this section I am making a distinction between X_h and D_h, but in the follow sections they will be integrated in the same matrix, which I will call X_h.

6 Location Choice Model

6.1 The Environment

The model presented in this project is a static location choice model, which considers two different but interconnected and simultaneous decisions. Given that the individual has already chosen the city he will live, he will decide in which residence he is going to live. In a city, there are $j = 1, ..., J$ different neighborhoods. Also, there are two different regions: inside a subnormal agglomerate and outside a subnormal agglomerate. Each neighborhood can be composed of only one of these regions or of both of them. However, the econometrician will only observe the individual neighborhood choice and the fraction of the population in each neighborhood who lives inside a subnormal agglomerate, i.e., the econometrician cannot observe the individual decision of the region the agent is going to live. Additionally, the particular characteristics of the household's physical structure (e.g., the number of bedrooms) are also being chosen, so we are indexing each specific combination of physical characteristics and neighborhood as h.

6.2 Choice of neighborhood and location to live

The indirect utility function of the individual i, living in a household with characteristics h is as following:

$$u_{i,h} = \beta_i^X X_h + \beta_i^p p_h + \xi_h + \epsilon_{i,h} \tag{3}$$

Following the BLP (1995) setting, X_h represents the observable characteristics of the residence that the individual is living, such as the neighborhood, the region, the number of bedrooms, number of bathrooms, the type (if it is a house, an apartment or just one room) and some other characteristics that I will explain in more detail in the data section. The rent of a residence h is given by p_h. ξ_h is a term that cannot be observed by the econometrician, but is common to everyone who lives in h. This term accounts for characteristics such as access to the labor market[7]. Moreover, $\epsilon_{i,h}$ is a mean-zero idiosyncratic component that follows a type-I extreme value distribution.

Similar to Bayer, Ferreira, and McMillan (2007) - from now on I will refer to their work as BFM -, the parameters associated with housing and neighborhood observable characteristics (X) and rent (p), β_i^r, where $r \in \{X, p\}$, are:

[7] According to the 2010 *Favela Census*, 87% of Rocinha's dwellers that were employed at the time worked in the same zone where they lived, while only 66% of Manguinhos' dwellers found themselves in the same situation.

$$\beta_i^r = \overline{\beta^r} + \sum_{k=1}^{K} \beta_k^r z_{i,k} M\times,\ r \in \{X, p\} \tag{4}$$

The individual characteristics vector z_i is composed of educational attainment, income, employment status, workplace, and family composition in the household. Its length will be defined as K. Besides that, M[8] is a vector, with Lx1 dimensions, whose values can only be zero or one (note that if M is a null vector we are not using any individual characteristics, and if M is a all-ones vector we are interacting every individual characteristic with every residence characteristic). This vector allows us to choose which specific residence's characteristics we want to interact with the household's characteristics.

7 Estimation and Identification

This section will closely follow the methods used in BFM (2007) and the ones used in Calder-Wang (2020) to estimate her Long-Term Rental Demand. However, I will clarify some modifications in relation to their approaches and some aspects of the instruments used.

7.1 Estimation Procedure: First Step

Just to avoid a heavy notation, let X_h now also include the vector of rental prices. The length of this vector is L, and each entry will be referred to as $X_{h,l}$. Hence, we have:

$$\begin{aligned}
u_{i,h} &= \beta_i X_h + \xi_h + \epsilon_{i,h} \\
&= \left(\overline{\beta} + \sum_{k=1}^{K} \beta_k z_{i,k} M\times\right) X_h + \xi_h + \epsilon_{i,h} \\
&= \overline{\beta} X_h + \sum_{k=1}^{K} \beta_k z_{i,k} M \times X_h + \xi_h + \epsilon_{i,h} \\
&= \underbrace{\overline{\beta} X_h + \xi_h}_{\delta_h} + \underbrace{\sum_{k=1}^{K} \beta_k z_{i,k} M \times X_h}_{\lambda_{i,h}} + \epsilon_{i,h} \\
&= \delta_h + \lambda_{i,h} + \epsilon_{i,h}
\end{aligned} \tag{5}$$

In the equation above, δ_h accounts for the part of the utility that is a common assessment by all individuals. δ_h will vary across residences and it includes the parameters that

[8] The symbol $\times$ here indicates a vector product between M and X_h.

are common to all the residences of the city[9]. Additionally, $\lambda_{i,h}$ represents the heterogeneous part of the utility that is observable by the econometrician.

