Residential patterns and local public goods in urban Brazil*

Mariaflavia Harari†
University of Pennsylvania

Preliminary and incomplete. Please do not cite or circulate.

This version: August 2024

Latest version downloadable here.

Abstract

Millions of migrants in developing countries move to urban areas in search of better prospects, but access to economic opportunities varies widely within cities. Yet, we know little about spatial inequalities within developing country cities. This paper studies patterns of residence and local public goods access by socio-economic status within Brazilian cities, making two contributions. First, I innovate on the measurement of residential segregation by computing spatial metrics of "distance segregation", capturing the physical proximity between low- and high-socio-economic status neighborhoods. I document descriptive patterns by income, race, and informality. Within cities, distance-segregated neighborhoods have lower access to local public goods. Second, I explore the causal link between residential patterns and access to public goods. Using an instrumental variables strategy based on within-city geography, I investigate how the levels and spatial allocation of public goods are affected by segregation. Cities where the poor and the rich live further apart have a lower share of households connected to sewerage and water, worse neighborhood quality, and lower access to public amenities. I consider mechanisms such as redistributive preferences, externalities across neighborhoods, and strategic underprovision of public services to deter the poor. Additionally, I provide suggestive evidence on some of the historical determinants of distance segregation, such as infectious diseases. 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, Vinicios Sant'Anna, Nick Tsivanidis, 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, and NBERSI 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 Vance 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 improved economic prospects. As cities accommodate in-migration by poor households, a variety of residential patterns emerge: in some cases, the urban poor concentrate in peripheral, low-accessibility areas, in contrast with wealthy households located near the city center. At the same time, in many cities, slums coexist with wealthy neighborhoods in highly accessible areas (Baker and Gadgil, 2017). While physical proximity to high-opportunity neighborhoods is valuable for the urban poor (Barnhardt et al., 2017), policy makers are often concerned about negative externalities and opportunity costs associated with poor neighborhoods located in high land value areas (Henderson et al., 2020). This tension lies at the heart of the debate on many consequential urban policies, such as slum upgrading on site (Harari and Wong, 2024) and the spatial targeting of public goods (Feler and Henderson, 2011).

While a large U.S.-based literature describes residential patterns by income and race and the impacts of segregation on a variety of outcomes, we know very little about residential patterns within developing country cities. One of the challenges is the lack of granular, geo-referenced data at the intra-urban level. This paper contributes to the debate on how to accommodate urbanization in developing countries by studying the spatial distribution of urban households by socio-economic status in Brazilian cities and its implications for public goods provision. Brazil is an ideal setting, being a rapidly urbanizing country with a large number of cities and featuring stark inequalities with respect to income and race.

There are three main contributions. First, I innovate on the measurement of residential segregation by developing spatial metrics that quantify "distance segregation" across groups. I document descriptive patterns by three significant markers of socio-economic status: income, race, and informal status. Second, I make progress on the identification of the causal impacts of residential patterns. I propose an instrumental variables strategy that isolates the variation in distance segregation due to geography and alleviates concerns associated with the endogeneity of residential patterns to public goods provision. Third, leveraging this strategy, I show that segregation leads to lower levels of access to public services and affects the spatial patterns of public goods access.

I begin by developing metrics to quantify "distance segregation" across groups which capture the extent to which poor and rich locate close to each other. These indexes innovate over conventional segregation measures used in the literature (e.g. the dissimilarity index) by considering *where* different neighborhoods are located and not just the proportion of different groups within each. The physical distance between neighborhoods matters for the spatial decay of externalities

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

(e.g. negative congestion externalities from slums), for access to jobs and amenities, and for the engineering costs of providing network goods (e.g. sewerage). These are all important inputs in policy decisions related to housing and local public goods provision (Kresch et al., 2020). Under my proposed metrics, a city can be considered distance segregated when the average distance between poor and rich neighborhoods is large relative to the average distance between any two neighborhoods. Intuitively, a city where neighborhoods are randomly distributed in space has a normalized index close to 1 and appears as a "checkerboard" of poor and rich neighborhoods.

Using these metrics, I document descriptive patterns of distance segregation across a sample of approximately 600 urban municipalities in Brazil using granular data from the 2010 Census, at the block (*setor censitario*) level. The median block in my sample contains around 200 households. The modal city in Brazil is distance segregated by income, meaning 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. There is also distance segregation by race, contrasting predominantly white with predominantly non-white blocks, but it appears less marked. Interestingly, slums are 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 (Baker and Gadgil, 2017). Turning to neighborhood-level evidence, I find that distance-segregated blocks have lower access to basic public services, conditional on own neighborhood characteristics and and distance to the Central Business District (CBD).

Motivated by the stylized facts above, I turn to the causal question of whether distance segregation affects public goods provision. A naïve regression of city-wide residential patterns on public goods access would be difficult to interpret due to many sources of endogeneity, which could bias the estimation in different directions. For example, households may sort within cities in response to public goods provision. Additionally, 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 go hand in hand with inefficient public goods provision.

To address these identification concerns, I develop an instrumental variables approach that leverages variation in a city's proneness to spatial segregation stemming from the distribution of geographic features within cities. The starting point for the instrument is noting that "first nature", or the spatial distribution of exogenous amenities, is an important determinant of residential patterns, in line with the intuition in Lee and Lin (2018) that natural amenities anchor rich neighborhoods in the U.S. I find that steep slopes and proximity to rivers are predictors of the location of poor neighborhoods, while rich neighborhoods tend to locate in areas with mild slopes and near the coast. In the next step, I replicate the construction of the distance segregation indexes using the location of geography-predicted poor neighborhoods instead of the actual ones. The ensuing

instrument is based on the distance between places with different combinations of slope and water bodies presence. Generic city-wide geographic characteristics such as average elevation, slope, ruggedness, or distance to the coast are included as controls in the first-stage regression. The main results are robust to a range of sample cuts, such as excluding coastal cities, mountainous cities, or state capitals.

With this instrument in hand, I turn to investigating how city-wide distance segregation by income affects public goods access, both in terms of overall levels and targeting of neighborhoods. I find that more distance segregated cities tend to have lower levels of basic urban services provision. As primary outcomes, I consider access to the public sewerage and water system, which are key urban services that are still not universally provided in Brazilian cities (Feler and Henderson, 2011) and a composite index of neighborhood public amenities, such as paved roads and sidewalks, both from the 2010 Census. Additionally, I also construct a measure of access to public amenities such as police stations and parks based on OpenStreetMap. For a one standard deviation increase in segregation, corresponding for the median city to 1.7 additional km between the average poor and rich neighborhood, the share of households with access to public sewerage and public water decreases by 3.4 and 3 percentage points 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.

A potential concern is that these results may be driven by higher costs of providing public goods in cities with more complex topography. I address this by showing that the results are robust to considering a residualized measure of public goods access where I control for block-level geographic characteristics. Moreover, these patterns are not limited to public goods that are delivered along a spatial grid but are detected also when considering other proxies, such as the share of municipal spending on health.

Having established that more segregation is associated with lower levels of public goods provision, I shed light on mechanisms by considering how the spatial pattern of access across neighborhoods and groups varies between segregated and integrated cities. First, I show that segregated cities tend to provide lower public goods access to both poor and rich neighborhoods. This suggests that residents in segregated cities may have a lower propensity towards redistribution and funding public goods in general. This plausibly stems from a combination of sorting and direct effects of residential patterns shaping preferences.[2]

Another potential explanation is related to the engineering costs of expanding the network for urban services such as sewerage or water connections. When poor neighborhoods are interspersed

[2]While inherently hard to disentangle the two, I find limited evidence of sorting by demographic characteristics: distance segregation does not appear to affect city-wide racial shares and the share of prime-age males (more likely to be migrants).

with rich ones, they may be more likely to be covered simply because they are along the network expansion path. However, I find that even poor neighborhoods that are located near rich ones appear under-served in segregated cities relative to integrated ones. This suggests that the engineering channel alone cannot explain the overall result.

Finally, I consider the relative provision of public goods in segregated versus integrated poor neighborhoods, conditional on the overall levels of public goods provision to poor areas. 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 underprovision of poor neighborhoods that are near rich ones (Feler and Henderson, 2011), in an attempt to deter the poor from settling. I find that, for positional public goods, the latter channel prevails in segregated cities: the gap in access to sewerage and water between the segregated and the integrated poor is less stark in segregated cities. In other words, while the segregated poor within cities tend to have worse access than the integrated poor, in segregated cities the integrated poor do not do much better. Interestingly, this pattern does not hold for access to amenities that residents can travel to, such as police stations and parks, where the first channel seems to be more relevant.

A remaining question is why certain cities are more distance-segregated than others. Beyond geography, "second nature" factors such as externalities and coordination are likely important drivers. I consider several potential historical determinants of present-day segregation. For example, I show that exposure to infectious diseases in the 19th century is associated with greater segregation today. This is in line with historical accounts documenting that disease outbreaks, commonly attributed to unsanitary conditions in dense tenements, were often followed by the demolition of poor neighborhoods in the city centers, effectively pushing the poor to the periphery. Along the same lines, I find that cities with a history of slavery are more race-segregated today. These results open interesting avenues for future research on the historical roots of segregation patterns.

