Residential Patterns and Local Public Goods in Urban Brazil*

Mariaflavia Harari†
University of Pennsylvania

This version: November 4, 2024
Please download latest version here.

Abstract

Millions of migrants in developing countries move to urban areas in search of better prospects, but access to public services varies widely within cities. Yet, we know little about spatial inequalities within cities in low- and middle-income countries. This paper investigates the spatial distribution of socio-economic status and public goods access within Brazilian cities, using high-resolution Census maps. I consider spatial metrics of "distance segregation", capturing the physical proximity between neighborhoods of different socio-economic status. I document descriptive patterns of segregation by income, race, and informality and disparities in access to public goods within cities. To make progress on the identification of the impacts of residential patterns on public goods provision, I develop an instrumental variables strategy that leverages within-city geography to predict where the poor and rich live. I find that cities with greater distance between rich and poor have fewer households connected to sewerage and water, worse neighborhood quality, and lower access to public amenities. Leveraging spatial variation in public goods provision within cities, I consider mechanisms that shape the allocation of urban services, including externalities across neighborhoods and preferences over public goods provision. These findings help inform the debate on policies such as slum clearance and relocations, social housing, and the spatial targeting of public goods.

JEL Classifications: O180, R120, H41

*I gratefully acknowledge the support of the Research Sponsors Program of the Zell/Lurie Real Estate Center and of Leonard J. Horwitz. I thank Marcella Alsan for extensive discussions on an earlier version of this project. I also thank Santosh Anagol, Gilles Duranton, Jessie Handbury, Fernando Ferreira, Thomas Fujiwara, Isabela Manelici, Paul Novosad, Vinicios Sant'Anna, Nick Tsivanidis, Shing-Yi Wang, Laura Weiwu, Maisy Wong, and seminar participants in the Wharton Real Estate Department, 13th European Meeting of the Urban Economics Association, 7th Workshop on Urban Economics at IEB – Universitat de Barcelona, NBERSI, PDRI, OSUS, Toronto, Warwick, UCL, UCLA, UC Irvine, USC Lusk, the Harvard Cities and Development Conference, and NEUDC for helpful comments and suggestions. Fahad Algifari, Heidi Artigue, Zach Babat, Joshua Boyd, Minahil Fatima, Muxin Li, Thiago Patto, Rafael Pucci, Fabio Schanaider, Nikhil Shankar, and Jorge Tello provided outstanding research assistance. All errors are my own.

†Wharton Real Estate. 428 Dinan Hall, 3733 Spruce Street, Philadelphia, PA 19104. Email: harari@wharton.upenn.edu.

1 Introduction

Developing country cities are experiencing massive growth[1], attracting millions of impoverished migrants every year in search of economic opportunity. One of the key challenges for policy makers is how to accommodate residents and provide public services in an inclusive way. As urban areas expand into different spatial configurations (Harari, 2020), physical access to amenities, jobs, and services can vary widely within and across cities, particularly among economically disadvantaged groups: in many cities, the poor concentrate in peripheral, low-accessibility areas, while in others poorer neighborhoods coexist with affluent ones in central areas (Gadgil and Baker, 2017).

The spatial settlement patterns of poorer and richer households underscore important policy tradeoffs related to the allocation of public services within cities. Proximity to central areas is valuable for the urban poor (Barnhardt et al., 2017) and providing basic services in these areas may be more cost-effective (Goksu et al., 2019). At the same time, policy makers are often concerned about negative externalities and opportunity costs associated with poor neighborhoods located in prime areas (Henderson et al., 2020).

Understanding where different socio-economic groups live is a necessary precursor to grappling with these tradeoffs and making consequential public investment decisions in areas such as sanitation, transportation, and public housing. Yet, systematic evidence on residential patterns in developing country cities is limited, largely due to a lack of representative, granular data for urban areas (Asher et al., 2024). Even when socio-demographic data is available at the neighborhood level, it is typically not geocoded, so researchers cannot observe where different socio-economic groups reside and their proximity to amenities and services.

This paper investigates the spatial distribution of socio-economic status and public goods access within Brazilian cities, using high-resolution Census maps. Brazil is an important setting: it is the world's fourth largest country by urban population (U.N., 2018) and features stark income inequality and persistent racial disparities. Brazilian cities also grapple with persistent deficiencies in the provision of public services and adequate housing, with 16 million people living in *favelas* (IBGE, 2020).

There are three main contributions. First, I characterize residential patterns and access to local public goods by socio-economic status at a granular level in a large, middle-income urban economy. My analysis focuses on income, race, and housing informality (that is, living in a slum) as three key dimensions of socio-economic inequality. I consider neighborhood goods such as sanitation, water, and sidewalks (from the Census) and proximity to public amenities that can be accessed across neighborhoods, such as post offices and park (from OpenStreetMap). With these

[1]The United Nations projects that by 2050, approximately 2.5 billion individuals will become urban residents, with 95% of this growth anticipated to occur in developing regions (UN Habitat, 2012).

rich data in hand, I document within-city inequality in access to public goods across space and across groups.

The second contribution is to shed light on an overlooked dimension of residential segregation, emphasizing physical distance between groups. In addition to affecting the degree of interaction between groups, distance mediates access to public services and amenities, determines the strength of externalities between neighborhoods, and impacts the cost of delivering many public services. Yet, the economics literature for both developed and developing countries has historically focused on non-spatial residential segregation indexes, which consider the proportion of groups within neighborhoods abstracting from their location. I consider spatial metrics of "distance segregation" that capture the extent to which poor and rich locate close to each other. Intuitively, a "distance-integrated" city is one where rich and poor neighborhoods are interspersed with each other in a "checkerboard" pattern, while a "distance-segregated" city is one where the poor tend to reside far away from the rich.

Third, I propose an instrumental variables strategy to make progress on the causal identification of the impacts of residential patterns on local public goods provision. I leverage variation in topographic features within cities to predict distance segregation, alleviating concerns associated with the endogeneity of residential patterns. Both OLS and IV analyses show that more distance-segregated cities are associated with lower levels of access to public goods.

My analysis considers a sample of 600 urban municipalities, using data and maps from the 2010 universal Census, at the block (*setor censitario*) level, a unit smaller than U.S. census block groups. Municipalities are the local administrative units responsible for the management and allocation of many important urban services, including water and sanitation. The segregation metric I employ is based on the average Euclidean distance between high- and low-socio-economic status neighborhoods. I consider various approaches to define neighborhoods, based on absolute and relative average household income, share of non-white residents, and whether the block is part of an *aglomerado subnormal* or slum.

The descriptive analysis establishes a number of new stylized facts. Across cities, there is significant variation in the degree of distance segregation. Most cities appear distance segregated by income, in the sense that the distance between blocks in the upper quartile in the city by average income and blocks in the bottom quartile is greater than the average distance between any two blocks. For the median sized-city, one standard deviation in distance is about 1.7 km. These patterns also hold using absolute income indicators (e.g. share of households with income below the minimum wage): greater absolute income disparities across neighborhoods are correlated with greater distances. There is also distance segregation by race, contrasting predominantly white with predominantly non-white blocks, but it appears less marked. Interestingly, turning to informality, slums appear to be more distance integrated with wealthy neighborhoods than poor (but formal) neighborhoods, consistent with the intuition that poor households locating in slums are trading off housing quality and tenure security with proximity to employment or public goods (Gadgil and Baker, 2017). City-level distance segregation displays a weak, positive correlation with the most commonly used metrics for inequality and segregation (the Gini and dissimilarity index, respectively).

Within cities, I document significant disparities in access to public services as a function of distance segregation. Public goods provision is highest in rich neighborhoods. Among poor neighborhoods, access declines monotonically with distance to rich neighborhoods. Distance-segregated poor blocks tend to have a lower share of households with access to public sewerage, water, paved roads, sidewalks, and a range of other neighborhood goods. They are also further away from amenities such as parks, post offices, police stations, and fire stations. These patterns hold conditional on own neighborhood composition, distance to the Central Business District (CBD), and predetermined geographic characteristics.

Motivated by the stylized facts above, I turn to the relationship between city-wide segregation and public goods provision. OLS regressions alone can be difficult to interpret due to many sources of simultaneity. For example, rich households may be sorting further away from poor ones in response to low public goods provision. At the same time, the proximity between poor and rich may be the byproduct of land market frictions, poor planning, or weak rule of law (Henderson et al., 2020), which could correlate with inefficient public goods provision.

To make progress on the identification, I develop an instrumental variables approach that leverages variation from the location of geographic features within cities. I exploit the fact that one of the determinants of residential sorting is the location of exogenous amenities (Lee and Lin, 2018). I find that steep slopes and proximity to riverbanks predict the location of poor neighborhoods, while rich neighborhoods tend to locate in areas with mild slopes and near the coast. I then reconstruct the distance segregation indexes using the location of geography-predicted poor neighborhoods instead of actual ones. The ensuing city-level instrument is based on the relative position of areas with desirable and undesirable combinations of slope and water bodies: when these areas are further away from each other, the city is more susceptible to a spatially segregated settlement pattern.

The IV analysis shows that more distance-segregated cities are associated with lower average access to basic urban services. OLS estimates are similar or attenuated in magnitude. For a one standard deviation increase in income segregation, corresponding for the median sized city to 1.7 additional km between the average poor and rich neighborhood, the share of households with access to public sewerage and water decreases by 3.4 and 3 percentage points (p.p.) respectively, corresponding to 6 and 3% of the sample mean. I find similar patterns when considering distance segregation by informal status and by race and other measures of local public goods.

I conduct a number of robustness tests addressing the concern that the IV results may be driven by higher costs of providing public goods in cities with more complex topography. Importantly, the results survive a specification in which the outcomes have been residualized by block-level topography and distance to the CBD. Additionally, the results are robust to a range of sample cuts.

These patterns are also robust to considering variations of the distance segregation metrics. For example, at baseline I consider distance between blocks defined as "poor" or "rich" according to a binary indicator, but the results are similar considering a continuous, weighted version which includes all blocks and weights the distances by the block-level shares of poor and rich residents. Furthermore, I also consider a version of the index that is calculated as a market access measure, with exposure to the rich weighted by inverse distance, following the quantitative spatial modeling literature on gentrification (e.g., Gechter and Tsivanidis (2023)). Interestingly, the patterns of lower public goods provision in more segregated cities cannot be replicated using the widely used dissimilarity index, which measures segregation in a non-spatial way.

Why is the provision of basic public goods lower in more segregated cities? To shed light on mechanisms, I consider how the spatial pattern of access across neighborhoods and groups varies between segregated and integrated cities. First, I find that segregated cities tend to provide lower public goods access to both poor and rich neighborhoods, with relatively more under-provision to the poor. This suggests that residents in segregated cities may have an overall lower propensity towards allocating municipal resources to basic public goods, particularly in poor areas. One potential channel is that preferences are shaped by a lack of exposure to poor neighborhoods due to distance, echoing evidence from the political economy and behavioral literature on how redistributive preferences are affected by social interactions (e.g., Alesina et al. (2018), Londoño-Vélez (2022)). At the same time there could be sorting of residents with these types of preferences in more segregation-prone cities. It is inherently difficult to disentangle the two channels given my sources of variation, but I find limited evidence of sorting by demographic characteristics: distance segregation does not appear to affect city-wide racial shares, education levels, and the share of prime-age males (more likely to be migrants).

Another potential explanation is related to the engineering costs of expanding the network for urban services such as sewerage or water connections, with lower costs associated to servicing the poor who reside nearby. While this is plausibly occurring within cities, the engineering channel alone cannot explain the overall cross-city pattern: I find that even poor neighborhoods that are located near rich ones appear under-served in segregated cities relative to integrated ones.

Finally, I consider the relative provision of public goods in segregated versus integrated poor neighborhoods. There are two competing mechanisms potentially at play. On the one hand, there may be fewer incentives to provide public goods in poor areas that are far from wealthy ones, since negative spillovers from those areas are likely limited (Xu, 2023). On the other hand, there could be strategic under-provision of neighborhood goods in poor areas that are near rich ones in an attempt to deter the poor from settling (Feler and Henderson, 2011). For public amenities, there is some evidence that the former mechanism could be stronger in segregated cities.

Taken together, this paper highlights previously undocumented ramifications of the spatial configuration of cities for the delivery of public services, pointing to the challenges of providing inclusive public services in spatially segregated urban areas. These findings have implications for the many contentious urban policies that affect where the urban poor live: slum upgrading on site versus slum clearance and relocations (Harari and Wong, forthcoming, Rojas-Ampuero and Carrera, 2023), public housing (Belchior et al., 2024), urban renewal (Gechter and Tsivanidis, 2023), and transit (Tsivanidis, forthcoming, Khanna et al., 2024, Gonzalez-Navarro and Quintana-Domeque, 2016). Beyond Brazil, these findings also suggest that incorporating space in the measurement of segregation provides useful insight into the functioning of cities and the economic livelihoods of disadvantaged groups. The descriptive analysis documents disparities in access within cities that would not be appreciated using conventional non-spatial metrics and that are useful inputs to policy makers considering how to spatially roll out public services. My approach also demonstrates how sources such as Open Street Map can be combined with surveys and maps to provide a more comprehensive picture of access within cities (Harari et al., 2024).

This paper contributes to a nascent literature on within-city residential patterns in developing countries (Deffebach et al., 2024).[2] The impact of neighborhoods on human capital and labor market access has been recently highlighted in Rojas-Ampuero and Carrera (2023) and Belchior et al. (2024). Recently, using uniquely granular data for India, Asher et al. (2024) have considered segregation by caste and religion, demonstrating lower public goods access in neighborhoods that are predominantly Muslim and low-caste.[3] My paper complements their analyses, with two differences. The first is in the measurement of segregation: while they employ the widely used dissimilarity and isolation indexes, which are based on relative group shares in each neighborhood, I measure segregation from a spatial angle leveraging the detailed maps provided by the Brazilian Census, and include measures of access to public goods outside the own neighborhood. The second difference is that my paper complements the descriptive analyses with an instrumental variables approach, making progress towards establishing causal evidence of how residential patterns affect public goods provision.

This paper is also connected to the literature on the provision of local public goods in de-

[2]In the U.S., a large literature has investigated residential patterns and documented the impact of place on socioeconomic outcomes. Prior research has focused on sorting by income (Lee and Lin, 2018, Couture et al., 2024) and segregation by race (see Boustan (2011) for an overview).

[3]In Brazil, recent descriptive work using block-level data from the 2010 Census includes Garcia-López and Moreno-Monroy (2018), Brueckner et al. (2019), and da Rocha Valente and Berry (2022).

veloping countries.[4] In Brazil, Feler and Henderson (2011) provide evidence of strategic under-provision of public goods to deter in-migration of poor households across Brazilian municipalities and Xu (2023) provides survey evidence from São Paulo showing that wealthy households living in proximity to slums are more likely to support the provision of "externalities correcting" public goods. The results in this paper echo both mechanisms linking public goods provision to residential patterns.

This paper is organized as follows. Section 2 discusses the background and data sources. Section 3 introduces the distance segregation metrics. Section 4 presents descriptive evidence within and across cities. Section 5 discusses the instrumental variables strategy and Section 6 presents the main city-level results. Section 7 concludes.

2 Background and Data

Brazil is an important setting for studying spatial inequalities within cities, being a large, middle-income, predominantly urban country facing extreme socio-economic disparities. Ranking 4th in the world by urban population, Brazil has over 250 cities above 100 thousand inhabitants (IBGE, 2010) and overall nearly 200 million urban residents (IBGE, 2020). Strikingly, it is the ninth most income-unequal country in the world with a Gini index of 0.53 (World Bank, 2024). It faces enduring racial inequality, with mixed-race and black Brazilians (approximately 58% of the total population) systematically lagging behind whites in a number of socio-economic indicators; for example, non-whites are approximately twice as likely to be below the poverty line than whites (IBGE, 2022). Additionally, it faces persistent informality with 16 million estimated slum residents (IBGE, 2020). The availability of basic urban services remains inadequate in many urban areas, particularly in the domain of sanitation, with only 51% of the total urban population having access to safely managed sanitation. While this figure has been increasing since 2010, an eight-times faster progress is required to reach the Sustainable Development Goal pertaining to sanitation by 2030 (SWA, 2022).

Uniquely for a developing country, the Brazilian Census provides both data and maps for very granular spatial units, allowing for a rich characterization of residential patterns across space. The primary source for this paper is the 2010 complete Census. Specifically, I employ the data released at the *setor* (census block) level. Block boundaries tend to coincide with those of neighborhoods (Belchior et al., 2024). The median block in my sample contains around 200 households.[5]

[4] In the U.S. context, previous work in political economy and local public finance has examined the link between municipal spending patterns and political polarization (Trounstine, 2016), income inequality (Boustan et al., 2013), and diversity (Alesina et al., 1999). In a historical context, Troesken (2001) and Beach et al. (2022) examine service provision and race segregation.