This step consists on estimating the vector of all δ_h's and all the parameters vectors β_k in $\lambda_{i,k}$. For this purpose, since $\epsilon_{i,k}$ follows a type-I extreme value distribution, we define the probability of individual i to choose the residence h:

$$P_h^i = \frac{N_h \ exp(\delta_h + \lambda_{i,h})}{\sum_{h'} N_{h'} \ exp(\delta_{h'} + \lambda_{i,h'})} \tag{6}$$

Where N_h is the weight given to residence h in the dataset (if all residences have the same weight, than $N_h = 1, \forall h$). The strategy applied by BFM (2007) is to maximize the likelihood of the accurate residential choice, and since we know - from the data - which residence each individual chose, we will follow this approach. Therefore, we also maximize the following log likelihood function:

$$\ell = \sum_i \sum_h \mathbb{1}_h^i \ \ln(P_h^i) \tag{7}$$

Where $\mathbb{1}_h^i$ is an indicator function that assumes value one if individual i has chosen residence h in the data, and zero otherwise.

The derivatives of the log likelihood function with respect to δ_h and $\beta_{k,l}$ are, respectively:

$$\frac{\partial \ell}{\partial \delta_h} = 1 - \sum_i P_h^i, \ \forall h \tag{8}$$

and

$$\frac{\partial \ell}{\partial \beta_{k,l}} = \sum_{i=h} \left(z_{i,k} M_l X_{h,l} - \frac{\sum_{h'} z_{i,k} M_l X_{h',l} N_{h'} \ exp(\delta_{h'} + \lambda_{i,h'})}{\sum_{h'} N_{h'} \ exp(\delta_{h'} + \lambda_{i,h'})} \right), \ \forall (k,l) \tag{9}$$

The optimal parameters can be found when these equations are equal to zero. In particular, using the fact that we will equal equation (8) to zero, we can construct a contraction mapping of the following form: $\delta_h^{t+1} = \delta_h^t - ln\left(\sum_i P_h^i\right)$, like in BFM (2007).

However, it is important to evaluate the hessians in these points in order to guarantee that this maximization will be treated adequately. To be more specific, the hessians are given by:

[9] In the framework used here, every residence in the database will correspond to an different h, allowing for distinct weights for each h.

$$\frac{\partial^2 \ell}{\partial \beta_{k,l} \beta_{\tilde{k},\tilde{l}}} = \sum_{i=h} \left(\frac{\left(\sum_{h'} z_{i,k} M_l X_{h',l} N_{h'} \; exp(\delta_{h'} + \lambda_{i,h'})\right) \left(\sum_{h'} z_{i,\tilde{k}} M_{\tilde{l}} X_{h',\tilde{l}} N_{h'} \; exp(\delta_{h'} + \lambda_{i,h'})\right)}{\left(\sum_{h'} N_{h'} \; exp(\delta_{h'} + \lambda_{i,h'})\right)^2} \right) - \sum_{i=h} \left(\frac{\sum_{h'} z_{i,k} M_l X_{h',l} z_{i,\tilde{k}} M_{\tilde{l}} X_{h',\tilde{l}} N_{h'} \; exp(\delta_{h'} + \lambda_{i,h'})}{\sum_{h'} N_{h'} \; exp(\delta_{h'} + \lambda_{i,h'})} \right) \tag{10}$$

7.2 Second Step and Instruments

The second step of the estimations procedure is way less computationally costly then the first one. Again, following BFM (2007), we will use the δ_h coefficients found in the first step to compute the homogeneous coefficients.

Remember that δ_h can be written as:

$$\delta_h = \overline{\beta^X} X_h + \overline{\beta^p} p_h + \xi_h \tag{11}$$

To identify the parameters related to the price and the observable characteristics that are common to all households in the city, I will need to construct an instrument to account for the simultaneity problem generated by the correlation between p_h and ξ_h. The construction of the instruments will be based on the BLP (1995) instruments, which in their case are the sum of the characteristics of the other products in other firms. In our case, the products are the residences and we can understand the interaction between the neighborhood and the adequacy level of the residence as the firms. To be more specific, the data we use in this instrument is the interaction between the *weighting areas*[10] and the level of adequacy[11].

