Income segregation in monocentric and polycentric cities: Does urban form really matter?

Miquel-Àngel Garcia-López[*][†]
Universitat Autònoma de Barcelona
Institut d'Economia de Barcelona

Ana I. Moreno-Monroy
OECD

May 2018

ABSTRACT: We estimate the effect of urban spatial structure on income segregation in Brazilian cities between 2000 and 2010. Our results show that, first, local density conditions increase income segregation: the effect is higher in monocentric cities and smaller in polycentric ones. Second, the degree of monocentricity-polycentricity also affects segregation: while a higher concentration of jobs in and around the CBD decreases segregation in monocentric cities, a higher employment concentration in and around subcenters located far from the CBD decreases segregation in polycentric cities. Third, results are heterogeneous according to city size: local density does not increase segregation in small (monocentric) cities, it increases segregation in medium size cities, and it decreases segregation in large (polycentric) cities. Finally, results also differ between income groups: while local density conditions increase the segregation of the poor, a more polycentric configuration reduces the segregation of the rich.

1. Introduction

Despite decreasing income inequality levels in the past decade, Brazil still ranks as one of the most unequal countries in the world. According to UN data, its ratio of the average income of the richest 10 percent to the poorest 10 percent was 40.6 in 2012, the 10th highest level in the world by this measure. At the same time, the country has gone through a profound urbanization process. Starting in the 1950s and for three decades, millions of lower-income people migrated from rural to urban areas. At that time, migration was directed disproportionally to already large metropolises, where location happened -mostly informally- in peripheral urban areas (Villaça, 1998, Telles, 1995). The following decades have been characterized by a shift towards high population growth in small and medium size cities, and stagnant population growth rates in large metropolises.

As a result, Brazil today has a highly complex urban system, with monocentric and polycentric metropolitan areas ranging from global megacities to upcoming regional urban centers (IPEA, IBGE, and UNICAMP, 2002). A fast-paced urbanization process under high income inequality has led to high asymmetries in the access to transportation, basic public services and urban amenities across income groups, related to different urban shapes. Given the diverse urban footprints that have resulted from an accelerated process of urbanization in Brazil, it may be worth going beyond considering an aggregate measure that carries no information about the actual use of urban space such as average population density to more informative measures of the disposition of population of different income levels and employment in space.

In this paper, we investigate the relationship between the urban spatial structure of Brazilian cities and their income segregation. Income segregation can be defined as the uneven sorting of households according to their income level within the urban space (Reardon and Bischoff, 2011). Urban spatial structure is the degree of spatial concentration of jobs and their intrametropolitan spatial distribution (Horton and Reynolds, 1971, Anas, Arnott, and Small, 1998). Specifically, we address two questions: Do local job density conditions affect income segregation? Does monocentricity-polycentricity foster income segregation? For these two questions, we also analyze how these relationships change with city size and income level.

Regarding our methodological approach, we first construct measures of income segregation for the years 2000 and 2010 for 121 urban agglomerations (UAs) in Brazil. Specifically, we use enumeration area level information to calculate rank-order measures of income segregation (Reardon and Bischoff, 2011, Monkkonen and Zhang, 2014). These measures are independent of shifts in the income distribution levels, and can be used to obtain consistent and comparable measures at different points of the income distribution. In this way, we also obtain measures of the segregation of the rich (90th percentile of the income distribution) and the segregation of the poor (10th percentile) (Reardon and Bischoff, 2011, Chetty, Hendren, Kline, and Saez, 2014).

On the other hand, we use detailed information on the intrametropolitan distribution of employment to identify and characterize the urban spatial structure of the urban agglomerations. First, our main explanatory variable is the city average job density in a 1-km radius surrounding each enumeration area. This 1-km density summarizes both the spatial concentration and the in-

trametropolitan distribution of jobs (Duranton and Turner, 2016). Second, we rely on the method developed by McMillen (2001) to identify the number of employment subcenters. According to this number, urban agglomerations are classified as monocentric (one center, the Central Business District or CBD) or polycentric (two or more centers, the CBD and subcenter/s). Third, we compute the degree of monocentricity-polycentricity with the urban centrality index proposed by Pereira, Nadalin, Monasterio, and Albuquerque (2013). Finally, we also characterize urban agglomerations computing the average distance from the CBD and the job population ratio.

Using data for 2000 and 2010 and Instrumental Variables (IV) techniques, we regress the log of our rank-order measure of income segregation on the log of urban spatial structure variables, while controlling for informality, inequality, geography, demography and industrial composition. Our results show that, first, local density conditions increase income segregation: the effect is higher in monocentric cities and smaller in polycentric ones. Second, the degree of monocentricity-polycentricity also affects segregation: while a higher concentration of jobs in and around the CBD decreases segregation in monocentric cities, a higher employment concentration in and around subcenters located far from the CBD decreases segregation in polycentric cities. Third, results are heterogeneous according to city size: local density does not increase segregation in small (monocentric) cities, it increases segregation in medium size cities, and it decreases segregation in large (polycentric) cities. Finally, results also differ between income groups: while local density conditions increase the segregation of the poor, a more polycentric configuration reduces the segregation of the rich.

To the best of our knowledge, there are no works analyzing the relationship between urban spatial structure and income segregation. The questions of whether and how urban spatial structure affects income segregation contribute to debates weighting the equity effects of urban policies such as zoning laws and urban plans affecting the distribution of employment within cities (OECD, 2018). Our results could shed light, for instance, on the expected effect of compact city interventions on income segregation in cities of different sizes and shapes. Previous related research has studied the relationship between income segregation and population density (Pendall and Carruthers, 2003, Wheeler, 2008). Population density is a city-level measure that does not take into account the intrametropolitan distribution of households of different income levels and firms. Furthermore, income segregation is measured with a dissimilarity index, which suffers from a number of limitations including scale sensitivity issues (Reardon and O'Sullivan, 2004). On the other hand, a number of works have analyzed the consequences of segregation on outcomes such as poverty and upward mobility (Ananat, 2011, Chetty *et al.*, 2014), but have not considered the spatial structure of cities. Moreover, much of the focus of the income segregation literature has been on metropolises (Villaça, 1998, Monkkonen and Zhang, 2014, Feitosa, Le, Vlek, Monteiro, and Rosemback, 2012), but not much is understood about the determinants of segregation across entire urban systems. We aim to provide a complete account of income segregation levels and changes (also by income groups), an account of the different urban spatial structures present in the Brazilian urban system, and econometric evidence on the relationship between the two.

After this introduction, Section 2 summarizes some predictions regarding the relationship between income segregation and urban spatial structure. Section 3 details data sources and the

measurement of income segregation. Section 4 presents the data and the different measures of urban spatial structure. Section 5 develops our econometric approach. Section 6 discusses the results, and Section 7 concludes.

2. Theoretical predictions

Models with residential mobility and firm immobility

In the the classical monocentric land use model by Alonso (1964), Mills (1967) and Muth (1969) (hereafter AMM), all firms have a fixed location and concentrate at a unique CBD where workers commute to work. As there are commuting costs per unit of distance, there is high demand for land near the CBD, making living space at those locations smaller and more expensive. In the spatial equilibrium, longer commutes are perfectly compensated by a lower price of living space. In a setting with two income groups, the rich and the poor, the rich locate far away from the CBD if their willingness to pay for housing when moving away from the CBD decreases more slowly than that of the poor. In models introducing two commuting modes with different speeds (e.g. the car and public transport), the rich commute longer distances faster by car, while the poor remain in central areas and commute using public transport, as long as the elasticity of housing demand with respect to income is greater than the elasticity of commuting costs per unit of distance with respect to income (LeRoy and Sonstelie, 1983, Glaeser, Kahn, and Rappaport, 2008).[1] This condition can be overturned by high congestion and its consequent high time costs, which can lead to a 'return to the city' process, where the rich move from the suburbs back to central areas, trading off living space for proximity to jobs.

The basic AMM model rests on the assumption that all intra-urban locations are homogeneous in terms of amenities, including the provision of public services. Including a spatial restriction on the provision of public services can lead to the prediction that the rich locate in well-provided areas, whereas the poor locate in under-provided areas (Griffin and Ford, 1980, Cheshire, Monastiriotis, and Sheppard, 2003). This pattern is reinforced by any other additional attribute of well-provided places, such as cultural amenities (Brueckner, Thisse, and Zenou, 1999). In cities with a high share of young people, preferences also play a role, for instance if young, rich small households trade off housing space for local amenities (Glaeser, 2008).

Lack of infrastructure provision across the whole city can also limit the supply of formal housing to a few specific areas in the city (Posada-Duque, 2015). Limited provision to central areas can lead the rich to segregate in more dense, vertical neighborhoods located near the CBD, while the poor would scatter in low-rise informal constructions around the periphery (Henderson, Regan, and Venables, 2016, Feler and Henderson, 2011). The residential mobility of the rich is spatially constrained to within the best served area in the city, where rental prices would be considerable higher because they integrate the valuation for proximity to public services and local amenities (Cheshire *et al.*, 2003). The rich can also influence zoning laws and the concentration

---

[1] Alternatively, the model by Brueckner, Thisse, and Zenou (2002), with worker and income heterogeneity but no agglomeration of firms, predicts that low-skill, low-income workers bear the longest commutes and reside in the periphery. This is because the low net wage earned by low-skilled workers (product of the training costs that their skill mismatch implies) imply a low valuation of time and a consequent higher tolerance for longer commutes.

of provision of public infrastructure and services, especially in contexts of high segregation (Cutler, Glaeser, and Vigdor, 1999, Tiebout, 1956), further reinforcing the segregation of those at the top of the income distribution. The eventual extension of the public infrastructure network to non-central areas can also be accompanied by the emergence of rich neighborhoods in the suburbs where the demand for single-family larger homes is met. Note that as long as these new neighborhoods remain homogeneous in terms of their income composition - that is, if there an exclusionary mechanism in place, such a substantial premium for exclusive local amenities -, the level of segregation of the rich would remain high.

Furthermore, in models that consider the durability of housing stock (Brueckner and Rosenthal, 2009), the prediction that the older hosing stock located near the CBD previously used by the rich filters down to the poor may have restricted validity for the bottom income groups who have tight residential mobility restrictions related to the difficulty of accessing credit for informal workers, the insecurity of tenure and poor definition of property rights that increase permanence in informal plots. More generally, housing demand for those at the bottom of the income distribution is also constrained by increasing housing demand, economic stagnation, deterioration of real income and persistent income inequality, and stricter regulations to build (Cavalcanti and Da Mata, 2013).

Models with residential mobility and firm mobility

So far we have not taken into account the location decisions of firms. The agglomeration of (formal) employment may be yet another element contributing to the spatial concentration of the rich in central areas. Without changing the assumption of firm immobility, Muth (1969) shows that allowing for multiple employment centers does not alter the main predictions of the monocentric model: the behaviour around each subcenter is the same as around the CBD. However, the number of subcenters and their location is exogenously given.

The seminal model of Fujita and Ogawa (1982) considers both residential and job location decisions of households, and the location decisions of firms. Firms agglomerate to take advantage of a locational potential (associated with positive agglomeration externalities that occur at short distances, such as knowledge spillovers). Under certain parameter values (including low commuting cost), a similar structure as in the AMM model arises, with all firms agglomerating in one location. However, other parameter specifications such as higher commuting costs lead to the emergence of employment subcenters and the consequent mixed used of land.

This model, and more sophisticated versions of it (Lucas and Rossi-Hansberg, 2002), although highly complex, do not explicitly consider the location decisions of households of different income levels. They also rest on the assumption of homogeneous provision across the city. Under asymmetric provision, besides agglomeration economies, the same factors pulling together rich households, such as the higher availability of amenities and better transport connectivity, could be pulling together (formal) firms and slowing down a process of employment decentralization.

