Optimal Provision of Urban Infrastructure: Theory and Evidence from Brazilian Cities*

Bruno Barsanetti†

Taric Latif Padovani‡

Valdemar Pinho Neto§

March 24, 2026

Abstract

We investigate if the distribution of local infrastructure within cities is efficient. Using high-resolution data on population, housing prices, and local urban infrastructure, we document that infrastructure provision consistently declines with distance from the city center, and that unconditional housing price and population density gradients are steeper than when conditioning on infrastructure. We interpret these facts through a monocentric city model in which infrastructure is a local public good, deriving a sufficient-statistics rule for its optimal spatial allocation. Under constant cost and elasticity conditions, population density and land prices should fully explain spatial differences in infrastructure. We find systematic deviations from this prediction, with infrastructure biased toward central neighborhoods. While aggregate losses in welfare or city size are modest, this bias shapes city structure, implying city centers 21% denser than optimal.

JEL Classification Codes: R14, H41, O18.
Keywords: urban infrastructure, urban spatial structure, local public goods.

*We appreciate feedback from audiences at FGV EPGE, the 2025 North American Meeting of the UEA, and LACEA 2025, and comments or conversations with João Andrade, Juan Pablo Chauvin, Adriano Borges Costa, Carlos E. da Costa, Alípio Ferreira, Nina Harari, and Ricardo C.A. Lima. Bruno Barsanetti appreciates funding for this study by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. All errors are our own.

†EPGE Brazilian School of Economics and Finance. E-mail: bruno.barsanetti@fgv.br
‡EPGE Brazilian School of Economics and Finance. E-mail: tariclatif@gmail.com
§EPGE Brazilian School of Economics and Finance. E-mail: valdemar.pinho@fgv.br

1 Introduction

Urban infrastructure, such as sanitation, paved streets, drainage, public lighting, or trash collection, plays an important role in the quality of life of city residents. Such infrastructure often varies widely across neighborhoods within cities, especially in developing economies (Pandey, Brelsford, and Seto 2022; Nicoletti, Sirenko, and Verma 2022; Harari 2024). Central neighborhoods tend to feature a pleasant environment with well-built sidewalks, clean streets, and trees, while peripheral areas sometimes lack even basic infrastructure like piped water (Feler and Henderson 2011). Since infrastructure contributes to local amenity values (Gonzalez-Navarro and Quintana-Domeque 2016, McIntosh et al. 2018), these spatial inequalities influence where people live and how efficiently cities are organized. In this paper, we study the optimal provision of urban infrastructure within cities, empirically investigating whether it is efficient in a large developing-economy context. Identifying deviations from optimal infrastructure provision is of particular policy relevance, as it indicates where public investment is mostly needed.

We combine urban theory and detailed data to evaluate the spatial distribution of urban infrastructure within cities. We construct a high-resolution dataset covering the universe of Brazilian cities and document two descriptive facts. First, infrastructure provision declines with distance to the city center, a pattern that holds across all sanitation and street infrastructure dimensions we observe. Second, unconditional gradients in housing prices and population density are 59% steeper than if conditioning on urban infrastructure. To interpret these facts, we consider a monocentric city model in which infrastructure is a local public good influencing the amenity value of neighborhoods. We derive a sufficient-statistics rule for the optimal allocation of infrastructure and use it to construct a simple diagnostic test of whether urban infrastructure is optimally allocated. The test is based on the implication that, under constant cost and elasticity conditions, population density and land prices should fully predict optimal infrastructure provision across space. In the data, however, these variables predict only about half of the observed infrastructure gradient, suggesting a sub-optimal underprovision of infrastructure in peripheral neighborhoods. We then calibrate a tractable version of the model to quantify the equilibrium consequences of this misallocation. Although aggregate effects on welfare and city size are modest, the infrastructure bias steepens the population density gradient by 27%, implying that cities are significantly more concentrated than they would be under efficient infrastructure provision.

At the heart of our investigation is a high-resolution dataset that integrates census and administrative data across urban Brazil. Our unit of observation is a grid cell of approximately 1.2 square kilometers, for which we measure multiple dimensions of urban infrastructure, including sanitation, trash collection, street paving, sidewalks, public lighting, drainage and trees, as well as key variables for characterizing spatial equilibrium: population density and housing prices. We construct the urban infrastructure and population density variables from census data, while we measure quality-adjusted housing prices by georeferencing Brazil's social registry (Cadastro Único), an extensive administrative registry of low income households that contains information on housing characteristics and rent expenditures. The resulting dataset spans the urban territory of Brazil at a spatial granularity rarely available for developing economies. For our analysis, we focus on 341 cities with at least 50,000 inhabitants, resulting in a sample with over 30,000 grid cells and a combined population of 115 million people (60% of the Brazilian population). This rich dataset allows us to document stylized facts about urban structure.

Two descriptive patterns emerge from our empirical analysis. First, infrastructure provision declines systematically with distance to the city center, a negative gradient that holds across all dimensions of infrastructure we measure. For example, neighborhoods located 10 kilometers from the center have sanitation and street infrastructure levels more than a quarter of a standard deviation below those of central neighborhoods. Second, the unconditional correlations between distance to the city center and either population density or housing prices are more than 50% larger than when conditioning on infrastructure. This suggests that a significant part of the density and price gradients is accounted for, in a linear regression sense, by spatial differences in infrastructure. Taken together, these stylized facts reveal substantial within-city inequality in infrastructure and point to a role for unequal infrastructure provision in shaping urban spatial structure.

To interpret and evaluate these patterns, we derive new theoretical results that build on a classical literature on optimal public good provision in cities (Kanemoto 1980, Fujita 1989). We consider a monocentric city in which infrastructure is a local public good that directly affects the amenity value of neighborhoods. In equilibrium, both population density and housing prices decline with distance to the city center and increase with neighborhood infrastructure. We derive a rule for the optimal spatial allocation of infrastructure, that is, the provision that maximizes the welfare of city residents given a city-wide budget for urban infrastructure. The rule shows that population density levels, the elasticity of density with respect to infrastructure, and the marginal cost of provision are sufficient statistics for the optimal infrastructure levels. If the planner also

assigns weight to landowner welfare, the level and elasticity of land prices also enter the rule. Although our main theoretical results are derived for a monocentric city model, we show in an extension to quantitative spatial models that the rule applies to residential equilibrium in more general city structures.

A key implication of the optimal rule is that, if the elasticities of population density and land prices with respect to infrastructure and the cost of provision are stable within the city, then population density and land prices should fully predict optimal infrastructure levels. In particular, after conditioning on these two spatial equilibrium variables, infrastructure should not systematically vary with distance to the city center. This prediction allows us to revisit our main stylized fact that infrastructure is concentrated in central areas, now interpreted through the lens of the theory.

We take this implication to the data. We find that infrastructure provision is more concentrated in central areas than would be predicted by population density or land prices. Together, these variables explain only about half of the observed spatial gradient in infrastructure, suggesting that infrastructure levels systematically deviate from the optimal rule, with a bias toward the city center.

Our empirical assessment relies on the auxiliary assumption that costs and elasticities do not vary across space. We discuss this hypothesis in detail and argue that it is unlikely that our results are driven by systematic spatial differences in these components. Our findings hold after controlling for geographic characteristics and across a range of alternative specifications. Moreover, the deviations from the optimal rule appear across nearly all types of infrastructure in our data, including those whose provision costs should vary little within cities, such as road paving, street signage, and trees. We discuss the elasticities of population density and land prices with respect to infrastructure and show that existing evidence does not indicate systematically higher elasticities in central areas. We also use the functional form of the optimal rule to conduct a sensitivity test of our findings to these auxiliary assumptions. In sum, the empirical patterns suggest that the provision of infrastructure in Brazilian cities is inefficiently biased toward central locations.

Finally, we quantify the equilibrium consequences of the bias documented above. We consider a tractable version of the monocentric city model featuring isoelastic preferences and technology. The counterfactual is a city in which the planner allocates infrastructure to maximize residents' welfare, holding the total infrastructure budget fixed. We derive closed-form expressions showing how deviations from this rule affects population density gradients, the welfare of urban residents, and city size. The model is amenable to a transparent calibration that targets the gradients observed in the data and is disciplined by the estimates from the

reduced-form test of optimality.

We find that unequal infrastructure provision meaningfully shapes the internal structure of Brazilian cities. The bias toward central areas generates a population density gradient that is 27% steeper than optimal. As a result, urban population is considerably more concentrated near the city center than in the counterfactual in which infrastructure would be optimally redistributed toward the periphery. For example, population density at the city center is 21% higher than optimal. These distortions in city structure arise despite relatively modest effects on aggregate welfare, which falls by 0.10%, and city size, which contracts by 0.44%, losses equivalent to a reduction of about 4% in the total urban infrastructure stock.

Importantly, we note that our analysis is based on within-city gradients with respect to distance to the city center. A focus on these gradients gives us theoretical and empirical discipline, even if Brazilian cities are not strictly monocentric. As discussed in the paper, the sufficient statistics optimal provision rule extends to non-monocentric cities. We emphasize that all our findings are descriptive, but since city center locations reflect historical decisions, the estimation of these gradients rules out reverse causality (i.e. infrastructure affecting distance to the city center). Moreover, by focusing on inequality along distance to the city center, our analysis only accounts for part of the total spatial inequality in urban infrastructure. For instance, differences in infrastructure provision between neighborhoods at the same distance to the city center do not impact our results. This makes our diagnostic test and our quantitative results somewhat conservative.

We contribute to three strands of the literature. First, we contribute to a classical theoretical literature on public goods, like urban infrastructure, in monocentric cities. Key references include Helpman, Pines, and Borukhov (1976), Arnott (1979), Kanemoto (1980), Wijkander (1984), Koide (1988), and Okamoto (2009). See Fujita (1989) for a textbook treatment. We build on the results in this literature and extend them in two ways. First, we derive a sufficient-statistics rule for the optimal spatial allocation of local public goods –"neighborhood goods" in the terminology of Fujita (1989). In doing so, we revisit classical results in urban models following modern approaches in public economics (Chetty 2009). Our optimal rule resembles results from a recent literature on optimal spatial policies, which are often characterized by equilibrium variables and elasticities (Fajgelbaum and Gaubert 2020; Fajgelbaum and Gaubert 2025). Second, we assess this rule empirically using newly constructed high-resolution data on infrastructure, population, and housing prices for a large sample of cities. To our knowledge, we are the first to empirically evaluate the efficiency of local urban infrastructure within cities.

Second, we contribute to a small but growing literature empirically documenting urban infrastructure and public good provision within cities. Several recent papers focus on the effects of infrastructure: Gonzalez-Navarro and Quintana-Domeque (2016) show that street pavements impact property values, McIntosh et al. (2018) find positive real estate returns from local infrastructure spending, Coury et al. (2024) estimate that land prices respond to the provision of sewage and water access, McCulloch et al. (2025) show that sewers affect population density, and Han et al. (2024) examine the effects of trees on property prices. There is also evidence on the determinants of infrastructure provision. Feler and Henderson (2011) examine the determinants of piped water access in Brazil, showing that unequal provision was used to restrict migration. Of particular relevance to us is Harari (2024), who shows that segregation impacts inequality in access to various urban public goods in Brazilian cities, including the infrastructure types we study. Rather than estimate such causal relationships, we complement this literature by documenting stylized facts about within-city infrastructure provision and interpreting them through the lens of an equilibrium model. This approach allows us to evaluate whether the observed allocation of infrastructure is consistent with optimal provision.

Third, we contribute to a long literature documenting and interpreting housing prices and population density gradients through the lens of the monocentric city model. Examples include Brueckner and Fansler (1983), Mieszkowski and Smith (1991), McMillen (1996), Baum-Snow (2007), Ahlfeldt (2011), Brinkman (2016), Combes, Duranton, and Gobillon (2019), and Liotta, Viguié, and Lepetit (2022). Duranton and Puga (2015) provide a comprehensive survey of the literature, emphasizing the importance of new data, new empirical findings, and a tight connection to the theory. We contribute to this literature by showing that neighborhood infrastructure varies systematically with distance to the city center, that this gradient is steeper than what is predicted by the equilibrium variables in the optimal rule, and that this bias implies that cities are more spatially concentrated than they would be under an optimal allocation. We thus also complement the literature on city structure in developing economies, which has shown the importance of factors such as informality (Posada 2018; Henderson, Regan, and Venables 2021; Guedes, Iachan, and Sant'Anna 2023) or non-homothetic preferences (Deffebach et al. 2025). Our findings point to unequal infrastructure provision as another important factor shaping city structure in developing-economy cities.

The rest of the paper is structured as: Section 2 describes the data; Section 3 presents the stylized facts; Section 4 derives the optimal rule; Section 5 presents a reduced-form assessment of it; Section 6 uses the model to quantify the equilibrium consequences.