[5] In terms of number of households, *setores censitarios* are smaller than U.S. census block groups and about twice as

Throughout the paper I will interchangeably refer to neighborhoods or blocks. Below, I detail my data sources and provide some context for the outcomes I consider.

2.1 Local public goods

Brazil is a federal, decentralized country in which municipalities (*municípios*) have a crucial role in providing basic public goods and services, particularly in the domains of health and infrastructure. I focus on public goods that are primarily managed at the local level by municipalities, following the literature on local public finance in Brazil (e.g. Feler and Henderson (2011)). Municipalities manage the local provision of urban services, including deciding which neighborhoods to serve (Kresch, 2020). The majority of municipal funding comes from intergovernmental transfers, either from the federal or state governments, that are primarily based on municipality population size, with a smaller component (on average, 10% of the total revenue) from local property taxes and other local revenue (Alves and Araujo, 2024).

I begin by considering neighborhood public goods from the 2010 Census block level-data. Among my primary outcomes, I consider the share of residents with access to the public sewerage and water networks. These are critical urban services: sanitation generates a large return to public spending by improving health and welfare and by producing positive externalities (Kresch et al., 2023). Approximately half of the urban population in Brazil lacks access to a public sewerage network. Access to a public water system is more prevalent, although not universal, with about 90% of urban households being connected.

I also consider neighborhood characteristics from the *entorno* Census module, where enumerators record the characteristics of a household's surroundings. These include the share of residents whose immediate vicinity has paved streets, sidewalks, no accumulated street garbage, no open sewer, street addresses, street lighting, curbs, manholes, ramps, and greenery. For many of the analyses, I use z-scores to aggregate these characteristics into a standardized neighborhood public goods index centered on 0 and expressed in standard deviations, with higher values corresponding to better neighborhood quality.

In addition to the Census measures, I code the location of a number of public amenities from OpenStreetMap (OSM). Specifically, I consider fire stations, police stations, post offices, and parks. These particular amenities were chosen because they appear to have comprehensive coverage in OSM (e.g. by validating parks from satellite imagery), attenuating concerns of selection in reporting associated with open-source data. For my main analysis, I construct a summary index by coding dummies for whether there is at least one of those amenities within 3 km of each block and averaging across the four types of amenities. I consider other measures - such as dis-

large as the neighborhoods considered for urban India by Asher et al. (2024).

tance to the nearest amenity - in the Appendix. Unlike the Census-based measures, which capture neighborhood-level public goods that are only valuable to those residing on site, the OSM data capture "super-neighborhood" public goods (Fujita, 1989) that residents from different neighborhoods can travel to.

All of the metrics above are calculated at the block level. For the cross-city analyses, they are averaged at the city level across all blocks (in the main analysis) or a subset of blocks (for the analysis of mechanisms).

2.2 Neighborhood characteristics

I consider segregation along three dimensions of socio-economic status: income, race, and housing informality (slum versus formal neighborhoods). For income, I classify neighborhoods as "rich" and "poor" based on the average household income in each block as reported by the Census. The baseline definition is based on whether a block ranks in the bottom or top quartile in the city by average income. I also show some descriptive results using absolute income definitions defined in reference to the minimum wage.

One of the advantages of employing a relative definition of income is that it ensures that poor and rich neighborhoods can be identified in all 608 cities in the full sample. Choosing an absolute income threshold for all cities can be problematic given that cities vary in their income distributions and some may not have any rich neighborhoods under some of the absolute definitions. However, the drawback of this approach is that the poor and rich are not directly comparable across cities.

I also explore segregation by race, where a block is defined as "predominantly non-white" if its population is more than 50% non-white.[6] It is important to consider race in its own right, to inform the debate on racial inequality in Brazil. As a marker of socio-economic status, race may be preferable to income because it is comparable across cities and tied to fixed individual characteristics (subject to the caveat of endogenous self-reporting (Francis and Tannuri-Pianto, 2013)). However, not all cities have meaningful variation in racial composition and the ensuing sample is reduced to 458 cities.

The third dimension is whether a block is part of what the 2010 Census defines as an *aglomerado subnormal*, which I refer to interchangeably as "slum", following the urban literature. The Census designation is based on a number of criteria, including (i) illegal occupation of land and (ii) at least one characteristic among presence of narrow and irregular roads, irregular buildings, and precarious basic public services. Because only the largest cities have reported slums, the sample size drops to 229 municipalities when considering this dimension.

[6] I do not consider finer race disaggregations noting that self-reported "mixed/brown" and black individuals are often considered together in Brazilian affirmative action policies.

Empirically, the three definitions are correlated. After presenting the descriptives for all three, I focus on income as the primary definition, but show that the results are qualitatively consistent using all three dimensions.

2.3 Definition of city

While municipalities are the natural unit of observation when considering local public service provision, municipal boundaries do not always coincide with the boundaries of actual cities. Often, the administrative boundaries grossly overestimate the geographic extent of the actual urban areas, by comprising large and largely uninhabited blocks. Including these areas would complicate the interpretation of measures that are based on physical distance.

As baseline units for the computation of my indexes I consider a sample of "cities" constructed as follows. First, I consider municipalities with more than 50,000 residents and a share of residents classified as "urban" by the Census greater than 50%. Second, within each municipality, I exclude blocks that fall in the bottom quartile of the country for population density. This removes any large and mostly uninhabited blocks outside urban areas. Third, based on visual inspection, I exclude any blocks that, while dense enough to meet the prior threshold, appear to be clearly disconnected from the core urban area in a municipality (e.g. because they belong to a bordering one).[7] To visually validate this procedure, I verify that the resulting spatial units largely overlap with the built-up footprint of cities as they appear from satellite imagery (e.g. the Global Human Settlement Layer (Commission et al., 2015)) and have blocks of relatively uniform area size. An example of the procedure for the city of Ourinhos (SP) is provided below. Throughout the paper I refer to these units as "cities". The final, full sample includes 608 cities with median number of residents of 93,000.

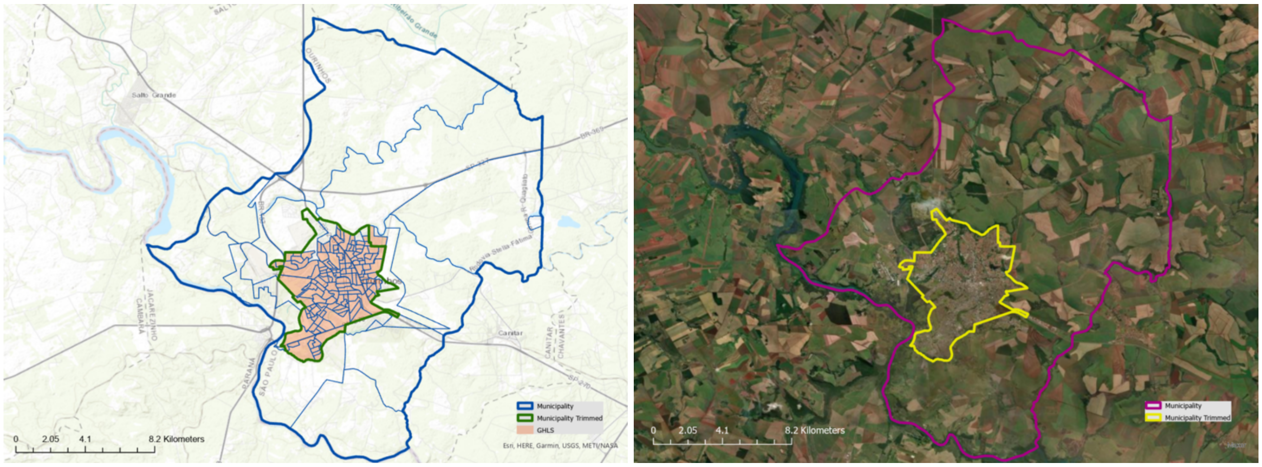
Another reason why the mapping between cities and municipalities is not one-to-one is that there are large urban agglomerations spanning across multiple municipalities. Because the public goods considered are mostly municipal, municipalities remain the unit of observation of choice. To account for correlated shocks within urban areas spanning multiple municipalities, standard errors are clustered at the meso-region (a Census unit in between municipality and state) throughout the analysis. I also verify that the baseline results are robust to controlling for whether a municipality is the suburb of a large agglomeration.[8]

For some of the analyses below, I consider distance to the Central Business District (CBD). I define it as the historical center of cities and identify it using a semi-automated procedure. I search

[7] I do not force blocks to be contiguous. Results are robust to controlling for disconnected cities that do not have a completely contiguous footprint as a result of these data preparation steps.

[8] The political economy implications of multiple jurisdictions within the same functional urban areas for the delivery of public goods are explored in Feler and Henderson (2011). The ramifications for distance segregation are a potential avenue for future research.

Figure 1: Example of delineation of cities



Notes: The map and satellite imagery show the city of Ourinhos (SP). In the left panel, the thin boundaries represent blocks. The thick blue boundary is the municipality boundary. The thick green boundary represents the boundary of the "trimmed municipality" used for the analysis. The shaded pink areas are built-up areas according to the satellite-derived GHLS dataset. The right panel shows boundaries for the municipality (in purple) and municipality trimmed (in yellow) overlaid with satellite imagery.

the keywords "centro historico" followed by the name of each municipality and state in Google Maps and then validate the resulting location by visually inspecting the map. Locations where the Google-returned location is implausible (e.g., outside the built-up area) are manually rectified by finding the location of prominent landmarks (e.g. city hall, main square). Results are robust to alternative definitions such as the centroid of the most lit up pixels as of 1998 in the DSM-OLS night lights dataset (NOAA National Geophysical Data Center, 2013).

2.4 Geography

Several block-level and city-level geographic variables are employed for the construction of the instrument and/or included as controls in the regressions. The latter are listed in Table A3. Slope, ruggedness, and elevation are calculated from the Topodata dataset (Valeriano and Rossetti, 2012). The raw source is Shuttle Radar Topography Mission (SRTM) data at the 90 x 90 resolution. The distance between blocks and water bodies is calculated using rivers, streams, and lakes from OSM. The extent of water bodies is drawn from the ASTER dataset, with resolution of 30 meters (NASA/METI). Soil characteristics and climate variables are included among the controls following Naritomi et al. (2012) and obtained from (NOAA, 2024) and (EMBRAPA, 2020), respectively.

3 Indexes of distance segregation

In this Section, I discuss the notion of segregation used in this paper. Anecdotally, many developing country cities display a pattern whereby wealthy neighborhoods are interspersed by pockets of urban poverty. This contrasts with the patterns observed in other cities, where poor neighborhoods tend to be peripheral and spatially isolated. Intuitively, these are the spatial patterns of segregation that this paper focuses on. There are several reasons why the distance between rich and poor neighborhoods is an interesting object of study. Physical distance affects the likelihood that two groups are exposed to each other (Athey et al., 2021), which is the aspect emphasized by conventional segregation indexes. Additionally, distance determines access to amenities and jobs (Belchior et al., 2024, Zenou, 2013). Moreover, there are externalities across neighborhoods that decay in space, including positive externalities from rich neighborhoods (e.g. from access to formal amenities (Gechter and Tsivanidis, 2023)) or negative congestion externalities from poor neighborhoods (e.g. crime, health (Khanna et al., 2024)). Finally, many public goods are delivered along contiguous spatial networks, where distance is an important determinant of cost.

This particular notion of segregation has received little attention in the economics literature, which has typically employed non-spatial segregation indexes. The most common, the dissimilarity index (Duncan and Duncan, 1955), focuses on the internal composition of neighborhoods without incorporating distance. In this paper, I quantify the patterns of proximity between neighborhoods using metrics of "distance segregation" in the spirit of White (1983)'s proximity index. Below, I describe my baseline metric, provide some examples, and I discuss how it differs from other indexes. Section 6.2.2 empirically explores variations of the index and comparisons with the dissimilarity index.

3.1 Definition

Consider $i \in \{1, ..., N\}$ neighborhoods within city c. Neighborhoods can be of two types, P ("poor") and R ("rich"). Define S_g to be the set of neighborhoods that are of type g. Let d_{ij} be a distance function between neighborhoods i and j. The distance segregation index between groups P and R for city c is the average distance between P and R neighborhoods, computed as follows:

$$D_c^{PR} \equiv \frac{\sum_i D_i^R \cdot \mathbf{1}_{i \in S_P}}{\sum_i \mathbf{1}_{i \in S_P}}$$

where

$$D_i^R \equiv \frac{\sum_j d_{ij} \cdot \mathbf{1}_{j \in S_R}}{\sum_j \mathbf{1}_{j \in S_R}}.$$

The computation involves two steps. First, for each neighborhood i, the average distance to R neighborhoods is computed (D_i^R). Second, this measure is averaged across all P neighborhoods in the city c, obtaining D_c^{PR}.

At baseline, I consider straight-line Euclidean distance between the centroids of two neighborhoods, abstracting from infrastructure and travel time.

The resulting indicator is in distance units, such as km, and is mechanically correlated with a city's land area. Several normalizations are possible. For the purposes of producing absolute rankings of cities of different area sizes and shapes, in some of the descriptive analyses below I normalize the index by the average distance between any two neighborhoods in a city.[9]

The index normalized in this manner takes a value of 1 when the average distance between P and R neighborhoods is the same as the average distance of any two neighborhoods, which would occur in a "checkerboard"-like city. Conditional on a city's shape and size, if neighborhoods were randomly tagged as P or R the index would be close to 1. The index takes values greater than one in a city where P and R neighborhoods are further away from each other than two random ones ("distance segregation"). Values lower than one imply "distance integration", which would occur if, for example, P neighborhoods are enveloped by R neighborhoods. I discuss examples in the next Section.

There are several possible extensions and generalizations. In my robustness checks below, I consider a weighted index where distances between blocks are weighted by R and P population densities, a version that averages distances in logs instead of levels, and an alternative exposure index inspired by market access measures (see Section 6.2.2). Extensions for future research may also consider richer travel cost functions which account for slope and other obstacles (Roberto, 2018) or transportation networks.

[9]Define the following normalizing distance N_c:

$$N_c \equiv \frac{1}{N^2} \sum_i \sum_j d_{ij}.$$

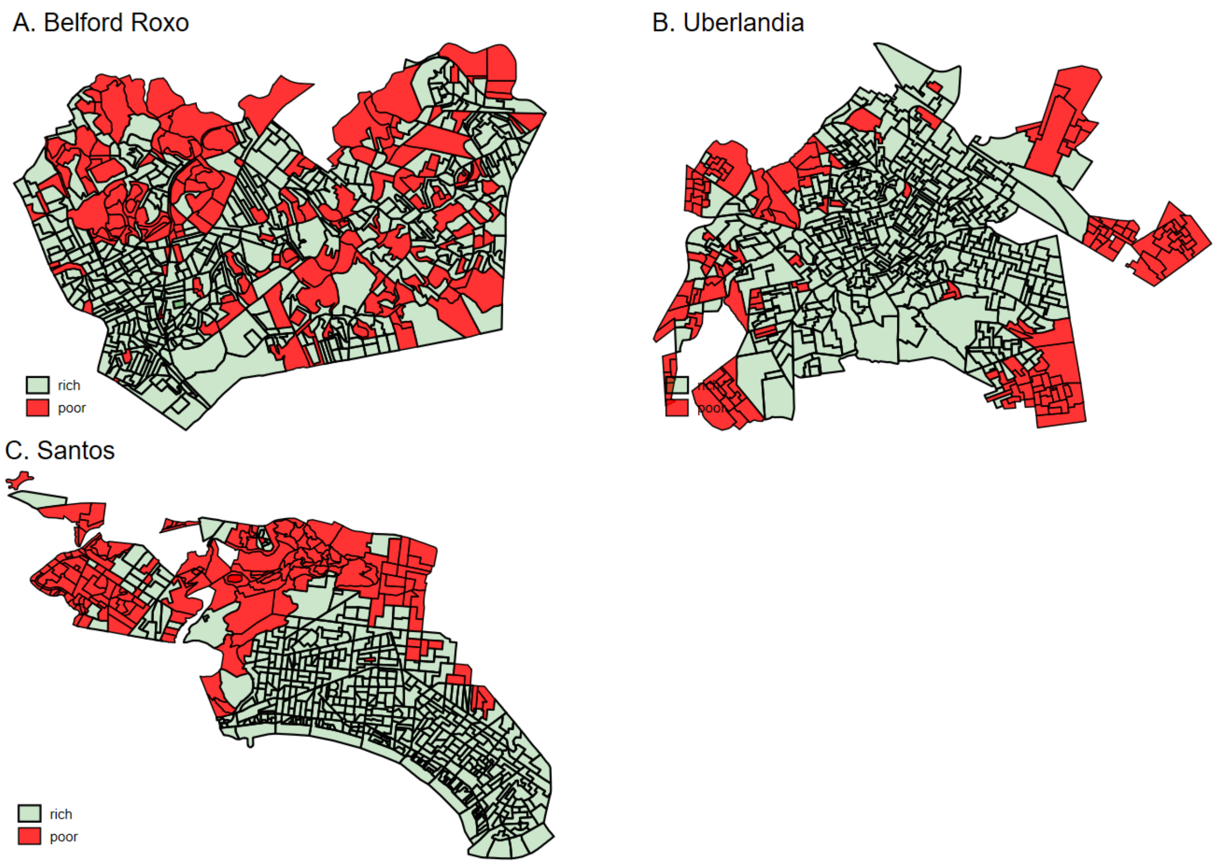
The normalized distance segregation index is:

$$\overline{D}_c^{PR} \equiv \frac{D_c^{PR}}{N_c}.$$

3.2 Examples and summary statistics

To illustrate the range of variation in spatial configurations that the index captures, Figure 2 shows examples of some of the most and least distance-segregated cities among the largest cities in Brazil. To facilitate comparisons across cities, I consider the index normalized as above. In this example, P neighborhoods are defined as blocks in the bottom quartile of the city distribution by average block income, whereas R neighborhoods are all other neighborhoods.