For each price p_h, the instrument used is going to be the sum of one specific attribute (e.g. number of bedrooms) of all the residences with the same adequacy level of h in almost every *weighting areas*, excluding only the area of the *weighting areas* to which residence h belongs. Therefore, suppose that residence h belongs to *weighting area* d and have the adequacy level a. Hence, the instruments will take the following form:

$$w_h^{d,a} = \sum_{d' \neq d} \sum_{a' \neq a} x_{h'} \tag{12}$$

[10] In *Rio de Janeiro* there are 162 city's divisions made by the executive power and 200 *weighting areas* design by IBGE. These *weighting areas* are usually composed by only one city's division or by a fraction of only one city's division.

[11] The IBGE's classification of an adequate residence consists of meeting all of these requirements: "*up to two residents per dorm; water supply through a general distribution network; sanitary sewage through a general sewage or pluvial network, or a septic tank; and collected directly by a cleaning service* ", according to the Methodology of the 2010 Population Census, 2nd ed., 2016.

The instruments follow the classical assumptions:
(i) $\mathbb{E}[\xi_h|w_{h,d}] = 0$, and
(ii) $cov\,(w_{h,d}, p_h) \neq 0$.
The main idea here is also similar to BFM (2007), in the sense that by isolating the area to which the residence belongs by construction, the instrument will not affect the housing demand in this location. Furthermore, the dimensions and number of residences in each of these city's divisions are high enough, so assumption (i) holds.

8 Results

In this section, I will present the results of the two-stage hedonic price regressions and some preliminary results of the estimation of the location choice model, in particular, of the heterogeneity coefficients. To begin with, the first stage of the hedonic price model includes a wide range of the residence characteristics, such as the number of bathrooms of exclusive use (the baseline variable here is one or none exclusive bathroom), number of bedrooms, if it has an energy meter[12], which neighborhood it is located, and what type of residence it is (the baseline variable, in this case, is that the type of residence is an apartment).

As seen in Table 6 below, all the neighborhoods that accommodate a poor region have average residence rents lower than in *Centro*[13]. On the other hand, a neighborhood like *Lagoa*, that even though is part of the CBD, has almost 10% of its residences located inside subnormal agglomerates, but have rents that are on average almost 90% more expensive than in *Centro* - which is the only neighborhood in the city that does not have a subnormal agglomerate. In addition, we can observe that even though the rents in *Rocinha*[14] are on average cheaper than in other PA's, if we dismember these in PA's in their respective neighborhoods, we find that there are neighborhoods like *Santa Cruz*, with less than 10% of its residences located inside subnormal agglomerates, that have lower rents than in *Rocinha*.

[12]According to the Brazilian 2010 Census, among all the characteristics of a residence, the lack of an energy meter was the most common feature between residences that belongs to a subnormal agglomerate.
[13]The baseline variable for the neighborhood's categorical variable is the main neighborhood in the city center, known as *Centro*.
[14]Even though this is one of the biggest subnormal agglomerates in *Rio de Janeiro*, but it is closely located to the most expensive neighborhoods in the city.

Table 6: Hedonic Price Regression, First Step

VARIABLES	(1) lnrent	(2) lnrent	(3) lnrent
2 bathrooms	0.508***	0.362***	0.351***
	(0.0109)	(0.0102)	(0.00913)
3 bathrooms	0.951***	0.743***	0.641***
	(0.0202)	(0.0190)	(0.0168)
4 or more bathrooms	2.087***	1.900***	1.682***
	(0.142)	(0.129)	(0.111)
2 dormitories	0.121***	0.177***	0.202***
	(0.00707)	(0.00656)	(0.00609)
3 dormitories	0.272***	0.328***	0.348***
	(0.0171)	(0.0155)	(0.0134)
4 or more dormitories	0.590***	0.597***	0.575***
	(0.0591)	(0.0521)	(0.0452)
energy meter	0.281***	0.286***	0.257***
	(0.00909)	(0.00879)	(0.00895)
house	-0.479***	-0.301***	-0.245***
	(0.00746)	(0.00726)	(0.00738)
tenement	-0.554***	-0.462***	-0.389***
	(0.0190)	(0.0185)	(0.0192)
Constant	5.907***	5.772***	5.874***
	(0.0105)	(0.0133)	(0.0207)
PA Dummies	No	Yes	No
Neighborhood Dummies	No	No	Yes
Observations	21,177	21,177	21,177
R-squared	0.538	0.621	0.681

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The table containing the PA and neighborhood dummies is available in the appendix section.