Taken together, these results exemplify the role of urban policies in shaping the spatial distribution of socio-economic status within cities and illustrate how residential patterns and public goods provision are jointly determined. My results highlight a key tradeoff that policy makers in developing countries consider when faced with spatial inequalities within their cities: one response is to remediate the negative externalities from poor neighborhoods by providing public goods, while the other is to deter the poor from settling in by strategically withdrawing public services. This tradeoff underscores the debate on important urban policies, such as on-site slum upgrading versus forced relocations and affordable housing provision. Additionally, the evidence on historical persistence suggest that systematic underprovision of public goods in segregated cities may be hard to break and raises the concern of potential spatial poverty traps for the distance-segregated poor.

This paper contributes to a nascent literature on residential segregation in developing countries.[3] Recently, Asher et al. (2023) have considered segregation by caste and religion in India, showing worse outcomes and lower public goods access in neighborhoods that are predominantly Muslims and low-caste.[4] My paper complements their analysis, with two key differences. The first is to introduce distance to the measurement of segregation, leveraging the detailed maps provided by the Brazilian Census. The vast majority of the literature measures segregation uses the conventional dissimilarity index (see Fossett (2017) for an overview), which does not incorporate space but only focuses on the internal composition of neighborhoods relative to the share of different groups in the city as a whole.[5] The second difference is that my paper makes progress towards establishing causal evidence of how residential patterns affect public goods provision, by developing an instrumental variables approach.

This paper is also connected to to the literature on the provision of local public goods in developing countries.[6] In particular, Feler and Henderson (2011) provide evidence of strategic underprovision 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 lend support to both mechanisms linking public goods provision to residential patterns, using an instrumental variables approach to mitigate endogeneity concerns.

Finally, this paper is also broadly connected with the literature in urban and development on spatial policies targeting the poor, such as relocations and public housing (Rojas-Ampuero and Carrera, 2023, Picarelli, 2019, Franklin, 2019), urban infrastructure (Tsivanidis, 2023, Khanna et al., 2024, Gonzalez-Navarro and Quintana-Domeque, 2016), and sanitation (Kresch et al., 2020).

2 Background and Data

Brazil is an ideal setting to investigate spatial inequalities within cities and public goods access. Ranking 5th in the world by urban population, Brazil has over 250 cities above 100 thousand

[3] 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).

[4] 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).

[5] Versions of the dissimilarity index that incorporate space in some form have been proposed by sociologists (Reardon and O'Sullivan, 2004) but have found very limited implementation in the economics literature. The distance-based metrics I employ are more closely related to measures of city compactness (Harari, 2020), clustering (Duranton and Overman, 2005), or access (Pereira et al., 2019).

[6] 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 (Abramitzky 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.

inhabitants (IBGE, 2010) and overall nearly 200 million urban residents (IBGE, 2020). A middle income economy, it is the ninth most income unequal country in the world with a Gini index of 0.53 (World Bank, 2023). It faces enduring racial inequality with mixed-race and black Brazilians (approximately 58% of the total population) systematically lagging behind white Brazilians 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 grapples with persistent informality with 11 million estimated slum residents in 2010 (UN Habitat, 2016).

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 population Census. Specifically, I employ the data released at the *setor* (census block) level. 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

I follow the literature on local public finance in Brazil (e.g. Feler and Henderson (2011)) and focus on public goods that are primarily managed at the local level by municipalities (*municípios*). Unless otherwise noted, the measures I consider are drawn 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 network. These are critical urban services as poor sanitation entails major public health costs. Approximately half of the urban population in Brazil lacks access to the sewerage network. Access to the water system is more prevalent, with about 90% of urban households being connected.

I also consider a number of 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 surroundings have paved streets, sidewalks, no accumulated street garbage, no open sewer, street addresses, street lighting, curbs, manholes, ramps, and greenery. I aggregate these characteristics into a standardized neighborhood public goods index.

In addition to the Census measures, I also code the location of a number of public amenities from OpenStreetMap (OSM). Specifically I consider fire stations, police stations, post offices, and parks, which appear to be well documented in OSM. 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 them. I consider other measures - such as distance to the nearest amenity - in the Appendix. Unlike the Census-based measures, which capture positional public goods to be enjoyed on site, the OSM data capture public amenities that people from different neighborhoods

can travel to.

Additional proxies I consider include the 2010 municipal expenditure, retrieved from the Secretaria do Tesouro Nacional (STN, National Treasury Secretariat), and data on number of hospital beds from DATASUS, the Brazilian Public Unified Health System (SU.S.) department.

2.2 Neighborhood characteristics

I consider segregation along three dimensions of socio-economic status: income, race, and informality.

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 the average income is in the bottom or the top quartiles of the municipality. I also show some descriptive results using absolute income definitions defined in reference to the minimum wage.

The advantage of employing a relative definition of income is that it ensures we can identify poor and rich neighborhoods in all cities in the 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 drawbacks of this approach is that the poor and rich may not be 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.[7] Considering race is important in and of itself and helps inform the debate on racial inequality in Brazil. Additionally, as a marker of low socio-economic status it may be preferable to income because it is comparable across cities and tied to fixed individual characteristics.

The third dimension is whether a block is a slum. I employ the Census designation of *aglomerado subnormal*. The latter 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 223 municipalities when considering segregation by informal status. These tend to be the largest cities.

2.3 Definition of city

In the context of Brazil, local public service provision decisions take place at the municipality level. However, municipality boundaries often grossly overestimate the geographic extent of the

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

actual urban areas, by comprising blocks that are large in size and largely uninhabited. Including these areas would make measures based on distance potentially misleading.

As baseline units for the computation of my indexes I consider "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).[8] The resulting spatial units are similar largely overlap with the built-up footprint of cities as they appear from satellite imagery (e.g. the Global Human Settlement Layer (GHLS)) 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.

There are also instances of large urban agglomerations spanning across multiple municipalities. The baseline results are robust to controlling for whether a municipality is the suburb of a larger agglomeration. To account for correlated shocks within broad urban areas, standard errors are clustered at the meso-region (a Census unit in between municipality and state) throughout the analysis. The political economy implications of having multiple jurisdictions within the same functional urban areas are to be explored in future research.

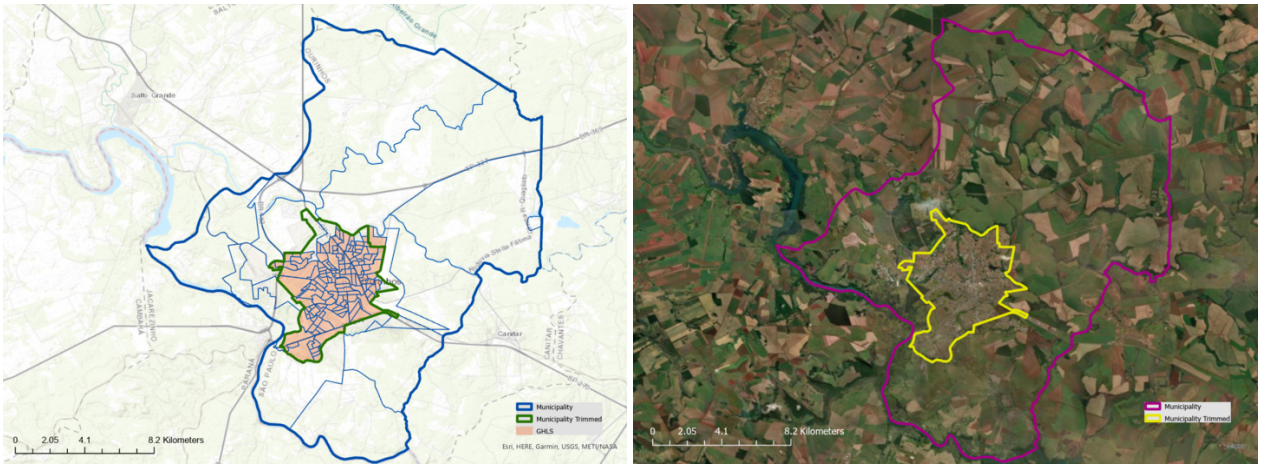
For some of the analyses below I consider distance to the Central Business District (CBD). I define it as the historical center of cities and identify it using a semi-automated procedure. I search the keywords "centro historico" followed by the name of each municipality and state in Google Maps and then 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 in the DSM-OLS night lights dataset.

2.4 Geography

Several block-level and city-level geographic variables are employed for the construction of the instrument and as controls (Table A1). Slope, ruggedness, and elevation are calculated from the Topodata dataset (Valeriano (2012)). The raw source is Shuttle Radar Topography Mission

[8] 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.

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.

(SRTM) data at the 90 x 90 resolution. The distance between blocks and water bodies is calculated using rivers, streams, and lakes are from OSM. The extent of water bodies is drawn from the ASTER dataset, with resolution of 30 meters (NASA/METI).Additional controls (such as soil type dummies and climate variables) were obtained from Naritomi et al. (2012).

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. A notable example is Rio de Janeiro where *favelas* coexist with high-rises in some of the highest-amenity areas along the coast. This contrasts with the patterns observed in other cities, such as São Paulo, where poor neighborhoods tend to be peripheral and spatially isolated.