Figure 2: Example cities



Notes: This figure shows examples of some of the most and least distance-segregated cities. Census blocks are classified as poor (in red) if the average income is in the bottom quartile of the city neighborhood income distribution. Blocks classified as rich (in green) are all other blocks.

Belford Roxo (A) is a striking example of one of the most distance-integrated cities among the top 50. The average distance between poor and rich neighborhoods is 4.86 km, whereas the distance between any two neighborhoods is 4.78 km, with a ratio of 1.02. The city appears visually as a "checkerboard". In contrast, in Uberlândia (B), poor and rich neighborhoods are on average 7.64 km apart relative to an average distance across neighborhoods of 6.19 km, resulting in a normalized index of 1.24. In Santos (C), the average distance between P and R is 5.25 km, relative to an average distance across neighborhoods of 3.87 km (normalized segregation index = 1.36). In the Brazilian context, distance-segregated cities will often be characterized by rich neighborhoods in the CBD and peripheral poor neighborhoods. Uberlândia displays this pattern in a roughly circular, monocentric city, whereas Santos is an example of a similar income gradient pattern for a coastal city.

3.3 Discussion

The prevailing notion of residential segregation in the economics literature is one that conceptualizes the degree of mixing between groups occurring within neighborhoods. The most commonly used metric is the dissimilarity index (e.g., Asher et al. (2024)), which measures the share of the minority group that would need to change neighborhoods for it to be evenly distributed within a city. The index ranges from 0 (perfect integration) to 1 (perfect segregation). Importantly, this notion of segregation is purely based on group shares in each neighborhood relative to the city-wide share and ultimately captures how uniform neighborhoods are internally. This class of indexes is subject to the so-called "checkerboard problem" (White, 1983, Reardon and O'Sullivan, 2004), whereby a city displaying a "checkerboard" of rich and poor neighborhoods (like Belford Roxo above) may have the same dissimilarity index as a city where rich and poor neighborhoods are completely clustered on either side of the city, as long as the internal compositions of the neighborhoods is similar across the two cities.[10] I compare the distance index with dissimilarity in Sections 4.1 and 6.2.2 below.

The dissimilarity index is also known to be subject to the Modifiable Area Unit Problem (MAUP), whereby cities will mechanically appear to be more integrated as the size of the neighborhoods increases. Distance-based metrics are believed to be less sensitive to this issue because they primarily use variation from distance across units (White, 1983). I discuss the sensitivity of my empirical results to MAUP in Section 6.2.2. I construct the index using coarser units and find it highly correlated with the baseline one. Moreover, the main empirical results are robust to the choice of unit.

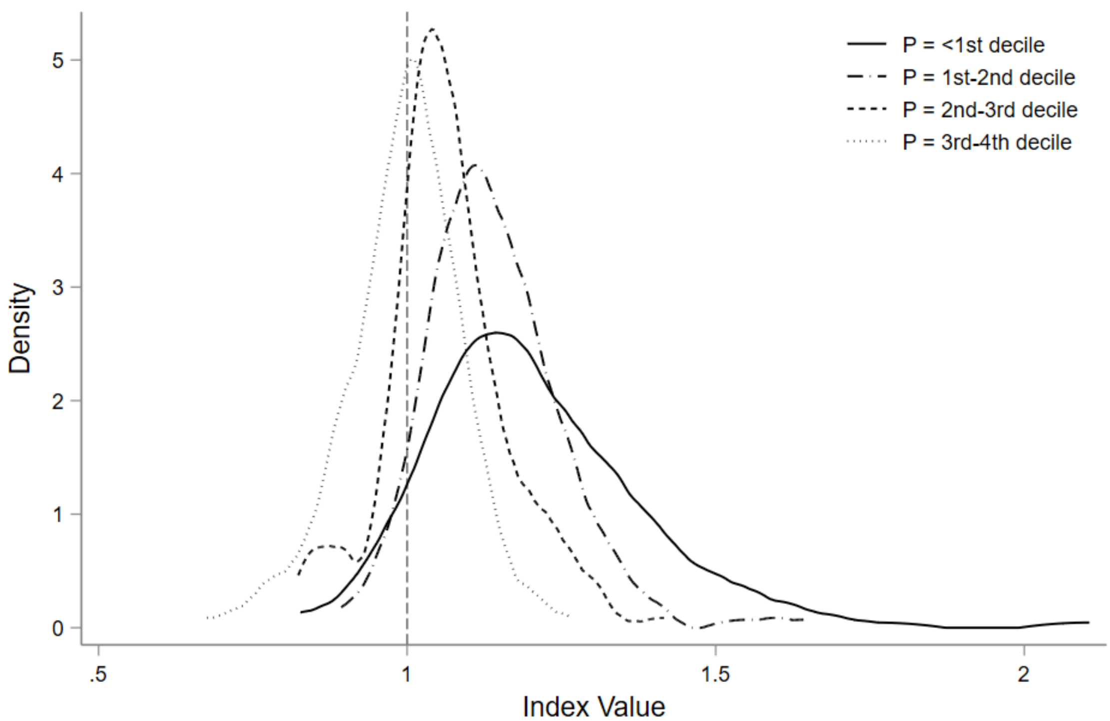
[10] Reardon and O'Sullivan (2004) develop spatial versions of the conventional indexes that are based on comparing the composition of each neighborhood with that of surrounding areas. For example, the spatial dissimilarity index can be interpreted as a measure of how different the social composition of neighborhoods is, on average, from the social composition of the study area. These indexes have been employed to a limited extent (e.g. Xu (2023)).

4 Descriptive patterns

I begin by describing how city-wide residential segregation varies across cities. I then consider correlates of distance segregation at the block level, including public goods provision.

4.1 City-level evidence

Figure 3: Distance segregation by relative income



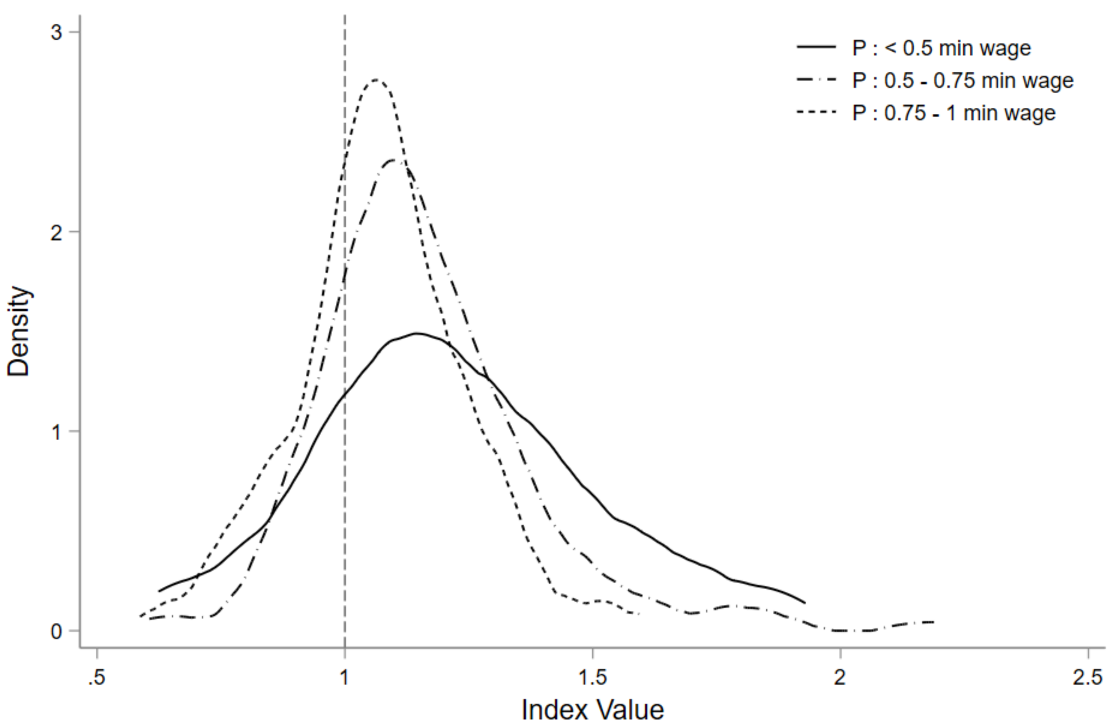
Notes: The figure shows the distribution of the distance segregation index across cities, for different definitions of poor and rich neighborhoods. Census blocks are classified by their rank in the city-wide distribution of neighborhoods by average income. A block is classified as "rich" if the average block income is above median for the city. The index is average distance of poor to rich blocks, normalized by average distance between any two blocks.

Figures 3 through 5 plot the distribution of the distance segregation index, normalized as described above so that 1 corresponds to a perfect checkerboard, values lower than 1 correspond to distance integration, and values higher than 1 correspond to distance segregation. All graphs show that most Brazilian cities are distance segregated, but there are differences depending on the dimension of segregation being considered.

Figures 3 considers distance segregation by income, defined by the relative position of each neighborhood in the distribution of blocks by average income in the city. The figure shows that the very poor are more distance segregated from the rich than the medium-poor. Each line corresponds to a different definition of a low-income neighborhood. The solid line corresponds to the most extreme definition of poverty (*P* = the block's average income in the first decile of the city's distribution) and the dotted line to the least extreme definition (between the third and fourth decile). The definition of high-income neighborhood is held constant for comparability across all definitions and is defined as having average income above the city median.

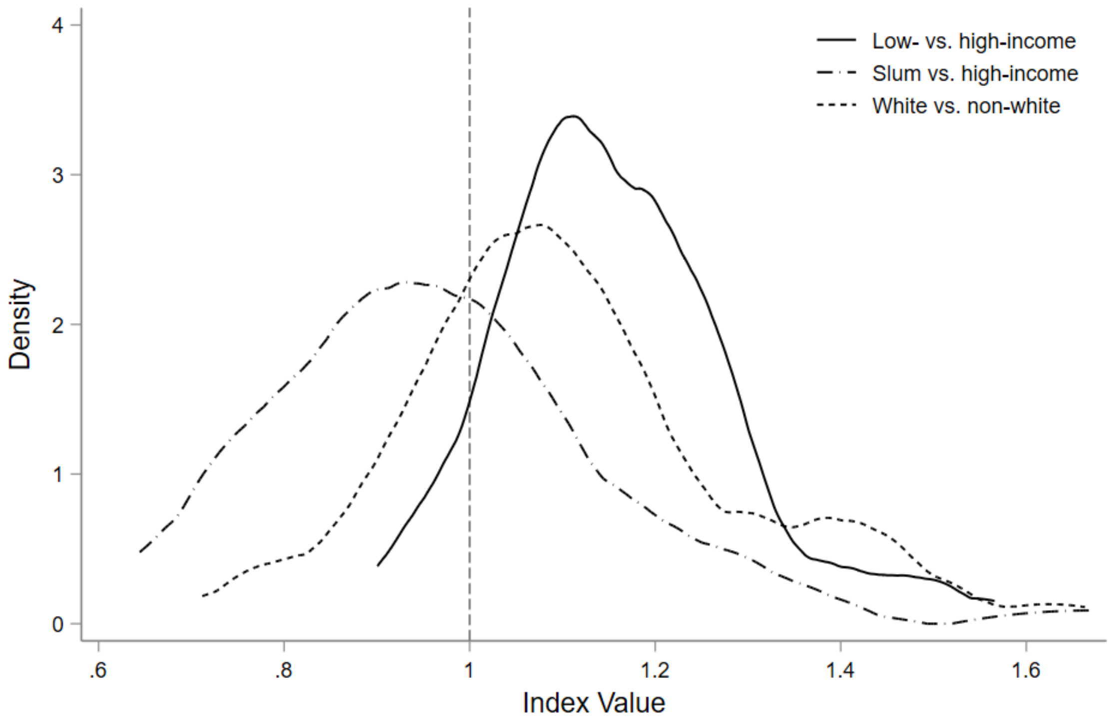
Figure 4 considers absolute measures of poverty. In the solid line graph, the *P* definition is whether the average income in the census block is below half the minimum wage. As a reference, the official government definition of poverty from the *Cadastro Único*, the government registry for Brazil's most vulnerable population, is whether households earn less than half of the minimum wage per capita (about USD $170 per month). The general pattern of the very low-income being more distance segregated than the medium-low-income holds also when considering absolute income.

Figure 4: Distance segregation by absolute income



Notes: The figure shows the distribution of the distance segregation index across cities, for different definitions of poor and rich neighborhoods. Census blocks are classified by average income relative to minimum wage. A block is classified as "rich" if the average block income is above two minimum wages. The index is average distance of poor to rich blocks, normalized by average distance between any two blocks.

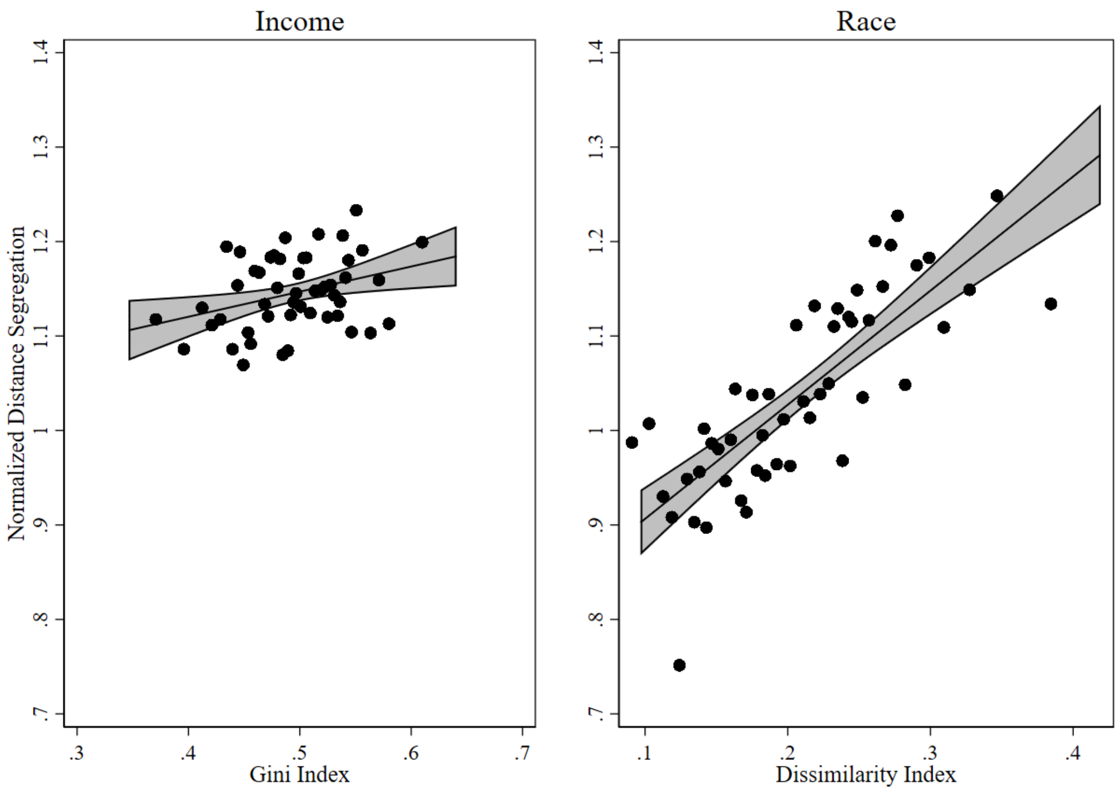
Figure 5: Distance segregation by income, race, and formality



Notes: The figure shows the distribution of the distance segregation index across cities, for different definitions of poor and rich neighborhoods. Low- vs. high-income corresponds to blocks with average income in the bottom quartile vs. top quartile. Slum vs high-income corresponds to the Census definition of a block as part of an "aglomerado subnormal" vs. a block with average income in the top quartile. The white vs. non-white index is defined based on whether the majority of residents are white or not. The index is average distance of poor to rich blocks, normalized by average distance between any two blocks.