The second stage of the hedonic model analyzes how enlarging the fraction of the residences in a specific neighborhood that are located inside a subnormal agglomerate affects, in average, the neighborhood rental prices. Our results shows that if the fraction of residences in a neighborhood which are located on subnormal agglomerates increases in ten percentage points, its neighborhood coefficient decreases its value by, on average, 0.054. For instance, it means that if *Botafogo*, which is currently 44% more expensive than the city center, experiences a ten percentage points increase in its fraction of residences located on a subnormal agglomerate, its average rent prices would be only 38,6% more expensive than in the city center.

However, is an increase of ten percentage points too much when we are dealing with an already densely populated city like *Rio de Janeiro*? According to the Brazilian 2010 and 2000 Censuses, the average growth of the population living inside subnormal agglomerates in the city was 19%, while the populational growth of dwellers who lived outside the slums was just 5%. Besides, some regions of the city experienced a slum's population growth of approximately 53% in just ten years. Therefore, ten percentage points appears to be a reasonable amount of growth for the subnormal agglomerates' population in particular neighborhood.

Also, a 5.42% decrease on the average rental prices of a neighborhood in comparison to the city center's rental prices is substantial for closely located neighborhoods. In fact, for neighborhoods that belongs to the same planning area, this decrease could make a neighborhood that was previously cheaper than another to become more expensive. This would happen, for example, with *Bangu* and *Guaratiba* and with *Meier* and *Iraja*.

Table 7: Hedonic Price Regression, Second Step (% residences)

VARIABLES	Neighborhood
pct res sub	-0.542***
	(0.129)
Observations	32
R-squared	0.362

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Alternatively, as a robustness check of the previous result, we do this same exercise but now taking into consideration the proportion of dwellers living in subnormal agglomerate. That is because the households inside subnormal agglomerates are usually bigger than the ones outside it. However, as it can be seen, the value of the fraction coefficient of dwellers

living in subnormal agglomerates is almost the same as the one of residences. Hence, the result remains valid.

Table 8: Hedonic Price Regression, Second Step (% population)

VARIABLES	Neighborhood
pct pop sub	-0.521***
	(0.126)
Observations	32
R-squared	0.354

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Tables 9 describes the heterogeneity coefficients. They account for the interaction between a particular housing characteristic and a particular household's dwellers characteristics. For instance, the housing characteristics are basically the planning area to which the residence belongs. Moreover, the demographic characteristics are composed of household income, its number of residents, the educational level of the householder, and the workplace of the householder. Since the willingness-to-pay (WTP) for a housing attribute l, while having a demographic characteristic k is given by $-\frac{\beta_{l,k}}{\overline{\beta^p}}$, and the coefficient $\overline{\beta^p}$ of the rent is estimated in the second step of the location choice model (described in table 10), we can calculate the willingness-to-pay of household with some particular characteristics for a residence in a specific planning area of the city.

In the city of *Rio de Janeiro*, wealthier households have higher willingness-to-pay to live in any of the PA's than in *Rocinha*. Moreover, larger households are willingness-to-pay more to live in a slum and in low-income planning areas than a smaller families. Therefore, how these coefficients can help in the understanding of the data described in section 6? Why households with the same size and same income choose to live in different locations?

One possible answer relies on the educational attainment and the workplace of the householder. The most common explanation for someone to live in a subnormal agglomerate find in the literature is the proximity with higher-paid jobs. Even though it can be true, the achievement of higher educational attainment - which is frequently related to higher paid jobs - interacted with the experience of living in a residence inside a slum (i.e., in *Rocinha*), leads to a reduction of the willingness-to-pay for these type of residence in comparison to having lower educational attainment, as can be seen in Table 9. The opposite happens with the willingness-to-pay for living in one of the low-income planning

areas of the city, in which the increase in the educational attainment generates an increase in the willingness-to-pay for living in these locations.

In fact, the higher willingness-to-pay to live in *Rocinha* comes from the household whose family chief does not have high school and works outside his home. On the other hand, the higher willingness-to-pay to live in poor area (PA 3 and PA 5) comes from the household whose family chief has completed the high school and works outside his home. Additionally, the willingness-to-pay to live in PA 2 ex-Rocinha (the wealthier PA in the city) is approximately 60% lower for the householders who do not have high school and work outside home in comparison to those who have high school and also work outside home.