A priori, there are several reasons why these patterns are interesting objects of study. First, physical distance from rich neighborhoods determines access to jobs and economic opportunities. Second, externalities typically exhibit spatial decay. Third, many public goods are delivered along contiguous spatial networks. Conventional segregation indexes fail to capture these aspects as they focus on the internal composition of neighborhoods, but typically do not incorporate distance.

I develop a metric of "distance segregation" capturing the extent to which two groups within a city (e.g. high- and low-income) are in physical proximity to each other. Below, I describe the simplest version of the index, which is used in this paper. I then discuss how this relates to other

indexes of segregation in the literature and how it can be extended.

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

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 ranking cities of different area sizes and shapes, in 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.[10] The index takes values greater than one in a city where P and R neighborhoods are further away from each other than two random

[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:

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

[10]I confirm this with a Monte Carlo exercise where I simulate random draws of neighborhoods in a city.

ones, a pattern which can be labeled as “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. Extensions for future research may also consider richer travel cost functions which account for slope and topographic obstacles.

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.

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 4860 m, whereas the distance between any two neighborhoods is 4780 m, 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 7647 m apart relative to an average distance across neighborhoods of 6189 m, resulting in a normalized index of 1.24. In Santos (C) the average distance between P and R is 5245 m, relative to an average distance across neighborhoods of 3870 m (normalized segregation index = 1.36). In the Brazilian context, distance segregation will often take the form of rich neighborhoods being in the CBD and poor neighborhoods being peripheral. 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 Relationship to existing segregation metrics

In the sociology and economics literature, the prevailing notion of residential segregation is one that conceptualizes the degree of mixing between groups occurring within neighborhoods. The indexes used are mostly a-spatial and do not account for distances between neighborhoods. The most commonly used metric of residential segregation in the is the dissimilarity index (e.g. Asher et al. (2023)), which measures the share of the minority group that would need to change neighborhood for it to be evenly distributed within a city. The index ranges from 0 (perfect integration)

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.

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. Below I show that the dissimilarity index correlates positively with the distance segregation one.

Sociologists have long recognized that this class of indexes is subject to the so-called "checkerboard problem" (Pisati, 2009), whereby a city that looks like a "checkerboard" of rich and poor neighborhoods 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. 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)).

The distance segregation metric employed in this paper is related to the city shape indexes used in Harari (2020). The disconnection index, in particular, measures the extent to which a city's footprint is compact, by computing the average distance between points in a city. Distance segregation can be viewed as a version of the disconnection index measured by group.

4 Descriptive patterns

4.1 City-level evidence

I begin by describing patterns of city-wide residential segregation across the sample of 608 cities.

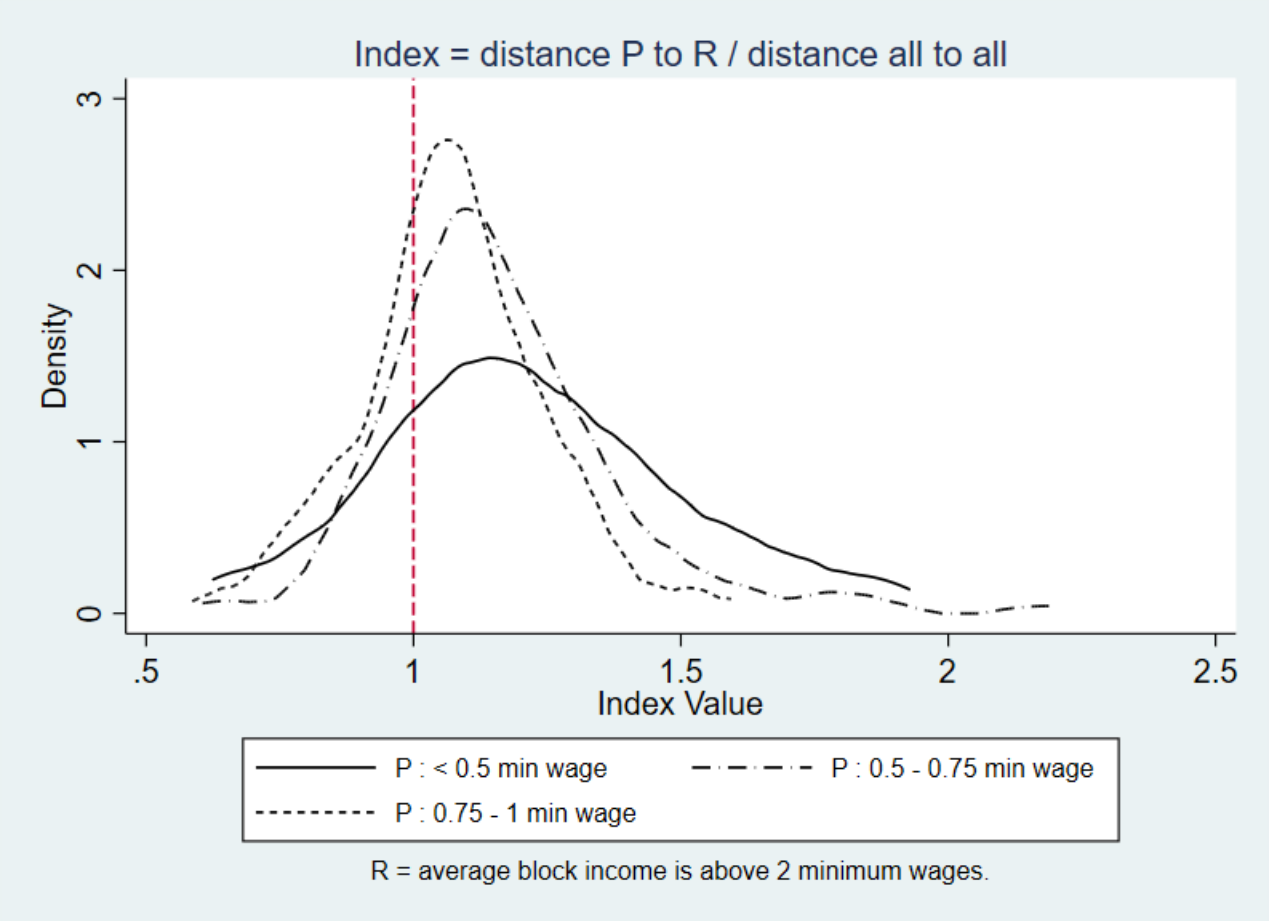
Figures 3 through 5 plot the distribution of the distance segregation index, normalized as described above so that 1 corresponds to a perfect checkerboard, values lower than 1 correspond to distance integration, and values higher than 1 correspond to distance segregation. All graphs show that the median Brazilian city is distance segregated, but there are significant 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 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 5 compares segregation by income, race, and slum status and highlights that slums tend to be better spatially integrated than poor, but formal neighborhoods. Additionally, distance segregation by race is more muted than distance segregation by income. The solid line corresponds to

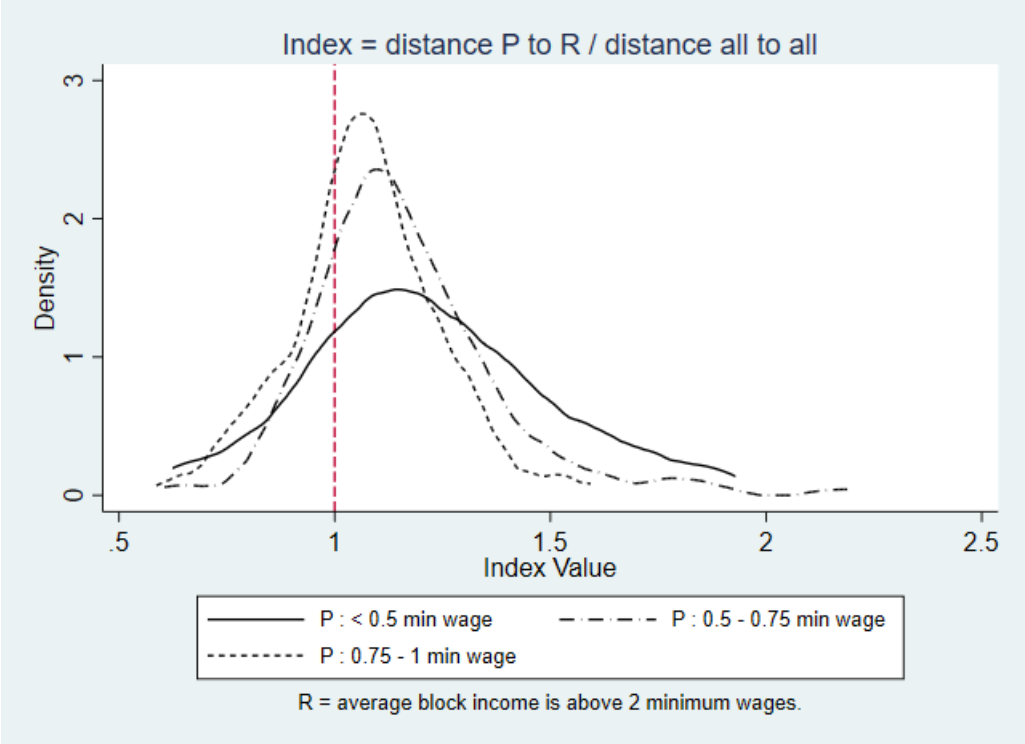
Figure 3: Distance segregation by relative income



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.