Figure 5 compares segregation by income, race, and slum status. A striking pattern is that slums tend to be better spatially integrated than the average poor neighborhood. Additionally, distance segregation by race is more muted than distance segregation by income. The solid line corresponds to distance segregation by relative income, with low-income neighborhoods defined as having average wage in the bottom quartile and high-income neighborhood defined as having average income in the top quartile. This is the income definition I will be focusing on for the rest of the paper. In the median city, poor and rich neighborhoods defined in this way are 13% further away from each other than any two neighborhoods (the median index is 1.13). The dotted line corresponds to segregation defined for neighborhoods that are majority white versus non-white. In the median city, non-white neighborhoods are 4% further away from white neighborhoods, reflecting less stark segregation by race than there is by income. The dashed line corresponds to distance segregation for slums versus high-income neighborhoods. In the median city, slums are 0.02 % closer to rich neighborhoods than any two neighborhoods (median index 0.98), indicating distance integration. This pattern holds also when comparing slums to non-slum neighborhoods., with slums only 2% further away from formal neighborhoods than any two neighborhoods are to each other. The within-city distance metrics corroborate this pattern: in cities that have slums, the median slum has a 7.86 km average distance to high-income blocks, whereas the median low-income block has an average distance to high-income blocks of 8.52 km. This finding is in line with the argument that slum households trade off housing quality and tenure security for access to economic opportunity (Celhay and Undurraga, 2022).

Figure 6: Distance segregation and conventional indexes



Notes: On the left is a binscatter representing the city-level correlation between the Gini income inequality index (Gastwirth, 1972) and the normalized distance segregation index between low- and high-income neighborhoods. On the right is the same binscatter for the dissimilarity index for whites and non-whites and the normalized distance segregation index between majority white and majority non-white neighborhoods.

Next, I present binscatter plots highlighting the correlations between the distance segregation index, normalized as above, and conventional indexes used in the literature. The left panel of Figure 6 shows a weak positive correlation between the Gini income inequality index and distance segregation by income, but the unconditional correlation between the two indexes in the raw data is statistically indistiguishable from zero. The right panel focuses on race and contrasts the distance segregation index with the widely used dissimilarity index. More distance-segregated cities by race also tend to be segregated in the conventional sense. The unconditional correlation between the two indexes is moderate but statistically significant (regression coefficient of 0.10 with a p-value of 0.000) indicating that for a one standard deviation increase in distance segregation conventional segregation increases by one third of a standard deviation. This correlation is not mechanical as the indexes use different variation. In particular, the baseline distance segregation index does not use information on the relative shares of the two groups in each neighborhood.

Across cities, normalized distance segregation appears uncorrelated with city size, decade of city foundation, and a dummy for the less economically advanced northern region of Brazil.

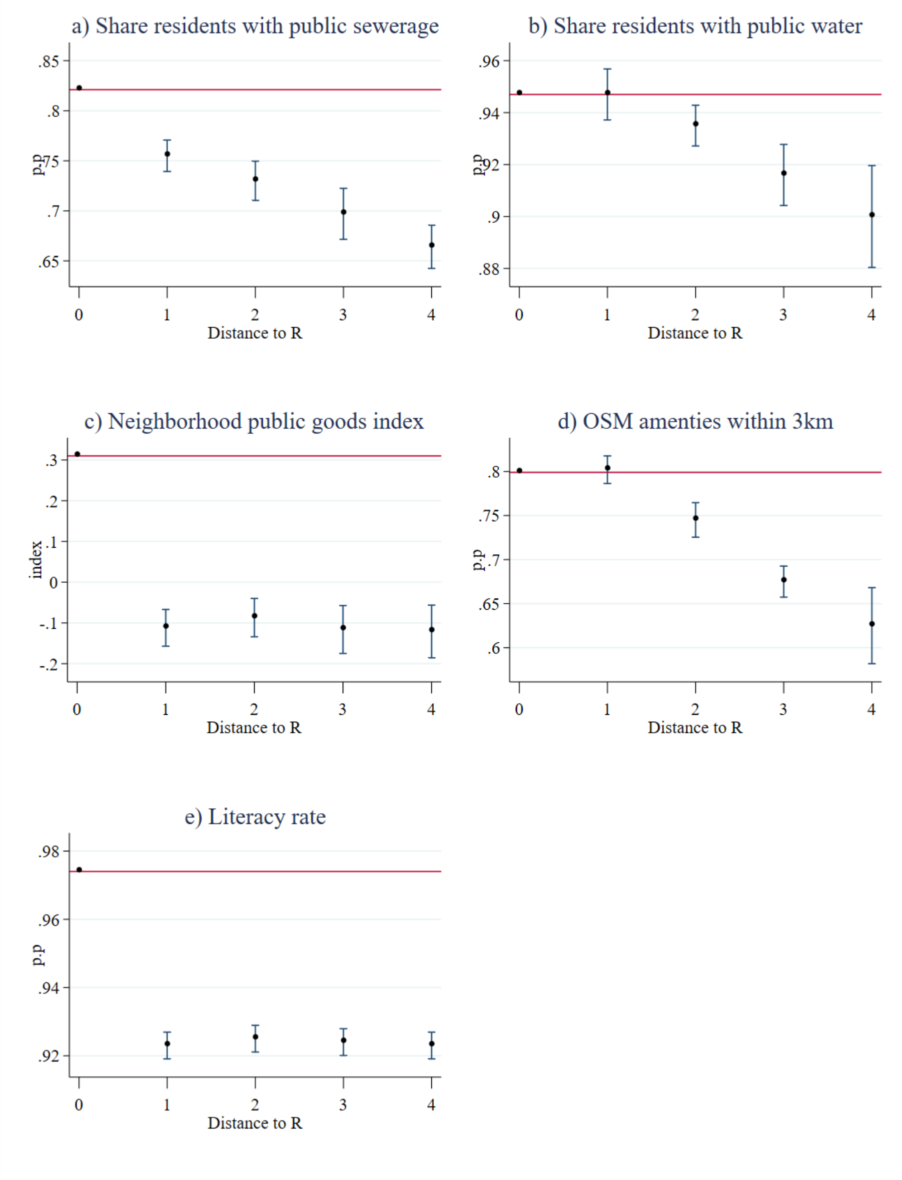
4.2 Neighborhood-level evidence

Having established that there is meaningful variation in the degree of distance segregation across cities, I turn to within-city patterns and provide descriptive evidence on the correlates of distance-segregated neighborhoods. In a given city, among similarly poor neighborhoods, how do distance-integrated areas differ from distance-segregated ones? The key stylized fact is that there is significant spatial inequality in access to public goods, not only between rich and poor neighborhoods, but also between the spatially integrated and spatially segregated poor. More distance-segregated neighborhoods tend to have lower access to a range of local public services, conditional on own neighborhood characteristics and distance to the CBD.

Figure 7 presents the results of block-level regressions documenting a monotonic pattern in access to local public goods as a function of distance segregation by income. I classify poor neighborhoods in four mutually exclusive categories depending on their average distance to rich neighborhoods. The key regressors are dummies for whether a block is poor and in a given quartile by average distance to the rich. Rich neighborhoods are the omitted category. All regressions include city fixed effects and a battery of block-level geographic characteristics that may correlate with the costs of providing public services. Additionally, I control for distance to the CBD, following the urban literature. I consider four primary measures of public goods access, measured at the block level. For each outcome, the figure reports the unconditional mean for the rich neighborhoods (distance to R=0) and conditional means in the poor neighborhoods at different distances, obtained using the regression coefficients.

Across all outcomes, poor neighborhoods have lower access than rich ones. Among the poor neighborhoods, there tends to be a monotonic pattern whereby the most distance-integrated among the poor (distance to R=1, in the first quartile) have better access than the distance-segregated ones (distance to R=4, in the fourth quartile). Panels a and b consider the share of households with access to the public sewerage and water network. In rich neighborhoods, on average 82% of res-

Figure 7: Correlates of distance-segregated neighborhoods by income.



Notes: The figures report conditional means for public goods outcomes (panels a through d) and literacy rate (e) in poor blocks in different quartiles by distance to the rich (distance to R = 1, 2, 3, and 4) and the unconditional sample mean for rich blocks (distance to R = 0). Conditional means are obtained from block-level regressions where the key regressors are dummies for whether a block is poor and in a given quartile by average distance to the rich. Rich blocks are the omitted category. Specifications include 608 city fixed effects, distance to the CBD, distance to the shoreline, rivers or streams, and lakes, land share covered by water bodies, average elevation, and average slope. Standard errors are clustered at the city level. 95% confidence intervals are reported. Poor (Rich) blocks are those with average income in the bottom (top) quartile for the city. The neighborhood public goods index includes: no accumulated street garbage, no open sewer, street addresses, street lighting, street paving, sidewalks, curbs, manholes, ramps, and greenery. The OSM amenities index includes fire stations, police stations, post offices and parks within 3 km.

idents have access to the public sewerage network. The most distance-integrated among the poor neighborhoods have 7 p.p. lower access, whereas the most distance-segregated have 16 p.p. lower access, a 9 p.p. difference. The average share of residents with access to the public water network is 95% among rich neighborhoods and distance-integrated neighborhoods, conversely it is 5 p.p. lower among the most distance-segregated poor neighborhoods. Panel c uses a standardized aggregate index of local public goods such as street paving, sidewalks, and street lighting (the full list is included in the figure notes). The index is expressed in standard deviations relative to the mean across all urban neighborhoods in the sample. Rich neighborhoods score 0.3 standard deviations higher than the average, and poor neighborhoods are 0.4 standard deviations below the rich. No monotonic pattern by distance can be detected here, but there is a gradient when considering a continuous measure of distance (see Table A1 below).[11] Panel d considers public amenities as measured in OSM, specifically whether there are fire stations, police stations, post offices, and parks within 3 km of a neighborhood. The individual dummies are averaged into an index ranging from 0 to 1. These patterns are qualitatively similar when considering distance segregation by race and informal status (Figures A1 and A2 in the Appendix), except that the neighborhood public goods index displays a negative monotonic pattern for race segregation. The results are very similar when controlling for population density, suggesting that lower access in segregated neighborhoods is unlikely to be explained by scale economies in public goods provisions associated with density.

Interestingly, there seem to be limited compositional differences among the distance-integrated and distance-segregated poor. Panel e shows no monotonic pattern for the share of residents who are literate. The results are similar for slum dwellers, although there is a gradient when considering segregation by race (Figures A1).

Tables A1 and A2 in the Appendix reach similar conclusions using a continuous measure of distance segregation (log distance to R blocks) and of own-neighborhood composition (share of residents who are poor or non-white).[12] Odd columns report a specification directly comparable with that in Asher et al. (2024), with access to public goods regressed on the share of minority households (lower-income in Table A1 and non-white in Table A2), conditional on number of residents and city fixed effects. Echoing their findings for minority neighborhoods in urban India, I find that access in decreasing in the share of the disadvantaged group in each neighborhoods. Even columns augment the specification adding distance to R neighborhoods, distance to the CBD, and

[11] When considering the individual components of the index, the rich always have higher access, but the pattern among poor neighborhoods varies. For segregation by income, there is a significant negative monotonic pattern for paved streets. There is a weak negative monotonic pattern for curbs, manholes, and sidewalks. There is no gradient for street addresses, street lighting, greenery, and not having open sewers. There is a positive monotonic pattern for ramps and for not having street garbage.

[12] I cannot estimate the same specification for informality because slums are coded as a dummy for each block.

geographic controls. For most outcomes, the coefficient for the share of the minority group drops substantively in magnitude once distance is accounted for.

The observed correlations likely arise from a combination of direct effects of segregation on public goods provision and endogenous sorting *in response to* public goods provision. Taking residential patterns as given, municipal governments may choose to differentially provide local public goods in distance-integrated and segregated poor neighborhoods, motivated by cost considerations or strategic targeting. At the same time households may sort in response to the availability of public services across neighborhoods, for example with richer households sorting into areas that are far away from low-public goods, impoverished ones.

In the next Section, motivated by the within-city descriptive evidence, I take the analysis to the city level and complement OLS evidence with an instrumental variables strategy to shed light on the link between residential patterns and public goods provision.

5 Instrumental variables approach

In this Section, I turn to the causal question of whether city-wide segregation affects public goods provision. There are several challenges in estimating a causal relationship. First, the OLS regression is likely subject to reverse causality: if low provision of public goods induces rich households to segregate away from the poor because of concerns of negative externalities, this will lead to a spurious negative correlation between segregation and public goods provision. Second, there could be omitted variables biasing the results in different possible directions. For instance, income inequality may make households more willing to segregate and less able to coordinate successfully on public goods provision (Boustan, 2011), leading to a spurious negative correlation between public goods provision and segregation. Local institutions may be an additional omitted variable: in developing country cities, irregular gradients can be the byproduct of weak land market institutions, poor planning, and poor enforcement of property rights (Henderson et al., 2020), in contrast with cities with smooth income gradients. Limited local state capacity may hinder effective provision of public goods and at the same time result in a mixed residential pattern with poor neighborhoods interspersed with rich ones. This would tend to bias the OLS coefficient towards a spurious positive correlation between segregation and public goods provision.

To shed light on how residential patterns affect public goods provision, I complement the OLS evidence with an instrumental variable strategy that leverages variation in distance segregation determined by geographic features. Below, I explain the construction of the instrument and discuss the exclusion restriction. Empirical OLS and IV results are presented in the next section.

5.1 Geographic determinants of residential patterns

The starting point for the construction of the instrument is the notion that the spatial distribution of rich and poor neighborhoods is likely to be in part shaped by the spatial distribution of underlying natural amenities (Lee and Lin, 2018, Deffebach et al., 2024). Figure 8 shows an example of geography anchoring the location of poor neighborhoods in the city of Caxias do Sul. A comparison between figures a and b shows that poorer neighborhoods tend to be in high-slope areas. Below, I generalize this intuition and show that slopes and distance to water bodies are predictors of the location patterns of high- and low-income households within cities and I leverage this finding to construct an instrument for distance segregation.

Table 1 shows that slope predicts the location of low-income neighborhoods (column 1), predominantly non-white neighborhoods (column 3) and slums (column 4) within cities. The relationship is non-linear, with steeper areas associated with lower-income neighborhoods and slums particularly associated with extreme slopes above 20 degrees.[13] These patterns are in line with anecdotal evidence from many Latin American cities and *favelas* in particular, where the poor tend to squat in areas that are unlikely to be ever considered for formal development due to unfavorable topography (Brueckner et al., 2019). High-income neighborhoods (column 2) are associated with lower slopes. In addition, proximity to water bodies explains the location of poor and rich neighborhoods. Poor neighborhoods tend to be closer to rivers and streams, consistent with poor settlements observed in higher-risk areas by riverbanks, while rich neighborhoods are associated with proximity to the coast, likely a landscape amenity. All regressions include city fixed effects.

5.2 Instrument construction

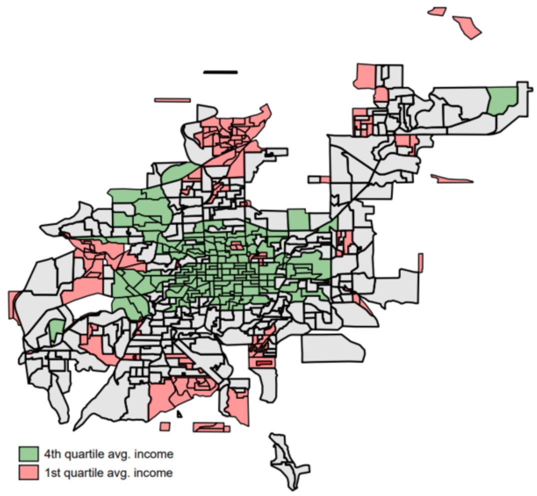
Having established that natural amenities anchor the location of poor and rich neighborhoods, I construct an instrument for distance segregation based on distance between predicted poor and predicted rich neighborhoods. First, I obtain the predicted likelihood of being poor for each neighborhood from the predictive regressions in Table 1. Second, I rank neighborhoods within each city by this predicted likelihood of being poor. Third, I assign a "predicted poor" dummy to the top n neighborhoods according to the ranking in step 2, setting n to match the actual number of poor neighborhoods in each particular city.[14] Figure 8c shows the result of this exercise for Caxias do Sul. The last step is to calculate the city-level distance segregation using predicted poor and rich neighborhood, obtaining "geography distance segregation".