Furthermore, the heterogeneity coefficients estimated in the location choice model reveal that the dwellers who work outside their homes and do not have completed high school are willing-to-pay 47.48% more of their monthly income to live in *Rocinha* instead of living in the city center. However, the ones who also work outside their home but have completed high school have are willing-to-pay 13.32% less of their monthly income to live in *Rocinha* instead of the city center. Hence, dwellers who work outside their houses have a higher willingness-to-pay for living in poor neighborhoods (that are not entirely composed of subnormal agglomerates) if they have completed their high school.

Moreover, we found that bigger and poorer households are more willing-to-pay to live in *Rocinha* than the other ones. Besides, the only *planning area* in the city with a lower willingness-to-pay from larger families than the city center is PA 2 ex-*Rocinha*, which is also the PA with the higher willingness-to-pay from the dwellers who have completed high school.

Table 9: Heterogeneity Coefficients - WTP (% Montlhy income)

Demographic Characteristics	Rocinha	PA 2 ex-Rocinha	PA 3	PA 4	PA 5
ln income	-0.1224	0.3253	-0.0118	0.1182	-0.0836
	(0.0468)	(0.0509)	(0.0508)	(0.0503)	(0.0497)
Number of dwellers	0.1236	-0.1517	0.0990	0.0197	0.1063
	(0.0410)	(0.0208)	(0.0078)	(0.0321)	(0.0367)
High School					
Does not work	-0.8119	0.1988	0.1395	0.0102	0.0985
	(0.0327)	(0.0594)	(0.0628)	(0.0488)	(0.0458)
Works outside home	-0.1332	0.0209	0.1511	0.0777	0.2423
	(0.0254)	(0.0426)	(0.0479)	(0.0376)	(0.0350)
No High School					
Does not work	-0.0256	-0.4289	0.1302	-0.1855	0.2201
	(0.0322)	(0.0463)	(0.0642)	(0.0444)	(0.0465)
Works outside home	0.4748	-0.5964	0.0138	0.1267	0.1655
	(0.0293)	(0.0417)	(0.0532)	(0.0417)	(0.0396)

Standard errors in parentheses

Table 10 illustrates the results of the second step of the location choice model, which provides the homogeneous coefficients of the residences' characteristics and the WTP for them. The dependent variable in equations (1) and (2) is δ_h and the standard errors were calculated via Delta Method, as suggested by Hole (2007). The naive is simply a non-instrumented OLS and the instrumented regression uses the instrument described in section 7. The $\overline{\beta^p}$ used to calculate the WTP coefficients from table 9 is from the instrumented version ($\overline{\beta^p} = -1.82$). Moreover, there is evidence that the use of the instrument helped to correct the bias in regression (1).

For instance, in regression (1) the coefficients for both 2 and 3 bathrooms are significant and negative, which is not in line with what we find in the data. In addition, the value of the tenement coefficient in this regression is not significant, even though we can find a considerable difference between the rent prices of regular apartments and tenements in the 2010 Brazilian Census. All this evidence suggests that the component of non-observable characteristics is not an idiosyncratic component and the use of an instrument

is indispensable in this case.

Consequently, the WTP coefficients in column (3) of Table 10 represent an adjustment of the results founded in column (1) of Table 6. A particularly interesting finding is that in the Location Choice Model, the house coefficient is positive, while all 3 Hedonic Price Regressions its value was negative. All the other coefficients just changed in proportion, not in direction. Also, the coefficient with the smaller variation between the two models is the one of the energy meter. We found that the presence of an energy meter in a residence enhance the rent price of a residence by 29.3%

Table 10: Homogeneous coefficients and WTP

	OLS	Instrumented	WTP (% of monthly income)
ln(rent)	-1.137***	-1.820***	
	(0.0201)	(0.145)	
2 bathrooms	-0.360***	-0.0255	-0.0140
	(0.0292)	(0.0769)	(0.0433)
3 bathrooms	-0.300***	0.317**	0.174**
	(0.0467)	(0.144)	(0.0658)
4 or more bathrooms	1.053***	2.359***	1.296***
	(0.172)	(0.362)	(0.1297)
2 dormitories	0.476***	0.590***	0.324***
	(0.0184)	(0.0253)	(0.0196)
3 dormitories	0.650***	0.936***	0.514***
	(0.0372)	(0.0518)	(0.0278)
4 or more dormitories	0.744***	1.270***	0.698***
	(0.0872)	(0.127)	(0.0514)
energy meter	0.330***	0.533***	0.293***
	(0.0263)	(0.0488)	(0.0150)
house	0.545***	0.244***	0.134***
	(0.0204)	(0.0719)	(0.0499)
tenement	0.0339	-0.315***	-0.173***
	(0.0557)	(0.0996)	(0.0445)
Constant	9.367***	13.44***	
	(0.122)	(0.856)	
Observations	21,177	21,177	21,177