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. The dashed line corresponds to distance segregation for slums versus high-income neighborhoods. The dotted line corresponds to segregation defined for neighborhoods that are majority white versus non-white. Distance segregation is strongest for low-income neighborhoods and weakest for slums. In the median city, slums are actually distance integrated, although there is also a tail of cities where slums are distance segregated. This pattern is in line with the argument that slum households trade off housing quality and tenure security for access to economic opportunity. This pattern holds also when comparing slums to any non-slum neighborhoods. Throughout the rest of the paper, I will be focusing on the income definition used in Figure 5.

Figure 4: Distance segregation by absolute income

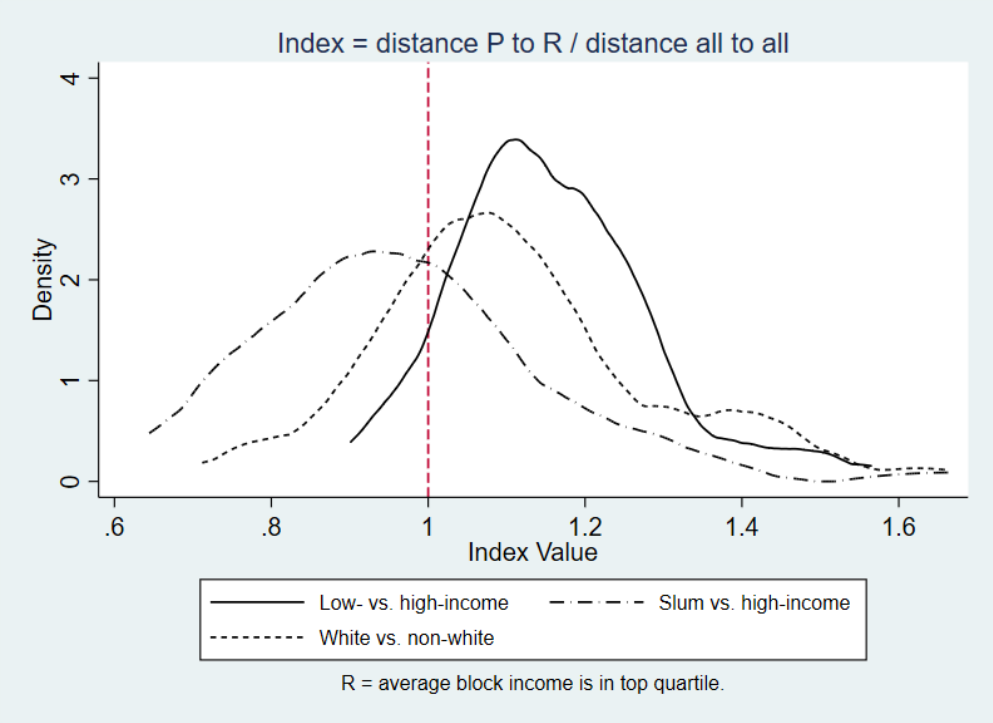


Notes: The figure shows the distribution of the distance segregation index across cities, for different definitions of poor and rich neighborhoods. Census blocks are classified by average income relative to minimum wage. A block is classified as "rich" if the average block income is above two minimum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

Next, I present correlations between the distance segregation index, normalized as above, and conventional indexes used in the literature. Figure 6 shows a positive correlation between the Gini income inequality index and distance segregation by income. More unequal cities are also more distance segregated, in line with the finding above that the very poor neighborhoods are more segregated than middle-income ones. This correlation is not mechanically driven by the index construction since the two indexes are not using the same information. In particular, the distance index only uses physical distance between the wealthiest and the poorest neighborhoods.

Figure 7 shows that more distance segregated cities by race are also more segregated in the conventional sense, as measured by the dissimilarity index. Here too, the correlation is not mechanical as the two indexes use different variation. In particular, the distance segregation index does not use information on the relative sizes of the two groups but merely distance, while the dissimilarity index does not incorporate distance.

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.

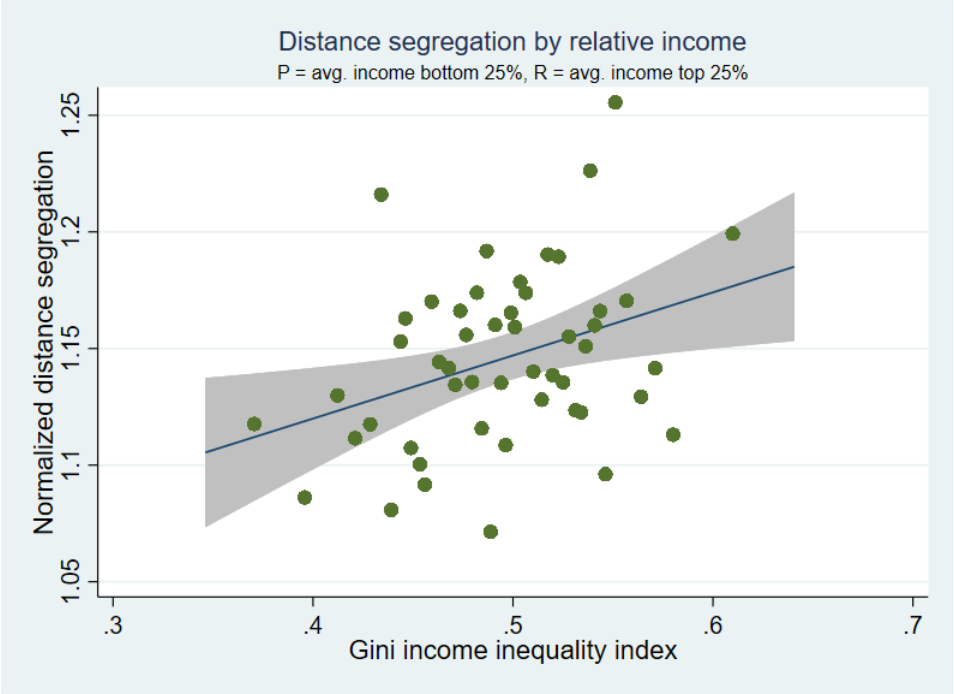
4.2 Neighborhood-level evidence

Below, I provide descriptive neighborhood-level evidence on distance-segregated neighborhoods.

Table 1 shows that distance-segregated neighborhoods are faring worse on a number of dimensions, conditional on their distance from the CBD and own-neighborhood characteristics (race shares, geographic characteristics such as elevation and slope, and the slum dummy). Low-income blocks that are physically distant from high-income neighborhoods have lower access to public goods, worse housing conditions, and lower literacy rates.

The key regressor is distance from high-income neighborhoods, defined with our baseline definition based on relative income (being in the upper quartile of the income distribution). Additionally, I include an interaction with whether the own neighborhood is low-income, distance to the CBD, and city fixed effects. Distance to the CBD is defined using the historical center but results are robust using other definitions (e.g. most lit-up pixel). In columns 1 and 2, I consider the share of households connected to public sewerage and public water networks. Access to public services

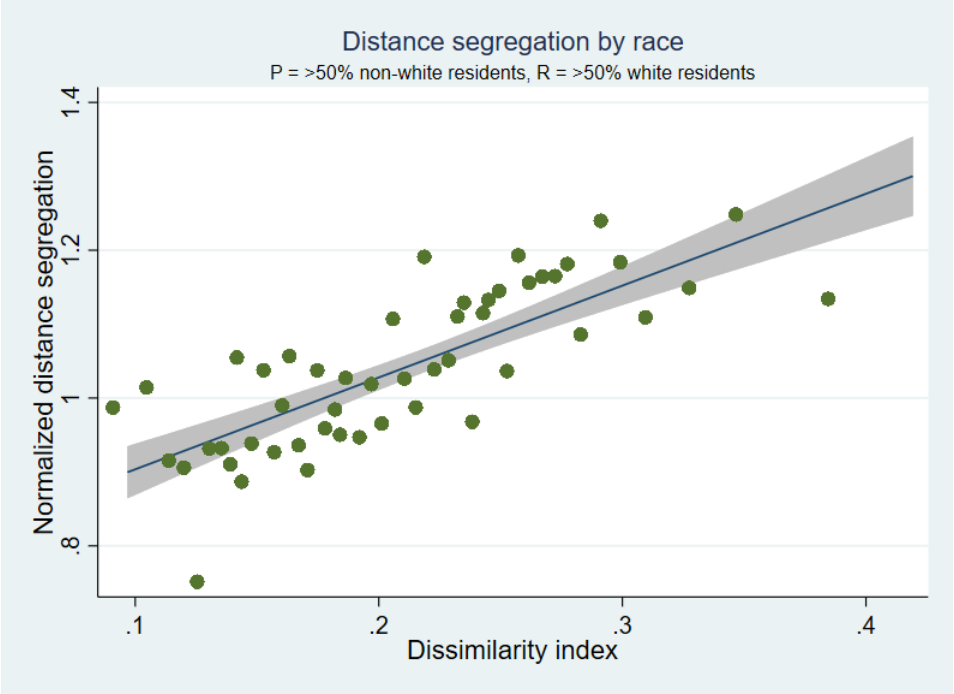
Figure 6: Distance segregation and income inequality