2 Data

We construct a high-resolution dataset combining information on urban infrastructure, housing prices, and population density for Brazilian cities. Our units are grid cells measuring $0.01^\circ \times 0.01^\circ$ (approximately 1.1 km at the Equator), cropped to Brazilian borders and restricted to cells overlapping urban areas. We assign each cell to a city, defined based on the classification of municipalities into *arranjos populacionais* by the Brazilian Institute of Geography and Statistics (IBGE). Based on commuting patterns, this city definition is equivalent to a metropolitan area. We restrict our sample to cells in the 341 cities with at least 50,000 urban inhabitants in 2010. Online Appendix A.2 contains details on the data construction. The resulting dataset provides unusually rich and spatially fine information on neighborhood infrastructure and spatial equilibrium object for a large sample of cities in an important developing economy.

We use census block data from the 2010 Census to measure neighborhood infrastructure. From the household questionnaire, we use information on access to three basic sanitation services (water, sewage, trash collection). From the *entorno* module of the census, we have information on 8 street-level infrastructure attributes (signs with a street name, public lighting, street pavement, sidewalks, curbs, drainage systems, accessibility ramps, trees) and two local public bads about the street (open sewage, uncollected trash). With a spatial crosswalk between grid cells and census blocks, we compute the share of households in each cell with access to each infrastructure item. For each of the two public bads, we measure instead the share of households free from the nuisance. To make the presentation of our results manageable, we often focus on aggregate indices of sanitation or street infrastructure. We construct each index by demeaning the infrastructure measures, dividing by their standard deviations, and averaging across all measures within each infrastructure category. In this way, the sanitation (street) infrastructure index indicates how much better the sanitation (street) infrastructure of the cell is in comparison with average neighborhoods, measured in standard deviations of the infrastructure distribution.[1]

We also incorporate data on two main equilibrium variables, housing prices and residential population density. Housing prices are estimated from Cadastro Único (CadUnico), an administrative database containing extensive demographic and housing information for low-income families. We georeference roughly

[1] Some census blocks lack infrastructure information due to censoring or coverage limitations in the 2010 Census, particularly in the *entorno* module. As a result, sanitation infrastructure data are missing in 6 grid cells, and street infrastructure data are missing in 1,423 grid cells.

104 million household-year observations between 2013 and 2020, of which about 19% report rent expenditures. Using these data, we estimate quality-adjusted log rent at the grid-cell level by regressing log rent on housing characteristics, year fixed effects, and grid fixed effects. To increase the cross-sectional coverage and quality of our log rent measures, our preferred measure pools observations from all years. This yields a high-resolution, harmonized, and precise database of housing prices across urban Brazil. We use census block data to measure population density. For consistency with the infrastructure data, our main results use the 2010 Census, although we also consider population density from the 2022 Census. Online Appendix A.2 contains details on the construction of these variables.

Although our housing price data are drawn from CadUnico, an administrative dataset focused on low-income households and thus not representative of the Brazilian population at large, the resulting rent measures are still informative about variations in housing prices. In the absence of strong segmentation in housing markets across income groups, spatial differences in rents paid by households in our sample tend to reflect common price differences across neighborhoods. In Online Appendix A.3, we validate our housing price measure by comparing it to municipal-level rents reported in the 2010 Census and to property sale prices within São Paulo, finding strong correlations in both cases. These validation tests show that our quality-adjusted log rent measures are good predictors of spatial variation in housing prices.

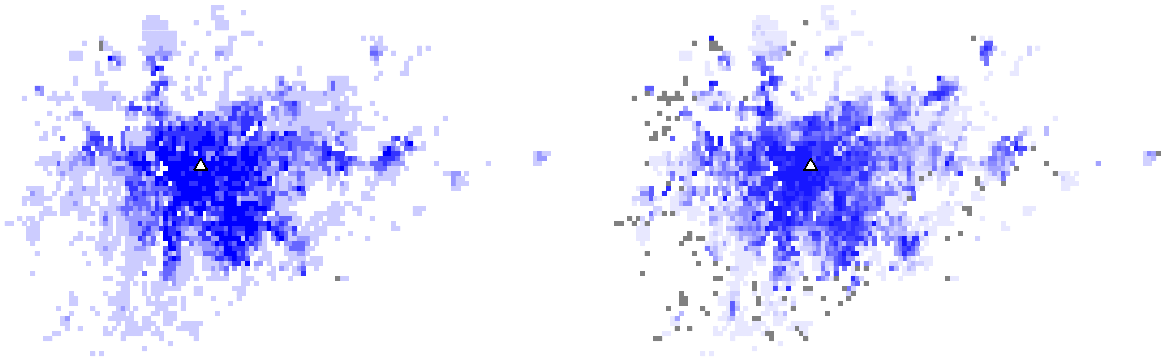
Our final database combines all the information above at the grid cell level. To preserve the anonymity of CadUnico respondents, we restrict our sample only to grid cells featuring at least 10 families with rent expenditure information in the data. This restriction mainly removes very low-density peripheral cells from the original grid. The resulting sample contains 32,610 grid cells, with a combined population of 115 million inhabitants in 2010. To improve precision and ensure that our estimates reflect conditions faced by the average urban household, we weight each grid cell by the number of resident households from the 2010 Census. As an illustration of our data, Figure 1 contains maps of infrastructure, population density and housing prices for São Paulo, the largest city in our sample.

Finally, we construct additional variables at the grid cell level. As geographic controls affecting amenities and construction costs, we record altitude, terrain ruggedness, whether the cell is located along the coast, and a series of soil characteristics. We identify *favelas* (informal settlements) using IBGE's classification of *aglomerados subnormais* for 2010. We measure employment density by georeferencing formal establishments in the *Relação Anual de Informações Sociais* (RAIS), a matched employer-employee administrative

Figure 1: Map of infrastructure, population density and housing prices – São Paulo

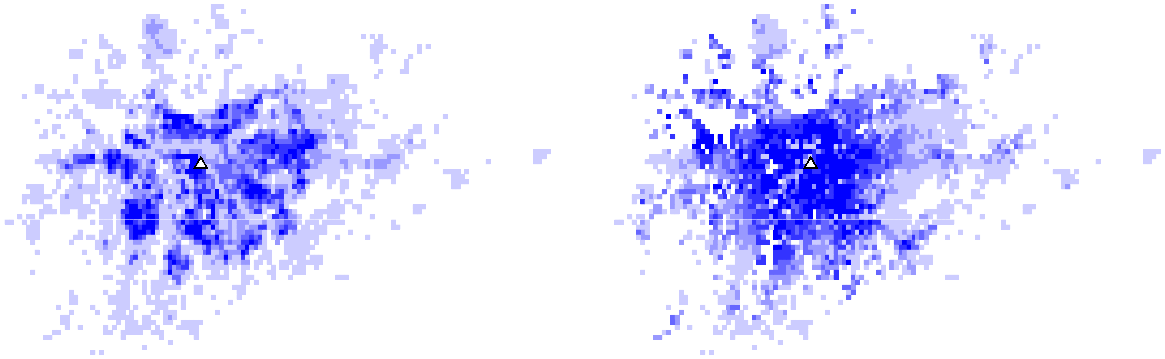
(a) Sanitation infrastructure

(b) Street infrastructure



(c) Population density

(d) Housing prices



Notes: The figure displays maps of our main variables for the city of São Paulo. The city center (Sé Square) is indicated by the white triangle. The maps show the sanitation infrastructure index (top left panel), street infrastructure index (top right panel), population density (bottom left panel) and housing prices (bottom right panel). Shades of blue represent the decile of each variable; darker colors indicate higher deciles. Gray grid cells on the top right map refer to those for which there is no street infrastructure data.

dataset. We incorporate data from property tax assessments and from locally taxed real estate transactions in São Paulo to construct alternative measures of land prices. To identify the location of city centers, we use the database of 2010 municipal seat locations from IBGE, complemented by Google Maps queries for selected large cities when necessary. Our main results use linear distances to measure how far a grid cell is from the city center; we also collect road distances and driving times. We winsorize distances at the top percentile to reduce the influence of peripheral outliers, and express distances in units of 10 kilometers to improve the readability of regression coefficients.

Appendix Table A.1 presents summary statistics of the infrastructure measures (top panel) and other

variables used in our analysis (bottom panel). The average household in our sample lives at a distance of 11.3 km from the city center, in cells with 2,912 households and with a population density of 79.7 inhabitants. Only 71.5% of the households have access to a sewage connection, 74.5% live in streets with sidewalks, 66.9% live in streets with some tree coverage, and 47.8% in places with drainage.

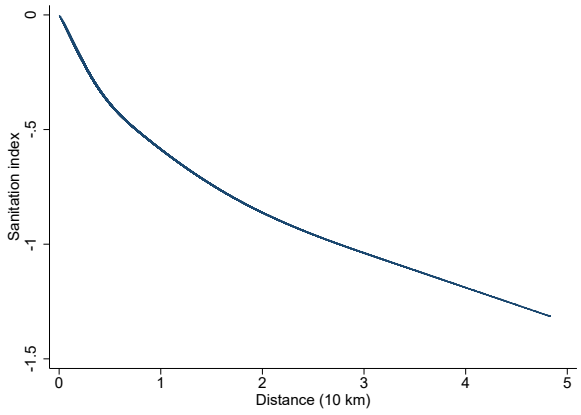
3 Stylized Facts

3.1 Fact 1: Infrastructure Gradients

Figure 2 shows that infrastructure provision declines with distance to the city center. We regress the sanitation (left panel) and the street (right panel) infrastructure indexes on city fixed effects and a cubic spline of distance to the city center. At the bottom of each figure, we also display the linear coefficient of the index on distance to the city center. The estimates indicate that, on average, a 10 km increase in distance is associated with a 0.264 (0.279) standard deviation decline in sanitation (street) infrastructures. These infrastructure gradients are the main stylized fact we document.

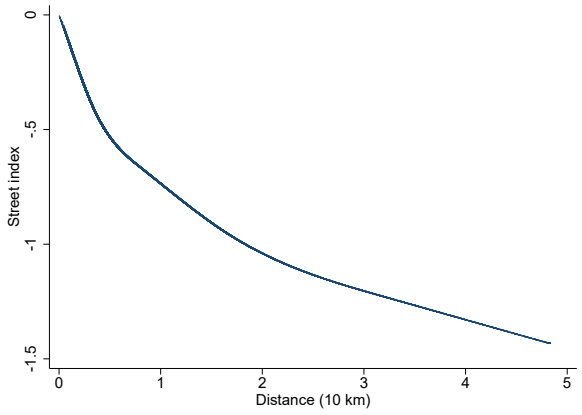
Figure 2: Gradients of infrastructure

(a) Sanitation infrastructure



-0.264 (0.031)

(b) Street infrastructure



-0.279 (0.017)

Note: The graphs display estimates from a regression of the sanitation (left figure) and street (right figure) infrastructure indices on city fixed effects and a cubic spline of distance to the city center. The coefficients on the bottom of the figures refer to the linear regression on the city fixed effects and a linear measure of distance to the city center, in units of 10 km. Robust standard errors in parentheses are clustered at the city level. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The left graph is estimated on the 32,604 grid cells with sanitation measures, while the right graph is estimated on the 31,187 grid cells for which there are data from the *entorno* module of the 2010 Census.

Table 1 reports estimated distance gradients for all three measures of sanitation infrastructure and all ten measures of street infrastructure. All coefficients are negative and statistically significant, highlighting that every aspect of neighborhood infrastructure is worse in more peripheral neighborhoods. Moreover, the coefficients are substantial relative to the sample means. For instance, each 10 km farther away from the city center reduces the share of households with sewage access by 11.9 p.p., out of a mean of 71.5%, or with access to drainage infrastructure by 10.1 p.p., out of a mean of 47.8%. In sum, Brazilian cities are characterized by unequal within-city distributions of urban infrastructure, which is systematically worse in more peripheral neighborhoods.