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

9 Conclusion

The urban fabric in developing countries is marked by the presence of slums. When these slums are located in many different neighborhoods across the same city, the housing location decision faced by its dwellers becomes less straightforward than just facing an income constraint. The focus of this dissertation is, therefore, to untangle the other motivations behind this decision.

Our results show that dwellers without high school and who work outside their residences prefer to live in a particular subnormal agglomerate closer to the wealthiest neighborhoods in the city of *Rio de Janeiro* than to live in a poor distant planning area. On the other hand, the dwellers that do have completed high school and work outside their residences prefer to live in these poor distant planning areas instead of living in *Rocinha*. This evidence suggests that perhaps a government intervention aimed at reducing the share of the population living in slums must take into account the educational level of the residents of each slum and how to enhance it. By doing that, the public policies focused on this population tend to be more effective.

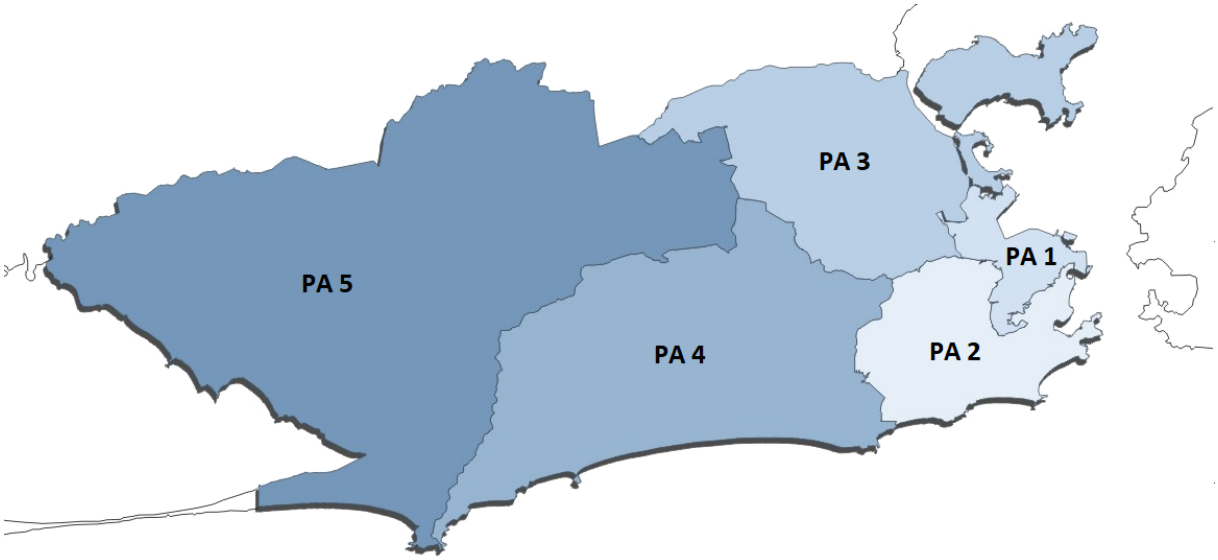
Besides, there is some limitation in our approach. The main one is our lack of data at a census tract level. Without this fine geographical characteristic, we are unable to affirm, in most cases, whether a specific residence belongs to a subnormal agglomerate or not. As a consequence, for any neighborhood, the penalty for amplifying the fraction of residences inside the slums is the same in our model.

However, the restricted-access sample of the 2010 Brazilian Census contains information about the individuals and their households at the census tract level. More specifically, it consists of all the information available in the unrestricted access version of the Census and whether or not each residence in the sample is located inside a subnormal agglomerate. This distinction in the restricted-access data set would allow us to individually estimate the different willingness-to-pay to live in a subnormal agglomerate in each neighborhood or subdistrict of *Rio de Janeiro*, which is our primary purpose for future research.