Notes: The figure shows a binscatter representing the city-level correlation between the Gini income inequality index and the normalized distance segregation index between low- and high-income neighborhoods.

declines with distance to the CBD, but declines even more steeply with distance to rich neighborhoods and this is more pronounced for poor neighborhoods, as shown by the negative coefficient of the interaction term. The patterns are similar for the other outcomes considered. Column 3 reaches a similar conclusion using a standardized aggregate index of local public goods including street paving, sidewalks, and garbage disposal (the full list is included in the table notes). Column 4 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. Column 5 reports results for the share of households that have "adequate" housing conditions, defined by the Census as having water supply via a general network, sanitary sewage disposal via a general network or septic tank, direct or indirect garbage collection and, at most, two residents per bedroom. Finally, column 6 reports results for the literacy rate - moderately lower in distance-segregated cities. These patterns are qualitatively similar when considering distance segregation by race and informal status (not reported in the table). Each observation is a census block and all regressions include city fixed effects. Block-level geography controls are also included (distance from the Atlantic coast, share covered by water bodies, average slope), to control for the fact that neighborhoods with unfavorable topography may be harder to reach by the public water and sewerage network. Means of the dependent variables are reported in

Figure 7: Distance segregation and dissimilariy



Notes: The figure shows a binscatter representing the city-level correlation between the dissimilarity index for whites and non-whites and the normalized distance segregation index between majority white and majority non-white neighborhoods.

this table and subsequent ones.

These correlations are qualitatively in line with those in Asher et al. (2023), who find lower access to public goods in Indian neighborhoods that have a higher share of marginalized groups, but are augmented by showing that, conditional on own neighborhood composition, distance also affects public goods provision.

The observed correlations are difficult to interpret as they arise from a combination of direct effects of segregation on public goods provision and endogenous sorting in response to public goods provision. On the one hand, taking residential patterns as given, policy makers may choose to provide public goods in distance-integrated poor neighborhoods in an attempt to alleviate negative externalities onto rich neighborhoods. On the other hand, households may sort in response to the (strategic) provision of public goods across neighborhoods, with richer households sorting into areas that are far away from the low-public goods, impoverished ones.

In the next Section, I shed light on the causal link between residential patterns and public goods provision by taking the analysis at the cross-city level and by developing an identification strategy based on geography.

5 Empirical strategy

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, a naïve 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: irregular gradients in developing country cities can be the byproduct of weak land market institutions Henderson et al. (2020), in contrast with "well functioning" cities with smooth income gradients. Limited local state capacity may thus 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.

In the next section, I develop an instrument that isolates variation in distance segregation that is determined by geographic features.

5.1 Geographic determinants of residential patterns

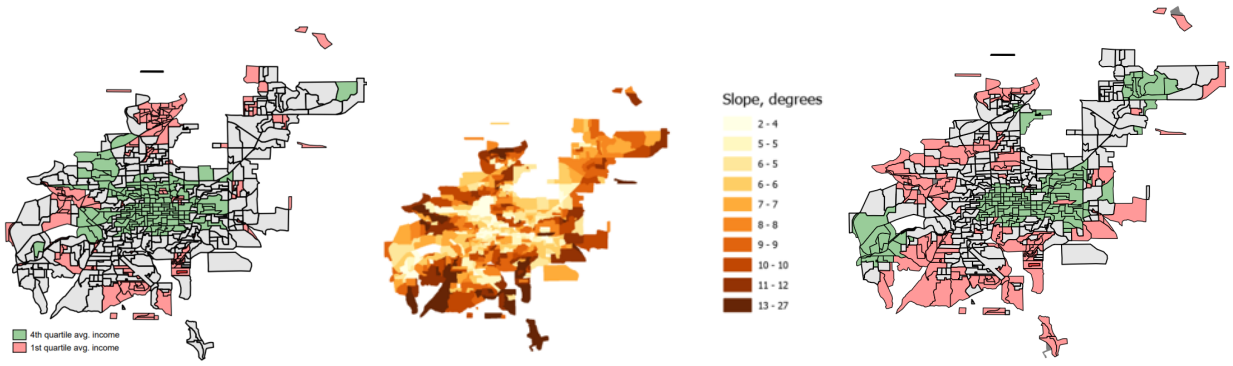
I begin by noting that the spatial distribution of rich and poor neighborhoods is likely to be in part shaped by the spatial distribution of underlying natural amenities, as suggested by Lee and Lin (2018). Figure 8 below shows an example of geography anchoring the location of poor neighborhoods in the city of Caxias do Sul. A comparison between panel 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 2 shows that slope is a strong predictor for the location of low-income neighborhoods (col. 1), predominantly non-white neighborhoods (col. 3) and slums (col. 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. These patterns are in line with anecdotal evidence from many Latin American cities and *favelas* in particular. High-income neighborhoods (col. 2) are associated with lower average 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, as in Lee and Lin (2018).

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. All regressions include city fixed effects.

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 2.

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 2. 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.[11] Figure 8c) shows the result of this exercise for Caxias do Sul. The last step is to calculate the city-level distance segregation for this particular poor definition can be computed using predicted poor and rich neighborhood, obtaining what I define "geography distance segregation".

[11] 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.

Table 3 shows that geography distance segregation is a strong predictor of actual distance segregation for the three dimensions of income, race, and informality. The dependent variable is actual distance segregation, i.e. 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).[12]

To account for the mechanical correlation with city size, I control for a city's equivalent area radius.[13] The regression additionally includes a battery of pre-determined city characteristics that could be associated with natural advantage or to 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.[14] The full list of controls is reported in table A1. 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 isolates 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. In other words, the location of geographic features makes cities more susceptible to being settled in a segregated manner.[15] 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 overall slope or elevation in the city, which may lead to a violation of the exclusion restriction.

Table A1 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 (12 meters for one additional km of distance segregation) and have higher precipitation (+1.4 mm). Reassuringly, the instrument is uncorrelated with average slope, ruggedness, and presence of water bodies. 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

[12]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).

[13]The equivalent area radius is the radius of a polygon with the same land area as that of the city, but circular shape. It is proportional to the average distance between any two points in the city, if the city were a circle.

[14]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.

[15]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.

in my robustness checks below, showing that the main results are robust to excluding state capitals, cities near state capitals, and coastal cities.

6 Results

With the instrument in hand, I 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 4 presents the main results, examining the impact of distance segregation by income on measures of public goods access. Columns 1 through 3 presents access to positional public goods reported in the Census, averaged at the city level. Residents in segregated cities are less likely to be connected to the public sewerage (column 1) and water networks (columns 2). Additionally, they have a lower neighborhood quality score based on the standardized index in column 3.[16] Column 4 considers the composite index of access to public amenities from OpenStreetmap: parks, fire stations, police stations, and post offices.

The table reports both IV and OLS coefficients on distance segregation measured in km. The controls listed in Table 3 are also included. To interpret magnitudes, note that, for the median-sized city, a one standard deviation in distance segregation by income is 1.7 km.[17] 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 percentage points respectively, 6 and 3% of the sample mean. The likelihood of having OSM amenities within 3 kilometers - roughly the distance reachable within a 30-45 minute walk - decreases by 3.4 percentage points or 5% of the mean.

Alternative outcome variables based on the individual components of the neighborhood index and the composite OSM index are reported in Table A5. For OSM amenities, Table A5 reports distance in meters to the nearest amenity as the dependent variable. Results are consistent.

[16]The index includes shares of residents in neighborhoods with the following characteristics: no accumulated street garbage, no open sewer, street addresses, street lighting, street paving, sidewalks, curbs, manholes, ramps, and greenery.

[17]A on standard deviation of the normalized index is multiplied by the median-radius city to obtain 1.7 km.

Table A3 shows qualitatively similar results when considering other dimensions of segregation - race and informal status.

The baseline results are similar when including the (non-predetermined) controls used in the literature on public goods provision (e.g. Trounstine (2016)) such as total expenditure, income inequality, the share of residents aged 65 or older, the share of white residents, and average years of education in 1998.

Comparing the estimates in tables 4 and A3, 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

The main 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. Below, I present additional tests that further probe these concerns.

In Table A2, panel B, 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 A4 shows results obtained residualizing sewerage and water access by block-level geographic characteristics. Before aggregating at the city level, I regress the block level outcomes on the baseline city-wide controls and block-level slope, elevation, and distance to water bodies. I then aggregate the residualized measure, which should be spurious from correlations with geography. Reassuringly, results are similar to the baseline ones.

Finally, 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 (i.e. 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.

Importantly, 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 A5 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 percentage point 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). It is reassuring that the findings are consistent using a cumulative outcome reflecting health investments over the course of the years and from a different source than the Census.