Table 1: Gradients of infrastructure

	(1)	(2)	(3)	(4)	(5)
Sanitation infrastructure:	water	sewage	trash		
Distance to city center	-0.038***	-0.119***	-0.009***		
	(0.008)	(0.015)	(0.002)		
Sample mean	0.931	0.715	0.985		
R-squared	0.388	0.610	0.206		
Street infrastructure (1):	sign	light	paved	sidewalk	curb
Distance to city center	-0.103***	-0.014***	-0.078***	-0.113***	-0.072***
	(0.008)	(0.002)	(0.007)	(0.013)	(0.010)
Sample mean	0.673	0.969	0.852	0.745	0.801
R-squared	0.545	0.128	0.368	0.408	0.405
Street infrastructure (2):	drainage	ramps	trees	open sew.	unc. trash
Distance to city center	-0.101***	-0.051***	-0.061***	-0.026***	-0.007**
	(0.008)	(0.007)	(0.008)	(0.005)	(0.003)
Sample mean	0.478	0.053	0.669	0.898	0.945
R-squared	0.545	0.203	0.527	0.527	0.101

Notes: The table display estimates from a regression of the infrastructure measure indicated on the top on city fixed effects and distance to the city center, in units of 10 km. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample for the top panel consists of the 32,604 grid cells with sanitation measures. The sample for the bottom panels consists of the 31,187 grid cells for which there are data from the *entorno* module of the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

3.2 Fact 2: Relation with Price and Density Gradients

The monocentric city model implies that housing prices and population density both decrease with distance to the city center due to commuting costs. Since we find that urban infrastructure also deteriorates with distance, this spatial inequality may reinforce these gradients if residents value infrastructure. As suggestive evidence on how much infrastructure contributes to observed housing price and density gradients, we compare the estimates from two linear regression models: one without infrastructure controls, and another that includes infrastructure measures as covariates. The amount by which the coefficient falls with the inclusion of the infrastructure measures reflects the portion of the overall gradient that we may attribute to observable differences in neighborhood infrastructure. We emphasize that this is a correlational decomposition: while informative about the spatial association between infrastructure and equilibrium outcomes, its goal is not to recover the causal effects of infrastructure on prices or density, which is beyond the scope of this paper.

We report the estimates in Table 2, using the logarithm of quality-adjusted rents (columns 1 to 3) and the logarithm of population density (columns 4 to 6) as dependent variables. In columns 1 and 4, we do not control for infrastructure. In columns 2 and 5, we control for infrastructure using the two composite indices. In columns 3 and 6, we include instead the full set of 13 individual infrastructure measures.[2] Conditioning on infrastructure substantially attenuates the estimated gradients: the rent gradient declines from –0.101 to –0.066 log points, and the density gradient from –0.450 to –0.297 log points. The unconditional gradients are thus 53–59% larger than the conditional ones, suggesting that a significant part of the overall gradients are predicted by differences in neighborhood infrastructure.

Figure 3 semi-parametrically illustrates this finding with a regression on a cubic spline of distance to the city center and city fixed effects. The two infrastructure indices are used as controls in the estimation of the dashed lines, but not of the solid ones. The gap between the two curves highlights how much of the spatial gradient in housing outcomes is predicted by infrastructure differences.

[2]To retain a consistent sample across specifications, we include indicators for missing sanitation and street infrastructure data.

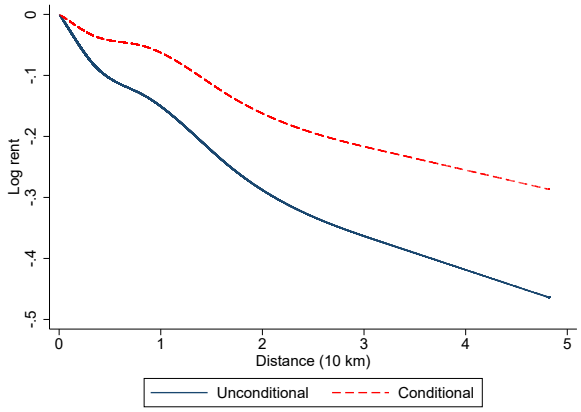
Table 2: Conditional and unconditional price and density gradients

	Log rent			Log population density		
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to city center	-0.101***	-0.066***	-0.064***	-0.450***	-0.297***	-0.282***
	(0.008)	(0.008)	(0.009)	(0.032)	(0.024)	(0.023)
Index of street infrastructure		0.093***			0.107***	
		(0.007)			(0.040)	
Index of sanitation infrastructure		0.036***			0.441***	
		(0.005)			(0.045)	
R-squared	0.729	0.765	0.770	0.472	0.564	0.593
Observations	32610	32610	32610	32610	32610	32610
Control for infrastructre?	no	indices	full	no	indices	full

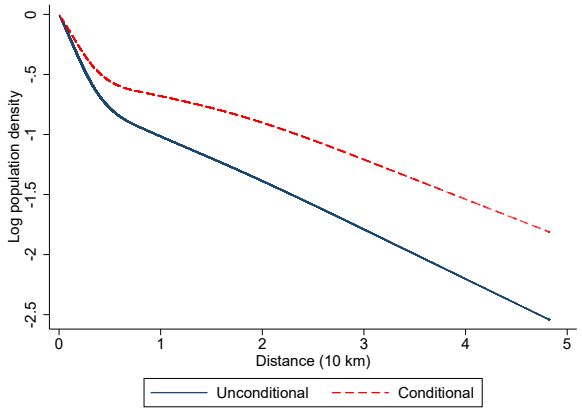
Notes: The table display estimates from a regression of quality-adjusted log housing rents (columns 1 to 3) or log population density in 2010 (columns 4 to 6) on city fixed effects and distance to the city center, in units of 10 km. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. In columns 2 and 5, the equation also includes the sanitation and street infrastructure indices, as well as dummy variables indicating whether each of these indices are missing. In columns 3 and 6, we replace the indices by a vector containing all 13 infrastructure variables. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Figure 3: Conditional and unconditional price and density gradients

(a) Housing prices



(b) Population density



Notes: The solid lines display estimates from a regression of quality-adjusted log housing rents (left figure) or log population density in 2010 (right figure) on city fixed effects and a cubic spline of distance to the city center. For the dashed lines, the equation also includes the sanitation and infrastructure indices, as well as dummy variables indicating whether each of these indices are missing. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census.

4 Optimal Infrastructure Provision

4.1 Economy

We consider an economy composed of a monocentric city and a rural hinterland. The total population in the economy is L_0, of which L live in the city in equilibrium. Rural residents receive an exogenous utility ω_0. Urban residents, on the other hand, derive utility from the consumption of private goods and from the infrastructure of the neighborhood they live in. The city features a continuum of neighborhoods, and each neighborhood is characterized by its distance $r \in [0, \overline{r}]$ to the city center, where $\overline{r} \leq \infty$. We consider symmetric equilibria in which every neighborhood at distance r has the same prices, density, and infrastructure. Denote by $\ell(r)$ the measure of land at distance r to the city center. Every urban resident works at the city center and, due to commuting costs, the income available for private good consumption is a strictly decreasing function of distance to the city center, $y(r)$.

Urban residents value both the infrastructure of the neighborhood they live, denoted as a composite public good $g(r)$, and their consumption of housing h and of a private tradable good x. We restrict $g(r)$ to be a continuous function. We take x as the numeraire and denote the (rental) price of housing as q. Each household i also draws an idiosyncratic urban taste shock, b_i, according to a continuous cumulative distribution function $F(b)$. The urban taste shock induces heterogeneity in the willingness to pay to live in the city, leading to an upward-sloped supply of residents to the city. The utility of resident i living in neighborhood r, as a function of their private consumption (h, x), is $V_i = a(g)u(h, x)b_i$. Note that neighborhood infrastructure enters as an amenity value associated with the neighborhood r. We assume that: the function $u(h, x)$ is twice continuously differentiable, increasing in both arguments, strictly quasi-concave, and the upper contour of any strictly positive private good bundle (h, x) does not intersects the axes; housing is a normal good; the function $a(g)$ is positive valued, strictly increasing, strictly concave, and twice continuously differentiable; $y(r)$ is continuously differentiable. These assumptions imply, through standard consumer theory arguments, that duality holds and that both the Marshallian demand for housing, $h^m(q, y)$, and the Hicksian demand $h^h(q, u)$ are continuously differentiable. Moreover, the consumer problem results in a continuously differentiable indirect utility function $v(q, y)$, so $V_i = a(g)v(q, y)b_i$. The indirect utility is increasing in income and decreasing in housing prices q. The Hicksian demand is decreasing in prices and, since housing is normal, so is the Marshallian demand.

Residents are perfectly mobile across neighborhoods, so the utility of living in each neighborhood is the same, conditional on the urban taste shock. Denote this utility level by ω. We write the indifference condition as:

$$\omega = \frac{V_i}{b_i} = a(g)v(q, y) \tag{1}$$

Residents are also mobile between the city and the rural hinterland. An individual moves to the city if $\omega b_i \geq \omega_0$. The average utility in the population is thus $W = E[max\{\omega_0, \omega b_i\}]$.

The economy features competitive markets for housing and urban land. Land is held by absent landowners who do not live in the city. We denote the equilibrium price of land by p. Profit-maximizing developers use land and capital (i.e. the tradable numeraire) to build housing according to a constant returns to scale technology. In equilibrium, prices adjust so developers make zero profits. These conditions imply that the total housing supply built per unit of land, $s(q)$, and the price of land, $p(q)$, are increasing functions of housing prices. We assume that $s(q)$ and $p(q)$ are continuously differentiable.

Finally, we consider a social planner that provides the public good $g(r)$ given an exogenous budget. We assume there is a citywide public budget constraint of the form

$$\int_0^{\bar{r}} c(g(r), r)\ell(r)dr = C_0, \tag{2}$$

where $c(g(r), r)$ indicates the weakly convex cost function of providing infrastructure to neighborhood r and C_0 indicates total public funds for infrastructure.

Comparison with the classical literature. The framework above builds on a classical theoretical literature examining optimal public good provision in a monocentric city (Kanemoto 1980, Fujita 1989). We model infrastructure as a neighborhood good in the sense of Fujita (1989): a purely local public good jointly consumed by residents of a given location. For tractability, we assume that the neighborhood good is non-congestible, so its amenity value does not depend on population density, and that the level of neighborhood goods does not affect the marginal rate of substitution between housing and the tradable consumption good. None of these assumptions are essential for our results.

Our framework departs from the canonical analysis in Kanemoto (1980) and Fujita (1989) in three important aspects. First, instead of focusing on a closed city (fixed population) or an open city (fixed utility), we consider an imperfectly open city where workers decide whether to live in the city or in the rural hinter-

land. By varying the distribution $F(b)$ of urban taste shocks, our environment nests both closed and open cities as limiting cases.[3] Second, we consider a parsimonious problem in which the central planner controls only the distribution of public goods, under a budget constraint, while the classical literature usually considers an extended maximization problem where the planner also chooses housing allocations and population density. Finally, while the earlier literature was focused on investigating the optimal city structure and the decentralization of public good provision, our goal is to find a sufficient-statistics rule that can be used to empirically evaluate the spatial allocation of infrastructure. To our knowledge, this type of condition has not been previously derived in this context.

4.2 Equilibrium

We first examine how the distribution of neighborhood infrastructure $g(r)$ affects the urban equilibrium. We denote equilibrium objects at the neighborhood level as a joint function of distance to the city center r and the neighborhood infrastructure $g(r)$. We consider housing prices $q(r, g(r))$, land prices $p(r, g(r))$, per capita housing consumption $h(r, g(r))$, housing supply $s(r, g(r))$ and population density $d(r, g(r))$. Aggregate equilibrium variables are the city size L and the welfare of urban residents ω.

The indifference condition 1 implies that, holding infrastructure constant, prices and density decrease with distance to the city center. The logic is the same as in the standard monocentric city model. The indifference condition also shapes the effects of improvements in neighborhood infrastructure on local prices and quantities. Since the utility of living in the neighborhood remains constant when its infrastructure improves, the larger g is capitalized in housing prices $q(r, g(r))$. As the neighborhood infrastructure does not affect income but increases housing prices, housing consumption reacts to the higher prices along the Marhsallian demand curve. Since housing is normal, then $h(r, g(r))$ decreases with g. The reaction on prices and housing consumption further implies that total housing supply $s(r, g(r))$, population density $d(r, g(r))$ and land prices $p(r, g(r))$ are increasing in infrastructure g. The proposition below formalizes the discussion above; the formal proof is in Appendix A.4.1.

Proposition 1. *Conditional on the level of public goods $g(r)$, housing prices $q(r, g(r))$, housing supply*

[3] For a similar imperfeclty open city approach, see Ahlfeldt, Baum-Snow, and Jedwab (2023). An advantage of this formulation is that it allows us to obtain results that apply to both canonical limiting cases; in contrast, Kanemoto (1980) has to present two model solutions.

$s(r, g(r))$, population density $d(r, g(r)$ and land prices $p(r, g(r))$ are decreasing in r, while per capita housing consumption $h(r, g(r))$ is increasing in r. Conversely, conditional on distance to the city center r, housing prices $q(r, g(r))$, housing supply $s(r, g(r))$, population density $d(r, g(r)$ and land prices $p(r, g(r))$ are increasing in $g(r)$, while per capita housing consumption $h(r, g(r))$ is decreasing in $g(r)$.