Once it occurs, the decentralization of employment could have an impact on residential segregation by levels of income depending on how distant the new employment subcenters are from the CBD and the area where the rich reside. For instance, if the provision of public infrastructure

and amenities is incremental from the historical CBD (towards an expanded CBD), and both rich households and firms de-concentrate in the proximities of the historical CBD as a result, the decentralization of employment would not be necessarily associated with lower residential segregation. However, new subcenters could also arise in peripheral locations from the need to benefit from lower rents and inter-city transport connections. These centers could release some of the pressure to locate in central areas for higher-income households, and favor a more even dispersion of households of different income levels across the city. Still, the segregation of the poor may not decrease if there is a large percentage of informal workers and informal employment centers are located far away from formal employment centers, for instance in lower income neighborhoods. In this case, informal workers may trade-off higher incomes in the formal sector for lower rents and lower commuting costs as long as they can find informal jobs in their local area (Moreno-Monroy and Posada, 2014).

3. Income segregation in Brazilian cities

3.1 Data

To calculate urban agglomeration-level measures of segregation[2], we use income distribution information at the *setor censitario* level from the 2000 and 2010 Population Census micro-data freely distributed by IBGE. The *setor censitario* level is equivalent to enumeration areas defined for surveying purposes. Each unit contains on average 400 households.

The individual unit of analysis is the head of household (i.e., a person responsible for the household who is older than 10 years-old) categorized as belonging to one of nine ordered income categories. This unit is chosen over households because, unlike the 2010 census, the 2000 census does not include income distribution information for households. Income is defined as the level of nominal monthly income from work or other sources measured in income bins relative to the number of minimum wages.[3] Table D.1 shows summary statistics of the frequency of income bins across UAs.

3.2 Dimensions of segregation

As explained by Reardon and O'Sullivan (2004), there are two conceptually different dimensions to spatial segregation: 1) spatial exposure (or spatial isolation), and 2) spatial evenness (or spatial clustering). As suggested by its name, spatial exposure (spatial isolation) captures how likely it is for a member of one group to encounter a member of another group (the same group) in their local environments. On the other hand, spatial evenness (clustering) captures how evenly (unevenly) distributed are groups in the urban space.

Figure 1, taken from Reardon and O'Sullivan (2004), helps illustrating the conceptual difference between exposure (isolation) and evenness (clustering). Consider families in two income groups, the rich (black dots) and the poor (white dots) and four different spatial arrangements of families

---

[2]Appendix A presents our definition of urban agglomerations, that is, our sample of Brazilian cities.

[3]The legal minimum wage was BRL 151,00 in 2000 and BRL 510,00 in 2010, approximately USD 206 and USD 287 respectively.

in four different cities, each represented in a quadrant of Figure 1. In which city are rich families more segregated? The answer will depend on the dimension of segregation we are considering. The rich in the two cities in the upper two quadrants are equally not segregated in terms of evenness/clustering, because they do not concentrate in any particular area within the city. However, they are more isolated in the city on the upper left quadrant, simply because there are less rich families in this city, making it less likely for a member of a rich family to encounter a member of a poor family in her neighborhood. If we are considering the evenness/clustering dimension, the answer will be that the rich are more segregated in the two cities at the bottom quadrants.

Figure 1: Dimensions of income segregation
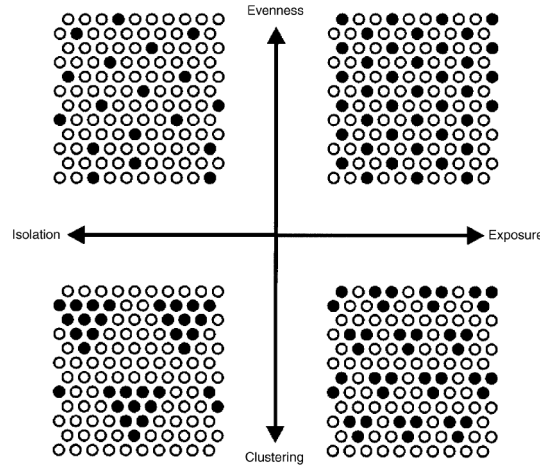
Note that while moving from a more dissimilar to a more similar percentage of people in different groups should result in higher spatial exposure, the level of spatial evenness can change in either direction, because it depends on the spatial arrangements of the different groups in space. Thus, unlike the exposure dimension, the evenness dimension of segregation is independent of the initial population composition, a desirable property for inter-city comparisons. We focus on the measurement of segregation measures that capture spatial evenness/clustering.

Lastly, we note that we are focusing in only one of many dimensions of segregation (Massey and Denton, 1988). In this respect, our measure does not capture the degree of *centralization*, or the spatial concentration of a certain income group in the central area of the city. Common interpretations related to "center-periphery" city structures (Ruiz-Rivera and van Lindert, 2016) are based in these type of indicators, which require the definition of a unique CBD as point of reference, and the definition of thresholds that divide income classes, which complicates comparisons over time and across cities. Moreover, in our approach we acknowledge the existence of polycentric city structures, which make it hard to interpret centralization measures based on a unique CBD. Our measure is also not meant to capture the concentration of income groups, that is, whether the proportion of land used by a certain income groups is disproportionally low or high compared to its relative share in the population.

3.3 Rank-order segregation index

Our choice of measure for income segregation, the rank-order information theory index, is based on income percentile ranks. The measure uses the full distribution of income, is independent of threshold choices and is insensitive to shape-preserving changes in the income distribution (Reardon and Bischoff, 2011). The measure then captures the extent of residential segregation by income levels, as opposed to capturing changes in income levels (resulting from changes in income inequality) even when no residential sorting takes place as a consequence in the change in income levels (Watson, 2009). As detailed in Reardon (2011), it satisfies a wide range of desirable properties for segregation indexes.

We dismiss the use of a "categorical approach", under which income information is split between two different income categories (Pendall and Carruthers, 2003), because it throws away valuable information on the distribution of income, and because it is sensitive to the choice of income threshold (Reardon and Bischoff, 2011, Watson, 2009).

The rank-order information theory index captures the ratio of within-unit income rank variation to overall income rank variation (Reardon and Bischoff, 2011). It is basically a weighted sum of all possible pair-wise income segregation indices. More specifically, let p denote percentile ranks in a given income distribution. The pair-wise information theory index H is defined as:

$$H(p) = 1 - \sum_j \frac{t_j E_j(p)}{TE(p)} \tag{1}$$

where t_j is the population in the local environment j, T is the total population in the urban area, and E is the entropy of the total population given by $E(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{(1-p)}$. The rank-order information theory index is defined as:

$$H = 2 \ln 2 \int_0^1 E(p) H(p) dp \tag{2}$$

The H index varies between 0 and 1, where 0 means there is no segregation (i.e., the income distribution of each local environment is exactly equal to that of the city), and 1 means there is maximum segregation (i.e., each local environment is composed of individuals of the same income category). An alternative interpretation of H is how much less income rank variation there is within neighborhoods than in the overall population. In the a-spatial version of the index (denoted hereafter as HA), which we use here, the local environment is defined by the area boundaries. The interpretation of HA is the same as that of H.[4]

We then construct an income profile for each urban agglomeration, that is, a curve describing the relationship between HA and the percentage of individuals in each income category. See Appendix C for methodological details. We use these profiles to define a measure of segregation as experienced by two given income percentiles (Reardon and Bischoff, 2011): the 10th percentile, representing the *segregation of the poor*, and the 90th percentile, representing the *segregation of the rich*. These cut-off points have been used in previous studies analyzing the segregation of the

---

[4]It is possible to construct a spatial version of the H index that allows for more flexible definitions of neighbourhoods based on different bandwidths. Here we prefer the a-spatial version, noting that it provides upper-boundary estimates for fine neighborhood definitions. See Appendix B for a detailed discussion.

poor and the rich using a similar methodology ([Bischoff and Reardon](), 2014, [Chetty *et al.*](), 2014). We also show results for a different cut-off (20 percent and 80 percent) that may better represent local poverty measures (i.e., those based on poverty lines).

Table 1 shows summary statistics for the rank-order HA index, as well as for our two different proxies of segregation of the poor (10th and 20th percentiles) and the rich (80th and 90th percentiles) for 2000 and 2010. On average, segregation of all income groups decreased from 0.139 to 0.133 between 2000 and 2010. In particular, average segregation increased in 38 cities and decreased in 83, with Cachoeiro de Itapemirim, Atibaia, Quadra, Teresópolis, João Pessoa and São Paulo showing the largest increases, and Cuiabá, Itaperuna, Umuarama, Vale do Aço, Jequié and Juazeiro do Norte showing the largest reductions. Those at the top of the income distribution (80th and 90th percentiles) are more segregated but, at the same time, experienced a reduction in their segregation index from 0.200-0.240 to 0.198-0.236. On the other hand, despite being less segregated, those at the bottom of the income distribution (10th and 20th percentiles) increased their segregation from 0.079-0.080 to 0.093-0.081.

Table 1: Summary statistics for rank-order segregation HA index

	All cities			
	Mean	S.D.	Min	Max
2000 HA	0.139	0.039	0.073	0.275
2010 HA	0.133	0.041	0.052	0.274
2000 HA 10th percentile	0.079	0.028	0.039	0.224
2010 HA 10th percentile	0.093	0.028	0.041	0.227
2000 HA 20th percentile	0.080	0.029	0.036	0.212
2010 HA 20th percentile	0.081	0.031	0.016	0.175
2000 HA 80th percentile	0.209	0.060	0.112	0.410
2010 HA 80th percentile	0.198	0.060	0.095	0.408
2000 HA 90th percentile	0.240	0.064	0.124	0.430
2010 HA 90th percentile	0.236	0.064	0.110	0.417

Notes: 121 observations (cities) in 'All cities' sample.

Compared to the US, the average value of HA for 2010 is smaller than the value of 0.148 reported in [Bischoff and Reardon]() (2014) for 117 metropolitan areas using the same methodology, even though the inequality level in Brazil is much higher. This can be partly explained by the fact that the poorer (10th percentile) experience higher levels of segregation in U.S. metropolitan areas with values of 0.146 and 0.163 in 2000 and 2010, respectively. On the other hand, segregation of the rich (percentile 90th) is lower in US cities with values of 0.185 and 0.200 in 2000 and 2010.[5]

---

[5]It is important to note that our estimates of the segregation of the poor possibly underestimate the true level of segregation at the bottom of the income distribution because we do not take into account heads of household with zero reported income. The reason to exclude them is that the number of such individuals is much larger in 2010 than in 2000, most likely due to a coding error in the original 2010 census.