6.2.2 Functional form

Table A2, panel C, explores different specifications and functional forms for the distance indexes. First, I consider an alternative instrument construction procedure that augments the block-level predictive regressions with distance to the CBD as an additional explanatory variable. In practice, this amounts to predicting cities to have a gradient with the rich in the center and the poor in the periphery whenever there is little variation in geographic features within a city. Results are similar to the baseline ones.

Next, 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, if anything, 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, 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.

Finally, 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, 2023) 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.

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.

6.2.3 Locally provided public goods

The interpretation of the main results rests on the assumption that the public goods considered are in fact provided at the municipality level. If the provision of these local public goods were entirely determined at a different level, then one might worry that the observed relationship is spurious.

In Table A2, panel C, I address the concern that the results may be affected by state-level factors (e.g. transfers) by showing robustness to including state fixed effects.

Table A7 provides an additional 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 2007 law re-established municipalities as the ultimate authority in this sector (Kresch, 2020). Currently, municipalities retaining regulatory authority even in cases in which the system is currently run by a state company (Kresch et al., 2023) and decisions on which areas to service are made at the municipal level (Feler and Henderson, 2011). I show that the negative impacts of segregation on access to water and sewerage are more pronounced in cities that have a municipal company (taken from (Kresch et al., 2023)), supporting the interpretation that indeed the

results are mediated by a municipality-level response. As a placebo, I also show that the interaction with the municipal company dummy is insignificant for local public goods that are unrelated to water.

6.3 Mechanisms

Overall, the results above point to lower levels of public goods provision in more segregated cities. I next turn to investigating mechanisms. There are three sets of explanations for how distance may affect public goods provision. First, many public goods - like sewerage or water - are delivered along contiguous spatial networks. Second, spatial segregation between poor and rich may be associated with lower propensity towards redistribution, through sorting or by directly shaping preferences. Third, distance between neighborhoods affects the decay of externalities and may affect public goods allocation decisions, conditional on redistributive preferences. Below, I exploit the granularity of the block-level data and develop empirical tests that leverage the spatial pattern of public goods provision across neighborhoods and groups.

Table 5 presents results analogous to those in Table 4, but instead of considering outcomes averaged at the city level reports outcomes averaged across specific set 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.

6.3.1 Network effects

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 neighborhoods 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.

To test this mechanism, Table 5 panel D considers public goods provision in poor neighborhoods that are close to rich ones (as measured by being in the top quintile in the city by average distance to rich neighborhoods). If the network expansion explanation above is the main mechanism, then the differences between segregated and integrated cities should be 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 to the integrated poor relative to integrated cities.

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. However, this test suggests that is not differential by high and low segregation cities in a way that can explain our baseline result.

6.3.2 Redistributive preferences

The second potential explanation is related to differences in redistributive preferences: residents in distance-segregated cities may have lower willingness to fund public goods in general, irrespective of distance. Under this explanation, we should see less redistribution across the board, and not only in far-away poor neighborhoods.

Panel A and B report the effects of segregation on public goods provision in poor (A) and rich (B) neighborhoods, showing coefficients that are close in magnitudes. This indicates that segregated cities feature lower public goods provision relative to integrated cities even in the rich neighborhoods. [18]

Panel C considers inequality in public goods provision as measured by the ratio between the outcome variables among the poor and among the rich. The sample means are all below 1, consistent with the poor being in general underserved. Segregated cities have a more unequal distribution of public goods when measured by sidewalks and OSM amenities (columns 4 and 5), suggesting stronger pro-poor redistribution in integrated cities. However, there is no difference in the degree of public goods inequality across segregated and integrated cities for the outcomes in columns 1 through 3.

Why would residents of segregated cities have weaker redistributive preferences? This could be because of sorting into cities that are geographically prone to segregation or because of segregation directly shaping residents' preferences through exposure. For example, Bjorvatn and Cappelen (2003) theorize that residential segregation may reduce the social attachment between societal groups and the propensity of the rich to redistribute to the poor and Alesina and Zhuravskaya (2011) empirically show that segregation is associated with lower trust.

My empirical strategy does not fully allow to disentangle sorting of residents from direct effects of segregation on preferences. Table A6 provides some suggestive evidence on compositional differences in the types of residents in segregated and integrated cities. 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

[18] One caveat in the interpretation of this result is that even neighborhoods that are classified as "rich" may include pockets of poverty and that these pockets could be the under-served areas. Absent more granular data, it is difficult to conclusively address this.

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 two dimensions, although this test is not conclusive.

6.3.3 Distance

The third set of explanations are related to public goods provision as a function of distance, conditional on overall levels of provision. The finding of lower public goods provision in segregated cities could be explained by policy makers choosing to under-provide poor neighborhoods that are far away.

Panel E investigates this by considering the ratio of the outcome variables in distance-segregated poor neighborhoods relative to poor neighborhoods in general. The sample means (in square brackets) indicate that, within cities, the distance-segregated poor indeed have lower access to public goods than their distance-integrated counterparts. This is consistent with an argument of negative externalities decaying in space (Xu, 2023): when poor neighborhoods are further away, there is less of an incentive to correct negative externalities from those neighborhoods by providing public goods (e.g. sanitation), since wealthier households are shielded by distance.

However, for the outcomes in columns 1 though 4, the gap in access between the segregated and the integrated poor is less stark in segregated cities, as reflected by the positive IV coefficients. In other words, in a city that is overall integrated, the integrated poor fare much better than the segregated ones, but a city that is overall segregated, the integrated poor are not faring that much better than their segregated counterparts. This is likely because an additional mechanism becomes relevant, operating in the opposite direction: strategic under-providing of poor neighborhoods that are near rich ones (Feler and Henderson, 2011). Policy makers may deliberately avoid providing public goods in areas that they do not want poor households to concentrate in, such as high land-values areas. It is plausible that these strategic motives may be more relevant in distance-segregated cities.

Interestingly, column 5 presents different results for public goods that are not positional (the OSM amenities). The coefficient has the opposite sign, suggesting that the externalities argument prevails. Unlike sewerage and public water, which need to be provided in the neighborhood where the poor live, OSM amenities are in locations that anyone can travel to. The strategic under-provision channel is thus likely muted in this case.

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, 2024).

6.4 Historical determinants of distance segregation

Thus far, I have focused on geography as a key determinant of segregation patterns. While this is helpful for establishing causality, there are of course many other determinants. The spatial distribution of income across neighborhoods can be conceptualized as an outcome of "first nature" forces, such as natural amenities, and "second nature" factors related to coordination, externalities, and path dependence. Table 7 provides a suggestive exploration of some of the historical factors affecting today's segregation patterns.

Column 1 investigates to what extent city-wide segregation is driven by peripheral areas that emerged in recent years, as opposed to the historical cores of cities. I compute the 2010 distance segregation index using only the historical portions of cities.[19] I find that segregation in the historical center is associated with segregation in the full contemporaneous extent of the city. For a one standard deviation increase in the historical index, the current one increases by 0.4 standard deviations. This suggests that the observed city-wide segregation patterns are explained both by cities' expansion on the extensive margin and by segregated patterns in the center.

Column 2 considers the role of infectious diseases, a key urban externality. A priori, negative externalities could be addressed through more inclusive public goods provision (as documented for the historical U.S. by Troesken (2001)) or by segregating the perceived sources of externalities away. I draw upon a historical example of the latter approach. During the 19th century, Brazilian cities were susceptible to outbreaks of yellow fever, which is caused by a mosquito but at the time was believed to originate from "miasma" emanating from overcrowded poor neighborhoods. Local urban governments often intervened by clearing tenements (*cortiços*) from the center of cities, justifying these interventions on grounds of *higienizaçao* or "hygienisation" planning principles (Garmany and Richmond, 2020). I find evidence of greater segregation in cities more exposed to outbreaks, as proxied by an interaction between mosquito suitability and whether the municipality was established after the first major outbreak in the state. Municipalities date of establishment are drawn from the IBGE Cidades portal. Suitability for the yellow fever mosquito is based on local environmental factors (Kraemer et al., 2015). Disease outbreaks were collected from a variety of historical records.

A prominent example of hygienist-inspired interventions was the creation of large boulevards or *avenidas* to replace "unsanitary" narrow streets in lower-income neighborhoods. Column 3 shows that cities featuring many streets labeled as *avenidas* are associated with greater segregation

[19] Specifically, I consider a circle around cities' historical centers with a radius that approximates the 1900 city extent. This radius is calculated for each city using the 1900 population reported in the Census and assuming a population density similar to that of São Paulo in 1900, as retrieved by historical maps showing the city's extent.

today, pointing to a potential role for urban planning in promoting segregation. The key explanatory variable is a dummy for whether a city is in the upper quartile of the distribution by share of road segments labeled as *avenidas* in Open Street Map. These findings echo those in Baruah et al. (2021) about the persistence of historical urban planning practices.

Finally, column 4 considers distance segregation by race and the legacy of slavery. Cities in areas with a high share of slaves in 1872 are more distance-segregated by race today, but this effect is reversed if they were established as municipalities after the abolition of slavery.

Taken together, these results point to a role of historical factors and persistence in shaping segregation patterns, suggesting a potential avenue for future research.