Proposition 1 characterizes the within-city distribution of prices and quantities. To characterize the aggregate equilibrium variables, urban welfare ω and city size L, we consider the supply and demand for city residents. The urban taste shocks imply that the supply of residents to the city is increasing in ω:

$$L = \left[1 - F\left(\frac{\omega_0}{\omega}\right)\right] L_0 \tag{3}$$

The demand for city residents is given by the integral of the population density curve $d(r, g(r))$:

$$L = \int_0^{\bar{r}} d(r, g(r))\ell(r)dr \tag{4}$$

Since the population density at each point is decreasing in ω, then total demand for city residents is also decreasing in ω. The equilibrium values of ω and L occur at the point where demand and supply are equal.

4.3 Optimal Infrastructure Provision Rules

We now characterize the optimal allocation of neighborhood infrastructure $g(r)$ across space. We begin by examining the allocation that maximizes average worker welfare in the economy, W. Since rural welfare is exogenous, this is equivalent to maximizing urban welfare ω. Moreover, since the supply of residents to the city is upward sloping in ω, this maximum occurs when the total city population, as expressed in the demand for residents equation 4, is maximized. Formally, the planner solves:

$$\max_{g(r)} \int_0^{\bar{r}} d(r, g(r))\ell(r)dr \quad \text{s.t.} \quad \int_0^{\bar{r}} c(g(r), r)\ell(r)dr = C_0.$$

This is a standard static isoperimetric problem in the calculus of variations. We form the Lagrangian $\mathcal{L}[g(\cdot)] = \int_0^{\bar{r}} \left[d(r, g(r)) - \lambda c(g(r), r)\right] \ell(r)dr$, where λ is the Lagrange multiplier on the budget constraint. The Euler-Lagrange condition implies that the partial derivative of the Lagranian with respect to g must be zero at the optimum, which yields the first-order condition $\frac{\partial d(r,g)}{\partial g} = \lambda c_g(g(r), r)$. To interpret this

condition in terms of sufficient statistics for optimal infrastructure provision, we rewrite it using elasticities. Let $\varepsilon_{d,g}(r) \equiv \frac{\partial d(r,g)}{\partial g} \cdot \frac{g(r)}{d(r,g)}$ be the elasticity of population density with respect to infrastructure and, to simplify the notation, let $c_g(r) = c_g(g(r), r)$ be the marginal cost of infrastructure provision in neighborhood r. By rearranging the condition, we obtain the optimal infrastructure rule:

$$g(r) = \frac{\varepsilon_{d,g}(r)}{\lambda c_g(r)} d(r, g) \tag{5}$$

This rule characterizes the optimal provision as proportional to local population density, scaled by the elasticity of density with respect to infrastructure and the local marginal cost. If both $\varepsilon_{d,g}$ and $c_g(r)$ are constant across space, then infrastructure should be optimally allocated in proportion to density. A remarkable feature of this rule is its simplicity: spatial equilibrium responses to neighborhood infrastructure imply that the goal of the planner is to maximize overall city size, which also leads to higher welfare. This explains why population density is a key statistic for optimal infrastructure provision.

The optimal rule 5 refers to the problem of maximizing workers' welfare. A social planner may also assign value to the land rents received by landowners. To allow for a more general rule, we consider an extended version of the planner problem in which the planner assigns weight θ to city demand for residents and weight $1-\theta$ to land prices.[4] This is again a standard isoperimetric problem in the calculus of variations, yielding first-order condition $\theta \frac{\partial d(r,g)}{\partial g} + (1-\theta)\frac{\partial p(r,g)}{\partial g} = \lambda c_g(r)$. Rearranging in terms of elasticities, we obtain the optimal rule:

$$g(r) = \frac{1}{\lambda c_g(r)} \Big[\theta \varepsilon_{d,g}(r) d(r, g) + (1-\theta) \varepsilon_{p,g}(r) p(r, g) \Big] \tag{6}$$

This extended optimal rule differs from the previous rule by including also land prices $p(r, g)$ and the elasticity of land prices with respect to neighborhood infrastructure, $\varepsilon_{p,g}(r)$. The proposition below summarizes the main takeaways from these optimal rules.

Proposition 2. *Equation 6 is a necessary condition held by any non-zero optimal level of public good provision $g(r)$. It implies that the levels of population density and land prices, the local elasticities of population density and land prices with respect to public infrastructure, and the marginal cost of infrastructure provision*

[4] We interpret θ as a reduced-form parameter capturing how land-rent considerations affect optimal provision. See Basso, Pezoa, and Silva (2024) for the relevance of land values to optimal policy in urban equilibrium.

fully explain the within-city distribution of infrastructure.

A relevant particular case occurs if the elasticities and costs do not vary over space. These constant elasticities would follow from standard Cobb-Douglas preferences and housing production technology, as we later show in Section 6, while the constant provision cost functions would be a reasonable assumption in homogeneous landscapes. In this case, Proposition 2 provides the even sharper prediction that the optimal provision can be explained fully by population density and land prices. The corollary below, which provides the basis for our empirical test, summarizes this prediction.[5]

Corollary 1. *If the elasticities of population density and land prices with respect to public infrastructure and the cost of infrastructure provision do not vary with distance to the city center, then neither does optimal infrastructure provision after conditioning on population density and land prices.*

4.4 Extension to Non-Monocentric Cities

While we consider a monocentric city model, our results extend to more general city structures. In recent years, a growing literature has used quantitative spatial models (QSMs) to examine the distribution of economic activity within cities (Ahlfeldt et al. 2015, Redding 2023). QSMs differ from the monocentric city model by allowing for dispersed employment, spatial heterogeneity in amenities and productivity, and location-specific preference shocks that imply households will not all be indifferent among neighborhoods. In Appendix A.4.2, we show that our sufficient-statistics rule extends to a class of QSMs. In particular, if wages are exogenous, optimal infrastructure provision will be proportional to population density. This extension shows that our rule is not specific to monocentric settings, but instead is a consequence of spatial equilibrium in residential markets.[6]

Note that the assumption of exogenous wages shuts down local agglomeration productivity spillovers. In a monocentric city, these agglomeration spillovers do not distort the optimal rule, as all employment is at the city center. However, in a non-monocentric city, they may bring a new consideration to the social

[5]The constant elasticity and cost assumption is sufficient but not necessary for density and prices to fully explain optimal infrastructure provision. This will also occur if these elasticities and marginal costs are a function of these variables. This is the case, for instance, if household preferences over housing and consumption have a constant elasticity of subsitution format, or if the costs of providing infrastructure depend on population density.

[6]The fact that our rule applies to a class of QSMs is consistent with a recent literature on optimal spatial policies in spatial economies, which also derives optimal policy as a function of equilibrium objects and elasticities (Fajgelbaum and Gaubert 2020, Fajgelbaum and Gaubert 2025).

planner: better infrastructure in a neighborhood increases the neighborhood population, which in turn may cause some households to change their employment locations. Hence, if these employment changes are towards locations with a larger agglomeration elasticity, there may be positive effects on the city aggregate productivity. These general equilibrium feedback effects are not captured by our sufficient-statistics rule, which focuses on spatial equilibrium in the residential market alone.

5 Diagnostic Test

5.1 Assessing the Optimal Rule

Proposition 2 implies a rule for the optimal spatial allocation of urban infrastructure, which should follow the pattern described in equation 6. Two observable components of this rule are population density and, when the planner assigns positive weight to landowners' welfare ($\theta < 1$), land prices. The rule also depends on three unobservable components: the elasticities of population density and land prices with respect to infrastructure, and the marginal cost of providing infrastructure.

The optimal rule naturally leads us to reconsider the infrastructure gradients with respect to distance to the city center, our main stylized fact from Section 3. Since central locations tend to have higher population density and more valuable land, the optimal rule implies that they would likely receive more infrastructure under an optimal allocation. However, this relationship should be fully accounted for by the components in the optimal rule. In what follows, we examine whether the infrastructure gradients persist after conditioning on population density and land prices, the two observable components in the rule. Any infrastructure gradient that cannot be accounted for these two variables will reflect either: **(i)** higher non-observable components of the optimal rule (elasticities and inverse marginal costs) near the city center, or **(ii)** inefficiencies in the within-city infrastructure provision. In fact, Corollary 1 implies that, if these elasticities and marginal costs were constant across space, the correlation between infrastructure and distance to the city center should be zero after conditioning on population density and land prices. This implication is what we use to test for optimality. We later argue against unobserved variation in costs or elasticities as the reason for our rejection of optimality in infrastructure provision.

Based on this logic, we estimate the following equation by ordinary least squares:

$$Infra_{i,c} = \beta_0 Dist_{i,c} + \beta_1 Density_{i,c} + \beta_2 Price_{i,c} + \gamma_c + u_{i,c}, \tag{7}$$

where $Infra_{i,c}$ is a measure of infrastructure in grid cell i of city c; $Dist_{i,c}$ is distance to the city center; $Density_{i,c}$ is population density; $Price_{i,c}$ is a proxy for land prices; γ_c are city fixed effects; and $u_{i,c}$ is the error term. The inclusion of density and prices in levels follows directly from the structure of the optimal allocation rule in equation 6; we show that the results hold under alternative specifications. As in the estimates reported in Section 3, we weight the observations by the number of residing households and compute robust standard errors clustered at the city level. We reject the optimality of the infrastructure distribution if distance retains predictive power after conditioning on population density and land prices.

We do not observe land prices directly. To circumvent this challenge, we assume a Cobb-Douglas housing production function and use our quality-adjusted housing rents from CadUnico to input land prices. In particular, we assume that land prices p are proportional to $q^{1/\alpha}$, where q is the quality-adjusted rent and α is the land share in housing production. We use a baseline value of $\alpha = 0.35$, consistent with estimates in Combes, Duranton, and Gobillon (2021), but also conduct robustness checks with alternative values. In robustness tests, we use alternative land price data available for São Paulo.

Our empirical diagnostic test follows directly from the planner's first-order condition for the optimal spatial allocation of infrastructure. Because infrastructure, density and land prices are jointly determined in spatial equilibrium, the estimated coefficients are not intended to identify causal effects, nor is a causal interpretation required for our objectives. The test is valid as long as, under optimal provision, the unobserved determinants of optimal infrastructure provision (marginal costs and elasticities) do not vary systematically with distance to the city center once density and prices are held constant. We later present a sensitivity test that examines violations of this assumption, and Section 5.4 discusses the assumption in detail.

5.2 Main Results

Table 3 presents estimates of equation 7. In columns 1 and 2, the dependent variable is the index of sanitation infrastructure, while in columns 3 and 4, it is the index of street infrastructure. In odd columns, we condition only on population density; these estimates refer to the restricted test that the allocation maximizes

residents' welfare, not weighting for landowners' welfare. Even columns include both population density and land prices. The latter provides our preferred results.

Table 3: Testing the optimality rule

	sanitation infrastructure		street infrastructure	
	(1)	(2)	(3)	(4)
Distance to city center	-0.180***	-0.147***	-0.234***	-0.157***
	(0.024)	(0.027)	(0.026)	(0.020)
Population density	0.003***	0.003***	0.001**	0.001***
	(0.001)	(0.001)	(0.001)	(0.000)
Land prices proxy		0.973***		2.227***
		(0.331)		(0.439)
R-squared	0.508	0.514	0.474	0.516
Observations	32604	32604	31187	31187

Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 2) or of the street infrastructure index (columns 3 and 4) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 (in inhabitants per hectare) and, in even columns, a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample for columns 1 and 2 consists of the 32,604 grid cells with sanitation measures. The sample for columns 3 and 4 consists of the 31,187 grid cells for which there are data from the *entorno* module of the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Across all specifications, distance to the city center retains strong, negative, and statistically significant correlations with sanitation and street infrastructure measures. Compared to the unconditional gradients reported at the bottom of Figure 2, the conditional correlations retain 56% of the original magnitude after conditioning on population density and land prices. These estimates reject the null hypothesis that the within-city infrastructure distribution can be fully explained by population density and land prices, suggesting that the gradients of infrastructure provision with respect to distance to the city center are steeper than what would be optimal. Strikingly, more than half of the overall infrastructure gradient cannot be explained by the equilibrium variables that should predict optimal provision.

To examine which specific dimensions of infrastructure drive the deviation from optimality, Table 4 reports estimates of equation 7 separately for each of the 13 infrastructure measures. With the exception of freedom from uncollected trash, all measures exhibit statistically significant negative correlations with distance to the city center after conditioning on population density and land prices. These results confirm that our main findings from Table 3 are not driven by few infrastructure dimensions, reflecting instead a broad pattern of underprovision of both sanitation and street infrastructures in peripheral neighborhoods.