Table 2: Income segregation (HA) in 35 UAs with population over half million, 2000–2010

UA	HA		HA (10th percentile)		HA (20th percentile)		HA (80th percentile)		HA (90th percentile)	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Aracajú	0.228	0.228	0.138	0.148	0.132	0.130	0.351	0.324	0.373	0.350
Baixada Santista	0.176	0.169	0.095	0.123	0.114	0.116	0.237	0.220	0.260	0.249
Barra Mansa/Volta Redonda	0.162	0.135	0.103	0.099	0.099	0.091	0.229	0.193	0.271	0.235
Belo Horizonte	0.214	0.202	0.094	0.102	0.105	0.109	0.329	0.312	0.371	0.361
Blumenau	0.089	0.079	0.039	0.041	0.045	0.029	0.139	0.122	0.169	0.169
Brasília	0.275	0.274	0.128	0.154	0.145	0.158	0.410	0.408	0.430	0.417
Campinas	0.162	0.160	0.067	0.084	0.083	0.081	0.254	0.250	0.307	0.308
Campo Grande	0.178	0.156	0.079	0.089	0.081	0.083	0.276	0.245	0.307	0.290
Cuiabá	0.175	0.145	0.097	0.113	0.092	0.101	0.264	0.216	0.299	0.260
Curitiba	0.188	0.166	0.074	0.085	0.097	0.084	0.286	0.256	0.309	0.291
Feira de Santana	0.124	0.120	0.090	0.097	0.083	0.057	0.173	0.177	0.197	0.205
Florianópolis	0.168	0.154	0.085	0.088	0.106	0.087	0.241	0.229	0.260	0.261
Fortaleza	0.202	0.208	0.126	0.144	0.121	0.175	0.326	0.309	0.383	0.376
Goiânia	0.178	0.163	0.082	0.097	0.082	0.091	0.288	0.260	0.328	0.315
Grande Vitória	0.222	0.204	0.101	0.107	0.113	0.114	0.329	0.310	0.362	0.353
Joinville	0.106	0.108	0.045	0.070	0.051	0.056	0.177	0.168	0.220	0.219
João Pessoa	0.241	0.261	0.115	0.086	0.130	0.029	0.372	0.369	0.398	0.395
Juiz de Fora	0.178	0.161	0.086	0.088	0.099	0.092	0.250	0.232	0.270	0.251
Jundiá	0.138	0.151	0.057	0.083	0.077	0.076	0.207	0.229	0.247	0.282
Limeira/Rio Claro	0.120	0.117	0.052	0.078	0.062	0.063	0.181	0.181	0.217	0.223
Londrina	0.166	0.149	0.078	0.095	0.079	0.085	0.257	0.232	0.296	0.286
Macapá	0.108	0.115	0.079	0.100	0.068	0.085	0.148	0.141	0.158	0.146
Maceió	0.214	0.228	0.132	0.151	0.126	0.148	0.329	0.340	0.366	0.377
Manaus	0.150	0.157	0.097	0.138	0.087	0.115	0.232	0.225	0.284	0.294
Natal	0.215	0.207	0.114	0.135	0.118	0.133	0.327	0.313	0.345	0.341
Porto Alegre	0.177	0.169	0.101	0.124	0.112	0.115	0.246	0.232	0.278	0.274
Recife	0.214	0.220	0.111	0.134	0.113	0.156	0.347	0.345	0.396	0.403
Ribeirão Preto	0.183	0.162	0.081	0.088	0.101	0.083	0.277	0.251	0.321	0.305
Rio de Janeiro	0.220	0.225	0.108	0.126	0.117	0.131	0.333	0.341	0.382	0.385
Salvador	0.232	0.222	0.118	0.137	0.123	0.150	0.368	0.355	0.411	0.407
Sorocaba	0.108	0.113	0.049	0.075	0.055	0.062	0.169	0.171	0.217	0.230
São José dos Campos	0.150	0.160	0.074	0.106	0.088	0.096	0.222	0.230	0.277	0.292
São Paulo	0.188	0.206	0.071	0.107	0.093	0.107	0.295	0.317	0.351	0.370
Teresina	0.186	0.193	0.124	0.147	0.114	0.138	0.279	0.277	0.331	0.329
Uberlândia	0.125	0.112	0.059	0.070	0.060	0.064	0.199	0.173	0.236	0.221

Table 2 shows the value of the HA indices for the 35 largest cities. When considering all incomes groups, Brasília is the most segregated city (a result already found by Telles (1995) in the 1980s), and Blumenau the least segregated. There is no clear trend in the changes in segregation in these larger cities between 2000 and 2010, as some cities decreased their segregation levels (e.g., Salvador and Brasília), whereas others increased them (e.g., Rio de Janeiro and São Paulo). For income groups, the much higher segregation of the rich is particularly salient in the largest cities, while there is no clear pattern in the relationship between city size and the segregation of the poor.

4. Urban spatial structure in Brazilian cities

4.1 Data

To calculate urban spatial structure measures, we use establishment-level data from the Annual Social Information Report (Relação Anual de Informações Sociais - RAIS) for 2000 and 2009. The RAIS is carried out by the Ministry of Labor of Brazil (Ministério do Trabalho e Emprego (MTE)), and covers an administrative report filled by all tax registered establishments. The data includes information on the number of employees and wage payroll, among others. The identified version of the database allocates a unique code to each firm, and provides the specific address of each establishment. In order to obtain employment totals at the enumeration area level, we use information on postal addresses and their corresponding 2010 enumeration area code for each municipality from the National Registry of Addresses (CNEFE) provided by the IBGE. For more details on the processing and geolocation of this data, see Appendix A.

Finally, it is important to notice that these datasets only consider formal jobs and, unfortunately, there is no other source to account informal employment at the enumeration area level. Since informality is an important feature of Brazilian cities and we can compute it at the city level using population census data, we will control for the informality rate in our empirical strategy. See Section 5 and Appendix A for further details.

4.2 Measuring urban spatial structure

Density

According to Anas *et al.* (1998), urban spatial structure refers to the degree of spatial concentration of population and employment within a city, and 'density' is usually used to measure such degree in terms of residents and/or jobs per unit of land. Traditionally this measure has been computed for the whole city, that is, by dividing the number of inhabitants and/or jobs by the area of the city. However, urban spatial structure also refers to the spatial distribution of firms and residences within the city (Horton and Reynolds, 1971), and the 'overall city density' does not take into account this feature.

Following Duranton and Turner (2016), we can consider both dimensions, the spatial concentration and the intrametropolitan distribution of jobs and/or residents, by computing 'density' at a more local geographical level. In particular, for the case of employment we can compute 'the

city average job density in a 1-km radius surrounding each enumeration area' (hereafter 1-km density):

$$\text{1-km ED} = \frac{1}{n} \sum_i \left(\frac{\text{UA Jobs in the surrounding 1-km radius}}{3.1416 \text{ sq km}} \right)$$

where n is the number of enumeration areas in each Brazilian UA.

While the 1-km density is our main explanatory variable, in some robustness checks we use this local measure using radii of 3, 5, 7 and 10 km.

Table 3 reports descriptive statistics for the 1-km density. For the sample of all 121 UAs, the average employment density within a radial distance of 1 km increased from 591 to 857 employments per square kilometer (45%) between 2000 and 2010. It is important to notice the important differences between the cities, as shown by the high standard deviations (448 and 591 jobs per sq km in 2000 and 2010, respectively).

Table 3: Summary statistics for urban spatial structure variables

	All cities				Monocentric cities				Polycentric cities			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
2000 Number of subcenters									2.9	4.1	1	22
2010 Number of subcenters									3.2	5.2	1	33
2000 1-km density	591	448	115	2,769	455	233	115	1,493	940	643	266	2,769
2010 1-km density	857	591	187	4,041	624	302	235	1,662	1,199	734	187	4,041
2000 UCI	0.382	0.165	0.003	0.702	0.389	0.161	0.203	0.702	0.366	0.177	0.003	0.686
2010 UCI	0.382	0.159	0.003	0.713	0.384	0.146	0.060	0.713	0.377	0.178	0.003	0.706
2000 ADC	9.9	13.2	0.9	76.0	8.0	10.6	0.9	76.0	14.5	17.5	2.0	72.1
2010 ADC	10.1	14.1	0.9	91.6	9.3	14.8	0.9	91.6	11.2	13.0	1.9	73.5
2000 JR	0.19	0.18	0.05	2.030	0.19	0.21	0.05	2.030	0.19	0.05	0.12	0.31
2010 JR	0.22	0.06	0.06	0.37	0.20	0.06	0.06	0.33	0.24	0.06	0.09	0.37

Notes: 121 observations (cities) in 'All cities' sample. 87 and 72 monocentric cities, and 34 and 49 polycentric cities in 2000 and 2010, respectively.

Table 4 reports individual descriptive statistics for the largest 35 UAs in Brazil. For the case of the 1-km density, there are important differences between cities: while São Paulo is the most dense (2,769 and 4,041 jobs per sq km in 2000 and 2010, respectively), Cuiabá is the less dense (339 and 521 jobs per sq km in 2000 and 2010, respectively).

Monocentricity vs. polycentricity

Taking into account the location patterns of jobs within cities, they can be classified as 'monocentric', when there is only one employment center or CBD, or as 'polycentric', when there are two or more employment centers (one CBD and one or more subcenters).

For the case of Brazil in 2000 and 2010, we identify the enumeration areas that make up the CBD and the subcenters. Since there is no official definition of CBDs in Brazilian cities, we identify the CBD in each of the UAs as the enumeration area (or group of enumeration areas) with the highest employment density and the highest employment count.

Table 4: Urban spatial structure in 35 UAs with population over half million, 2000–2010

UA	Subc. 2000	Subc. 2010	1-km density 2000	1-km density 2010	UCI 2000	UCI 2010	ADC 2000	ADC 2010	JR 2000	JR 2010
Aracaju	1	1	822	1,231	0.51	0.46	3.11	3.72	0.20	0.21
Baixada Santista	1	3	933	1,316	0.38	0.39	15.08	15.13	0.15	0.19
Barra Mansa/Volta Redonda	2	1	613	658	0.16	0.15	15.79	15.47	0.21	0.18
Belo Horizonte	9	13	2,476	2,901	0.48	0.42	6.59	7.92	0.29	0.32
Blumenau	0	1	360	507	0.20	0.18	11.81	11.58	0.34	0.37
Brasília	4	6	343	710	0.43	0.45	20.83	19.04	0.12	0.21
Campinas	2	3	858	1,206	0.01	0.01	19.59	20.03	0.20	0.26
Campo Grande	1	2	531	664	0.49	0.48	3.45	3.58	0.19	0.21
Cuiabá	1	0	339	521	0.36	0.28	72.06	91.59	0.17	0.20
Curitiba	4	4	1,485	2,194	0.44	0.41	6.96	7.07	0.28	0.36
Feira de Santana	1	1	530	890	0.44	0.42	5.04	5.15	0.14	0.18
Florianópolis	1	2	1,700	2,269	0.48	0.43	4.82	5.98	0.31	0.37
Fortaleza	4	4	1,130	1,677	0.46	0.45	8.50	8.71	0.16	0.20
Goiânia	4	5	834	1,402	0.50	0.53	7.34	6.58	0.20	0.27
Grande Vitória	1	2	987	1,748	0.28	0.27	8.99	9.36	0.18	0.28
Joinville	2	2	420	657	0.33	0.32	13.63	13.74	0.28	0.33
João Pessoa	0	2	407	1,375	0.40	0.41	5.26	3.09	0.10	0.20
Juiz de Fora	0	1	1,493	2,187	0.40	0.38	2.95	3.10	0.18	0.24
Jundia	0	1	627	958	0.31	0.26	5.28	6.25	0.23	0.29
Limeira/Rio Claro	1	1	533	723	0.04	0.04	17.88	17.02	0.20	0.24
Londrina	1	2	701	1,034	0.34	0.31	12.07	12.83	0.20	0.27
Macapá	1	1	266	749	0.67	0.67	9.70	7.11	0.12	0.19
Maceió	1	1	806	1,042	0.08	0.18	6.33	5.66	0.15	0.17
Manaus	1	3	627	1,045	0.69	0.65	12.45	13.96	0.15	0.20
Natal	0	2	686	1,653	0.16	0.25	6.55	6.38	0.15	0.25
Porto Alegre	5	7	1,034	1,057	0.26	0.25	54.36	57.72	0.27	0.25
Recife	3	5	1,240	1,818	0.22	0.18	6.36	7.10	0.18	0.23
Ribeirão Preto	0	0	909	1,195	0.32	0.31	6.92	7.06	0.24	0.26
Rio de Janeiro	12	15	2,222	2,318	0.35	0.49	11.89	9.40	0.21	0.20
Salvador	3	6	2,294	2,784	0.40	0.37	9.16	10.48	0.22	0.26
Sorocaba	0	0	464	685	0.20	0.21	15.70	15.58	0.18	0.23
São José dos Campos	2	6	673	1056	0.00	0.00	29.13	29.25	0.18	0.24
São Paulo	22	33	2,769	4,041	0.32	0.30	11.39	11.82	0.27	0.35
Teresina	1	1	426	722	0.57	0.62	5.18	4.66	0.15	0.21
Uberlândia	2	2	835	1,196	0.12	0.12	71.80	73.49	0.22	0.26

Similarly, there is no official definition of employment subcenters. An employment subcenter is a place with a significantly larger employment density than nearby locations that has a significant effect on the overall spatial distribution of jobs. We identify subcenters using the method first developed by McDonald and Prather (1994) and improved by McMillen (2001). The main idea is to estimate densities following a monocentric spatial pattern. The predicted densities obtained are subtracted from the corresponding real densities. From these residuals, those that are positive and statistically significant are selected.