6.5 Conclusion

This paper provides one of the first systematic investigations of within-city residential patterns in a developing country context. I develop novel 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. Leveraging a novel instrumental variables approach that isolates the variation in segregation determined by geography, I provide causal evidence 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. Additionally, I provide suggestive evidence on some of the historical determinants of distance segregation such as exposure to infectious diseases and slavery.

Taken together, 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), urban renewal programs (Gechter and Tsivanidis, 2024), and investments in transit infrastructure (Tsivanidis, 2023, 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 points to some of the political economy factors that may complicate the provision of public goods, particularly in poor areas.

Coupled with the evidence on historical determinants of segregation, these results also suggest that segregation and public goods provision can be viewed as two competing responses to urban congestion externalities. This could lead to different city configurations: on one end of the spectrum, segregated cities with low levels of public goods provision, on the other end, more integration

and higher levels public goods provision.

Beyond developing countries, this paper also illustrates that accounting for space and distance when considering residential segregation provides additional insight into the economic livelihoods of disadvantaged groups in cities.

References

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2013): "Have the Poor Always Been Less Likely to Migrate? Evidence from Inheritance Practices During the Age of Mass Migration," *Journal of Development Economics*, 102, 2–14.

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

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

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 (2023): "Residential Segregation and Unequal Access to Local Public Services: Evidence from 1.5m Neighborhoods in India," .

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

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.

BARUAH, N., J. V. HENDERSON, AND C. PENG (2021): "Colonial Legacies: Shaping African Cities," *Journal of Economic Geography*, 21, 29–65.

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.

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

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.

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.

DURANTON, G. AND H. G. OVERMAN (2005): "Testing for Localization Using Micro-Geographic Data," *The Review of Economic Studies*, 72, 1077–1106.

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.

FOSSETT, M. (2017): "New Methods for Measuring and Analyzing Segregation," *The Reivew of Economic Studies*, 663–694.

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

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.

GECHTER, M. AND N. TSIVANIDIS (2024): "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills," .

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. AND M. WONG (2024): "Slum Upgrading and Long-run Urban Development: Evidence from Indonesia," .

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," .

KRAEMER, M. U., M. E. SINKA, K. A. DUDA, A. Q. MYLNE, F. M. SHEARER, C. M. BARKER, C. G. MOORE, R. G. CARVALHO, G. E. COELHO, W. VAN BORTEL, G. HENDRICKX, F. SCHAFFNER, I. R. ELYAZAR, H.-J. TENG, O. J. BRADY, J. P. MESSINA, D. M. PIGOTT, T. W. SCOTT, D. L. SMITH, G. W. WINT, N. GOLDING, AND S. I. HAY (2015): "The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus," *eLife*, 4, e08347.

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.

KRESCH, E. P., M. LIPSCOMB, AND L. SCHECHTER (2020): "Externalities and Spillovers from Sanitation and Waste Management in Urban and Rural Neighborhoods," *Applied Economic Perspectives and Policy*, 42.

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.

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

PEREIRA, R., D. BANISTER, T. SCHWANEN, AND N. WESSEL (2019): "Socio-spatial inequalities in access to opportunities in Brazilian cities, 2019 [PORT] | Access to Oportunities," .

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

PISATI, M. (2009): "Spatial Indices of Residential Segregation," .

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

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.

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. (2023): "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio," Working paper, American Economic Review.

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

——— (2016): "World Cities Report," .

WORLD BANK (2023): "World Bank open data: Gini index," .

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.

Table 1: Correlates of distance-segregated neighborhoods

	(1)	(2)	(3)	(4)	(5)	(6)
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km	Share hhs with adequate housing	% Literate
Log distance to CBD	-0.024***	-0.004	0.057***	0.002***	-0.008	-0.002***
	(0.006)	(0.005)	(0.006)	(0.001)	(0.006)	(0.000)
Log distance to R blocks	-0.133***	-0.050***	0.218***	0.005***	-0.117***	-0.019***
	(0.018)	(0.014)	(0.033)	(0.001)	(0.017)	(0.001)
Log distance to R x P	-0.011***	-0.001***	0.032***	-0.001***	-0.011***	-0.005***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Constant	2.156***	1.293***	0.604*	0.209***	1.962***	1.128***
	(0.326)	(0.149)	(0.353)	(0.013)	(0.283)	(0.026)
Observations	161,460	161,460	161,460	161,460	152,097	161,460
R-squared	0.571	0.438	0.557	0.160	0.522	0.615
Mean dep. var.	0.720	0.930	3.229	0.278	0.788	0.938

Notes: Each observation is a census block. The neighborhood public goods index includes the following characteristics: no accumulated street garbage, no open sewer, street addresses, street lighting, street paving, sidewalks, curbs, manholes, ramps, greenery. The OSM amenities index includes fire stations, police stations, post offices and parks within 3 km. All variables are from the 2010 Census except for the OSM ones. All specifications include city fixed effects and geography controls (distance to the shoreline, distance to rivers or streams, distance from lakes, the share of a block covered by water bodies, average elevation, average slope). Additional controls include share white residents, share black residents, and a slum dummy. Standard errors clustered at the city level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table 2: 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	-0.035***	0.044***	-0.046***	-0.033***
	(0.011)	(0.010)	(0.010)	(0.008)
Slope^2	0.007***	-0.006***	0.006***	0.005***
	(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	-0.076***	0.087**	-0.054*	-0.061***
	(0.025)	(0.035)	(0.030)	(0.010)
Distance^2 to rivers	0.015***	-0.013	0.005	0.009***
	(0.005)	(0.006)	(0.006)	(0.002)
Distance to lakes	0.051***	0.063***	0.047***	0.010***
	(0.007)	(0.010)	(0.011)	(0.003)
Distance^2 to lakes	-0.001***	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)
Constant	0.107**	0.440***	0.355***	0.152***
	(0.050)	(0.041)	(0.030)	(0.020)
Observations	162,464	162,464	162,064	162,064
R-squared	0.058	0.063	0.516	0.158
F statistic	1202	377	1282	165

Notes: Each observation is a census block, from 605 cities. All specifications include city fixed effects. Standard errors clustered at the city level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table 3: First stage
Dependent variable: distance segregation, km

	(1)	(2)	(3)
	Income	Race	Slum
Geography distance-segregation, km	0.610***	0.549***	0.394***
	(0.046)	(0.072)	(0.076)
Equivalent area radius, km	0.512***	0.515***	0.677***
	(0.073)	(0.090)	(0.095)
Average elevation, m	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Ruggedness, m	0.000	0.001	-0.001
	(0.001)	(0.001)	(0.001)
Average slope, degrees	-0.026	0.093	0.043
	(0.048)	(0.127)	(0.111)
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	-0.651*	0.144	1.078
	(0.372)	(0.608)	(0.692)
Landslide risk	0.064	-0.516	0.070
	(0.223)	(0.531)	(0.609)
Slope adjustment factor for distance	-2.592	-6.371**	2.498
	(2.227)	(3.073)	(5.882)
Distance to state capital, km	-0.000	-0.001	-0.000
	(0.000)	(0.001)	(0.001)
Distance to Atlantic, km	-0.000	-0.001***	0.001
	(0.000)	(0.000)	(0.001)
Observations	597	447	223
R-squared	0.903	0.828	0.863
F statistic	174	59	27

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

Table 4: 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***	-0.017***	-0.051***	-0.021***
	(0.008)	(0.005)	(0.013)	(0.004)
OLS	-0.015**	-0.016***	-0.037***	-0.020***
OLS SE	(0.006)	(0.004)	(0.009)	(0.003)
Observations	597	597	597	597
R-squared	0.582	0.460	0.607	0.182
IV F statistic	174	174	174	174
Mean dep. var.	0.571	0.899	0.010	0.630

Notes: Each observation is a city. All specifications include the controls in Table 3. 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)
IV estimates for distance -segregation by income, km	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	-0.022**	-0.020***	-0.012**	-0.022***	-0.022***
	(0.009)	(0.005)	(0.006)	(0.007)	(0.004)
	[0.479]	[0.867]	[0.610]	[0.439]	[0.556]
A. Rich	-0.019***	-0.017**	-0.014***	-0.021***	-0.017***
	(0.007)	(0.007)	(0.003)	(0.005)	(0.005)
	[0.667]	[0.922]	[0.909]	[0.831]	[0.696]
C. Poor / Rich	-0.024	-0.000	-0.002	-0.015**	-0.015**
	(0.024)	(0.012)	(0.006)	(0.007)	(0.006)
	[0.759]	[0.945]	[0.652]	[0.503]	[0.828]
D. Poor close to Rich	-0.021***	-0.018***	-0.015***	-0.025***	-0.017***
	(0.008)	(0.007)	(0.004)	(0.006)	(0.006)
	[0.629]	[0.920]	[0.845]	[0.735]	[0.698]
E. Poor far from Rich / Poor	0.026	0.028***	0.050**	0.113***	-0.032***
	(0.039)	(0.011)	(0.023)	(0.040)	(0.009)
	[0.885]	[0.991]	[0.982]	[1.054]	[0.776]
Observations	597	597	597	597	597

Notes: This table reports IV coefficients for distance-segregation by income from a specification similar to that in Table 4. 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. Panel E considers the ratio between averages among the poor neighborhoods that are far from the rich and poor neighborhoods in general. The table reports IV coefficients, standard errors in parentheses, and sample averages of each dependent variable in square brackets. Each observation is a city. All specifications include the controls in Table 3. F statistics are all above 100 and available upon request. Standard errors are clustered at the meso-region level. *** p>0.01, ** p>0.05, * p>0.1.