The table also reports the sample mean for each indicator, which allows us to compute the conditional semi-elasticity of infrastructure with respect to distance. On average, a 10 km increase in distance is associated with a 9.0% decline in infrastructure levels relative to the sample mean; the median semi-elasticity across the 13 measures is 4.6%. We later use these semi-elasticities in the quantitative analysis of Section 6.[7]

Table 4: Results for each infrastructure measure

	(1)	(2)	(3)	(4)	(5)
Sanitation infrastructure:	water	sewage	trash		
Distance to city center	-0.018***	-0.077***	-0.004**		
	(0.003)	(0.013)	(0.002)		
Sample mean	0.931	0.715	0.985		
R-squared	0.410	0.629	0.222		
Street infrastructure (1):	sign	light	paved	sidewalk	curb
Distance to city center	-0.073***	-0.009***	-0.039***	-0.068***	-0.034***
	(0.008)	(0.003)	(0.005)	(0.010)	(0.007)
Sample mean	0.673	0.969	0.852	0.745	0.801
R-squared	0.563	0.139	0.406	0.439	0.433
Street infrastructure (2):	drainage	ramps	trees	open sew.	unc. trash
Distance to city center	-0.069***	-0.028***	-0.038***	-0.013**	0.001
	(0.012)	(0.008)	(0.006)	(0.006)	(0.005)
Sample mean	0.478	0.053	0.669	0.898	0.945
R-squared	0.561	0.257	0.555	0.534	0.121

Notes: The table display estimates from a regression of the infrastructure measure indicated on the top on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample for the top panel consists of the 32,604 grid cells with sanitation measures. The sample for the bottom panels consists of the 31,187 grid cells for which there are data from the *entorno* module of the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

5.3 Robustness

Sensitivity analysis. The validity of the test depends on the auxiliary assumption that elasticities and marginal costs of infrastructure provision are the same across space within each city, two assumptions dis-

[7] In Appendix Table A.2, we report qualitatively similar results when conditiong only on population density, but not on land prices.

cussed in the next sub–section. To assess how deviations from this assumption affect our results, we exploit the fact that in the optimal rule 6 the ratios of elasticities to marginal costs multiply population density and land prices. We thus re-estimate the test replacing these variables by the product of each variable with the factor $e^{-s\ Dist_{i,c}}$, where $Dist_{i,c}$ is the distance to the city center and s indicates how the ratios of the elasticities to marginal costs decrease with distance. Our baseline results correspond to $s = 0$. By setting $s > 0$ ($s < 0$), we allow instead for infrastructure provision to be uniformly cheaper (more expensive) or for its impacts on density and land values to be higher (lower) near the city center.[8] The parameter s indicates the semi-elasticity of the ratio of elasticity to marginal costs with respect to distance, in units of 10 km. This robustness test provides a tractable way to gauge the sensitivity of our main findings. The estimates, reported in Figure 4, show that for a wide range of values for s the coefficient on distance to the city centers remains negative and statistically significant. These results suggest that violations of the constant–elasticity and cost assumptions do not explain our rejection of optimality.

Specification tests. Appendix Table A.3 shows that our rejection of optimality is robust to several alternative specifications. In column 1, we include a long list of geographic controls: altitude, ruggedness, a coastal dummy, soil type fixed effects, soil bulk density, and soil available water content.[9] Although these geographic variables would plausibly absorb much of the unobserved variation in marginal costs or elasticities, the estimates barely change. In columns 2 and 3, we construct land prices by using alternative values for the share of land in housing production, $\alpha = 0.25$ and $\alpha = 0.45$, again rejecting optimality. In columns 4 and 5, we condition more flexibly on population density and land prices by controlling for, respectively, quintile dummies and cubic splines of these variables. These estimates show that our results are not due to the linear assumption of our model. Finally, in column 6, we allow the coefficients on population density and land prices to be city specific. This robustness specification allows for some parameters in the optimal rule, such as the elasticities or the weight assigned to landowners' welfare, to vary across cities. Our estimates again show that population density and land prices do not explain the density gradient.

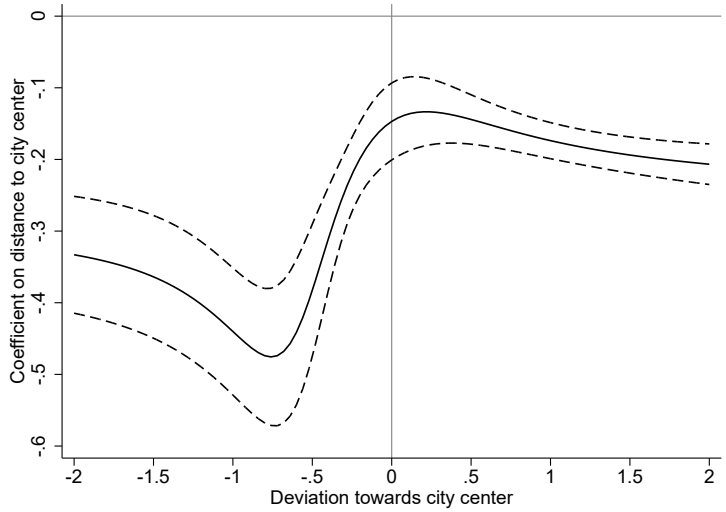
Weighting. Appendix Table A.4 presents additional robustness tests related to the weighting of observations. First, we use equal weights. By doing so, we give more weight to peripheral areas, even if sparsely

[8] For instance, $s = 1$ ($s = 2$) implies that the ratio of elasticities to marginal costs at a distance of 10 km is only 37% (14%) of the ratio at the city center, which would hold only under extreme differences in provision costs or in local valuations of infrastructure.

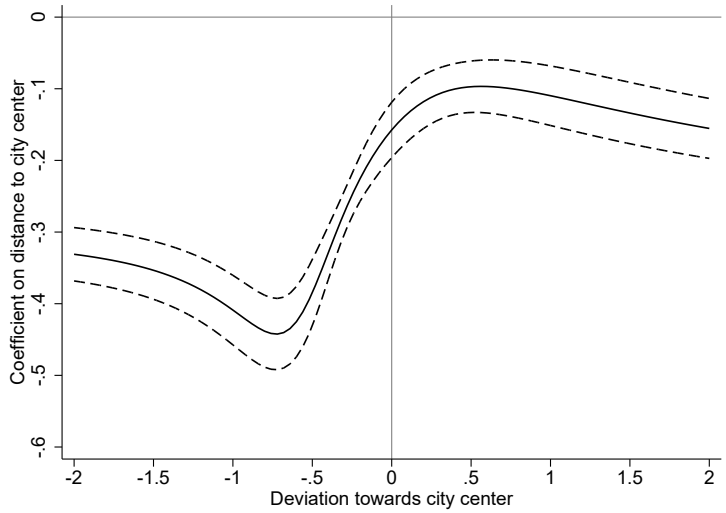
[9] Since there are missing soil data for some grid cells, we also include missing soil data indicators as controls to keep the sample size.

Figure 4: Sensitivity of the results to the constant elasticity and cost assumption

(a) Sanitation infrastructure



(b) Sreet infrastructure



Notes: The solid curves display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. Population density and land prices are multiplied by $e^{-s \times Dist_i}$, where $Dist_i$ is the distance to the city center and s indicates how the ratio of elasticities to marginal cost decreases linearly with distance to the city center. The values of s differ from -2 to 2 and are represented by the horizontal axis. The dashed lines display the 95% confidence interval, based on robust standard errors clustered at the city level. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census.

populated. Second, we weight by the number of CadÚnico observations used to estimate housing (and thus land) prices. This gives greater weight to grid cells with a bigger sample used to estimate housing prices,

directly addressing concerns that our conclusions are driven by measurement in our land price proxy. In both cases, the coefficients on distance to the city center remain similar.

Sample restriction by distance. Appendix Table A.5 shows that our findings are not due infrastructure differences between very central and very peripheral areas. We consider restricted samples in which we either remove grid cells too close to the city center, using cutoffs of 2 and 5 km, or too far away, using cutoffs of 10 or 20 km. In all cases, we keep finding negative and significant coefficients. Note that this test indicates that our rejection does not occur due to other motives leading to the infrastructure provision of city centers, such as the concentration of jobs and government facilities.

Other sample restrictions. Appendix Table A.6 shows that our results are robust to alternative sample restrictions. We exclude São Paulo and Rio de Janeiro, the two Brazilian cities with more than 10 million people, and we also restrict the sample to cities with fewer than 1 million inhabitants, showing that our findings are not explained solely by large cities. We also verify that they are not driven by smaller cities by excluding all cities with fewer than 100,000 inhabitants. Finally, to reduce the influence of extreme values in population density and land prices, we drop grid cells in the bottom and top 1 percentiles of these variables.

Alternative distances. Appendix Table A.7 shows that our findings hold when using alternative measures of distance to the city center. In column 1, we use the logarithm of distance, instead of levels. In columns 2 and 3, we redefine the city center as, respectively, the population and employment centroids of the city. In column 4, we use the driving distance to the city center, as opposed to the geodesic distance. In column 5, we use instead the driving time, in hours.

Land prices from tax assessment values. In Appendix Table A.8, we use land prices from the property tax assessment from São Paulo, the largest municipality in our sample, as an alternative measure of land prices. The table shows that our results hold when restricting the sample to grid cells within São Paulo, and that the results are similar regardless of the land price proxy used. Although the sample is smaller here due to data availability, this test shows that our results are robust to different data sources for land prices.

Land prices from real estate sales. For the municipality of São Paulo, we also have data from the *Imposto sobre Transmissão de Bens Imóveis* (ITBI), a municipal tax on real estate transactions. The data contain the transaction values, floor areas, and land prices on all property transactions subject to the tax. We use the ITBI for 2010 to derive two alternative measures of land prices. The first uses the average floor price of all residential unit sales in the grid cell and transforms this average into a proxy of land prices by

assuming a land share of $\alpha = 0.35$. The second measures uses the average land prices of all vacant land sales in the grid cell. The results, reported in Appendix Table A.9, show that our results are again robust to these alternative land price measures.

Housing prices for early years. In our baseline results, we measure housing prices by pooling all CadUnico observations from 2013 to 2020. This improves the precision and spatial coverage of our proxy for land prices, but also relies on rents observed several years after 2010, when infrastructure is measured. Appendix Table A.10 shows that our findings are robust to using only CadUnico observations from 2013 or from 2014, the earliest years available in our data.

Alternative city defintion. For our main results, we define a city as an *arranjo populacional*, a city definition for Brazil which is based in commuting patterns. A city in our data may thus encompass more than one municipality. In Appendix Table A.11, we show that our rejection of the optimal rule holds when considering municipalities as cities. The specification includes municipality fixed effects. We find that distance to either municipality or *arranjo populacional* centers are strong predictors of infrastructure provision after including municipality fixed effects and controlling for density and land prices.

Distribution across cities. To investigate how common our findings are across cities, we estimate, for each of the 341 cities, a regression of infrastructure on distance to the city center, population density, and land prices. Appendix Figure A.1 shows the distribution of these city-specific coefficients. When weighting cities by population, 90.7% of cities show a negative coefficient on distance for sanitation, and 86.5% do so for street infrastructure. The bias toward central areas is therefore widespread across Brazilian cities.

5.4 Discussion of Auxiliary Assumptions

Our test of whether infrastructure is optimally provided, based on the rule derived in Section 4, relies on the auxiliary assumption that the cost of infrastructure provision and the elasticities of population density and land prices do not vary substantially across space. Since our rejection of optimality is due to the under-provision of infrastructure in peripheral neighborhoods relative to central ones, the main concern is that these marginal costs might increase or the elasticities might decrease with distance to the city center. Although the sensitivity tests suggests our rejection is robust to this possibility, elasticities and marginal costs are not observed in the data so we cannot directly assess this assumption. Below, we discuss it in detail.

We begin by discussing provision costs. In Table 4, we show that the conditional gradients are negative

for almost all infrastructure categories, with 12 out of the 13 measures exhibiting steeper-than-optimal gradients. While some types of infrastructure, especially network infrastructure such as sewers, may be costlier to expand into peripheral neighborhoods, this is unlikely to apply to all infrastructure types in our data. For instance, the marginal cost of paving a road or installing street signs should not vary much within a city. Moreover, if our findings were driven by systematic differences in marginal costs, the estimated coefficients would likely change once we control for geographic variables. Soil characteristics and terrain ruggedness, in particular, would plausibly account for substantial variation in these costs. However, the robustness test in column 1 of Table A.3 shows no attenuation of the coefficients. Overall, it seems unlikely that spatial variation in marginal costs can account for all the observed deviations from the optimal rule.[10]

We now consider the elasticities of population density, $\varepsilon_{d,g}$, and of land prices, $\varepsilon_{p,g}$, with respect to infrastructure provision. Following the logic used in the proof of Proposition 1, these elasticities may be decomposed as $\varepsilon_{d,g} = \left(\varepsilon_{s,q} - \varepsilon_{h,q}\right) \times \varepsilon_{q,g}$ and $\varepsilon_{p,g} = \varepsilon_{p,q} \times \varepsilon_{q,g}$. In words, the elasticity of infrastructure on population density depends on how housing prices are affected by infrastructure ($\varepsilon_{q,g}$), and how these housing prices in turn affect population density through the housing supply elasticity ($\varepsilon_{s,q}$) and the elasticity of housing consumption with respect to prices ($\varepsilon_{h,q}$). The elasticity of infrastructure on land prices similarly depends on the elasticity of housing prices with respect to infrastructure and on how housing prices are capitalized into land prices ($\varepsilon_{p,q}$).