Appendix

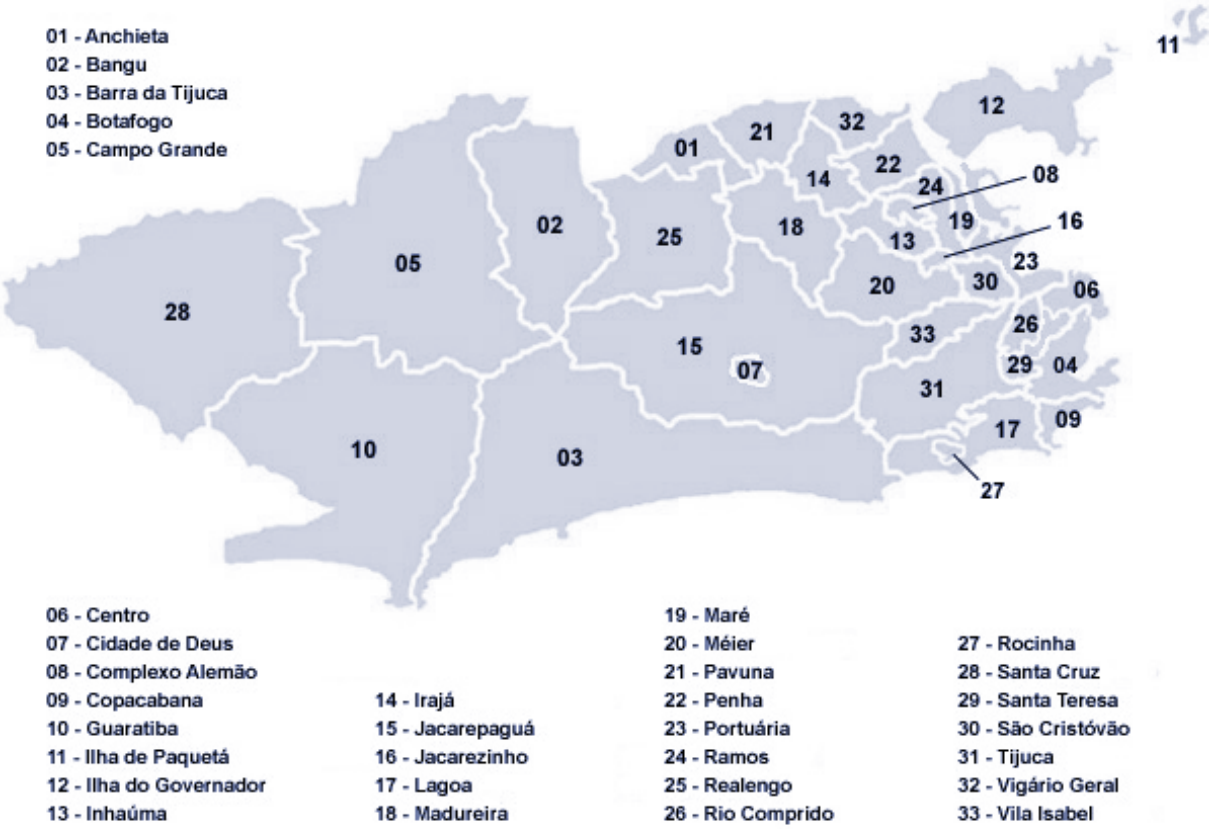
A. Figures

Figure 1: Rio de Janeiro's Planning Areas



Source: *Secretaria Municipal de Planejamento Urbano - SMPU*

Figure 2: Rio de Janeiro's Subdistricts



Source: *Instituto Brasileiro de Geografia e Estatística - IBGE*

B. Tables

Table 11: Hedonic Price Regression, First Step - PA and Neighborhood Dummies

VARIABLES	(1) lnrent	(2) lnrent	(3) lnrent
Portuaria			-0.270***
			(0.0299)
Rio Comprido			-0.183***
			(0.0276)
Sao Cristovao			-0.182***
			(0.0282)
Ilha de Paqueta			0.0321
			(0.0374)
Santa Teresa			-0.00275
			(0.0308)
Botafogo			0.440***
			(0.0227)
Copacabana			0.547***
			(0.0244)
Lagoa			0.872***
			(0.0333)
Tijuca			0.00503
			(0.0249)
Vila Isabel			-0.0333
			(0.0243)
Ramos			-0.211***
			(0.0257)
Penha			-0.277***
			(0.0258)
Vigario Geral			-0.321***
			(0.0271)
Inhauma			-0.258***
			(0.0257)
Meier			-0.148***
			(0.0216)
Iraja			-0.139***