While McDonald and Prather (1994) estimate by OLS a two dimensional density function (log of density on the distance to CBD), McMillen (2001) suggests estimating a three-dimensional density function (log of density on north-south and east-west distances to CBD) with a Locally Weighted Regression (LWR). Both improvements allow taking into account geographical differences, which, in terms of the spatial pattern of densities, can occur in any direction from the CBD (e.g. steeper density gradients on the north side than on the south side of the city). They additionally allow us to define any type of monocentric spatial pattern: concave, convex or linear.

We therefore estimate the following population/employment density function:

$$\ln(\text{Density}) = \gamma_0 + \gamma_1 \times \text{North-south distance to CBD} + \gamma_2 \times \text{East-west distance to CBD} \quad (3)$$

where density is measured as jobs per square kilometer, and distances are in kilometers. Since our estimates will be based on LWR, we need to define a bandwidth. As McMillen (2001) points out, this is a critical choice because we need a monocentric benchmark. We experiment with alternative window sizes ranging from 1% to 9% and from 10% to 90%. Based on visual inspection, a 50% window size shows the first monocentric spatial configuration.

Second, for each site, we compute the residual as the difference between the log of real density and the estimated log of density. We then select those that are significantly positive at a 10% level, according to their own standard errors that can vary over space (McMillen, 2001). Finally, we will group the selected sites in subcenters when they are contiguous. We use a 'queen' criterion for contiguity: two sites are contiguous if they share at least one point in their boundaries.

We detect the existence of subcenters in 34 and 49 UAs (Table 3). On average, the number of subcenters increased from 2.9 to 3.2 and São Paulo is the city with more subcenters (22 and 33 in 2000 and 2010, respectively) (Table 4).

According to their urban spatial structure, Table 3 shows that polycentric cities are more dense (940 and 1,199 jobs per sq km in 2000 and 2010, respectively) than their monocentric counterparts (455 and 624 jobs per sq km in 2000 and 2010, respectively). Furthermore, standard deviations show important differences within both urban forms.

Table 4 show that (1) most populated cities are mainly polycentric (only Ribeirão Preto and Sorocaba are monocentric in 2000 and 2010, and Cuiabá becomes monocentric in 2010), (2) most polycentric cities increase the number of subcenters (only 8 UAs have the same subcenters in 2000 and 2010), and (3) the higher the 1-km density, the higher the number of subcenters.

Other measures

In order to fully characterize the urban spatial structure we also compute other measures. First, Pereira *et al.* (2013) propose the 'urban centrality index (UCI)' to measure the degree of monocentricity-polycentricity:

$$\text{UCI} = LC \times PI = LC \times (1 - \frac{SSI}{SSI_{max}})$$

Where LC is the traditional 'location coefficient' that measures the unequal distribution of jobs within each UA, that is, how disproportionately jobs are clustered in a few locations or dispersed (Galster, Hanson, Ratcliffe, Wolman, Coleman, and Freihage, 2001):

$$LC = \frac{1}{2}\sum_i(\frac{\text{Jobs}_i}{\text{UA jobs}} - \frac{1}{n})$$

where n is the number of enumeration areas in each UA. The range of the LC is zero to $1 - 1/n$. If LC equals zero, then economic activity is evenly distributed, while values close to $(1 - 1/n)$ indicate that employment is concentrated in a few areas. In its conventional form, this coefficient captures only the nonspatial inequality of job distribution.

PI is a 'proximity index' built on the 'spatial separation index (SSI)' proposed by Midelfart-Knarvik, Overman, Redding, and Venables (2002) to overcome its problems: We also consider the Spatial Separation Index (SSI) proposed by Midelfart-Knarvik *et al.* (2002):

$$\text{SSI} = \sum_i\sum_j(\frac{\text{Jobs}_i}{\text{UA jobs}} \times \text{Distance from i to j} \times \frac{\text{Jobs}_j}{\text{UA jobs}})$$

The minimum value of SSI is zero and indicates that all jobs are concentrated in just one enumeration area. By normalizing with the maximum attainable value of the SSI (SSI_{max}), the PI can be compared across UAs. The range of PI is zero to one. If PI is zero, all jobs are spatially separated as possible and the distance between them is at its maximum.

The UCI combines the advantages of the LC and the PI: it controls for differences in size and shapes and, as result, allow for comparison across UAs. UCI values range between 0 and 1: values close to zero are related to a more polycentric spatial configuration, and values close to one to a more monocentric urban spatial structure.

Table 3 shows average UCI values of 0.381 in both years and, as a result, the 121 cities are more close to a polycentricity than to monocentricty. When considering the monocentric sample, the UCI value reduces indicating that these monocentric cities became more disperse between 2000 and 2010. On the other hand, polycentric cities increase their UCI value and, as a result, these cities are becoming more concentrated around their employment subcenters and around their CBD (with (new) subcenters near the CBD).

Table 4 also shows big differences between the largest cities: while the most polycentric city is São José dos Campos (its jobs are highly concentrated around its unique subcenter and its CBD), the least polycentric UA is Manaus (with 1 and 3 subcenters in 2000 and 2010, respectively). Among the monocentric cities, jobs are more concentrated (around the CBD) in Ribeirão Preto than in Sorocaba.

We also use the weighted average distance from the CBD (ADC) to measure the degree of centralization, that is, the extent to which employment is concentrated near the CBD (Galster et al., 2001):

$$\text{ADC} = \sum_i \left(\frac{\text{Jobs}_i}{\text{UA jobs}} \times \text{Distance to CBD}_i \right)$$

Table 3 shows that all 121 cities are highly centralized: on average, jobs are located 9.9 and 10.1 km from the CBD in 2000 and 2010, respectively. According to their urban form, monocentric cities are more centralized and, at the same time, they decentralized between 2000 and 2010 (from 8 to 9.3 km from the CBD). On the contrary, polycentric UAs are more decentralized but became more centralized between 2000 and 2010 (from 14.5 to 11.2 km). These trends are in line with UCI trends: monocentric cities are becoming more disperse/decentralized, and polycentric cities are becoming more concentrated/centralized with (new) subcenters located closer to the CBD. Once again, it is important to notice the important differences between cities within the monocentric and polycentric samples as shown by the high standard deviations in ADC values.

Table 4 also highlights these big differences between the largest cities: while Aracaju (polycentric with one subcenter) is the most centralized city, Cuiabá (which became monocentric in 2010) is the most decentralized.

Finally, we consider the traditional job ratio to measure the balance between employment and population:

$$\text{JR} = \frac{\text{UA jobs}}{\text{UA inhabitants}}$$

On average, the number of jobs per inhabitant increased from 0.19 to 0.22 between 2000 and 2010. According to their urban configuration, the increase was smaller in monocentric cities (from 0.19 to 0.20) than in polycentric cities (from 0.19 to 0.24) (Table 3). Among the largest cities (Table 4), the JR is smaller in Baixada Santista (0.15 and 0.19 in 2000 and 2010, respectively) and higher in Blumenau (0.34 and 0.37 in 2000 and 2010, respectively).

5. Econometric approach

5.1 Expectations on empirical estimates

Previous studies have attempted to establish a relationship between observed city-level values of income segregation and city characteristics, such as city size, population density and income inequality (Pendall and Carruthers, 2003, Wheeler, 2008). Based on the theoretical predictions outlined in Section 2, our aim is to establish whether urban spatial structure has an independent role in determining income segregation levels. According to the models discussed, the main driving force behind this relationship is the competition for proximity to existing economic opportunities among individuals with varying income levels. Lower commuting costs, preference for more housing space, higher service provision, more decentralized amenities, and a more decentralized pattern of employment are opposing forces to the expected effect of urban spatial structure on income segregation.

Clearly, the scale of existing economic activity is a key factor mediating the relationship between urban spatial structure and income segregation. Access to economic centers typically decreases as population expands while economic activity remains concentrated and transport provision remains constant. For this reason, it can be expected that for small cities in which commuting costs to a unique center remain relatively low, there is a weaker impact of urban spatial structure on income segregation levels. For medium-sized cities, the expectation is that larger local density levels are related to larger levels of income segregation, as there is tighter competition for access to economic opportunities between firms and households of different income levels. The positive impact of local employment density on income segregation can reach its maximum level in cities where the tight locational competition is exacerbated by policies restricting the availability of housing at accessible locations, and/or a lack of provision of public services, transport and amenities outside the main center. After a certain scale is reached, congestion effects at the unique center start dominating, and the emergence of new economic (sub)centers acts as a release on the clustering of households of similar incomes in particular neighborhoods.

The way in which urban spatial structure impacts segregation levels at the top and bottom of the income distribution is determined by a combination of varying degrees of residential mobility and the availability of housing catering different income groups. The fact that higher income levels are associated with a higher degree of residential mobility implies that many of the dynamics underlying the impact of urban spatial structure on income segregation levels are driven by residential mobility decisions of higher income households. In other words, concentrations of high-income households in a few neighborhoods is likely to be the outcome of locational bidding, while concentrations of lower income households may be the result of low residential mobility. As cities get larger, the very same high degree of residential mobility at the top of the income distribution can be behind a fall in income segregation levels, as higher income households disperse across multiple locations in the city, while the low residential mobility of lower-income households can have the opposite effect, as lower income households remain clustered in few areas.

The extent to which these relationships hold can help understand what kind of city configurations are more conducive to more inclusive cities, and serve as a framework to understand the possible role of policies affecting accessibility and availability of housing across income groups on income segregation. Given that predictions of the effect of urban spatial structure on income segregation are expected to vary with the degree of monocentricity, city size and income group, in the empirical application we present separately the results for these dimensions, after presenting the result for the base average case.

5.2 Specification

After obtaining our measures of income segregation and urban spatial structure, we turn to our research questions: holding other factors constant, which is the effect of the urban spatial structure on income segregation?

To answer this question, we use data for 2000 and 2010 to regress the log of HA on the log of urban form variable/s:

$$
\begin{aligned}
\ln(\mathrm{HA}_{i,t}) = \delta_0 + \delta_1 &\times \ln(\text{Urban spatial structure}_{i,t}) \\
&+ \sum_j (\delta_{2,j} \times \text{Informality}_{i,j,t} \ \& \ \text{Inequality}_{i,j,t-10}) \\
&+ \sum_s (\delta_{3,s} \times \text{Geography}_{i,s}) \\
&+ \sum_k (\delta_{4,k} \times \text{Demography}_{i,k,t-10}) \\
&+ \sum_l (\delta_{5,l} \times \text{Industrial composition}_{i,l,t-10})
\end{aligned}
\tag{4}
$$

Since urban spatial structure variables are computed for formal jobs (because of data availability), we first control for the share of informal jobs (%) in the UAs in 2000 and 2010 (Informality) (see Appendix A for further details). Furthermore, we also control for differences in the degree of income Inequality in the cities by including the log of Gini index[6] and the log of per capita income in 1990 and 2000.