We know of no heterogeneity results in the literature on how housing prices respond to improvements in local infrastructure. For this channel to explain our findings, the elasticity $\varepsilon_{q,g}$ would need to be systematically lower in peripheral neighborhoods, that is, housing prices in those areas would have to respond less to infrastructure improvements. We see no clear reason why this would be the case, especially given that infrastructure deficits are larger in these areas, and marginal willingness to pay for improvements may possibly be even higher.

The other elasticities are better known to urban economists. Consider first the housing supply elasticity, $\varepsilon_{s,q}$. Baum-Snow and Han (2024) provide spatially detailed estimates of this elasticity for cities in the United States and find that it is higher in more peripheral areas. It is plausible that a similar pattern holds

[10]Note that marginal costs differences would not be a concern for our empirical test if they are explained by population density or land prices. For instance, density and land prices could impact the cost of providing infrastructure through construction or land acquisition costs. But since such cost differences are already accounted for by conditioning on population density or land prices, even some possible forms of cost heterogeneity do not invalidate our normative conclusions.

for Brazilian cities, as the same underlying forces shaping housing supply elasticities, including more land availability and less binding regulatory constraints in the urban fringe, may also be at play. Importantly, if $\varepsilon_{s,q}$ were increasing with distance to the city center, then the optimal rule would call for even more infrastructure in peripheral neighborhoods, since housing supply there would respond more to infrastructure improvements. This would only reinforce our rejection of optimality, as the patterns we document go in the opposite direction.

The Marshallian demand elasticity $\varepsilon_{h,q}$ follows from household preferences. Under Cobb-Douglas preferences, it would be constant across space and equal to -1. Other preferences, like CES, will allow this elasticity to vary with housing prices. In this case, since our land price measure is constructed based on housing prices, our test already conditions on equilibrium variables that would predict this elasticity.

Finally, consider the elasticity of land prices with respect to housing prices, $\varepsilon_{p,q}$. Under a Cobb-Douglas housing production function, this elasticity is equal to the inverse of the land share, and therefore constant across space. In practice, however, $\varepsilon_{p,q}$ may vary due to physical or regulatory constraints. In particular, the same forces constraining housing supply in central areas would also mute the response of land prices to housing demand. As with the housing supply elasticity, this would imply a lower $\varepsilon_{p,q}$ near the city center, and hence a lower $\varepsilon_{p,g}$. This again goes in the opposite direction of our empirical findings.

In sum, while we cannot rule out the possibility that the unobserved components of the optimal rule (elasticities and marginal costs) are themselves biased toward the city center, this appears unlikely to overturn our rejection of optimality, given our results, sensitivity test, the institutional context, and existing evidence on elasticities. Moreover, since population density and land prices vary substantially across space and are key determinants of the optimal rule, they should plausibly explain much of the infrastructure gradients observed in Figure 2 if infrastructure were optimally allocated. The fact that conditioning on these variables leaves more than half of the gradient unexplained (Table 3) is compelling evidence that infrastructure provision in Brazilian cities is biased toward central areas relative to the optimal rule.

5.5 Investigation of Possible Drivers

We now extend our empirical analysis to investigate possible explanations for why infrastructure remains biased toward central areas after conditioning on population density and land prices. To do so, we collect data on urban characteristics that, although not included in our model, may influence infrastructure provision.

A useful exercise, reported in Appendix Table A.12, is to examine how the residual gradient of infrastructure changes after including grid-cell variables that capture possible drivers of deviations from the optimal rule. In column 1, we control for a dummy if the grid cell is located within the main municipality of each city. The bias toward the city center falls a bit but remains significant, suggesting that administrative decentralization does not fully explain the underprovision of infrastructure in peripheral areas. In column 2, we control for the share of households living in *favelas*, which feature worse infrastructure (Brueckner, Mation, and Nadalin 2019). The residual gradient actually increases a little, indicating that the bias is not due to informal settlements or unclear property rights obstructing public investment.[11] In column 3, we use remotely sensed land use data from MapBiomas to control for the share of the urban area in each cell that was already urbanized by 1985. This test accounts for potential time-to-build frictions in infrastructure provision. Although the bias falls substantially, it remains significant, indicating that it may be partly related to historical infrastructure investments in older neighborhoods, but is not fully explained by them. In column 4, we include forward-looking controls: yearly population growth (2010–2022) and rental price growth (2013–2020). This specification accounts for the possibility that planners allocate infrastructure based on expected future population density and land prices.[12] The bias changes little, suggesting that such dynamic considerations do not explain the deviation from the optimal rule. In column 5, we control for employment density in the grid cell, allowing for the possibility that infrastructure also responds to the daytime population of the neighborhood. Once again, the coefficients on distance to the city center remain similar to the baseline ones. Finally, column 6 shows that our findings remain after including the average speed of a car trip to the city center as a control variable. This specification accounts for the possibility that underprovision of infrastructure is a policy response to traffic congestion externalities (Wijkander 1984). The coefficients decline somewhat in this specification, but remain sizeable and statistically significant

The empirical results above are exploratory, revealing correlates of the residual infrastructure gradient, and should thus be interpreted cautiously. Yet they suggest that the residual gradient may be partly, but not uniquely, related to historical urbanization, administrative decentralization, employment concentration, and traffic congestion considerations.

[11] In our sample, informal settlements are actually more frequent near the city center, explaining why their inclusion increases the coefficient on distance to the city center.

[12] Since anonymity restrictions imply that only rental prices for cell-years with at least 10 families reporting rent in the CadUnico are recorded in our data, the rental price growth rate is missing for part of the sample. To preserve sample size, we include a missing-data indicator.

6 Quantifying Equilibrium Consequences

6.1 Quantitative Model

We turn to a quantitative exploration of the equilibrium implications of the deviations from the optimal provision rule documented above. The goal of this section is not to estimate a full structural model of infrastructure provision, but to use a tractable version of the economy in Section 4 to obtain transparent estimates of the equilibrium importance of these distortions. We impose the following structural assumptions:

1. Preferences over private goods are iso-elastic according to utility $u(h,x) = \frac{h^{\mu}x^{1-\mu}}{\mu^{\mu}(1-\mu)^{1-\mu}}$, where $\mu \in (0,1)$ refers to the share of housing in total expenditure. These preferences imply Marshallian demand $h^m(q,y) = \mu q^{-1} y$, Hicksian demand $h^h(q,v) = \mu q^{\mu-1} v$, and indirect utility $v(q,y) = q^{-\mu} y$.

2. The neighborhood infrastructure is measured by a composite good g that enters iso-elastically into utility: $a(g) = g^{\gamma}$. The parameter $\gamma \in (0,1)$ refers to the elasticity of utility with respect to g. The cost of providing one unit of infrastructure g is constant across locations and equal to one.

3. Disposable income declines exponentially with distance to the city center: $y(r) = y_0 e^{-\tau r}$. The parameter τ indicates the commuting costs per unit of distance as a percentage of the wages.

4. The taste shock b_i for living in the city follows a Pareto distribution with scale parameter 1 and shape parameter $\beta > 1$. This assumption implies that the supply of residents to the city is iso-elastic in ω:

$$log\ L = c_L^s + \beta log\, \omega, \tag{8}$$

   where $c_L^s = log\ L_0 - \beta log\, \omega_0$ is a constant. The parameter β is the migration elasticity with respect to real wages.

5. Housing is produced according to the Cobb-Douglas production function $s = \varphi(1-\alpha)^{\alpha-1} m^{1-\alpha}$, where m is capital (with price 1) and φ is an exogenous productivity shifter of housing production. The parameter α refers to the share of land in housing production. From profit maximiziation, housing supply satisfies $s = \varphi^{\frac{1}{\alpha}} q^{\frac{1-\alpha}{\alpha}}$ and, from the zero profit condition, land prices follow $p = \alpha \varphi^{\frac{1}{\alpha}} q^{\frac{1}{\alpha}}$.

6. The city extends infinitely on a two dimensional plane: $\overline{r} = \infty$ and, for every $r \geq 0$, $\ell(r) = 2\pi r$.

We assume that $\mu\alpha > \gamma$. Since the amenity value of optimally provided infrastructure creates a local agglomeration force, this assumption is needed for a realistic optimal city structure. Otherwise, the optimal equilibrium would concentrate all the population in the city center, which would feature an infinite population density. This restriction is akin to the "no black hole" conditions limiting the strength of agglomeratiom forces in new economic geography models (Fujita, Krugman, and Venables 1999).

Turning to the equilibrium, note that, by taking logarithms of the indifference condition $\omega = g^{\gamma} q^{-\mu} y_0 e^{-\tau r}$ and rearranging, we obtain an equation for log housing prices:

$$log\ q = \frac{1}{\mu} log\ y_0 - \frac{1}{\mu} log\ \omega + \frac{\gamma}{\mu} log\ g - \frac{\tau}{\mu} r \tag{9}$$

Housing prices are decreasing in distance and increasing in the logarithm of neighborhood infrastructure. Using the housing supply and the Marshallian demand for housing yields a similar equation for log population density,

$$log\ d = c_d + \frac{1-\alpha\mu}{\alpha\mu} log\ y_0 - \frac{1}{\alpha\mu} log\ \omega + \frac{\gamma}{\alpha\mu} log\ g - \frac{(1-\alpha\mu)\tau}{\alpha\mu} r, \tag{10}$$

where $c_d = \frac{1}{\alpha} log\varphi - log\ \mu$.

In our counterfactual, we model a social planner which aims to maximize resident welfare, without considering landowners in their objective function. Equation 10 implies that the elasticity of population density with respect to infrastructure is constant and equal to $\frac{\gamma}{\alpha\mu}$. As a result, using equation 5, the optimal allocation of infrastructure g^* follows $log\ g^* = -log\ \lambda + log\ \frac{\gamma}{\alpha\mu} + log\ d$, where λ is the Lagrangian multiplier associated with the budget constraint. However, as documented in in Section 5, actual infrastructure provision deviates systematically from the predictions in the optimal rule. To account for this possibility, we parameterize the actual distribution of infrastructure as

$$log\ g = c_g + log\ d - \kappa r, \tag{11}$$

where c_g is a city-wide constant.[13] The parameter κ measures the semi-elasticity of infrastructure provision with respect to distance to the city center r, after conditioning on the variables that would predict an optimal

[13] Note that c_g is endogenous, being determined by the budget constraint.

allocation, and so it is naturally related to our reduced-form test. When $\kappa > 0$, there is a sub-optimal provision bias towards the city center.

To investigate the equilibrium effects of the distortions in infrastructure provision, we begin by examining how these distortions impact the infrastructure and population density gradients. Combining equations 11 and 10 yields the following expressions for d and g:

$$\log d = \frac{\alpha\mu}{\alpha\mu-\gamma}c_d + \frac{\gamma}{\alpha\mu-\gamma}c_g + \frac{1-\alpha\mu}{\alpha\mu-\gamma}log\ y_0 - \frac{1}{\alpha\mu-\gamma}log\ \omega - \frac{\kappa\gamma+(1-\alpha\mu)\tau}{\alpha\mu-\gamma}r$$

$$log\ g = \frac{\alpha\mu}{\alpha\mu-\gamma}(c_d+c_g) + \frac{1-\alpha\mu}{\alpha\mu-\gamma}log\ y_0 - \frac{1}{\alpha\mu-\gamma}log\ \omega - \frac{\kappa\alpha\mu+(1-\alpha\mu)\tau}{\alpha\mu-\gamma}r$$

These equations feature closed-form expressions for the population density and infrastructure gradients, as expressed in the proposition below:

Proposition 3. *The logarithms of population density and infrastructure decline linearly with distance to the city center r, according to the respective gradients:*

$$\Delta_d = \frac{\gamma\kappa+(1-\alpha\mu)\tau}{\alpha\mu-\gamma}, \quad \Delta_g = \frac{\alpha\mu\kappa+(1-\alpha\mu)\tau}{\alpha\mu-\gamma}$$

These gradients are strictly increasing in the bias parameter κ.