Table 6: Historical determinants of distance segregation
Dependent variable: distance segregation index, normalized

	(1)	(2)	(3)	(4)
	Income	Income	Race	Race
Distance segregation 1900 radius	0.318***			
	(0.051)			
Mosquito suitability		0.001		
		(0.091)		
Mosquito suitability		0.040**		
x Established after outbreak in state		(0.018)		
Avenidas			0.039***	
			(0.013)	
% Slaves in 1872				0.291**
				(0.134)
Established after slavery abolished				0.064
				(0.039)
% Slaves in 1872				-0.435***
x Established after slavery abolished				(0.138)
Observations	516	600	600	451
R-squared	0.233	0.169	0.076	0.367

Notes: Each observation is a city. Average distance between P and R neighborhoods is normalized by average distance between any two neighborhoods. All specifications include geography controls and dummies for period of settlement. Standard errors in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A1: Correlates of geography distance-segregation

	(1)	(2)
	OLS	Sample mean
Average elevation (m)	-12.449**	414.300
	(4.831)	
Ruggedness (m)	2.632	93.980
	(2.661)	
Average slope (degrees)	0.010	5.006
	(0.078)	
Water bodies within 30 km (sqkm)	-1.440	56.070
	(1.818)	
Low-fertility soil within 30 km (sqkm)	-29.759**	1517.000
	(14.451)	
% land available within 30km	0.004	0.845
	(0.003)	
Landslide risk	0.002	1.455
	(0.010)	
Slope adjustment factor for distance	0.000	1.190
	(0.000)	
Distance to state capital (km)	-9.936***	181.100
	(3.312)	
Distance to Atlantic (km)	-12.230**	216.400
	(5.383)	
Latitude of municipality seat	0.048	-16.740
	(0.174)	
Longitude of municipality seat	0.193*	-45.840
	(0.098)	
Precipitation annual avg. (mm/day)	1.372**	124.400
	(0.654)	
Sunshine annual avg. (wh/m2.day)	-2.463	4935.000
	(10.176)	
Soil type: neosoils	0.004	0.352
	(0.009)	
Soil type: planosoils	-0.002	0.183
	(0.006)	
Soil type: argisoils	-0.008	0.734
	(0.009)	
Soil type: luvisoils	-0.007*	0.063
	(0.004)	
Soil type: gleisoils	0.014**	0.170
	(0.007)	
Soil type: latosoils	-0.007	0.656
	(0.008)	
Soil type: nitosoils	-0.005**	0.071
	(0.002)	
Soil type: plintosoils	-0.002	0.066
	(0.004)	
Soil type: cambisoils	0.008	0.339
	(0.008)	
Soil type: espondosoils	0.006	0.071
	(0.005)	
Observations		605

Notes: Column 1 reports coefficients and standard errors 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 A2: Robustness

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.
Panel A: Sample cuts						
Exclude state capitals	–0.021**	–0.018***	–0.054***	–0.020***	570	135
	(0.010)	(0.005)	(0.015)	(0.004)		
Exclude near state capitals	–0.012*	–0.016**	–0.042***	–0.017***	452	94
	(0.007)	(0.006)	(0.013)	(0.005)		
Exclude elevated	–0.021**	–0.018***	–0.046***	–0.022***	542	135
	(0.009)	(0.006)	(0.013)	(0.005)		
Exclude coastal	–0.028***	–0.004	–0.045**	–0.022***	507	104
	(0.009)	(0.008)	(0.020)	(0.007)		
Exclude top largest	–0.021**	–0.018***	–0.049***	–0.022***	566	152
	(0.009)	(0.006)	(0.015)	(0.005)		
Exclude bottom largest	–0.023***	–0.021***	–0.055***	–0.022***	566	370
	(0.009)	(0.004)	(0.014)	(0.004)		
Panel B: Additional proxies for city disconnectedness						
Disconnected city	–0.021**	–0.018***	–0.052***	–0.022***	597	133
	(0.008)	(0.006)	(0.013)	(0.004)		
Nr water basins	–0.025***	–0.018***	–0.054***	–0.022***	597	161
	(0.009)	(0.006)	(0.014)	(0.005)		
Nr polygons, available land within 30km	–0.023***	–0.017***	–0.054***	–0.021***	597	173
	(0.008)	(0.005)	(0.013)	(0.004)		
Perimeter/area ratio, available land within 30km	–0.022***	–0.017***	–0.052***	–0.022***	597	171
	(0.008)	(0.006)	(0.013)	(0.004)		
Panel C: Specification and functional form						
Include distance to CBD in prediction	-0.023***	-0.013**	-0.060***	-0.023***	597	113
	(0.008)	(0.007)	(0.015)	(0.004)		
Population density-weighted distance segregation	-0.032***	-0.025***	-0.075***	-0.031***	597	311
	(0.011)	(0.009)	(0.017)	(0.007)		
Index based on log distance	-0.164***	-0.105**	-0.233*	-0.0962	597	187
	(0.063)	(0.047)	(0.139)	(0.060)		
State FE	-0.02***	-0.02***	-0.06***	-0.02***	597	151
	(0.01)	(0.01)	(0.02)	(0.00)		

Notes: this table reports IV coefficients of distance segregation by income on the four primary outcomes. Specifications are analogous to those in Table 3. Panel A reports specifications that exclude certain cities from the sample. Panel B reports specifications adding additional controls. Panel C considers different functional forms and specifications

Table A3: Other dimensions of segregation

	(1)	(2)	(3)	(4)
	Share residents with public sewerage	Share residents with public water	Neighborhood public goods index	OSM amenities within 3km
Panel A: distance-segregation by race				
IV	-0.0356***	-0.0136***	-0.0762***	-0.0262***
	(0.00962)	(0.00461)	(0.0179)	(0.00745)
OLS	-0.017***	-0.015***	-0.037***	-0.017***
	(0.006)	(0.005)	(0.012)	(0.005)
Observations	447	447	447	447
R-squared	0.494	0.399	0.478	0.117
IV F statistic	59	59	59	59
Mean dep. var.	0.647	0.928	0.151	0.645
Panel B: distance-segregation by slum status				
IV	-0.0568***	-0.0228*	-0.0898**	-0.0139
	(0.0157)	(0.0117)	(0.0349)	(0.0111)
OLS	-0.025***	-0.014*	-0.026	-0.010
	(0.008)	(0.008)	(0.020)	(0.008)
Observations	223	223	223	223
R-squared	0.569	0.413	0.475	0.216
IV F statistic	27	27	27	27
Mean dep. var.	0.602	0.900	0.05	0.662

Notes: Each observation is a city. All specifications include the controls in Table 3. Standard errors clustered at the meso-region level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A4: Additional robustness for network goods

	(1)	(2)	(3)	(4)
	Share residents with public sewerage		Share residents with public water	
Residualized by	topography	topography + distance	topography	topography + distance
Distance-segregation by income, km				
IV	-0.021**	-0.019**	-0.017***	-0.017***
	(0.008)	(0.008)	(0.005)	(0.005)
Observations	597	597	597	597
R-squared	0.127	0.139	0.129	0.128
IV F statistic	174	174	174	174
Mean dep. var.	-0.070	-0.075	-0.020	-0.020
StDev dep. var.	0.234	0.236	0.133	0.133

Notes: Each observation is a city. All specifications include the controls in Table 3. Standard errors clustered at the meso-region level in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.

Table A5: Other public goods measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance-segregation by income, km	Share municipal spending on health	Nr hospital beds per 1000 ppl	Park	Post office	Police station	Fire station	Paved streets, % residents	Sidewalks, % residents
IV	-0.007***	-0.190***	142.373***	580.913***	414.868***	613.625***	-0.013***	-0.024***
	(0.002)	(0.045)	(34.034)	(95.628)	(90.738)	(145.500)	(0.004)	(0.006)
Observations	593	595	593	419	496	332	597	597
R-squared	0.156	0.098	0.336	0.311	0.155	0.357	0.387	0.492
Mean dep. var.	0.239	2.362	619.2	2510	2031	3111	0.771	0.645

Notes: Each observation is a city. All specifications include the controls 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 A6: Human capital and demographics

	(1)	(2)	(3)	(4)	(5)
Distance-segregation by income, km	% Literacy rate	% Employment in service sector	% High school degree or higher	% non-White	% Prime age males
IV	-0.003***	-0.006***	-0.006***	0.005	-0.000
	(0.001)	(0.001)	(0.002)	(0.003)	(0.000)
Observations	597	597	597	597	597
R-squared	0.757	0.468	0.547	0.782	0.617
IV F statistic	174	174	174	174	174
Mean dep. var.	0.908	0.412	0.366	0.516	0.208

Notes: Each observation is a city. All specifications include the controls in Table 3. Standard errors clustered at the meso-region in parenthesis. *** p>0.01, ** p>0.05, * p>0.1.