Table 2 shows that geography distance segregation predicts actual distance segregation for

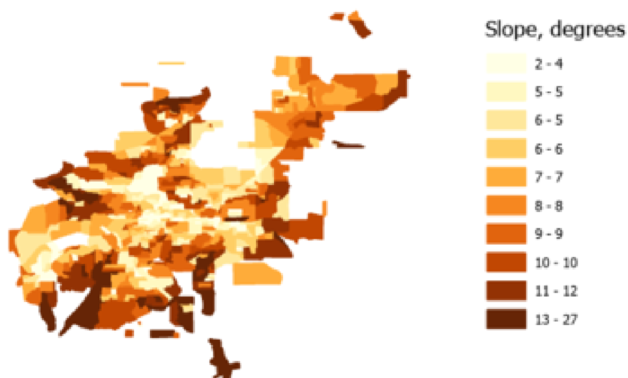
[13] The functional forms of these predictive regressions was selected by considering different combinations of polynomials in slope and distance to water bodies and selecting those that yielded the strongest F statistic.

[14] For the baseline "poor" definition there will always be by construction 25% of neighborhoods that are poor and 25% that are rich, but that is not the case for the slum and majority white dummies.

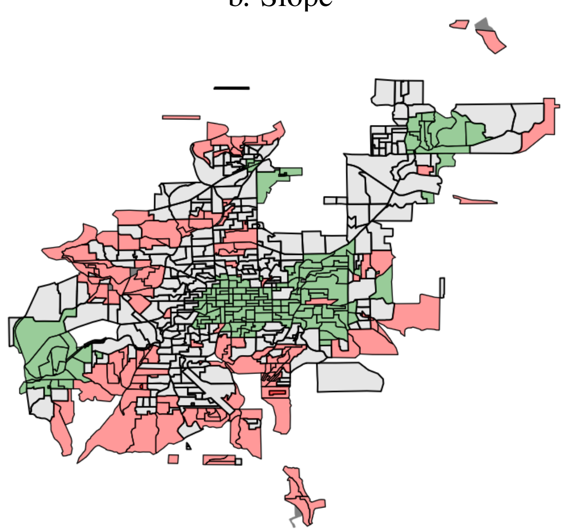
Figure 8: Using geography to predict residential patterns



a. Actual



b. Slope



c. Predicted

Notes: These maps show the city of Caxias do Sul. Figure a maps the actual distribution of poor and rich neighborhoods. Figure b maps the average slope in each block. Figure c maps the predicted poor and rich neighborhoods based on the regression in Table 1. Poor (Rich) blocks are those with average income in the bottom (top) quartile for the city.

the three dimensions of income, race, and informality. The dependent variable is actual distance segregation, that is, the average distance in km between poor and rich neighborhoods (column 1), between majority non-white neighborhoods and other neighborhoods (column 2), and between slums and non-slum neighborhoods (column 3).[15]

To account for the mechanical correlation with city size, I control for a city's equivalent area radius.[16] The regression additionally includes a battery of pre-determined city characteristics (Naritomi et al., 2012) that could be associated with natural advantage or differences in the engineering costs of public goods delivery according to the urban literature. These include average altitude, ruggedness, slope, presence of water bodies, share of developable land within 30 km from the city center (Saiz, 2010), climate and soil controls, and a slope-adjustment factor for travel costs.[17] The full list of controls is reported in Table A3. First-stage F statistics are above conventional levels. Standard errors are clustered at the meso-region, a Census unit that groups together various municipalities in proximity with each other and with common characteristics.

Geography-predicted distance segregation exploits the variation in segregation that is driven by the relative positioning of high- and low- natural amenity areas in the city. A city has low geography distance segregation when its topography provides poor households with an opportunity to settle fairly close to the high-amenity locations settled by rich households.[18] Importantly, similar to Harari (2020), the ensuing cross-city variation stems from the *relative location* of areas with different slopes and water access, and not from the presence or extent of topographic obstacles. While the latter geographic characteristics are likely correlated with public goods provision through channels such as cost or tax revenues, leading to a violation of the exclusion restriction, the distance between favorable and unfavorable topography areas is less likely to directly impact the provision of services. Threats to the identification are discussed in Section 6.2 below.

Table A3 shows how geographic distance segregation by income correlates with the pre-determined controls. Most of the characteristics are uncorrelated or have very small effect sizes. For example, cities that are distance segregated according to geography tend to have slightly lower elevation (13 meters for one additional km of distance segregation), they are slightly more fertile (on average, from a 54% to 53% share of low-fertility soil), and have higher precipitation (+1.2 mm). Reassuringly, the instrument is uncorrelated with average slope, ruggedness, presence of

[15] The differences in sample sizes are due to the different definitions. Not all cities have neighborhoods that have a majority of non-white residents (column 2) and the "slum" Census designation is only recorded for large cities (column 3).

[16] The equivalent area radius is the radius of a circle with the same land area as that of the city.

[17] This is computed as the ratio of the distance between any two neighborhoods using plain Euclidean distance and the same distance adjusted by slopes along the trajectory. The adjustment is implemented applying the Tobler hiker rule, used to determine how much slower travel is along a slope (Tobler, 1993).

[18] This is similar in spirit to the instrumental variables approach in Ananat (2011), where the positioning of railroad tracks is used to predict racial segregation in U.S. cities.

water bodies, and the share of developable land around the city (Saiz, 2010). There is a moderate negative correlation with distance to the state capitals (10 km out of a sample mean of 181) and to the coast (12 km out of a mean of 216). I address this in Table A4, showing that the main results are robust to excluding state capitals, cities near state capitals, coastal cities, and high-soil fertility cities.

6 City-wide distance segregation and local public goods

I now turn to examining the impact of distance segregation on city-wide outcomes. I primarily focus on segregation by income and report additional results with race and informality in the Appendix, cautioning that the instrumental variables approach relies on a similar source of variation for predicting all three dimensions of segregation. I begin by considering the link between residential patterns and average levels of public goods provision at the city level and present robustness tests. Next, I turn to mechanisms by considering patterns of public goods provision across groups and neighborhoods.

6.1 Distance segregation and city-wide public goods access

Table 3 presents the main results, examining the impact of distance segregation by income (in km) on measures of public goods access. I report both IV and OLS coefficients on distance segregation measured in km. The controls listed in Table 2 are also included. Columns 1 through 4 consider access to the four primary public goods metrics, averaged at the city level across all neighborhoods. Across all metrics, I find lower access in more distance-segregated cities.

To interpret magnitudes, note that, for the median-sized city, a one standard deviation in distance segregation by income is 1.7 km. The IV estimates in column 1 indicate that, for a one standard deviation increase in segregation, the share of households with access to public sewerage and public water decreases by 3.4 and 3 p.p. respectively, 6 and 3% of the sample mean. The neighborhood quality index decreases by 7% of a standard deviation. The likelihood of having OSM amenities within 3 kilometers - roughly the distance reachable within a 30-45 minute walk - decreases by 3.4 p.p. or 5% of the mean.

These results are robust to defining public goods in different ways. Alternative outcome variables based on the individual components of the neighborhood index and the composite OSM index are reported in Table A8. For OSM amenities, Table A8 reports distance in meters to the nearest amenity as the dependent variable.

Tables A5 show similar results when considering other dimensions of segregation - race and informal status. For some of the outcomes, the point estimates appear larger for segregation by slum status than by income, but this is mainly driven by differences in the sample: we only observe slums for a sample of approximately 200 larger cities. When restricting to the same sample, the magnitudes for the effect of segregation by income become similar.

Comparing the IV and OLS estimates in Tables 3 and A5, OLS coefficients tend to be similar or attenuated relative to IV ones. Smaller OLS coefficients are consistent with omitted variable bias from local institutions: if residential integration is partly a byproduct of weak land market institutions and limited state capacity, this will result in both lower levels of public goods provision and a more irregular income gradient within the city.

6.2 Threats to identification and robustness

Below, I discuss potential violations of the exclusion restriction and present robustness exercises.

6.2.1 Direct effects of geography on public goods

A key threat to the identification is that geographic configurations explaining distance segregation may directly affect the costs of providing public goods, particularly those delivered through a spatial network, such as sewerage and water. As discussed above, the baseline regression controls for average slope, ruggedness, and an adjustment factor to address the higher travel costs associated with changing slopes. However, there could be other functions of geography that are correlated with both the costs of providing public goods and with the instrument.

First, it is reassuring to note that this type of violation of the exclusion restriction would plausibly tend to bias the results against the main findings. The concern is that uneven topography may hinder the provision of public goods. However, according to the instrument prediction, cities with "uneven" topographies (that is, large spatial variations in slope within small areas), are cities with lower distance segregation, which have, if anything, higher levels of public goods access according to the IV estimates. This argument also assuages concerns that cities with uneven topographies may be poorer and less able to fund public goods provision. Below, I present additional tests to probe these concerns.

In Table A6 I show that the results are robust to controlling for a number of additional proxies for irregular topography, such as the number of distinct water basins (which could affect the costs of water drainage and sewerage provision). I also show robustness to controlling for different proxies for city shape (Harari, 2020): the perimeter-to-area ratio and the number of distinct contiguous polygons obtained from considering developable land around 30 km of each center.

Table A7 shows that the results are robust to accounting for block-level geographic characteristics that may affect provision of public goods in that particular block. I construct a residualized version of my primary outcomes as follows. Before aggregating at the city level, I regress each public goods outcome measured at the block level on the baseline city-wide controls and block-level slope, elevation, and distance to water bodies (odd columns). I then aggregate the residualized measure, which should now be uncorrelated with geography. Even columns repeat the same exercise but the block-level regression additionally controls for distance to the CBD to further account for the engineering costs of expanding the network. Reassuringly, results are similar to the baseline ones.

The finding that more distance-segregated cities have lower public goods provision is not limited to public goods delivered along spatial networks. The conclusion continues to hold when considering other measures related to municipal spending, which are less susceptible to potential correlations with geography. Table A8 considers the share of municipal spending in health in 2010 (according to municipal finance records, column 1). The share of municipal spending on health is 1 p.p. lower (4% of the baseline mean) in cities that are more distance segregated by one standard deviation according to the income definition. This pattern is confirmed in column 2 where I consider the number of hospital beds per capita in 2010 (from the Ministry of Health and DATASUS).

6.2.2 Alternative metrics and specifications

Table 4 explores different specifications and functional forms for the distance segregation metrics. Each panel reports IV estimates of the distance segregation metric for a specification similar to that of Table 3.

In panel A, I explore a population-weighted version of the baseline segregation index. Denote the number of group g residents in block i as L_i^g and the block's land area as a_i. The population density-weighted distance segregation index D_{Wc}^{PR} for city c is computed as:

$$D_{Wc}^{PR} = \frac{\sum_i D_{Wi}^R \left(\frac{L_i^P}{a_i}\right)}{\sum_i \left(\frac{L_i^P}{a_i}\right)} \text{ where } D_{Wi}^R = \frac{\sum_j d_{ij} \left(\frac{L_j^R}{a_j}\right)}{\sum_j \left(\frac{L_j^R}{a_j}\right)}.$$

Intuitively, D_{Wi}^R is the average distance of block i from all other blocks j, weighted by the density of R households in all destination blocks j. This block-level measure is then averaged among all origin blocks i, weighing by the density of P households. This measure differs from the baseline one in that it utilizes information on the internal composition of neighborhoods (as opposed to a simple binary indicator for P or R blocks). Weighting by population density (as opposed to counts) addresses the concern that blocks can have different area sizes. Using the population-weighted index confirms the baseline results and yields larger coefficients. This is reassuring that the main results are not driven by areas with low population density (which may disproportionally be in high-slope areas).

Turning to different distance functions, in panel B I consider an index based on average log distances instead of linear distances. Intuitively, under this distance function, less weight is placed on variation in distance for very far away neighborhoods. Results are similar to the baseline ones.

In panel C I consider an alternative index, inspired by the market access literature, which measures the average city-wide exposure of group P to group R, weighted by an exponential distance function. Define $\tilde{d_{ij}} = \exp\left(k\frac{d_{ij}}{s}\right)$ with $k = 0.013$ (Tsivanidis, forthcoming) and $s = 30$ km/h. The exposure-based index for city c is computed as

$$E_c^{PR} = \frac{\sum_i E_i^R \left(\frac{L_i^P}{a_i}\right)}{\sum_i \left(\frac{L_i^P}{a_i}\right)} \text{ where } E_i^R = \frac{\sum_j \frac{1}{\tilde{d_{ij}}} \left(\frac{L_j^R}{a_j}\right)}{\sum_j \frac{1}{\tilde{d_{ij}}}}.$$

Intuitively, E_i^R measures the exposure of block i to R households, weighting each destination block j by an inverse exponential distance function. This block-level measure is then averaged within the city weighing each origin block i by its P population density. This functional form echoes the way in which the quantitative spatial modeling literature defines gentrification spillovers from high-skill residents (e.g. Gechter and Tsivanidis (2023)). The resulting index is expressed in households per squared km and takes higher values in cities where P and R are more exposed to each other, corresponding to more integration. For ease of interpretation, I report standardized coefficients corresponding to the effects of a one standard-deviation increase in exposure. Qualitatively, the results are similar to the baseline ones: cities where P and R are *less* exposed to each other have lower provision of public goods.

Panel E shows robustness of the results to MAUP. For this exercise, I construct distance segregation metrics for units coarser than census blocks. Absent an intermediate-sized Census unit, I construct ad hoc units encompassing on average four *setores*, resulting in units similar in size to U.S. Census tracts. I consider clusters of adjacent blocks with similar incomes, to replicate the way in which neighborhoods are delineated in administrative datasets. I use a K means clustering algorithm using latitude and longitude and average income, the former being double weighted.[19] For all three dimensions of segregation, the normalized distance segregation index obtained with the coarser units is positively and significantly correlated with the block-level one. The IV results are also very similar to the baseline ones when using the coarser units.

While the focus of this paper is on distance segregation, a question may arise of whether the conventional dissimilarity index affects public goods provision in the same way. Table A9 correlates the dissimilarity index (calculated using households in the bottom quartile by city-wide income distribution as the minority) with city-wide public goods provision. There is no IV strategy

[19] As an alternative ad hoc unit, I also consider grid pixels obtained overlaying an arbitrary fishnet of size 1.21 square kilometers on the census block maps. The IV and OLS results remain robust.

for dissimilarity, so only OLS estimates are reported. Columns 1 and 4 show that, unconditionally, higher values of dissimilarity are associated with greater provision of public goods. This is likely driven by the correlation between dissimilarity and city income. Columns 2 and 5 include controls for average household income and total municipal expenditure, which are not predetermined but are sometimes used in the literature on public goods provision (e.g. Trounstine (2016)). Once these additional variables are included, the magnitude of the coefficient on dissimilarity drops and becomes statistically insignificant or negligible in magnitude. Columns 3 and 6 conversely show that the IV on distance segregation remains significant and similar in magnitude to the baseline specification once income and expenditure are controlled for. Understanding the implications of dissimilarity for public goods provision is left for future research.

6.3 Mechanisms

Overall, the results above point to lower levels of public goods provision in more segregated cities. I next turn to mechanisms and consider three sets of explanations for how distance may affect public goods provision. The first channel is engineering costs: many public goods - like sewerage or water - are delivered along contiguous spatial networks. The second is preferences: segregation between poor and rich may be associated with different preferences towards the provision of basic neighborhood public goods, either through sorting or directly through exposure effects. The third channel is externalities: conditional on the overall preferences for public goods provision, there are stronger incentives to service neighborhoods that are nearby because of more pronounced negative externalities that need correction. Below, I exploit the granularity of the block-level data to develop suggestive empirical tests that leverage the spatial pattern of public goods provision across neighborhoods and groups.