We use these gradients to derive the effects of κ on the welfare of urban residents, ω, and on the city size, L. Since we are interested in the welfare losses due to the distortions, our counterfactual is a city with the same budget constraint for public goods (same C_0) but where infrastructure is optimally provided ($\kappa = 0$). In this economy, the infrastructure and density gradients are equal to $\Delta_g^* = \Delta_d^* = \frac{(1-\alpha\mu)\tau}{\alpha\mu-\gamma}$. Denote by ω^* and by L^*, respectively, the maximum urban welfare and the optimal city size. The proposition below contains closed-form expressions for the welfare and city size losses in comparison with the counterfactual city. Its proof, in Appendix A.5.1, is based on solving for the welfare of residents in the city center.

Proposition 4. *The losses in the welfare of urban residents, given a budget constraint for urban infrastructure provision and in comparison with a counterfactual city where infrastructure is optimally allocated, are:*

$$log\ \frac{\omega}{\omega^*} = -\frac{2}{1+\beta\alpha\mu}\left(\alpha\mu\ log\ \frac{\Delta_d}{\Delta_d^*} - \gamma\ log\ \frac{\Delta_g}{\Delta_g^*}\right)$$

The equilibrium city size is also lower than optimal in equilibrium, following:

$$log\ \frac{L}{L^*} = -\frac{2\beta}{1+\beta\alpha\mu}\left(\alpha\mu\ log\ \frac{\Delta_d}{\Delta_d^*} - \gamma\ log\ \frac{\Delta_g}{\Delta_g^*}\right)$$

The expressions in Proposition 4 are intuitive. When urban infrastructure is allocated with a bias toward the city center ($\kappa > 0$), the equilibrium gradient of population density Δ_d becomes steeper than optimal. This is costly for the welfare of residents living in the city center, which is now denser and features higher housing costs. These negative effects on housing affordability will be stronger when the expenditure share on housing μ is higher and when land is more important for housing production (higher α). Part of this negative effect is compensated by the greater provision of infrastructure at the center, captured by the steeper gradient Δ_g. This compensation will be larger the more the residents value the public good, that is, the higher is γ. However, because the steepness of Δ_g reflects a deviation from the optimal allocation rule, it does not fully offset the welfare losses caused by the excessive concentration of the population near the city center. Moreover, these welfare losses translate into inefficiently small city sizes, as more individuals choose to remain in the rural hinterland. The migration elasticity β governs how much of the deviation in gradients results in reductions in welfare and in population.

6.2 Calibration

We calibrate six model parameters: the share of housing in consumption μ, the share of land in housing production α, the per-distance commuting costs τ, the elasticity of amenity-values with respect to infrastructure γ, the elasticity of population with respect to utility β, and the excess semi-elasticity of infrastructure provision with respect to distance to the city center κ. Table 5 summarizes our calibration. We later discuss the robustness of our findings to the calibrated values.

We calibrate three of these parameters with external sources. We calculate the housing share from the 2017–2018 Brazilian Expenditure Survey (*Pesquisa de Orçamento Familiares*, POF), considering the average share of rent expenditures among urban residents. This results in $\mu = 0.19$. We assign the land share in housing production as $\alpha = 0.35$, using estimates from Combes, Duranton, and Gobillon (2021). For the migration elasticity, we use $\beta = 4.5$ following the estimates from Morten and Oliveira (2024), who study the migration responses to increased market access in Brazil.

Table 5: Calibrated parameters

Parameter	Value	Source
μ	0.19	POF 2017–2018 (rent share)
α	0.35	Combes, Duranton, and Gobillon (2021)
β	4.5	Morten and Oliveira (2024)
τ	0.012	From rent gradient, conditional on infrastructure (Table 2)
κ	0.090	Average excess semi-elasticity (Table 4)
γ	0.034	Calibrated to match $\Delta_d = -0.45$ (Table 2)

We calibrate the other three parameters to match the data. To find τ, we use equation 9, which expresses the gradient of housing prices conditional on infrastructure as $\frac{\tau}{\mu}$, and the equivalent empirical gradient of 0.064 from column 3 of Table 2. This results in $\tau = 0.012$. The remaining two parameters, κ and γ, are more challenging to calibrate as they refer to the composite urban infrastructure measure. We do not observe the composite good directly, only the 13 sanitation and street infrastructure variables separately, and we also do not have exogenous variation in infrastructure to directly estimate by how much residents value g. Instead, we direct our calibration to be consistent with the results from our optimality test in Section 5 and with the density gradients observed in Brazilian cities, which we document in Section 3. We believe this approach is parsimonious, transparent, and tightly connected with our empirical approach. First, we calibrate the size of the distortion as $\kappa = 0.09$, according to the average semi-elasticity across all 13 infrastructure measures from Table 4.[14] This number indicates the bias of infrastructure towards the city center identified from our optimality test, and it is the distortion whose consequences we analyze. Second, we choose the parameter γ in order to match the overall population density gradient (-0.450) we observe in column 4 of Table 2. Using Proposition 3, we choose γ so that $\Delta_d = \frac{k\gamma+(1-\alpha\mu)\tau}{\alpha\mu-\gamma} = 0.45$. That is, we choose the value of γ that, together with the other parameters and the observed infrastructure bias κ, rationalizes the spatial slope of population density. This results in $\gamma = 0.034$, so urban residents would have to be compensated in 0.34% of their income for every 10% decline in local infrastructure.[15]

[14] Since our infrastructure measures are all shares, neither is amenable to logarithm transformations, so we use the linear coefficients divided by the sample averages as these semi-elasticities. Note also that the mean is one of many ways of aggregating these thirteen measures; since we do not observe how each infrastructure type contributes to g, we use the mean as a transparent choice. In robustness tests, we consider other values for κ.

[15] Note that, from the condition that $\alpha\mu > \gamma$, the parameter γ is no greater than 0.067, so our calibration assigns a value that is half of the theoretically highest value for the parameter.

6.3 Quantitative Results

Table 6 presents the results from our quantitative exercise on the welfare of urban residents, ω, city sizes, L, the equivalent variation in the budget for total infrastructure (C_0) that would result in the same welfare losses, and how much steeper the population density gradient is because of the sub-optimal bias of infrastructure provision towards the city center.

Table 6: Quantitative results

Aggregate consequences on:		Equivalent variation	Add. steepness due to
urban welfare	city sizes	in infrastructure budget:	bias in infrastructure:
$log\ \frac{\omega}{\omega^*}$	$log\ \frac{L}{L^*}$	$-\frac{2}{\gamma}\left(\alpha\mu\ log\ \frac{\Delta_d}{\Delta_d^*} - \gamma\ log\ \frac{\Delta_g}{\Delta_g^*}\right)$	$\frac{\Delta_d - \Delta_d^*}{\Delta_d^*}$
-0.10%	-0.44%	- 3.71%	27.27%

In comparison with a counterfactual city where the provision of infrastructure is optimal, the welfare of urban residents is only 0.10% lower and the city population is 0.44% smaller. The closed form expressions in Proposition 4 rationalize why the welfare losses are modest. The formula evaluates resident welfare at the city center. These residents are harmed by the increases in housing costs due to the higher density, which is captured by the term that multiplies $log\ \frac{\Delta_d}{\Delta_d^*}$. These effects through density reduce welfare by 1.07% and city sizes by 4.82%. However, city center residents are also compensated by the increased infrastructure. This compensation is expressed by the term that multiplies $log\ \frac{\Delta_g}{\Delta_g^*}$. It almost fully compensate city center residents, by increasing welfare by 0.97% and city sizes by 4.38%.

These modest welfare effects cannot be attributed to a low valuation of infrastructure by city residents. The equilibrium conditions allow us to derive the elasticity of urban welfare ω with respect to the total budget for stock of infrastructure, C_0, as $\frac{\partial log\ \omega}{\partial log\ C_0} = (1 + \beta\alpha\mu)^{-1}\gamma$. According to our calibration, this elasticity equals 0.026, so a 10% increase in infrastructure expenditures would be equivalent to an increase of 0.26% in residents' incomes. This is roughly three times larger than the effects of optimally readjusting infrastructure given a fixed budget. In fact, the losses from misallocated infrastructure, conditional on the bias, are equivalent to a modest decrease of 3.71% in the total stock of infrastructure.[16]

In contrast with the modest welfare effects, the bias of neighborhood infrastructure towards the city center has important consequences for the spatial distribution of population within the city, which ends up

[16]It is important to highlight that our quantitative results on welfare are conservative, as we only consider deviations from optimality that are systematically related with distance to the city center.

substantially more concentrated near the city center than what is optimal. By comparing the actual density gradient, $\Delta_d = 0.450$, with the optimal gradient, $\Delta_d^* = 0.359$, we find that the excessive spatial inequality in infrastructure increases the population density gradient by 27%. This contribution is conceptually distinct from the decomposition in Table 2, where the unconditional density gradient is 59% larger than the gradient conditional on infrastructure. That regression compares observed outcomes to a naive counterfactual of constant infrastructure across space. However, this is not the relevant counterfactual, as the optimal provision should be proportional to density. From this perspective, the observed bias in infrastructure provision toward the city center accounts for nearly half of the total linear-regression contribution of infrastructure to the steepness of the density gradient. This is a substantial effect, indicating that spatial misallocation of public infrastructure has meaningful consequences for the internal structure of cities. Our quantification implies that population density is 21% larger at the city than in the optimal equilibrium.[17]

Appendix A.5.2 shows the robustness of our quantitative results to alternative calibrations. We also consider the inclusion of agglomeration elasticities, and show results for cities with a bounded geography ($\bar{r} < \infty$). Across all these robustness exercises, we find welfare effects of up to -0.032%, effects on city sizes of up to -1.92%, and increases in the density gradient ranging from 14.09% to 45.24%.

In sum, although the city-level aggregate consequences are modest, the consequences for the spatial distribution of population within the city are large. This finding illustrates a key insight from spatial equilibrium: differences in amenities can meaningfully reshape the geography of cities even if aggregate welfare losses are small. This reflects the fact that urban residents respond to unequal infrastructure by relocating within the city, which in turn helps to mitigate welfare losses, especially given that the distortion is budget balanced.

7 Conclusion

Urban infrastructure, such as trees, sidewalks and drainage, is a key element of the built environment, determining how pleasant it is to live in a given neighborhood. In this paper, we investigate its within-city distribution. Exploring spatially granular data for hundreds of Brazilian cities, we document that infrastructure is unequally distributed across space, with better provision near city centers. To interpret and normatively

[17] This number follows from equation $log\ \frac{d(0)}{d^*(0)} = -\frac{1}{\alpha\mu} log\ \frac{\omega}{\omega^*} + 2\frac{\gamma}{\alpha\mu} log\ \frac{\Delta_g}{\Delta_g^*}$.

assess these patterns, we include urban infrastructure into a monocentric city as a local public good that determines the amenity value of neighborhoods. We then derive a sufficient-statistics rule for its optimal spatial allocation. Based on this result, we empirically evaluate the optimality of neighborhood infrastructure by assessing whether population density and land prices explain the observed distribution of infrastructure. We find systematic deviations from this rule, as these variables explain only half of the infrastructure gradients. Calibrating a tractable version of the model, we quantify the equilibrium impacts of this bias of infrastructure towards the city center: aggregate welfare and city size losses are modest, but infrastructure misallocation implies a 27% steeper population density gradient.

Our paper has concrete policy implications. Since neighborhood infrastructure has an amenity value, it impacts housing and land prices and shapes the spatial distribution of the population within the city. Therefore, it is essential to assess public investments in local infrastructure through a spatial-equilibrium lens. This is what we do in our paper, providing a rule characterizing optimal infrastructure provision. Our results offer a simple diagnostic test of optimality based on population density, land prices and distance to the city center, and that policymakers can use to assess infrastructure provision.

References

Ahlfeldt, Gabriel (2011). "If Alonso was right: modeling accessibility and explaining the residential land gradient". *Journal of Regional Science* 51.2, pp. 318–338.

Ahlfeldt, Gabriel M, Nathaniel Baum-Snow, and Remi Jedwab (2023). "The skyscraper revolution: Global economic development and land savings".

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf (2015). "The economics of density: Evidence from the Berlin Wall". *Econometrica* 83.6, pp. 2127–2189.

Arnott, Richard (1979). "Optimal city size in a spatial economy". *Journal of Urban Economics* 6.1, pp. 65–89.

Basso, Leonardo J, Raúl Pezoa, and Hugo E Silva (2024). "Maximization of equilibrium utility vs minimization of resources in the urban equilibrium". *Economics Letters* 240, p. 111772.

Baum-Snow, Nathaniel (2007). "Suburbanization and transportation in the monocentric model". *Journal of Urban Economics* 62.3, pp. 405–423.