We add controls for Geography such as the total area (km^2) of the city, a dummy for UAs located on the coast, and a dummy for cities located on semi-arid regions. Following Da Mata, Deichmann, Henderson, Lall, and Wang (2007), we control for planning policies by adding the share of population in municipalities within the UA with land zone law (%).

We also include controls for Demography such as the share of population above 55 years old (%), the share of population below 25 years old (%), and the share of migrants (%) in 1990 and 2000.

Finally, we control for Industrial composition with the share of jobs in manufacturing (%) and the share of jobs in services (%) in 1990 and 2000.

Summary statistics for the segregation index and the urban spatial structure variables were previously discussed and are in Tables 1 and 3. Descriptive statistics for our controls variables are in Appendix A Table D.2.

5.3 Endogeneity

We fear for some sources of endogeneity in the relationship running from urban spatial structure to segregation, in particular for our main explanatory variable, the 1-km employment density.

To address these concerns, our empirical strategy rely on instrumental variables (IV) techniques in which we instrument the 1-km job density with the overall city job density. Following Ciccone and Hall (1996) and Wheeler (2008), our instrument uses 30 years lagged values: the UA job density for 1970 and 1980.

This instrument is valid because of its significant first-stage coefficients (available upon request). This instrument is also exogenous because, as Da Mata *et al.* (2007) highlight, Brazil and its cities have undergone significant changes in their economy and society since the 1970s and the 1980s. Similar to Duranton and Turner (2012), Garcia-López (2012), Garcia-López, Holl,

---

[6]Alternatively, we use the log of Theil index and results hold.

and Viladecans-Marsal ([2015]) and Garcia-López, Hemet, and Viladecans-Marsal ([2017a,b]), the exogeneity of historical instruments hinges on having an appropriate set of controls, geography and history variables in particular. In our case, we add controls for geography, but also inequality, demography and industrial composition variables computed using historical values (1990 and 2000).

6. Results

6.1 Does employment density affect income segregation?

To study the impact of urban spatial structure on income segregation, we first investigate whether higher local employment densities increase or reduce income segregation levels. To do so, we use Eq. ([4]) to estimate the effect of 1-km job density on our HA index.

Table [5] presents OLS and TSLS estimates. Conditional on the full set of control variables, OLS results are in column 1 and the estimated coefficient of interest show that a 10% increase in the local density increases income segregation index by 2% (around a 0.02-0.03 points increase in the index).

Table 5: The effect of urban spatial structure on income segregation: Employment density

Dependent variable:	ln(HA index)							
Years:	00-10	00-10	00-10	00-10	00-10	00-10	2000	2010
Method:	OLS [1]	TSLS [2]	TSLS [3]	TSLS [4]	TSLS [5]	TSLS [6]	TSLS [7]	TSLS [8]
ln(1-km employment density)	0.197^a (0.037)	0.321^a (0.039)	0.304^a (0.075)	0.339^a (0.070)	0.368^a (0.074)	0.446^a (0.100)	0.421^a (0.087)	0.485^a (0.155)
Adjusted R^2	0.720							
First-stage F-statistic		84.95	49.89	59.91	58.79	46.48	35.09	16.65
Instrument				ln(t-30 years employment density)				
Informality & Inequality	Y	N	Y	Y	Y	Y	Y	Y
Geography	Y	N	N	Y	Y	Y	Y	Y
Demography	Y	N	N	N	Y	Y	Y	Y
Industrial composition	Y	N	N	N	N	Y	Y	Y
Observations	242	242	242	242	242	242	121	121
UA	121	121	121	121	121	121	121	121

Notes: All regressions include state and year fixed-effects. Robust standard errors are in parenthesis (clustered by UA in columns 1-5). [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

In columns 2-8, we instrument local density with the overall UA density lagged $t - 30$ years. Column 2 includes the 1-km density and state and year fixed-effects, column 3 adds controls for income and inequality, column 4 adds geography, column 5 adds demography, and column 6 includes industrial composition controls. Columns 7 and 8 are cross-section estimates for

years 2000 and 2010. With estimated coefficients[7] that are quite stable across the different specifications, TSLS results clearly show that local density conditions have a significant effect on income segregation. In particular, results in our preferred specification in column 6 indicate that a 10% increase in local job density increases income segregation index by 4.5% (around a 0.06 points increase in the index). First-stage statistics for the TSLS regressions are above the Stock and Yogo (2005) critical values.

Now we consider alternative computations of job density and alternative measures of urban spatial structure. The idea is to check the robustness of the above results and, in particular, to test whether other aspects of the urban spatial structure also matter. Table 6 reports TSLS results when we compute the local job density for different radii (columns 1-4) and when we combine our preferred local density measure (1-km) with other urban spatial structure variables used in the empirical literature (see Section 4): the weighted average distance from the CBD (columns 5 and 8), the job population ratio (columns 6 and 8), and the urban centrality index (columns 7 and 8). Table 6 also reports their first-stage statistics and all of them are above the Stock and Yogo (2005) critical values.

Table 6: The effect of urban spatial structure on income segregation: Other measures

Dependent variable:	ln(HA index)							
Years:	00-10	00-10	00-10	00-10	00-10	00-10	00-10	00-10
Method:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
Radius:	3-km	5-km	7-km	10-km	1-km	1-km	1-km	1-km
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
ln(X-km employment density)	0.356^a	0.296^a	0.253^a	0.214^a	0.457^a	0.361^a	0.340^a	0.351^a
	(0.072)	(0.053)	(0.043)	(0.036)	(0.095)	(0.083)	(0.074)	(0.069)
ln(Weighted average distance from CBD)					-0.014			-0.036
					(0.024)			(0.031)
ln(Job ratio)						-0.145		-0.132
						(0.092)		(0.085)
ln(Urban centrality index)							-0.039^c	-0.069^b
							(0.023)	(0.029)
First-stage F-statistic	85.27	125.7	174.6	237.5	59.55	68.12	58.20	119.6
Instrument				ln(t-30 years employment density)				
Informality & Inequality	Y	Y	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y	Y	Y
Demography	Y	Y	Y	Y	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y	Y	Y	Y	Y
Observations	242	242	242	242	242	242	242	242
UA	121	121	121	121	121	121	121	121

Notes: All regressions include state and year fixed-effects. Robust standard errors clustered by UA are in parenthesis. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

For the case of alternative local density variables, their estimated coefficients decrease with

---

[7]Alternatively, we also estimate a 2000–2010 changes on 2000 levels version of Eq. (4) to account for potential path dependence issues and the results hold. Similarly, as explained in Section 3, our segregation index is computed using information for the head of household because income of other household members is only available for 2010. Table E.1 show results for 2010 cross-sections when HA is computed with income information of all household members. Results are similar to those in column 8 Table 5.

the radius of the density variable, but, in general, they are in line with our preferred estimate in column 6 Table 5: higher local densities increase income segregation.

When combining the 1-km density with alternative urban form variables, the estimated coefficient for the 1-km density remains positive and similar to our preferred estimate in column 6 Table 5. Among the alternative urban spatial structure variables, only the urban centrality index is significant (columns 7 and 8). Its negative estimated coefficient indicates that a higher degree of monocentricity decreases income segregation. In the following analyses we study the impact of urban spatial structure on income segregation using both significant variables: the 1-km job density, to consider the effect of local density conditions, and the urban centrality index, to consider the effect of the intrametropolitan distribution of jobs in a monocentric or polycentric context.

As a whole, TSLS results in Tables 5 and 6 show that density increases income segregation. At the same time, a more monocentric configuration, that is, an employment location pattern more centralized around the CBD, reduces income segregation.

6.2 *Does monocentricity-polycentricity foster income segregation?*

Descriptive statistics in Section 4 show the existence of monocentric and polycentric cities, and, in particular, the clear relationship between higher densities and a polycentric location pattern. Since the type of urban spatial structure and the above mentioned relationship might affect the estimated coefficients for the 1-km density and the urban centrality index, we now turn our attention to study whether previous results hold when we separately consider both urban forms.

Table 7 reports TSLS results when we separately study monocentric (columns 1-2) and polycentric (columns 3-4) cities. For both types of cities and conditional on the full set of control variables, columns 1 and 3 includes the 1-km job density, and columns 2 and 4 adds the urban centrality index. Their first-stage statistics are above the Stock and Yogo (2005) critical values (columns 2-4) or near the Stock and Yogo (2005)'s rule of thumb (F>10).

Table 7: The effect of urban spatial structure on income segregation: Mono vs. Polycentricity

Dependent variable:	ln(HA index)			
Urban spatial structure:	Monocentric		Polycentric	
Years:	00-10	00-10	00-10	00-10
Method:	TSLS	TSLS	TSLS	TSLS
	[1]	[2]	[3]	[4]
ln(1-km employment density)	0.819^a	0.676^a	0.286^a	0.312^a
	(0.284)	(0.186)	(0.087)	(0.072)
ln(Urban centrality index)		-0.069		0.073^b
		(0.046)		(0.031)
First-stage F-statistic	8.46	14.74	49.21	52.77
Instrument	ln(t-30 years employment density)			
Informality & Inequality	Y	Y	Y	Y
Geography	Y	Y	Y	Y
Demography	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y
Observations	159	159	83	83

Notes: All regressions include state and year fixed-effects. Robust standard errors clustered by UA are in parenthesis. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

According to the type of employment location pattern, the local density still show a positive and significant impact on income segregation. However, the effect is significantly smaller for polycentric cities than for monocentric ones: a 10% increase in the 1-km job density increases income segregation by 7% in monocentric cities and only by 3% in polycentric cities.

Similarly, the effect of the urban centrality index differs between urban forms: while it is no significant for monocentric cities, it is positive and significant for polycentric cities. This latter result implies that a higher degree of polycentricity (i.e., a lower value for the urban centrality index) reduces income segregation. In other words, in polycentric cities like Sao Paulo the more the employment is clustered around the subcenters and far from the CBD, the more income segregation decreases.

Jointly, the 1-km job density and the urban centrality index results clearly show that urban spatial structure plays an important role on city's income segregation.

6.3 Does city size matter?

Since our sample has a high degree of heterogeneity in city population size and, in particular, since descriptive statistics in Sections 3 and 4 also show a clear relationship between city size and income segregation and between city size and urban spatial structure, we now investigate the effect of the 1-km density and the urban centrality index for different city sizes.

Table 8 shows TSLS results for different city size subsamples and urban spatial structures. Column 1 considers only cities with less than 100,000 inhabitants. These smaller cities are all monocentric. The estimated coefficients for the 1-km density and the urban centrality index are insignificant.

Table 8: The effect of urban spatial structure on income segregation: Population size

Dependent variable:	ln(HA index)				
Population:	<100,000	100,000-700,000			>700,000
Urban spatial structure:	Mono	Both	Mono	Poly	Poly
Method:	TSLS	TSLS	TSLS	TSLS	TSLS
	[1]	[2]	[3]	[4]	[5]
ln(1-km employment density)	0.138	0.591[a]	0.745[a]	0.436[a]	-0.092[b]
	(0.274)	(0.151)	(0.174)	(0.098)	(0.033)
ln(Urban centrality index)	0.098	-0.036	-0.092[c]	0.149[a]	0.092[a]
	(0.119)	(0.033)	(0.050)	(0.037)	(0.013)
First-stage F-statistic	8.20	17.00	13.47	4.51	94.38
Instrument		ln(t-30 years employment density)			
Informality & Inequality	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y
Demography	Y	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y	Y
Observations	41	159	118	41	42

Notes: Robust standard errors clustered by UA are in parenthesis. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

In columns 2-4, we study UAs with population between 100,000 and 700,000 inhabitants. When we jointly consider both urban forms (column 2), only the 1-km density is significant and positive (0.591). According to their urban spatial structure, the effect of 1-km density increases to 0.745 in monocentric cities (column 3), and reduces to 0.436 in polycentric UAs (column 4). Furthermore, both urban forms also differ in their effect of the urban centrality index. For monocentric cities, the estimated coefficient is negative and significant (-0.092), and show that an increase in their urban centrality index (i.e., an employment location pattern more centralized around the CBD than disperse) reduces income segregation. On the other hand, the estimated coefficient is positive and significant (0.149) for polycentric cities and show that a reduction in the urban centrality index (i.e., employment concentrated around subcenters and located far from the CBD) reduces income segregation.