Table 5 presents results analogous to those in Table 3, but instead of considering outcomes averaged at the city level, it reports outcomes averaged across specific sets of neighborhoods only. For ease of interpretation, instead of the aggregate neighborhood quality index, I report two among the most salient components of the index, corresponding to street paving and sidewalks. The table reports IV coefficients and the sample mean of each outcome variable in square brackets. Panels A and B consider public goods access averaged among poor and rich neighborhoods in each city, respectively. Panel C considers the ratio between the two, as a proxy for inequality between groups in public goods provision. Panel D considers distance-integrated poor neighborhoods, defined as those in the bottom quartile of each city by distance to rich ones. Interestingly, lower public goods access in segregated cities occurs across the board, not only in neighborhoods that are poor or distance-segregated, with some indication that segregated cities provide a more unequal redistribution in favor of the rich. Panel E considers the ratio between public goods access among the segregated poor (poor neighborhoods in the top quartile by distance to rich neighborhoods) and the poor in general. This is a measure of spatial inequality in provision of public goods among the poor. Below, I interpret these patterns through the lens of the mechanisms listed above.

6.3.1 Engineering cost

The first potential explanation is related to the costs of providing public goods along spatial networks. Poor neighborhoods that are close to rich neighborhoods are more likely to be along the sewerage or water expansion network. Thus, according to this channel, a city with poor neighborhoods interspersed with rich neighborhoods is more likely to serve the poor than a similar city where the poor are at the edges of the network. While this argument applies primarily to network goods, other types of public goods are to some extent complementary and may follow a similar spatial pattern.

If the network expansion explanation above is the main mechanism, then the differences between segregated and integrated cities should be primarily driven by segregated poor neighborhoods, while integrated poor neighborhoods should be equally well provided in both types of cities. This is not what I find. The negative estimates in panel D indicate that segregated cities under-provide even to the integrated poor relative to integrated cities. The conclusions of this test are similar considering definitions of integrated poor neighborhoods that are based on absolute, instead of relative distance from the rich (e.g. within 500 or 1000 m).

Note that the sample means in panel A (poor) are lower than in panel B (integrated poor), consistent with the within-city descriptive results: indeed, within cities, the integrated poor are better served than the segregated poor, consistent with an engineering cost channel amongst others. However, the test in panel D suggests that this is not differential by high- and low-segregation cities in a way that can explain the cross-city differences.

These results are corroborated with those in Table A7 (even columns), where the public goods metrics have been residualized by distance to the CBD to account for the engineering costs of expanding the network.

6.3.2 Preferences

The second potential explanation is related to differences in preferences for public goods provision: residents in distance-segregated cities may have overall lower willingness to direct municipal funding towards these basic public goods, unrelated to the spatial targeting of specific neighborhoods. Under this explanation, we should see less provision of these public goods across the board, and not only in far-away poor neighborhoods. Consistent with this prediction, panels A and B show that the under-provision of public goods in segregated cities occurs both in poor (A) and rich (B) neighborhoods.[20] Panel C considers the ratio between poor and rich neighborhoods, showing that segregated cities tend to have a more unequal distribution of public goods between poor and rich neighborhoods, as evidenced by the negative coefficients. However, the pattern is noisy and only the coefficients for public water and sidewalks are significant.

One hypothesis is that segregated residential patterns may be shaping residents' preferences by limiting the exposure of richer households to poor neighborhoods ("Out of sight, out of mind"). This contact hypothesis has been proposed in the political economy and behavioral literature in reference to conventionally measured segregation shaping perceptions of inequality (Davidai et al., 2024), societal attachment between groups (Bjorvatn and Cappelen, 2003) and trust (Alesina and Zhuravskaya, 2011). This is also in line with a literature showing that interactions with poorer people can shift preferences towards becoming more redistributive (Alesina et al., 2018, Londoño-Vélez, 2022).

Another potential channel is sorting: residents who prefer not to allocate local funding toward these basic public goods may choose cities that are prone to segregation. My empirical strategy does not allow to disentangle the two, but Table A10 provides suggestive evidence on sorting patterns by demographic composition. Columns 1 through 3 show that residents in distance-segregated cities have lower literacy rates, a lower share of high-school or college graduates and a lower share of employment in the service sector, but these effects are small. Columns 4 and 5 consider immutable individual characteristics, such as race and being a prime age male (the category that migrants are the most likely to belong to). There are no differences across cities in these two characteristics. This points to limited sorting along these dimensions, although this test is not conclusive.

6.3.3 Externalities

The third set of explanations are related to targeting neighborhoods as a function of distance, conditional on overall levels of public goods provision. There are two competing channels that may lead to a differential allocation of public goods to the spatially integrated versus segregated poor. Because (true or perceived) negative externalities from poor neighborhoods decay with distance, there may be stronger incentives to provide externalities-correcting public goods in poor areas that are close by (Xu, 2023). At the same time, there could also be a deterrence motive operating in the opposite direction: policy makers may attempt to discourage the poor from settling near the rich

[20] While segregated cities feature lower public goods provision relative to integrated cities even in rich neighborhoods, these results do not imply that the urban rich are relatively deprived in segregated cities: for many of the public goods considered, there are private alternatives available to wealthier households. For example, gated communities or large luxury condominiums may have private treatment plants (von Sperling, 2016) when not connected to the public sewerage.

by strategically under-providing neighborhood public goods in these areas (Feler and Henderson, 2011).

In Panel E, for the outcomes in columns 2 through 4, the positive (albeit insignificant) coefficients suggest that the difference in access between the segregated and the integrated poor is attenuated in segregated cities. In other words, in a city that is overall segregated, the segregated poor are not faring that much worse than their integrated counterpart. This may be because of deterrence being more relevant in distance-segregated cities, where residents may have stronger preferences to stay segregated.

The deterrence channel should only be relevant for neighborhood public goods. Withdrawing the provision of public amenities that residents can travel to is likely to be ineffective as a deterrent, as the spatially integrated poor can already access many of these amenities in nearby areas. Consistent with this logic, I find no evidence of the deterrence mechanism for "super-neighborhood" public goods (the OSM amenities). In column 5, the coefficient is negative and significant, suggesting that in integrated cities, the integrated poor have relatively better access than in segregated cities. This is consistent with an externalities correcting mechanisms being relatively stronger in integrated cities.

These competing mechanisms highlight a key tradeoff that policy makers in developing countries consider when debating the spatial distribution of public goods: on the one hand, public goods can alleviate negative externalities; on the other hand, policy makers want to avoid further in-migration and congestion in central areas of their cities, which may hinder redevelopment and depress land values (Harari and Wong, forthcoming).

6.3.4 Heterogeneity by the degree of local autonomy

All the mechanisms considered above, except the engineering cost one, rely on the assumption that the public goods considered are, in fact, allocated at the municipality level. If decisions on the provision and spatial allocation of these local public goods were entirely determined at a different level (e.g. the state), then we should expect a much attenuated relationship between city-level characteristics and public goods provision. Conversely, we should expect stronger results in cities where there is more local autonomy in public goods provision decisions.

Table A11 provides a heterogeneity test leveraging differences in local governance across cities as far as water and sanitation are concerned. Historically, this sector was a prerogative of municipal governments, but there was an attempt to centralize the system through the creation of state-run companies. A period of uncertainty in governance followed, until a 2007 law re-established municipalities as the ultimate authority in this sector, retaining regulatory power even when a state company is present (Kresch, 2020). Columns 1 and 2 show that the negative impacts of segregation on access to water and sewerage are approximately twice as strong in cities that have a municipal company, as per the indicator variable coded by Kresch et al. (2023). This supports the interpretation that the main results are mediated by a municipality-level response. The remaining columns show that the interaction with the municipal company dummy is insignificant for local public goods that are unrelated to water, which provides a placebo test.

In line with these results, Table 4, panel D, shows that the results are robust to including state fixed effects, assuaging the concern of confounding by state-level factors (e.g., transfers, state spending mandates, or state-run companies).

7 Conclusion

This paper provides one of the first systematic investigations of within-city residential patterns in a middle-income country context. I present metrics of "distance segregation", capturing the physical proximity between low- and high-socio-economic status households, and document several stylized facts. Most cities are segregated by income, but slums tend to be more spatially integrated. Additionally, segregated poor neighborhoods tend to have lower access to public goods conditional on their own composition and distance to the center. Leveraging an instrumental variables approach that isolates the variation in segregation determined by geography, I show that more segregated cities are associated with lower public goods provision. I explore potential mechanisms such as engineering costs, redistributive preferences, negative externalities from poor neighborhoods, and strategic targeting of neighborhoods.

This paper highlights residential patterns as important determinants of the level and spatial allocation of local public goods in a developing country context. These results help inform many contentious urban policies that directly and indirectly affect where the poor and rich live in cities, including slum relocations (Rojas-Ampuero and Carrera, 2023), public housing (Picarelli, 2019, Franklin, 2019), urban renewal programs (Gechter and Tsivanidis, 2023), and investments in transit infrastructure (Tsivanidis, forthcoming, Khanna et al., 2024). These findings also help shed light on the determinants of spatially targeted urban investments. Many urban areas still lag behind in the provision of basic services, such as sewerage and water. This paper helps rationalize why it may be difficult to mobilize consensus to invest in public goods provision in disadvantaged areas.

Several avenues are open for future research. One is to examine the determinants of segregation patterns beyond "first nature", such as externalities, coordination, and path dependence (Lin, 2015). Investigating the historical roots of segregation in Brazilian cities, particularly during colonial times when infectious diseases were a key urban externality (Garmany and Richmond, 2020), could provide valuable insights. Another potential direction is exploring the dynamics of distance-based segregation over time, focusing on one of the larger cities for which data may be available at different points in time. Additionally, it would be interesting to investigate the implications of distance segregation for human capital outcomes, like upward mobility and labor markets (Barza et al., 2024, Belchior et al., 2024).

Beyond Brazil, this paper suggests that incorporating space and distance in the study of residential segregation offers insights into the functioning of cities and the economic livelihoods of disadvantaged urban groups, suggesting a promising line of inquiry for future research in both developing and developed contexts (Harari and Stuart, 2024, Davis et al., 2024).

References

ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): "Public Goods And Ethnic Divisions," *The Quarterly Journal of Economics*, 114, 1243–1284.

ALESINA, A., E. LA FERRARA, S. STANTCHEVA, AND E. TESO (2018): "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 108, 521–554.

ALESINA, A. AND E. ZHURAVSKAYA (2011): "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 101, 1872–1911.

ALVES, P. J. H. F. AND J. M. ARAUJO (2024): "The Effects of Intergovernmental Transfers on the Local Fiscal Incentives of Brazilian Municipalities," *Journal of Government and Economics*, 13, 100104.

ANANAT, E. 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, 34–66.

ASHER, S., K. JHA, P. NOVOSAD, A. ADUKIA, AND B. TAN (2024): "Residential Segregation and Unequal Access to Local Public Services: Evidence from 1.5m Neighborhoods in India," Working Paper.

ATHEY, S., B. FERGUSON, M. GENTZKOW, AND T. SCHMIDT (2021): "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 118, e2026160118.

BARNHARDT, S., E. FIELD, AND R. PANDE (2017): "Moving to Opportunity or Isolation? Network Effects of Randomized Housing Lottery in Urban India," *American Economic Journal: Applied Economics*, 9, 1–32.

BARZA, R., E. L. GLAESER, C. A. HIDALGO, AND M. VIARENGO (2024): "Cities as Engines of Opportunities: Evidence from Brazil," Working Paper.

BEACH, B., J. PARMAN, AND M. SAAVEDRA (2022): "Segregation and the Initial Provision of Water in the United States," *AEA Papers and Proceedings*, 112, 193–98.

BELCHIOR, C. A., G. GONZAGA, AND G. ULYSSEA (2024): "Who Benefits from Social Housing? Experimental Evidence from a Large-Scale Program in Brazil," Working Paper.

BJORVATN, K. AND A. W. CAPPELEN (2003): "Inequality, Segregation, and Redistribution," *Journal of Public Economics*, 87, 1657–1679.

Boustan, L., F. Ferreira, H. Winkler, and E. M. Zolt (2013): "The Effect of Rising Income Inequality on Taxation and Public Expenditures: Evidence from U.S. Municipalities and School Districts, 1970–2000," *The Review of Economics and Statistics*, 95, 1291–1302.

Boustan, L. P. (2011): "Racial Residential Segregation in American Cities," in *Handbook of Urban Economics and Planning*, ed. by K. D. Nancy Brooks and G. Knaap, Oxford University Press.

Brueckner, J., L. Mation, and V. Nadalin (2019): "Slums in Brazil: Where are they located, who lives in them, and do they 'squeeze' the formal housing market?" *Journal of Housing Economics*, 44, 48–60.

Celhay, P. and R. Undurraga (2022): "Location Preferences and Slums Formation: Evidence from a Panel of Residence Histories," *Regional Science and Urban Economics*, 97, 103816.

Commission, E., J. R. C. (JRC), and Columbia University, Center for International Earth Science Information Network (CIESIN) (2015): "Global Human Settlement Layer (GHSL)," `https://ghsl.jrc.ec.europa.eu/`.

Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2024): "Income Growth and the Distributional Effects of Urban Spatial Sorting," *Review of Economic Studies*, 91, 858–898.

da Rocha Valente, R. and B. J. L. Berry (2022): "Residential Segregation by Skin Color: Brazil Revisited," *Latin American Research Review*, 55.

Davidai, S., D. Goya-Tocchetto, and M. A. Lawson (2024): "Economic segregation is associated with reduced concerns about economic inequality," *Nature Communications*, 15, 5655.

Davis, D., M. Easton, and S. Thies (2024): "Segregation, Spillovers, and the Locus of Racial Change," Working Paper.

Deffebach, P., D. Lagakos, and Y. Miyauchi (2024): "Economic Development and the Spatial Distribution of Income in Cities," Working Paper.

Duncan, O. D. and B. Duncan (1955): "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, 20.

EMBRAPA (2020): "Brazilian Soil Vector Data," `https://www.embrapa.br/en/international`, shapefile, vector data, scale 1:5,000,000, projection EPSG:4326, WGS-84 geographic CRS.

FELER, L. AND J. V. HENDERSON (2011): "Exclusionary Policies in Urban Development: Urban-Servicing Migrant Households in Brazilian Cities," *Journal of Urban Economics*, 69, 253–272.

FRANCIS, A. AND M. TANNURI-PIANTO (2013): "Endogenous race in Brazil: Affirmative action and the construction of racial identity among young adults," 61, 731–753.

FRANKLIN, S. (2019): "The Demand for Government Housing: Evidence from Lotteries for 200,000 Homes in Ethiopia," Working Paper.

FUJITA, M. (1989): "Local Public Goods," in *Urban Economic Theory: Land Use and City Size*, Cambridge: Cambridge University Press, chap. 6, 177–225.

GADGIL, G. U. AND J. L. BAKER (2017): *East Asia and Pacific Cities: Expanding Opportunities for the Urban Poor*, Washington, DC: World Bank.

GARCIA-LÓPEZ, M. AND A. I. MORENO-MONROY (2018): "Income segregation in monocentric an d polycentric cities: Does urban form really matter?" *Regional Science and Urban Economics*, 62–79.

GARMANY, J. AND M. A. RICHMOND (2020): "Hygienisation, Gentrification, and Urban Displacement in Brazil," *Antipode*, 52, 124–144.

GASTWIRTH, J. L. (1972): "The Estimation of the Lorenz Curve and Gini Index," *The Review of Economics and Statistics*, 54.

GECHTER, M. AND N. TSIVANIDIS (2023): "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills Redevelopment," Working Paper.

GOKSU, A., A. BAKALIAN, B. KINGDOM, G. SALTIEL, Y. MUMSSEN, G. SOPPE, J. KOLKER, AND V. DELMON (2019): *Reform and Finance for the Urban Water Supply and Sanitation Sector*, Washington, DC: World Bank.

GONZALEZ-NAVARRO, M. AND C. QUINTANA-DOMEQUE (2016): "Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects," *Review of Economics and Statistics*, 98, 254–267.

HARARI, M. (2020): "Cities in Bad Shape: Urban Geometry in India," *American Economic Review*, 10, 2377–2421.

HARARI, M., V. ALATAS, E. SATRIAWAN, S. SUMARTO, AND M. WONG (2024): "Big Data and Poverty Targeting in Indonesia," Work in progress.

HARARI, M. AND B. STUART (2024): "Racial Segregation and Inequality in Access across U.S. Neighborhoods," Work in progress.

HARARI, M. AND M. WONG (forthcoming): "Slum Upgrading and Long-run Urban Development: Evidence from Indonesia," *The Review of Economic Studies*.