		(0.0237)
Madureira		-0.187***
		(0.0216)
Ilha do Governador		-0.128***
		(0.0244)
Anchieta		-0.273***
		(0.0256)
Pavuna		-0.382***
		(0.0258)
Cidade de Deus		-0.321***
		(0.0449)
Jacarepagua		-0.108***
		(0.0205)
Barra da Tijuca		0.284***
		(0.0265)
Realengo		-0.258***
		(0.0248)
Bangu		-0.324***
		(0.0229)
Campo Grande		-0.260***
		(0.0222)
Santa Cruz		-0.418***
		(0.0259)
Guaratiba		-0.316***
		(0.0300)
Rocinha		-0.378***
		(0.0249)
Jacarezinho		-0.467***
		(0.0369)
Complexo do Alemao		-0.459***
		(0.0307)
Mare		-0.354***
		(0.0234)
PA 2 ex Rocinha	0.460***	
	(0.0136)	
PA 3	-0.102***	
	(0.0107)	

PA 4		0.126***	
		(0.0128)	
PA 5		-0.170***	
		(0.0120)	
Rocinha		-0.267***	
		(0.0192)	
Observations	21,177	21,177	21,177

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

References

Altonji, Joseph G., and Richard K. Mansfield. "Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects." *American Economic Review* 108.10 (2018): 2902-46.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. "A unified framework for measuring preferences for schools and neighborhoods." *Journal of political economy* 115.4 (2007): 588-638.

Bergman, Peter, et al. Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. No. w26164. *National Bureau of Economic Research*, 2019.

Berry, Steven, James Levinsohn, and Ariel Pakes. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society* (1995): 841-890.

Berry, Steven, James Levinsohn, and Ariel Pakes. "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* 112.1 (2004): 68-105.

Brueckner, Jan K. "Urban squatting with rent-seeking organizers." *Regional Science and Urban Economics* 43.4 (2013): 561-569.

Calder-Wang, Sophie. "The Distributional Impact of the Sharing Economy on the Housing Market." *Job Market Paper* (2020).

Cavalcanti, Tiago, Daniel Da Mata, and Marcelo Santos. "On the determinants of slum formation." *The Economic Journal* 129.621 (2019): 1971-1991.

Chyn, Eric. "Moved to opportunity: The long-run effects of public housing demolition on children." *American Economic Review* 108.10 (2018): 3028-56.

Davis, Donald R., et al. "How segregated is urban consumption?." *Journal of Political Economy* 127.4 (2019): 1684-1738.

Davis, Peter. "Spatial competition in retail markets: movie theaters." *The RAND Journal of Economics* 37.4 (2006): 964-982.

Diamond, Rebecca. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review* 106.3 (2016): 479-524.

Dingel, Jonathan I., Antonio Miscio, and Donald R. Davis. "Cities, lights, and skills in developing economies." *Journal of Urban Economics* (2019): 103174.

Ferreira, Fernando V., and Maisy Wong. Estimating Preferences for Neighborhood Amenities Under Imperfect Information. No. w28165. *National Bureau of Economic Research*, 2020.

Ferreira, Pedro C., Alexander Monge-Naranjo, and Luciene Torres de Mello Pereira. "Of cities and slums." *FRB St. Louis Working Paper* 2016-22 (2016).

Fu, Chao, and Jesse Gregory. "Estimation of an Equilibrium Model With Externalities: Post-Disaster Neighborhood Rebuilding." *Econometrica* 87.2 (2019): 387-421.

Hole, Arne Risa. "A comparison of approaches to estimating confidence intervals for willingness to pay measures." *Health economics* 16.8 (2007): 827-840.

Neri, Marcelo Côrtes. "The New Middle Class in Brazil: The Bright Side of the Poor." (2010).

Redding, Stephen J., and Esteban Rossi-Hansberg. "Quantitative spatial economics." *Annual Review of Economics* 9 (2017): 21-58.

Tucker, Catherine, Juanjuan Zhang, and Ting Zhu. "Days on market and home sales." *The RAND Journal of Economics* 44.2 (2013): 337-360.