Finally, column 5 includes cities with more than 700,000 inhabitants. With the exception of Cuiabá and Sorocaba, these cities are polycentric. Now the estimated coefficient for the local density is negative and significant: a 10% increase in the 1-km density decreases income segregation by 0.9%. Furthermore, an increase in the degree of polycentricty (measured by a lower urban centrality index), also helps to reduce income segregation.

To sum up, these results show that the effects of local density and of the urban centrality index depend on the size of the city (inhabitants) and on the urban form. For the case of the 1-km density: (1) there is no effect related to smaller (monocentric) cities, (2) it appears in medium size cities by increasing income segregation and it is smaller in polycentric cities, (3) an increase of local densities reduces income segregation in large (polycentric) cities. For the case of the urban centrality index: (4) medium and large cities may reduce their income segregation by increasing their degree of monocentricity (i.e., with jobs more clustered in and around the CBD) or by increasing their degree of polycentricity (i.e., with jobs clustered in and around their subcenters

and located far from the CBD).

6.4 Do income groups matter?

We turn our attention to the different income groups. As shown in Section 3, the level of segregation increases with income levels, so that in most cities the rich are far more segregated than the poor. At the same time, while there was a reduction in the segregation of the rich between 2000 and 2010, there was an increase in the segregation of the poor.

Table 9 shows TSLS results for the poor (Panel A) and the rich (Panel B). Column 1 considers all cities in the sample and, in both cases, shows that, on average, only local density is significant: a 10% increase in the 1-km density of jobs increases segregation of the poor and of the rich by a 4%.

Column 2 reports results for the smaller (monocentric) cities. While the urban centrality index remains not significant, job density significantly affects segregation of the rich, but not of the poor.

Results for medium size cities (columns 3-5) show that, on average, a 10% increase in the 1-km density increases segregation by a 5-6% (column 3). However, this effect is higher in monocentric cities (column 4) than in polycentric UAs (column 5). When comparing the poor and the rich, the density effect is similar between the two groups in monocentric cities, whereas it is higher and only significant for the poor in polycentric cities. The urban centrality index is only significant for polycentric UAs and it shows that an increase in the degree of polycentricity reduces segregation of both income groups.

Finally, urban spatial structure seems to differently affect the segregation of the rich and of the poor in the large (polycentric) cities: while local density conditions increase the segregation of the poor, a more polycentric configuration reduces the segregation of the rich.

As a robustness check, Table E.2 in Appendix E reports results when considering a different definition of the poor (20th percentile) and of the rich (80th percentile). While on average results holds, there are a couple of differences that need to be commented: (1) the density effect is always higher for the poor than for the rich in medium size and large cities; (2) an increase of density reduces segregation of the rich and it does not affect segregation of the poor in large cities; (3) medium size monocentric cities can reduce segregation of the poor by increasing their degree of monocentricity (with more jobs centralized in and around the CBD).

Table 9: The effect of urban spatial structure on income segregation: Poor vs. Rich

Dependent variable:	ln(HA index)					
Population:	All	<100,000	100,000-700,000			>700,000
Urban spatial structure:	Both	Mono	Both	Mono	Poly	Poly
Years:	00-10	00-10	00-10	00-10	00-10	00-10
Method:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A: Poor (10th percentile)						
ln(1-km employment density)	0.425^a	0.031	0.486^b	0.769^a	0.442^b	0.196^a
	(0.117)	(0.390)	(0.201)	(0.256)	(0.183)	(0.033)
ln(Urban centrality index)	-0.029	0.140	-0.043	-0.091	0.115^c	-0.005
	(0.029)	(0.197)	(0.038)	(0.056)	(0.065)	(0.013)
Panel B: Rich (90th percentile)						
ln(1-km employment density)	0.417^a	0.371^c	0.582^a	0.729^a	0.178	-0.010
	(0.081)	(0.222)	(0.132)	(0.161)	(0.132)	(0.022)
ln(Urban centrality index)	-0.016	0.070	-0.025	-0.075	0.145^a	0.076^a
	(0.019)	(0.123)	(0.031)	(0.050)	(0.045)	(0.010)
First-stage F-statistic	64.40	8.20	17.00	13.47	4.51	94.38
Instrument	ln(t-30 years employment density)					
Informality & Inequality	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y
Demography	Y	Y	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y	Y	Y
Observations	242	41	159	118	41	42

Notes: All regressions include state and year fixed-effects. Robust standard errors clustered by UA are in parenthesis. ^{a}, ^{b}, and ^{c} indicates significant at 1, 5, and 10 percent level, respectively.

7. Conclusions

We have analyzed the relationship between income segregation, or the degree of unevenness in the distribution of households by levels of income within cities, and urban spatial structure, or the degree of spatial concentration of employment and its distribution in the urban space. Coming back to our first question, 'Does employment density affect income segregation?', we find that higher local employment densities lead to higher levels of average income segregation. This effect depends on city size and the level of monocentricity: the effect is not significant in small cities with a unique center, strongest in medium sized cities with a high degree of monocentricity, and opposite in larger cities that are more polycentric.

Regarding our second question, 'Does monocentricity-polycentricity foster income segregation?', we find that the positive effect of local employment density on income segregation is smaller in cities with a polycentric structure. What is more, in polycentric cities with employment subcenters located far from the CBD, we find that local employment density actually decreases income segregation. This result holds when we study the relationship for the segregation of the rich, but it is actually opposite when considering the segregation of the poor.

These results can be interpreted in the light of urban theoretical models. Small cities are characterized by relatively low levels of income segregation and high levels of employment density and monocentricity, reflecting the fact that at small sizes commuting costs are relatively

low, making competition for location near the unique center less intense both for households and firms.

As cities grow, the competition for proximity to an existing employment center intensifies, raising land prices for both firms and households. How households of certain income levels and firms react to this increased competition will depend on their valuation of local amenities (which may be highly concentrated around the unique employment center), and how much firms value proximity to these local amenities and to other firms. According to our results, the effect of local employment density on income segregation has its peak in medium sized cities with a monocentric structure, meaning that under these conditions, a higher local density of employment leads to more homogeneous neighborhoods in terms of their income composition.

In larger cities, high rental prices and congestion costs in central locations ultimately lead to a deconcentration of employment. This process is also accompanied by a relocation of households of different income levels that adjust their residential location to their valuations of proximity to old and new employment centers and local amenities. When cities reach a polycentric structure with subcenters also present in areas outside the historical CBD -reflecting for instance the availability of public and transport infrastructure outside central areas -, increases in local employment density lead to more heterogeneous neighborhoods in terms of income composition. According to our results, this increased heterogeneity that accompanies the emergence of subcenters far away from the CBD occurs because of a lower segregation of households at the top of the income distribution. For those at the bottom of the income distribution, increased levels of local employment density under a polycentric structure still lead to higher segregation levels, perhaps reflecting the low residential mobility of those at the bottom of the income distribution and their inability to bid for locations where households in higher income levels locate, and the presence of alternative sources of informal employment.

From a policy point of view, our results suggest that income segregation is directly related to the intrametropolitan location pattern of firms. Holding other factors constant (e.g., the benefits of agglomeration economies), policies fostering more concentration of jobs in and around the CBD and, in particular, in and around employment subcenters located far from the CBD (i.e., a more polycentric spatial structure) might help to alleviate income segregation in Brazilian cities. The question that arises is how to modify the urban spatial structure and, in particular, how to promote polycentricity. While the literature on this topic is still scarce, recent research shows that the emergence of local job density peaks that eventually become employment subcenters is related to the intrametropolitan location of transportation infrastructure (Garcia-López *et al.*, 2017*a,b*), a key variable also related to income segregation (Glaeser *et al.*, 2008).

References

Alonso, William. 1964. *Location and Land Use. Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Ananat, Elizabeth O. 2011. The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics* 3(2): 34–66.

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36(3): 1426–1464.

Bischoff, Kendra and Sean F. Reardon. 2014. Residential segregation by income, 1970–2009. In John Logan (ed.) *Diversity and Disparities America Enters a New Century*. Rusell Sage Foundation, 208–233.

Brueckner, Jan K and Stuart S Rosenthal. 2009. Gentrification and neighborhood housing cycles: WillAmerica's future downtowns be rich? *The Review of Economics and Statistics* 91(4): 725–743.

Brueckner, Jan K., Jacques-François Thisse, and Yves Zenou. 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review* 43(1): 91–107.

Brueckner, Jan K, Jacques-François Thisse, and Yves Zenou. 2002. Local labor markets, job matching, and urban location. *International Economic Review* 43(1): 155–171.

Cavalcanti, Tiago and Daniel Da Mata. 2013. On the determinants of slum formation. *Unpublished manuscript, Department of Economics, University of Cambridge* .

Cheshire, Paul, Vassilis Monastiriotis, and Stephen Sheppard. 2003. Income inequality and residential segregation: labour market sorting and the demand for positional goods. In Paul Cheshire and Vassilis Monastiriotis (eds.) *Geographies of Labour Market Inequality*. Routledge: London, 83–109.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics* 129(4): 1553–1623.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *The American Economic Review* 86(1): 54–70.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. The rise and decline of the American ghetto. *Journal of Political Economy* 107(3): 455–506.

Da Mata, Daniel, Uwe Deichmann, J Vernon Henderson, Somik V Lall, and Hyoung Gun Wang. 2007. Determinants of city growth in Brazil. *Journal of Urban Economics* 62(2): 252–272.

Duranton, Gilles and Matthew A. Turner. 2012. Urban growth and transportation. *The Review of Economic Studies* 79(4): 1407–1440.

Duranton, Gilles and Matthew A. Turner. 2016. Urban form and driving: Evidence from US cities. Processed, Wharton School, University of Pennsylvania.

Feitosa, Flávia F., Quang Bao Le, Paul L.G. Vlek, Antônio Miguel V. Monteiro, and Roberta Rosemback. 2012. Countering urban segregation in Brazilian cities: Policy-oriented explorations using agent-based simulation. *Environment and Planning B: Planning and Design* 39(6): 1131–1150.

Feler, Leo and J. Vernon Henderson. 2011. Exclusionary policies in urban development: Underservicing migrant households in Brazilian cities. *Journal of Urban Economics* 69(3): 253–272.

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2): 161–196.

Galster, George, Royce Hanson, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage. 2001. Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate* 12(4): 681–717.

Garcia-López, Miquel-Àngel. 2012. Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics* 72(2): 176–190.

Garcia-López, Miquel-Àngel, Camille Hemet, and Elisabet Viladecans-Marsal. 2017*a*. How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail. *Journal of Regional Science* 57(5): 758–780.

Garcia-López, Miquel-Angel, Camille Hemet, and Elisabet Viladecans-Marsal. 2017*b*. Next train to the polycentric city: The effect of railroads on subcenter formation. *Regional Science and Urban Economics* 67(C): 50–63.

Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics* 85: 52–67.

Glaeser, Edward L. 2008. *Cities, agglomeration, and spatial equilibrium*. Oxford University Press.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics* 63(1): 1–24.

Griffin, Ernst and Larry Ford. 1980. A model of Latin American city structure. *Geographical Review* : 397–422.