HENDERSON, J. V., T. REGAN, AND A. J. VENABLES (2020): "Building the City: From Slums to a Modern Metropolis," *The Review of Economic Studies*, 88, 1157–1192.

IBGE (2010): "Instituto Brasileiro de Geografia e Estatística, Population Census," .

——— (2020): "Instituto Brasileiro de Geografia e Estatística, Population Census," .

——— (2022): "Instituto Brasileiro de Geografia e Estatística, Desigualdades sociais por cor ou raça no Brasil," .

KHANNA, G., C. MEDINA, A. NYSHADHAM, D. RAMOS, J. TAMAYO, AND A. TIEW (2024): "Spatial Mobility, Economic Opportunity, and Crime," Working Paper.

KRESCH, E., M. WALKER, M. C. BEST, F. GERARD, AND J. NARITOMI (2023): "Sanitation and Property Tax Compliance: Analyzing the Social Contract in Brazil," *Journal of Development Economics*, 160.

KRESCH, E. P. (2020): "The Buck Stops Where? Federalism, Uncertainty, and Investment in the Brazilian Water and Sanitation Sector," *American Economic Journal: Economic Policy*, 12.

LEE, S. AND J. LIN (2018): "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income," *The Review of Economic Studies*, 663–694.

LIN, J. (2015): "The Puzzling Persistence of Place," *Business Review*, Q2, 1–8.

LONDOÑO-VÉLEZ, J. (2022): "The Impact of Diversity on Perceptions of Income Distribution and Preferences for Redistribution," *Journal of Public Economics*, 214, 104732.

NARITOMI, J., R. SOARES, AND J. ASSUNCAO (2012): "Institutional Development and Colonial Heritage within Brazil," *The Journal of Economic History*, 72.

NOAA (2024): "University of Delaware Terrestrial Precipitation," Available: `https://psl.noaa.gov/data/gridded/data.UDel_AirT_Precip.html`.

NOAA NATIONAL GEOPHYSICAL DATA CENTER (2013): "DMSP-OLS Nighttime Lights Time Series," `https://ngdc.noaa.gov/eog/dmsp.html`, accessed: 2024-09-29.

PICARELLI, N. (2019): "There Is No Free House," *Journal of Urban Economics*, 111, 35–52.

REARDON, S. F. AND D. O'SULLIVAN (2004): "Measures of Spatial Segregation," *Sociological methodology*, 34, 121–162.

ROBERTO, E. (2018): "The Spatial Proximity and Connectivity Method for Measuring and Analyzing Residential Segregation," *Sociological Methodology*, 48, 182–224.

ROJAS-AMPUERO, F. AND F. CARRERA (2023): "Sent Away: The Long-Term Effects of Slum Clearance on Children," .

SAIZ, A. (2010): "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 125, 1253–1296.

SWA (2022): "Country Overview: Brazil," Sanitation and Water for All.

TOBLER, W. (1993): "Three Presentations on Geographical Analysis and Modeling," Tech. Rep. 93, National Center for Geographic Information and Analysis.

TROESKEN, W. (2001): "Race, Disease, and the Provision of Water in American Cities, 1889-1921," *Journal of Economic History*, 61, 750–776.

TROUNSTINE, J. (2016): "Segregation and Inequality in Public Goods," *American Journal of Political Science*, 60, 709–725.

TSIVANIDIS, N. (forthcoming): "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio," *American Economic Review*.

U.N. (2018): *World Urbanization Prospects: The 2018 Revision*, New York: United Nations, Department of Economic and Social Affairs, Population Division.

UN HABITAT (2012): "State of the World's Cities 2012/2013," .

VALERIANO, M. D. M. AND D. D. F. ROSSETTI (2012): "Topodata: Brazilian full coverage refinement of SRTM data," *Applied Geography*, 32, 300–309.

VON SPERLING, M. (2016): "Urban Wastewater Treatment in Brazil," Technical Note IDB-TN-930, Inter-American Development Bank.

WHITE, M. J. (1983): "The Measurement of Spatial Segregation," *American Journal of Sociology*, 88, 1008–1018.

WORLD BANK (2024): "World Bank Open Data: Gini Index," https://data.worldbank.org/indicator/SI.POV.GINI, accessed: July 2024.

XU, A. Z. (2023): "Segregation and the Spatial Externalities of Inequality: A Theory of Interdependence and Public Goods in Cities," *American Political Science Review*, 1–18.

ZENOU, Y. (2013): "Spatial versus social mismatch," *Journal of Urban Economics*, 74, 113–132.

Table 1: Predicting the location of poor and rich neighborhoods

	(1)	(2)	(3)	(4)
	Avg. income bottom quartile	Avg. income top quartile	>50% non-white residents	Slum
Slope	-0.032***	0.041***	-0.043***	-0.031***
	(0.011)	(0.011)	(0.010)	(0.008)
Slope^2	0.006***	-0.006***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Slope^3	-0.000***	0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Distance to rivers	-0.077***	0.087**	-0.054*	-0.061***
	(0.025)	(0.035)	(0.030)	(0.010)
Distance^2 to rivers	0.015***	-0.012**	0.005	0.009***
	(0.005)	(0.006)	(0.006)	(0.002)
Distance to lakes	0.051***	-0.064***	0.047***	0.010***
	(0.007)	(0.010)	(0.011)	(0.003)
Distance^2 to lakes	-0.002***	0.002***	-0.002***	-0.000
	(0.000)	(0.001)	(0.001)	(0.000)
Distance to shore	0.018***	-0.037***	0.031***	-0.003
	(0.006)	(0.007)	(0.007)	(0.003)
Distance^2 to shore	-0.000***	0.000***	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	162,297	162,297	162,297	162,297
R-squared	0.058	0.063	0.516	0.157
F statistic	1178	386	1237	169

Notes: Each observation is a census block, from 608 cities. All specifications include city fixed effects. Standard errors clustered at the city level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table 2: First stage

Dependent variable: distance segregation, km

	(1)	(2)	(3)
	Income	Race	Slum
Geography distance-segregation, km	0.613***	0.560***	0.399***
	(0.046)	(0.071)	(0.078)
Equivalent area radius, km	0.510***	0.504***	0.671***
	(0.072)	(0.092)	(0.093)
Average elevation, m	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Ruggedness, m	0.000	0.001	-0.000
	(0.001)	(0.001)	(0.001)
Average slope, degrees	-0.031	0.084	0.027
	(0.048)	(0.116)	(0.110)
Water bodies within 30 km, sqkm	-0.000	-0.001	0.001**
	(0.000)	(0.001)	(0.001)
Low-fertility soil within 30 km, sqkm	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
% Land available within 30km	-0.630*	0.202	0.984
	(0.377)	(0.545)	(0.720)
Landslide risk	0.074	-0.434	0.159
	(0.222)	(0.511)	(0.620)
Slope adjustment factor for distance	-2.647	-5.527**	1.825
	(2.141)	(2.609)	(5.593)
Distance to state capital, km	-0.000	-0.001	-0.000
	(0.000)	(0.000)	(0.001)
Distance to Atlantic, km	-0.000	-0.001**	0.001
	(0.000)	(0.000)	(0.000)
Observations	608	458	229
R-squared	0.905	0.834	0.863
F statistic	174	62	27

Notes: Each observation is a city. All specifications include as additional controls latitude, longitude, precipitation, sunshine, and soil type dummies. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table 3: City-wide public goods

	(1)	(2)	(3)	(4)
Distance-segregation by income, km	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km
IV	-0.022***	-0.016***	-0.049***	-0.020***
	(0.008)	(0.005)	(0.013)	(0.004)
OLS	-0.015**	-0.015***	-0.036***	-0.020***
	(0.006)	(0.004)	(0.009)	(0.003)
Observations	608	608	608	608
R-squared	0.578	0.468	0.605	0.179
IV F statistic	174	174	174	174
Mean dep. var.	0.567	0.898	0	0.629

Notes: Each observation is a city. All specifications include the controls listed in Table 2. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table 4: Specification and functional form

IV estimates on distance segregation	(1)	(2)	(3)	(4)		
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km	Obs.	IV F stat.
A. Population density-weighted distance segregation	-0.032***	-0.024**	-0.072***	-0.030***	608	302
	(0.011)	(0.009)	(0.018)	(0.007)		
B. Index based on log distance	-0.170***	-0.094**	-0.253*	-0.084	608	202
	(0.064)	(0.045)	(0.141)	(0.058)		
C. Exposure index	0.062***	0.028***	0.160***	0.025***	608	841
	(0.017)	(0.011)	(0.038)	(0.006)		
D. State FEs	-0.021***	-0.016***	-0.058***	-0.022***	608	153
	(0.008)	(0.005)	(0.016)	(0.004)		
E. Coarser neighborhood units	-0.023***	-0.014	-0.050***	-0.019***	607	47
	(0.008)	(0.009)	(0.015)	(0.006)		

Notes: Each observation is a city. This table reports IV estimates for a specification similar to that in Table 3. Each panel considers an alternative versions of the distance segregation metrics, as detailed in Section 6.2.2. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table 5: Public goods by neighborhood type

	(1)	(2)	(3)	(4)	(5)
	Share residents with public sewerage	Share residents with public water	Share residents with paved streets	Share residents with sidewalks	OSM amenties within 3km
A. Poor	-0.035***	-0.022***	-0.019**	-0.031***	-0.022***
	(0.009)	(0.007)	(0.008)	(0.009)	(0.006)
	[0.492]	[0.880]	[0.619]	[0.449]	[0.564]
A. Rich	-0.025***	-0.013*	-0.020***	-0.023***	-0.014*
	(0.010)	(0.007)	(0.005)	(0.007)	(0.007)
	[0.684]	[0.930]	[0.912]	[0.841]	[0.706]
C. Poor / Rich	-0.009	-0.015*	-0.005	-0.023**	-0.012
	(0.024)	(0.008)	(0.008)	(0.009)	(0.009)
	[0.726]	[0.950]	[0.661]	[0.513]	[0.827]
D. Poor close to Rich	-0.037***	-0.018**	-0.022**	-0.039***	-0.018**
	(0.010)	(0.007)	(0.009)	(0.010)	(0.008)
	[0.563]	[0.911]	[0.696]	[0.524]	[0.681]
E. Poor far from Rich / Poor	-0.000	0.010	0.014	0.013	-0.020**
	(0.019)	(0.016)	(0.017)	(0.019)	(0.009)
	[0.785]	[0.961]	[0.848]	[0.759]	[0.771]
Observations	580	580	580	580	580

Notes: Each observation is a city. This table reports IV coefficients for distance segregation by income from a specification similar to that in Table 3. Each column corresponds to a different outcome variable. Each panel shows results for different aggregations of the outcome variable at the city level. In Panel A, the dependent variables are outcomes averaged among the poor neighborhoods in the city. Panel B corresponds to averages among the rich neighborhoods. In Panel C, the dependent variables are the ratios between the average among the poor and rich. In Panel D, the outcomes are averaged among poor neighborhoods that are close to rich neighborhoods (bottom quartile by distance). Panel E considers the ratio between averages among the poor neighborhoods that are far from the rich (top quartile by distance) and poor neighborhoods in general. The table reports IV point estimates, standard errors in parentheses, and sample averages of each dependent variable in square brackets. F statistics are all above 100. Standard errors are clustered at the meso-region level. *** p>0.01, ** p>0.05, * p>0.1.

Table A1: Correlates of neighborhood-level public goods access: income.

	(1) % Residents pub. sewerage	(2) % Residents pub. sewerage	(3) % Residents pub. water	(4) % Residents pub. water	(5) Neighborhood pub. goods index	(6) Neighborhood pub. goods index	(7) OSM amenities within 3km	(8) OSM amenities within 3km
Share P	-0.554***	-0.325***	-0.108***	-0.012	-1.897***	-1.594***	-0.537***	-0.088***
	(0.035)	(0.026)	(0.018)	(0.013)	(0.094)	(0.093)	(0.019)	(0.015)
Log distance to R blocks		-0.112***		-0.050***		-0.119***		-0.237***
		(0.018)		(0.014)		(0.016)		(0.018)
Log distance to CBD		-0.021***		-0.004		-0.039***		-0.032***
		(0.006)		(0.005)		(0.007)		(0.007)
Observations	152,305	152,305	152,305	152,305	152,305	152,305	152,305	152,305
R-squared	0.571	0.589	0.445	0.458	0.556	0.570	0.428	0.577
Geography controls	N	Y	N	Y	N	Y	N	Y
Mean dep. var.	0.728	0.728	0.932	0.932	-0.0010	-0.00100	0.698	0.698

Notes: Each observation is a census block. All specifications include 608 city fixed effects and controls for total number of residents. Geographic controls are listed in Figure 7. Standard errors clustered at the city level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A2: Correlates of neighborhood-level public goods access: race.

	(1) % Residents pub. sewerage	(2) % Residents pub. sewerage	(3) % Residents pub. water	(4) % Residents pub. water	(5) Neighborhood pub. goods index	(6) Neighborhood pub. goods index	(7) OSM amenities within 3km	(8) OSM amenities within 3km
Share P	-0.418***	-0.172***	-0.058***	0.034**	-1.523***	-1.180***	-0.496***	-0.138***
	(0.039)	(0.028)	(0.014)	(0.016)	(0.085)	(0.086)	(0.021)	(0.020)
Log distance to R blocks		-0.178***		-0.071***		-0.219***		-0.228***
		(0.016)		(0.015)		(0.019)		(0.017)
Log distance to CBD		-0.027***		-0.006		-0.042***		-0.055***
		(0.006)		(0.005)		(0.008)		(0.006)
Observations	162,297	140,415	140,415	140,415	140,415	140,415	140,415	140,415
R-squared	0.107	0.522	0.558	0.407	0.429	0.479	0.507	0.552
Geography controls	N	Y	N	Y	N	Y	N	Y
Mean dep. var.	0.719	0.719	0.937	0.937	0.0200	0.0200	0.700	0.700

Notes: Each observation is a census block. All specifications include 458 city fixed effects and controls for total number of residents. Geographic controls are listed in Figure 7. Standard errors clustered at the city level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A3: Correlates of geography distance-segregation

	(1)	(2)
	OLS	Sample mean
Average elevation (m)	-13.230***	415
	(4.874)	
Ruggedness (m)	2.360	94
	(2.577)	
Average slope (degrees)	-0.004	5
	(0.078)	
Water bodies within 30 km (sqkm)	-1.379	55
	(1.811)	
Low-fertility soil within 30 km (sqkm)	-31.439**	1522
	(14.271)	
% Land available within 30km	0.005	0.842
	(0.004)	
Landslide risk	0.000	1.460
	(0.010)	
Slope adjustment factor for distance	0.000	1.190
	(0.000)	
Distance to state capital (km)	-9.838***	181
	(3.306)	
Distance to Atlantic (km)	-12.384**	216
	(5.413)	
Latitude	0.066	-16.76
	(0.173)	
Longitude	0.189*	-45.83
	(0.098)	
Precipitation annual avg. (mm/day)	1.204*	124.4
	(0.632)	
Sunshine annual avg. (wh/m2.day)	-2.045	4934
	(10.128)	
Soil type: neosoils	0.004	0.350
	(0.009)	
Soil type: planosoils	-0.002	0.183
	(0.006)	
Soil type: argisoils	-0.009	0.734
	(0.008)	
Soil type: luvisoils	-0.007*	0.063
	(0.004)	
Soil type: gleisoils	0.016**	0.169
	(0.007)	
Soil type: latosoils	-0.008	0.658
	(0.008)	
Soil type: nitosoils	-0.005**	0.071
	(0.002)	
Soil type: plintosoils	-0.002	0.066
	(0.005)	
Soil type: cambisoils	0.007	0.339
	(0.008)	
Soil type: espondosoils	0.006	0.071
	(0.005)	
Observations		608

Notes: Column 1 reports coefficients and standard errors (in parentheses) from OLS regressions of each control variable on geography distance segregation by income, controlling for city radius, with standard errors clustered at the meso-region level. Column 2 reports sample means of each control variable.