Henderson, J. Vernon, T. Regan, and Anthony Venables. 2016. Building the city: Sunk capital and sequencing. Technical report, Discussion paper 11211, Centre for Economic Policy Research, London.

Henley, Andrew, G. Reza. Arabsheibani, and Francisco G. Carneiro. 2009. On defining and measuring the informal sector: Evidence from Brazil. *World Development* 37(5): 992–1003.

Hong, Seong-Yun, David O'Sullivan, and Yukio Sadahiro. 2014. Implementing spatial segregation measures in R. *PloS one* 9(11): e113767.

Horton, Frank E. and David R. Reynolds. 1971. Effects of urban spatial structure on individual behavior. *Economic Geography* 47(1): 36–48.

IPEA, IBGE, and UNICAMP. 2002. *Caracterização e tendências da rede urbana do Brasil: Configuração atual e tendências da rede urbana*, volume 1. IPEA.

Jonasson, Erik. 2011. Informal employment and the role of regional governance. *Review of Development Economics* 15(3): 429–441.

LeRoy, Stephen F. and Jon Sonstelie. 1983. Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics* 13(1): 67–89.

Lucas, Robert E. and Esteban Rossi-Hansberg. 2002. On the internal structure of cities. *Econometrica* 70(4): 1445–1476.

Massey, Douglas S. and Nancy A. Denton. 1988. The dimensions of residential segregation. *Social Forces* 67(2): 281–315.

McDonald, John F. and Paul J. Prather. 1994. Suburban employment centres: The case of Chicago. *Urban Studies* 31(2): 201–218.

McMillen, Daniel P. 2001. Nonparametric employment subcenter identification. *Journal of Urban Economics* 50(3): 448–473.

Midelfart-Knarvik, Karen H., Henry G. Overman, Stephen Redding, and Anthony J. Venables. 2002. Integration and industrial specialisation in the European Union. *Revue Économique* 53(3): 469–481.

Mills, Edwin S. 1967. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review* 57(2): 197–210.

Monkkonen, Paavo and Xiaohu Zhang. 2014. Innovative measurement of spatial segregation: Comparative evidence from Hong Kong and San Francisco. *Regional Science and Urban Economics* 47: 99–111.

Moreno-Monroy, Ana Isabel and Héctor M Posada. 2014. On the effect of transport subsidies on informality rates. Technical report, CEPESP discussion paper, Fundação Getulio Vargas.

Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago, IL: University of Chicago Press.

OECD. 2018. Divided cities: Understanding intra-urban inequalities. Technical report, Organisation for Economic Co-operation and Development.

Pendall, Rolf and John I Carruthers. 2003. Does density exacerbate income segregation? Evidence from US metropolitan areas, 1980 to 2000. *Housing Policy Debate* 14(4): 541–589.

Pereira, Rafael H.M., Vanessa Nadalin, Leonardo Monasterio, and Pedro H.M. Albuquerque. 2013. Urban centrality: A simple index. *Geographical analysis* 45(1): 77–89.

Posada-Duque, Héctor M. 2015. *Informality in the City: a Theoretical Analysis*. Ph.D. thesis, Universidad del Rosario.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reardon, Sean F. 2011. Measures of income segregation. Stanford Center for Education Policy Analysis.

Reardon, Sean F and Kendra Bischoff. 2011. Income inequality and income segregation. *American Journal of Sociology* 116(4): 1092–1153.

Reardon, Sean F and David O'Sullivan. 2004. Measures of spatial segregation. *Sociological Methodology* 34(1): 121–162.

Ruiz-Rivera, Naxhelli and Paul van Lindert. 2016. Urban segregation in Latin America. *Habitat International* (54): 1–2.

Stock, Jame H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In Donald W.K. Andrews and James H. Stock (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

Telles, Edward E. 1995. Structural sources of socioeconomic segregation in Brazilian metropolitan areas. *American Journal of Sociology* : 1199–1223.

Tiebout, Charles M. 1956. A pure theory of local expenditures. *The Journal of Political Economy* : 416–424.

Villaça, Flávio. 1998. *Espaço intra-urbano no Brasil*. Studio Nobel.

Watson, Tara. 2009. Inequality and the measurement of residential segregation by income in American neighborhoods. *Review of Income and Wealth* 55(3): 820–844.

Wheeler, Christopher H. 2008. Urban decentralization and income inequality: Is sprawl associated with rising income segregation across neighborhoods? *Federal Reserve Bank of St. Louis Regional Economic Development* 4(1): 41–57.

Appendix A. Data sources and processing

Definition of urban agglomerations

Urban agglomerations (UAs) include metropolitan regions, non-metropolitan urban agglomerations (resulting from conurbation), and subregional urban centers. Most urban agglomerations extent beyond the boundaries of a single municipality[8], and may include peri-urban areas and small towns that fall under the influence of a proximate urban center. Besides the definition of 68 metropolitan regions, there is no official consistent definition of city boundaries from the Census. We use the grouping of municipalities by Da Mata *et al.* (2007), based on the definition of functional urban areas of IPEA *et al.* (2002), to define 121 UAs in 2000 and 2010. These add up to 171,393 enumeration areas in 2000 (out of which 158,307 are classified as urban) with 104'813,949 inhabitants and 176,056 areas in 2010 (out of which 158,307 are classified as urban) with 109'343.196 inhabitants.

Geoprocessing of census data

The Brazilian Institute of Geography and Statistics (IBGE) freely distributes the Census microdata, which can be found at `ftp://ftp.ibge.gov.br/Censos/`. It also distributes the digital networks containing the boundaries of the enumeration areas for 2000 and 2010. The original digital networks can be found at `ftp://geoftp.ibge.gov.br/malhas_digitais/`. We transformed them to SIRGAS 2000 projected data (UTM South hemisphere zone 24 projection). The geoprocessing was done in the R statistical environment (R Core Team, 2015) using the *maptools*, *sp*, *rgdal*, *rgeos* and *cleangeo* packages. Figures A.1 and A.2 illustrate the geographical coverage and layout.

Figure A.1: Urban agglomerations, 2010

---
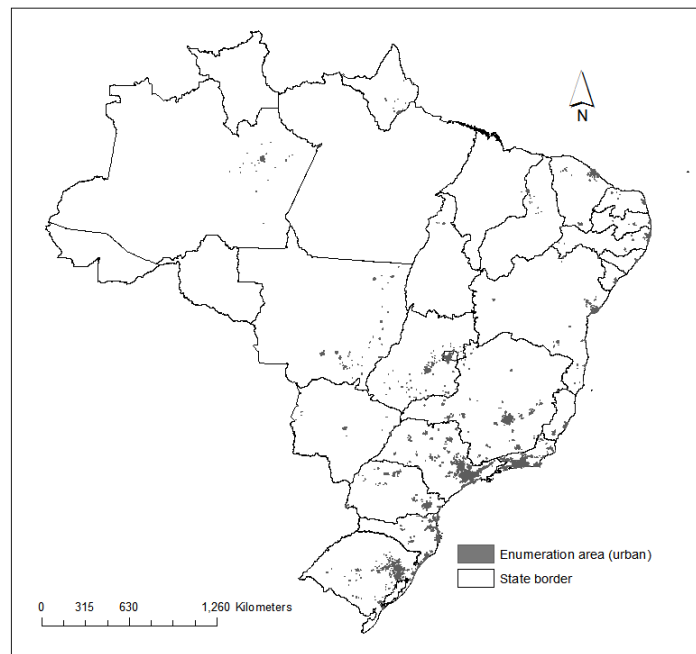
[8] For instance, the Metropolitan region of São Paulo extends over 39 different municipalities.

Figure A.2: Detail of urban agglomeration: Fortaleza, 2010
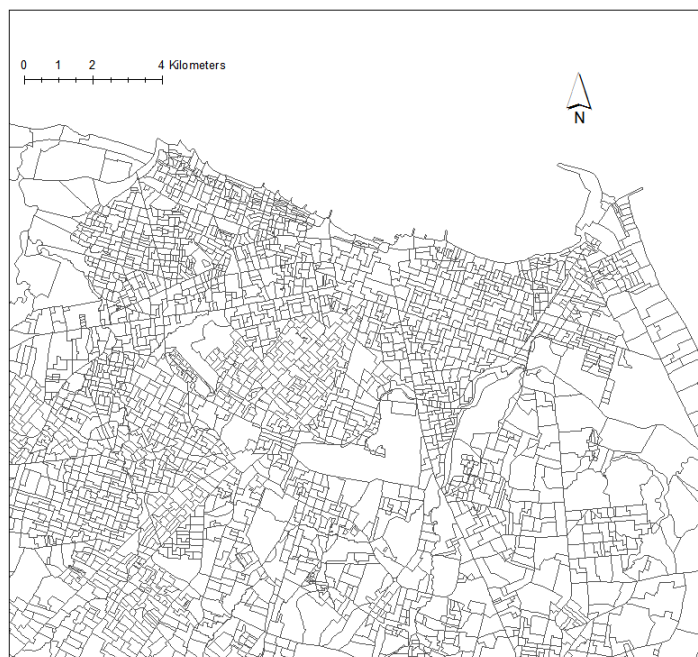
Geoprocessing of RAIS data

In order to geolocate firms, we find the best match of the postal address of each firm to a corresponding address and in this way allocate each firm to an enumeration area, where possible. We then sum the total number of employees in each enumeration area.

In more detail, we start by cleaning and fixing typos and errors, and creating standard denominations for street types, abbreviations, and the like in the original RAIS database. We then create a string for each firm containing the street name and number (i.e., we remove information related to apartment/house/office/floor numbers and other specifics). To match the addresses in the CNEFE database and these strings, we then use the *amatch* function from the package *stringdist* in R (R Core Team, 2015) using an OSA matching algorithm for optimal string alignment distance, with a tolerance of 30% difference with the total number of characters in each case.

After assigning one employee to firms with zero employees, we could geolocate 18'582,545 workers of a total of 22'163,453 employees in the original RAIS 2000 database located in municipalities belonging to 121 UAs (i.e., 83,8% of the total), and 26,733,918 workers in 121 UAs, out of 33'582,123 (79,4%) in 2009. The performance by UAs is between 75% and 90% in most individual cases, with the exception of Brasília, for which we could geo-locate approximately 50% of employees of the RAIS. This is due to exceptionally low quality in address data in both the RAIS and CNEFE databases.

Informality rate by UA

We use microdata from the census sample (*amostra*) to construct the informality rate by UA. To calculate the rate, we aggregate the number of informal workers in each UA, and divide it

by the total number of workers (i.e., the sum of formal and informal workers). A worker is classified as informal if he or she is an unregistered employee (*empregado sem carteira assinada*), or a self-employed individual not contributing to social security, or an employer not contributing to social security (Jonasson, 2011, Henley, Arabsheibani, and Carneiro, 2009). A formal worker, by contrast, is a registered employee (*empregado com carteira assinada*), or self-employed individual contributing to social security, or an employer contributing to social security. This definition corresponds to the 'no signed labor card' criteria of Henley *et al.* (2009).

Appendix B. A-spatial versus spatial segregation indices

The spatial rank-order segregation indices $\tilde{H}$ rely on surface-based smooth density approximations that allow adjusting for the spatial extend of local neighborhoods, instead of relying on *ad hoc* boundaries (Reardon and O'Sullivan, 2004, Hong, O'Sullivan, and Sadahiro, 2014). A plausible population density surface can be obtained using interpolation techniques. Basically, discrete enumeration-level data is converted to a population density surface (implicitly assuming that the population of the census sector is uniformly distributed inside the sector), and the true distribution is approximated with a Gaussian kernel density estimator. The value of the kernel bandwidth can be varied to reflect more ample neighborhood definitions. The larger the neighborhood definition, the lower the resulting segregation measure, because a larger geographical reach implies more heterogeneity.

Although part of the literature suggests the use of spatial measures over a-spatial ones, spatial measures reflect the assumptions made on interpolation, which may not best reflect the actual connectivity between places. For instance, enumeration areas are often defined based on existing natural and man-made barriers, such as major roads and rivers. For Brazilian cities, the average enumeration area falls within a radius of 100 to 500 meters, so in principle, a-spatial measures and spatial measures at this spatial range should be more or less equivalent. However, spatial measures in a way "blur" existing delimitations between areas that may correspond to actual barriers, and in this way are likely to under-estimate the actual level of separation between two places. On the other hand, it is possible that by taking into account that neighborhoods extend beyond administrative boundaries, spatial measures correct for a possible upward bias in a-spatial measures of segregation. To what extent these cases apply is unknown to the researcher. We use a-spatial measures to keep out results comparable to the rest of the literature that largely relies on a-spatial measures. The results using spatial measures for alternative neighborhood definitions are available upon request. Both spatial and a-spatial measures were calculated using the R package *seg* (Hong *et al.*, 2014, R Core Team, 2015).

Appendix C. Income profile estimation

To illustrate how we obtain the rank-order indices, Figure C.1 shows a fractional polynomial fitted to the values of the a-spatial rank-order HA index by income percentiles for three selected cities. Clearly, the level of segregation increases with income. In the three cases, the segregation of the

rich (HA(0.9)), that is, the value of the fitted line for $p = 0.9$, more than doubles the value of segregation of the poor (HA(0.1)).

Figure C.1: Income profile for selected cities, 2010
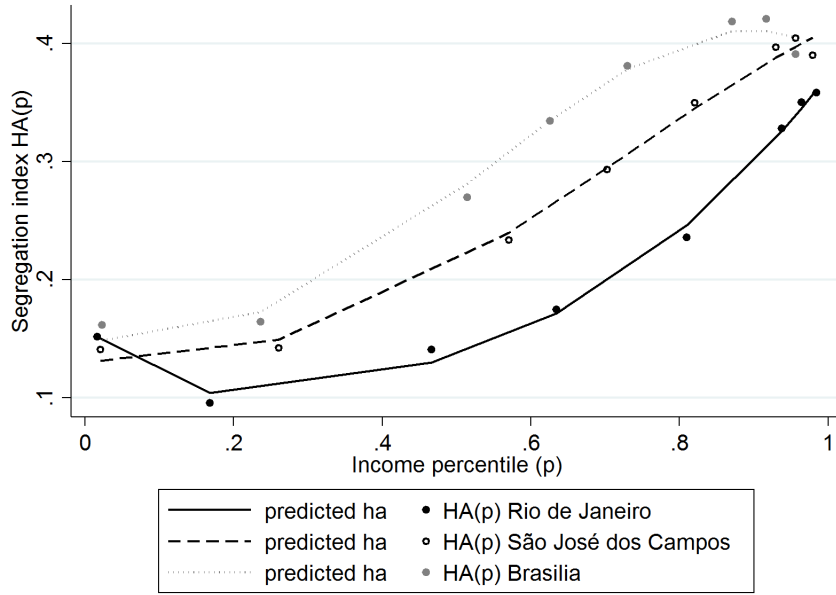
The objective is to summarize the income profile into a single city-level measure of income segregation. This is equivalent to finding the best fitting polynomial in a regression of the segregation index for each income bin (e.g. $HA(p)$) against the cumulative proportion of population in each income bin p in each city. Following Reardon (2011) and Reardon and Bischoff (2011), we estimate the function $H(p)$ in the following way. First, we calculate the pair-wise index H_k for those above and below each $k-1$ income threshold for each census sector. For each city, we then run the following WLS regression of the $k-1$ values of the segregation measures against p_k, the cumulative proportions of the population with incomes equal to or below k, and its M order polynomial terms:

$$H_k = \beta_0 + \beta_1 p_k + \beta_2 p_k^2 + ... + \beta_M p_k^M + \epsilon_k$$

with weights calculated as $E(p) = -[p \log_2 p + (1-p) \log_2(1-p)]$. The estimated city-level values of H, are then calculated as $\hat{H} = \Delta \hat{\mathbf{B}}$, where $\hat{\mathbf{B}}$ is the vector of estimated coefficients $\hat{\mathbf{B}} = (\beta_0, \beta_1, \beta_2 ... \beta_M)$ and Δ is a vector of scalars for the corresponding polynomial case, as detailed in Reardon (2011).[9]

---

[9] H can be then interpreted as a weighted mean of the pair-wise indices, where the weights are constructed so as to give more importance to observations in the middle of the income distribution (since this range is more informative about the segregation experienced by two randomly selected individuals) (Reardon, 2011).

Appendix D. Summary statistics

Table D.1: Summary statistics for frequency of income bins

	2000				2010			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
< 1/2 m.w.	0.011	0.011	0.001	0.063	0.035	0.031	0.004	0.148
1/2 − 1 m.w.	0.187	0.091	0.063	0.441	0.267	0.099	0.094	0.511
1 − 2 m.w.	0.209	0.041	0.124	0.276	0.314	0.045	0.201	0.417
2 − 3 m.w.	0.138	0.023	0.077	0.206	0.14	0.038	0.059	0.228
3 − 5 m.w.	0.172	0.04	0.079	0.258	0.118	0.031	0.051	0.19
5 − 10 m.w.	0.169	0.048	0.072	0.274	0.087	0.022	0.031	0.153
10 − 15 m.w.	0.045	0.013	0.017	0.078	0.016	0.006	0.003	0.045
15 − 20 m.w.	0.03	0.01	0.009	0.058	0.013	0.005	0.003	0.04
> 20 m.w.	0.038	0.016	0.007	0.1	0.01	0.005	0.002	0.043

Notes: 121 observations (cities) in 'All cities' sample. m.w. = minimum wage(s). < 1/2 m.w. category does not include heads of household with zero income.

Table D.2: Summary statistics for control variables

	All cities				Monocentric cities				Polycentric cities			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Informality and Inequality												
2000 Informality (%)	0.49	0.10	0.26	0.76	0.49	0.11	0.26	0.76	0.48	0.09	0.32	0.67
2010 Informality (%)	0.34	0.09	0.18	0.62	0.35	0.09	0.18	0.62	0.33	0.08	0.18	0.48
1990 Gini	0.55	0.04	0.43	0.65	0.56	0.05	0.43	0.65	0.55	0.04	0.48	0.63
2000 Gini	0.57	0.04	0.48	0.66	0.57	0.03	0.50	0.63	0.58	0.04	0.48	0.66
1990 Income cap. (2000 R$)	246	79	99	463	233	75	99	423	280	80	140	463
2000 Income cap. (2000 R$)	320	94	129	520	307	87	129	489	338	100	146	520
Geography (in 2000)												
Total area (km^2)	10,949	42,376	305	413,516	4,842	10,316	305	70,839	26,578	76,828	586	413,516
Dummy for coast	0.17	0.37	0	1	0.09	0.29	0	1	0.35	0.49	0	1
Dummy for semi-arid	0.10	0.30	0	1	0.10	0.31	0	1	0.09	0.29	0	1
Pop under land zone law (%)	0.77	0.34	0.00	1.00	0.73	0.38	0.00	1.00	0.89	0.10	0.64	1.00
Demography												
1990 Pop > 55 yo (%)	0.10	0.02	0.05	0.15	0.11	0.02	0.05	0.15	0.09	0.02	0.06	0.13
2000 Pop > 55 yo (%)	0.12	0.02	0.06	0.17	0.13	0.02	0.07	0.17	0.11	0.02	0.06	0.15
1990 Pop < 25 yo (%)	0.68	0.05	0.56	0.85	0.67	0.05	0.56	0.85	0.70	0.06	0.56	0.82
2000 Pop < 25 yo (%)	0.48	0.04	0.42	0.61	0.47	0.04	0.42	0.57	0.49	0.04	0.43	0.61
1990 Migrants (%)	0.43	0.11	0.15	0.68	0.41	0.11	0.15	0.65	0.48	0.09	0.26	0.68
2000 Migrants (%)	0.43	0.10	0.16	0.64	0.42	0.10	0.23	0.62	0.45	0.09	0.16	0.64
Industrial composition												
1990 Manufacturing (%)	0.27	0.08	0.14	0.54	0.27	0.09	0.14	0.54	0.27	0.08	0.15	0.50
2000 Manufacturing (%)	0.24	0.07	0.12	0.48	0.24	0.07	0.12	0.48	0.24	0.07	0.14	0.47
1990 Services (%)	0.59	0.08	0.37	0.74	0.57	0.07	0.37	0.73	0.63	0.07	0.43	0.74
2000 Services (%)	0.65	0.07	0.42	0.80	0.63	0.06	0.42	0.80	0.68	0.07	0.49	0.79

Notes: 121 observations (cities) in 'All cities' sample. 87 and 72 monocentric cities, and 34 and 49 polycentric cities in 2000 and 2010, respectively.

Appendix E. Robustness checks

Results for 2010 HA computed with households information

Table E.1: The effect of urban spatial structure on income segregation: Alternative 2010 HA

Dependent variable:	ln(HA index)				
Percentile:		10th	20th	80th	90th
Method:	TSLS	TSLS	TSLS	TSLS	TSLS
	[1]	[2]	[3]	[4]	[5]
ln(1-km employment density)	0.568^a	0.372^a	0.521^a	0.612^a	0.537^a
	(0.145)	(0.114)	(0.140)	(0.160)	(0.148)
First-stage F-statistic	16.65	16.65	16.65	16.65	16.65
Instrument		ln(t-30 years employment density)			
Inequality & Inequality	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y
Demography	Y	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y	Y
Observations	121	121	121	121	121

Notes: Robust standard errors clustered by UA are in parenthesis. ^{a}, ^{b}, and ^{c} indicates significant at 1, 5, and 10 percent level, respectively.

Results for alternative definitions of the poor and the rich

Table E.2: The effect of urban spatial structure on income segregation: Income 20% vs. 80%

Dependent variable:	ln(HA index)					
Population:	All	<100,000	100,000-700,000			>700,000
Urban spatial structure:	Both	Mono	Both	Mono	Poly	Poly
Years:	00-10	00-10	00-10	00-10	00-10	00-10
Method:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A: Poor (20th percentile)						
ln(1-km employment density)	0.529^a	-0.023	0.824^b	1.224^a	0.643^a	-0.010
	(0.158)	(0.413)	(0.331)	(0.378)	(0.213)	(0.111)
ln(Urban centrality index)	-0.110^b	0.174	-0.154^b	-0.283^a	0.131^c	0.123^a
	(0.052)	(0.180)	(0.072)	(0.092)	(0.073)	(0.046)
Panel B: Rich (80th percentile)						
ln(1-km employment density)	0.468^a	0.371	0.646^a	0.786^a	0.338^a	-0.089^a
	(0.086)	(0.259)	(0.145)	(0.166)	(0.109)	(0.025)
ln(Urban centrality index)	-0.020	0.054	-0.030	-0.081	0.156^a	0.097^a
	(0.022)	(0.139)	(0.035)	(0.054)	(0.040)	(0.010)
First-stage F-statistic	64.40	8.20	17.00	13.47	4.51	94.38
Instrument			ln(t-30 years employment density)			
Inequality & Inequality	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y
Demography	Y	Y	Y	Y	Y	Y
Industrial composition	Y	Y	Y	Y	Y	Y
Observations	242	41	159	118	41	42

Notes: All regressions include state and year fixed-effects. Robust standard errors clustered by UA are in parenthesis. ^{a}, ^{b}, and ^{c} indicates significant at 1, 5, and 10 percent level, respectively.