Table A4: Sample cuts

IV estimates on distance segregation	(1)	(2)	(3)	(4)		
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km	Obs.	IV F stat.
A. Exclude state capitals	-0.021**	-0.018***	-0.053***	-0.020***	581	136
	(0.010)	(0.005)	(0.016)	(0.004)		
B. Exclude near state capitals	-0.013*	-0.016***	-0.044***	-0.017***	461	95
	(0.007)	(0.006)	(0.013)	(0.005)		
C. Exclude elevated	-0.022**	-0.017***	-0.046***	-0.022***	549	135
	(0.010)	(0.006)	(0.014)	(0.005)		
D. Exclude coastal	-0.028***	-0.003	-0.043**	-0.022***	517	101
	(0.009)	(0.008)	(0.020)	(0.007)		
E. Exclude largest	-0.021**	-0.017***	-0.048***	-0.022***	577	152
	(0.009)	(0.006)	(0.015)	(0.004)		
F. Exclude smallest	-0.024**	-0.020***	-0.053***	-0.022***	577	370
	(0.009)	(0.004)	(0.014)	(0.004)		
G. Exclude most fertile	-0.031***	-0.013*	-0.058***	-0.020***	548	134
	(0.009)	(0.008)	(0.016)	(0.007)		

Notes: Each observation is a city. Specifications are analogous to those in Table 3, but exclude certain cities from the sample. Cities near state capitals are defined as within 42 km of a state capital. Elevated cities are above 850m. Coastal cities have a CBD within 5 km from the coast. The largest and smallest cities are in the top and bottom 5% by total population. Most fertile cities are the bottom 10% of cities by the share of low-fertility soil.

Table A5: City-wide public goods

Panel A: race				
	(1)	(2)	(3)	(4)
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km
IV	-0.034***	-0.013***	-0.073***	-0.025***
	(0.010)	(0.005)	(0.018)	(0.007)
OLS	-0.019***	-0.015***	-0.040***	-0.017***
	(0.006)	(0.006)	(0.013)	(0.005)
Observations	458	458	458	458
R-squared	0.495	0.413	0.482	0.120
IV F statistic	62	62	62	62
Mean dep. var.	0.642	0.926	0.137	0.644

Panel B: slum status				
	(1)	(2)	(3)	(4)
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km
IV	-0.054***	-0.021*	-0.086**	-0.013
	(0.016)	(0.012)	(0.036)	(0.011)
OLS	-0.024***	-0.012	-0.025	-0.009
	(0.009)	(0.008)	(0.021)	(0.008)
Observations	229	229	229	229
R-squared	0.566	0.418	0.488	0.220
IV F statistic	27	27	27	27
Mean dep. var.	0.598	0.897	0.0399	0.659

Notes: Each observation is a city. Specifications are similar to those in Table 3. Panel A (B) considers distance segregation by race (slum status). Standard errors are clustered at the meso-region level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A6: Additional proxies for city disconnectedness

IV estimates on distance segregation	(1)	(2)	(3)	(4)		
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km	Obs.	IV F stat.
A. Disconnected city	–0.021**	–0.017***	–0.051***	–0.022***	608	133
	(0.009)	(0.006)	(0.013)	(0.004)		
B. Nr water basins	–0.024***	–0.018***	–0.053***	–0.022***	608	162
	(0.009)	(0.006)	(0.014)	(0.005)		
C. Nr polygons, available land within 30km	–0.023***	–0.016***	–0.052***	–0.021***	608	173
	(0.009)	(0.006)	(0.013)	(0.004)		
D. Perimeter/area ratio, available land within 30km	–0.022***	–0.016***	–0.051***	–0.021***	608	171
	(0.008)	(0.006)	(0.013)	(0.004)		

Notes: Each observation is a city. Specifications are analogous to those in Table 3, but include additional controls. Standard errors are clustered at the meso-region level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A7: Robustness: residualized outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share residents with public sewerage		Share residents with public water		Neighborhood public goods index		OSM amenties within 3km	
Residualized by	topography	topography + distance	topography	topography + distance	topography	topography + distance	topography	topography + distance
Distance-segregation by income, km								
IV	-0.021**	-0.019**	-0.015***	-0.016***	-0.059***	-0.052***	-0.019***	-0.013***
	(0.009)	(0.008)	(0.006)	(0.005)	(0.016)	(0.015)	(0.004)	(0.004)
Observations	608	608	608	608	608	608	608	608
R-squared	0.211	0.143	0.219	0.129	0.184	0.224	0.169	0.198
IV F statistic	174	174	174	174	174	174	174	174
Mean dep. var	-0.013	-0.075	0.041	-0.019	0	0	-0.0588	-0.0732
StDev dep. var.	0.251	0.239	0.144	0.133	0.583	0.592	0.206	0.210

Notes: Each observation is a city. Specifications are analogous to those in Table 3, with different versions of the dependent variable. Standard errors clustered at the meso-region level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A8: Other public goods measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Distance to nearest, m					
Distance-segregation by income, km	Share municipal spending on health	Nr hospital beds per 1000 ppl	Park	Post office	Police station	Fire station	Share residents with paved streets	Share residents with sidewalks
IV	-0.008***	-0.191***	142.172***	591.432***	416.979***	606.851***	-0.013***	-0.023***
	(0.002)	(0.046)	(34.084)	(96.068)	(92.684)	(143.025)	(0.004)	(0.006)
Observations	604	605	604	426	507	336	608	608
R-squared	0.154	0.100	0.323	0.312	0.152	0.360	0.381	0.487
Mean dep. var.	0.238	2.344	623.5	2505	2023	3121	0.767	0.641

Notes: Each observation is a city. Specifications are analogous to those in Table 3. The dependent variable in columns 3 though 6 is distance in meter to the nearest OSM amenity. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table A9: City-wide public goods: comparison with dissimilarity

Panel A

	(1)	(2)	(3)	(4)	(5)	(6)
	Share residents with public sewerage			Share residents with public water		
	Dissimilarity OLS	Dissimilarity OLS	Distance segregation IV	Dissimilarity OLS	Dissimilarity OLS	Distance segregation IV
Index	0.562***	0.025	-0.018**	0.312***	0.148**	-0.016***
	(0.148)	(0.187)	(0.009)	(0.098)	(0.074)	(0.005)
Log municip. expenditure		0.050***	0.126***		0.000	0.041***
		(0.015)	(0.022)		(0.007)	(0.014)
Log avg. income		0.133**	0.136***		0.038	0.053**
		(0.058)	(0.049)		(0.023)	(0.022)
R-squared	0.584	0.607	0.628	0.464	0.462	0.495
Observations	604	604	604	604	604	604

Panel B

	(1)	(2)	(3)	(4)	(5)	(6)
	Neighborhood public goods index			OSM amenities within 3km		
	Dissimilarity OLS	Dissimilarity OLS	Distance segregation IV	Dissimilarity OLS	Dissimilarity OLS	Distance segregation IV
Index	1.797***	-0.035	-0.034***	0.637***	0.223*	-0.017***
	(0.351)	(0.274)	(0.011)	(0.082)	(0.117)	(0.004)
Log municip. expenditure		0.074***	0.282***		0.009	0.039**
		(0.025)	(0.035)		(0.011)	(0.015)
Log avg. income		0.669***	0.673***		0.202***	0.226***
		(0.102)	(0.083)		(0.034)	(0.028)
R-squared	0.630	0.691	0.729	0.205	0.251	0.271
Observations	604	604	604	604	604	604

Notes: Each observation is a city. Specifications are similar to the OLS and IV regressions reported in Table 3. The segregation index considered in columns 1, 2, 4, and 5 is the dissimilarity index computed using households in the bottom quartile of the city income distribution as the minority group. Columns 3 and 6 report the same IV as in Table 3, augmented controlling for log municipal expenditure and log average income in 2010. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Table A10: Human capital and demographics

	(1)	(2)	(3)	(4)	(5)	(6)
Distance segregation by income, km	Literacy rate	% Employment in service sector	% High school degree or higher	% Non-White	% Prime age males	% Aged 65+
IV	-0.003***	-0.006***	-0.007***	0.005	0.000	-0.004***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.000)	(0.001)
Observations	608	608	608	608	608	608
R-squared	0.753	0.455	0.541	0.777	0.412	0.392
IV F statistic	174	174	174	174	174	174
Mean dep. var	0.908	0.412	0.366	0.519	0.001	0.138

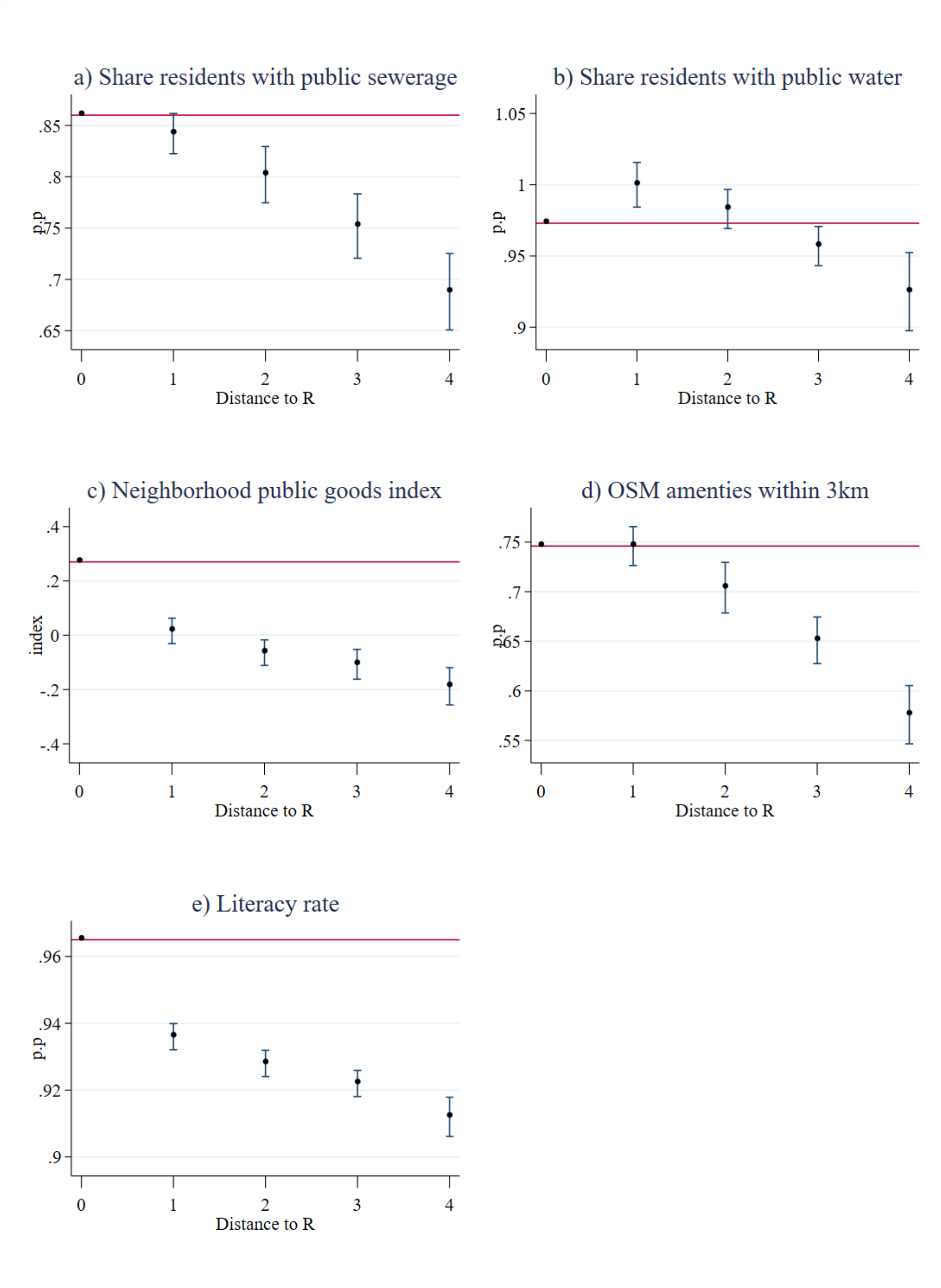
Notes: Each observation is a city. All specifications include the controls listed in Table 2. Standard errors clustered at the meso-region in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A11: Role of municipal companies

	(1)	(2)	(3)	(4)
IV	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenties within 3km
Distance-segregation	-0.018**	-0.016***	-0.045***	-0.021***
	(0.008)	(0.004)	(0.013)	(0.004)
Distance-segregation X municipal company	-0.018**	-0.012**	-0.021	-0.005
	(0.008)	(0.005)	(0.019)	(0.007)
Municipal company	0.165***	0.107***	0.287***	0.046
	(0.038)	(0.032)	(0.090)	(0.034)
Observations	593	593	593	593
R-squared	0.593	0.499	0.620	0.175
IV F statistics	225	225	183	183
	183	183	225	225
Mean dep. var.	0.575	0.903	0.0185	0.632

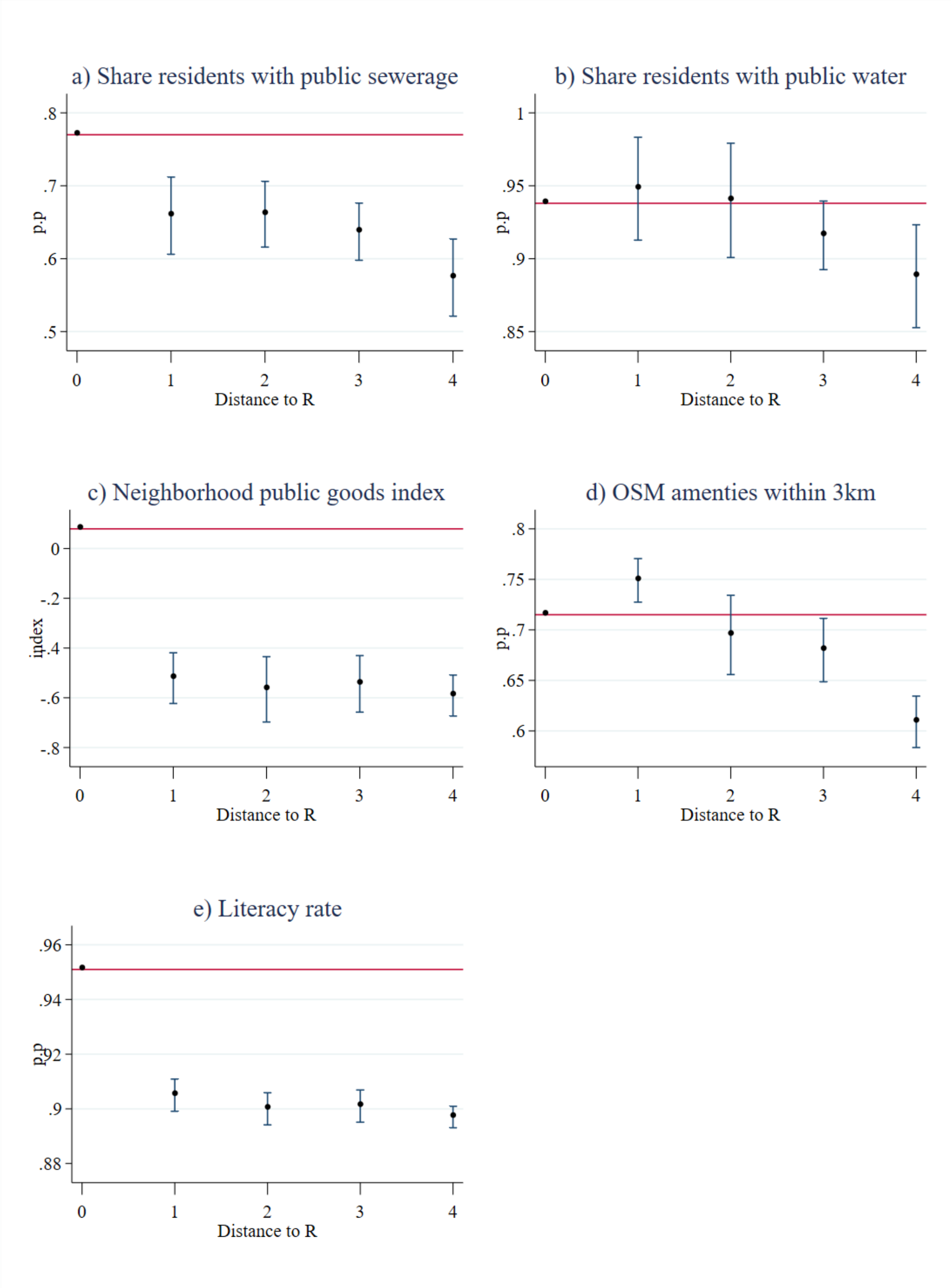
Notes: Each observation is a city. All columns report IV estimates. All specifications include the controls listed in Table Table 2. Standard errors clustered at the meso-region level in parentheses. *** p>0.01, ** p>0.05, * p>0.1.

Figure A1: Correlates of distance segregated neighborhoods by race



Notes: this figure is similar to Figure 7 but P (R) blocks are defined as those that are predominantly non-white (white).

Figure A2: Correlates of distance segregated neighborhoods by slum status



Notes: this figure is similar to Figure 7 but P (R) blocks are defined as slums (non-slums).