Baum-Snow, Nathaniel and Lu Han (2024). "The microgeography of housing supply". *Journal of Political Economy* 132.6, pp. 1897–1946.

Brinkman, Jeffrey C (2016). "Congestion, agglomeration, and the structure of cities". *Journal of Urban Economics* 94, pp. 13–31.

Brueckner, Jan K and David A Fansler (1983). "The economics of urban sprawl: Theory and evidence on the spatial sizes of cities". *The review of Economics and Statistics*, pp. 479–482.

Brueckner, Jan K, Lucas Mation, and Vanessa G 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, pp. 48–60.

Chetty, Raj (2009). "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods". *Annu. Rev. Econ.* 1.1, pp. 451–488.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2019). "The costs of agglomeration: House and land prices in French cities". *The Review of Economic Studies* 86.4, pp. 1556–1589.

— (2021). "The Production Function for Housing: Evidence from France". *Journal of Political Economy* 129.10, pp. 2766–2816. DOI: 10.1086/715076.

Coury, Michael, Toru Kitagawa, Allison Shertzer, and Matthew A Turner (2024). "The value of piped water and sewers: Evidence from 19th century chicago". *Review of Economics and Statistics*, pp. 1–47.

Deffebach, Peter, David Lagakos, Yuhei Miyauchi, and Eiji Yamada (2025). "The Spatial Distribution of Income in Cities: New Global Evidence and Theory".

Duranton, Gilles and Diego Puga (2015). "Urban land use". *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 467–560.

Fajgelbaum, Pablo D and Cecile Gaubert (2020). "Optimal spatial policies, geography, and sorting". *The Quarterly Journal of Economics* 135.2, pp. 959–1036.

Fajgelbaum, Pablo D and Cecile Gaubert (2025). *Optimal spatial policies*. Tech. rep. National Bureau of Economic Research.

Feler, Leo and J Vernon Henderson (2011). "Exclusionary policies in urban development: Underservicing migrant households in Brazilian cities". *Journal of urban economics* 69.3, pp. 253–272.

Fujita, Masahisa (1989). *Urban Economic Theory: Land Use and City Size*. Cambridge, UK: Cambridge University Press.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Gonzalez-Navarro, Marco and Climent Quintana-Domeque (2016). "Paving streets for the poor: Experimental analysis of infrastructure effects". *Review of Economics and Statistics* 98.2, pp. 254–267.

Guedes, Ricardo, Felipe S Iachan, and Marcelo Sant'Anna (2023). "Housing supply in the presence of informality". *Regional Science and Urban Economics* 99, p. 103875.

Han, Lu, Stephan Heblich, Christopher Timmins, and Yanos Zylberberg (2024). *Cool cities: The value of urban trees*. Tech. rep. National Bureau of Economic Research.

Harari, Mariaflavia (2024). *Residential Patterns and Local Public Goods in Urban Brazil*. Tech. rep. Mimeo, Wharton School.

Helpman, Elhanan, David Pines, and Eli Borukhov (1976). "The interaction between local government and urban residential location: Comment". *The American Economic Review* 66.5, pp. 961–967.

Henderson, J Vernon, Tanner Regan, and Anthony J Venables (2021). "Building the city: from slums to a modern metropolis". *The Review of Economic Studies* 88.3, pp. 1157–1192.

Kanemoto, Yoshitsugu (1980). *Theories of Urban Externalities*. Amsterdam: North-Holland.

Koide, Hideo (1988). "Local public goods and spillover effects". *Journal of Urban Economics* 24.2, pp. 177–193.

Liotta, Charlotte, Vincent Viguié, and Quentin Lepetit (2022). "Testing the monocentric standard urban model in a global sample of cities". *Regional Science and Urban Economics* 97, p. 103832.

McCulloch, Sean E, Matthew P Schaelling, Matthew Turner, and Toru Kitagawa (2025). *Sewers and Urbanization in the Developing World*. Tech. rep. National Bureau of Economic Research.

McIntosh, Craig, Tito Alegría, Gerardo Ordóñez, and René Zenteno (2018). "The neighborhood impacts of local infrastructure investment: Evidence from urban Mexico". *American Economic Journal: Applied Economics* 10.3, pp. 263–286.

McMillen, Daniel P (1996). "One hundred fifty years of land values in Chicago: A nonparametric approach". *Journal of Urban Economics* 40.1, pp. 100–124.

Mieszkowski, Peter and Barton Smith (1991). "Analyzing urban decentralization: The case of Houston". *Regional Science and Urban Economics* 21.2, pp. 183–199.

Morten, Melanie and Jaqueline Oliveira (2024). "The effects of roads on trade and migration: Evidence from a planned capital city". *American Economic Journal: Applied Economics* 16.2, pp. 389–421.

Nicoletti, Leonardo, Mikhail Sirenko, and Trivik Verma (2022). "Disadvantaged communities have lower access to urban infrastructure". *Environment and Planning B: Urban Analytics and City Science* 50.3, pp. 831–849. DOI: 10.1177/23998083221131044. URL: https://doi.org/10.1177/23998083221131044.

Okamoto, Ryosuke (2009). "The optimal provision of local public goods in a metropolitan area with flexible jurisdictional boundaries". *Journal of Regional Science* 49.2, pp. 349–359.

Pandey, Bhartendu, Christa Brelsford, and Karen C. Seto (2022). "Infrastructure inequality is a characteristic of urbanization". *Proceedings of the National Academy of Sciences of the United States of America* 119.15, e2119890119. DOI: 10.1073/pnas.2119890119. URL: https://doi.org/10.1073/pnas.2119890119.

Posada, Héctor M (2018). "Informal housing, spatial structure, and city compactness". *Journal of Regional Science* 58.4, pp. 822–836.

Redding, Stephen J (2023). “Quantitative urban models: from theory to data”. *Journal of Economic Perspectives* 37.2, pp. 75–98.

Wijkander, Hans (1984). “Provision of public goods in congested cities”. *Journal of Public Economics* 25.1-2, pp. 127–141.

Online Appendix

A.1 Appendix Tables

Table A.1: Summary statistics of grid sample

	Sample mean (1)	Standard deviation (2)	Obs. (3)
Panel A: infrastructure (share of households)			
Piped water	0.931	(0.160)	32604
Sewage connection	0.715	(0.335)	32604
Trash collection	0.985	(0.045)	32604
Sign with street name	0.673	(0.282)	31187
Public lighting	0.969	(0.064)	31187
Paved street	0.852	(0.229)	31187
Sidewalks	0.745	(0.291)	31187
Curbs	0.801	(0.254)	31187
Drainage system	0.478	(0.296)	31187
Accessibility ramps	0.053	(0.123)	31187
Trees	0.669	(0.284)	31187
Free from open sewers	0.898	(0.188)	31187
Free from uncollected trash	0.945	(0.087)	31187
Panel B: other variables			
Distance to city center (km)	11.297	(10.255)	32610
Population density (per ha), 2010	79.726	(64.425)	32610
Population density (per ha), 2022	77.807	(59.212)	32610
Quality-adjusted log rent	6.038	(0.294)	32610
Rent observations from CadUnico	769.225	(962.846)	32610
Number of households in 2010 Census	2912.428	(2510.039)	32610
Altitude (m)	419.472	(369.901)	32610
Terrain ruggedness index (m)	18.069	(12.479)	32610
Coastal dummy	0.059	(0.235)	32610
Share of households in favelas, 2010	0.087	(0.187)	32610
Average individual income (BRL), 2010	961.289	(763.546)	32604

Note: Each observation is a grid cell in our sample All statistics are weighted by the number of households residing in the grid cell according to the 2010 Census.

Table A.2: Results for each infrastructure measure (density only)

	(1)	(2)	(3)	(4)	(5)
Sanitation infrastructure:	water	sewage	trash		
Distance to city center	-0.023***	-0.092***	-0.006***		
	(0.003)	(0.015)	(0.001)		
Sample mean	0.931	0.715	0.985		
R-squared	0.408	0.624	0.220		
Street infrastructure (1):	sign	light	paved	sidewalk	curb
Distance to city center	-0.094***	-0.012***	-0.054***	-0.090***	-0.054***
	(0.010)	(0.003)	(0.005)	(0.011)	(0.006)
Sample mean	0.673	0.969	0.852	0.745	0.801
R-squared	0.547	0.130	0.392	0.422	0.415
Street infrastructure (2):	drainage	ramps	trees	open sew.	unc. trash
Distance to city center	-0.086***	-0.044***	-0.066***	-0.022***	-0.006
	(0.011)	(0.010)	(0.007)	(0.006)	(0.006)
Sample mean	0.478	0.053	0.669	0.898	0.945
R-squared	0.551	0.211	0.527	0.528	0.102

Notes: The table display estimates from a regression of the infrastructure measure indicated on the top on city fixed effects, distance to the city center, in units of 10 km, and population density in 2010. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample for the top panel consists of the 32,604 grid cells with sanitation measures. The sample for the bottom panels consists of the 31,187 grid cells for which there are data from the *entorno* module of the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.3: Specification tests

	Add geog. controls	Alt. land share (α): 0.25	0.45	Condition on quint.	Cubic splines	City-specif. coeffs.
	(1)	(2)	(3)	(4)	(5)	(6)
Sanitation infrastructure						
Distance to city center	-0.141***	-0.163***	-0.134***	-0.105***	-0.180***	-0.113***
	(0.027)	(0.026)	(0.028)	(0.033)	(0.028)	(0.020)
Observations	32602	32604	32604	32604	32604	32604
Street infrastructure						
Distance to city center	-0.156***	-0.182***	-0.143***	-0.124***	-0.154***	-0.130***
	(0.017)	(0.020)	(0.020)	(0.026)	(0.020)	(0.020)
Observations	31185	31187	31187	31187	31187	31187

Notes: The table display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where usually $\alpha = 0.35$. In column 1, we further control for altitude, ruggedness, whether the grid cell is coastal, and soil characteristics. In columns 2 and 3, our proxy for land prices assumes, respectively, $\alpha = 0.25$ and $\alpha = 0.45$. In column 4, we account flexibly for population density and land prices by controlling for the quintiles of these variables. In columb 5, we control instead for cubic splines of these variables. In column 6, we allow the coefficients on the population density and land price variables to be city specific. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.4: Robustness to alternative weights

	Equal weights		Obs. used for log rents	
	sanit.	street	sanit.	street
	(1)	(2)	(3)	(4)
Distance to city center	-0.209***	-0.105***	-0.156***	-0.121***
	(0.037)	(0.015)	(0.025)	(0.020)
Observations	32604	31187	32604	31187

Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 3) or of the street infrastructure index (columns 2 and 4) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. In columns 1 and 2, all grid cells are equally weighted. In columns 3 and 4, grid cells are weighted by the number of CadUnico observations used to estimate the quality-adjusted log rents for that grid cell. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.5: Exclusion of cells that are too central or too peripheral from the sample

	Excludes if less than:		Excludes if more than:	
	2 km	5 km	10 km	20 kmm
	(1)	(2)	(3)	(4)
Sanitation infrastructure				
Distance to city center	-0.133***	-0.102***	-0.446***	-0.308***
	(0.025)	(0.023)	(0.062)	(0.054)
Observations	29234	21453	17667	25303
Street infrastructure				
Distance to city center	-0.137***	-0.107***	-0.646***	-0.370***
	(0.018)	(0.017)	(0.060)	(0.029)
Observations	27822	20205	17140	24342

Notes: The table display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where usually $\alpha = 0.35$. We restrict the sample across columns. In columns 1 and 2, the sample consists only of grid cells at a distance to the city center greater than, respectively, 2 or 5 km. In columns 3 and 4, the sample consists only of grid cells at a distance to the city center lower than, respectively, 10 or 20 km. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.6: Robustness to sample restrictions

	Sample excludes cities if:			Excludes top &
	$>$ 10 million	$>$1 million	$<$ 100 ths	bottom percentiles
	(1)	(2)	(3)	(4)
Sanitation infrastructure				
Distance to city center	-0.202***	-0.212***	-0.141***	-0.102***
	(0.023)	(0.025)	(0.028)	(0.032)
Observations	28054	19702	22310	26044
Street infrastructure				
Distance to city center	-0.169***	-0.141***	-0.161***	-0.121***
	(0.022)	(0.022)	(0.022)	(0.026)
Observations	26772	18746	21417	25785

Notes: The table display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where usually $\alpha = 0.35$. We restrict the sample across columns. In columns 1 and 2, the sample consists only of grid cells in cities with less than 10 million or 1 million inhabitants, respectively, in the 2010 Census. In column 3, we consider instead grid cells in cities with at least 100 thousand inhabitants. In column 4, we remove from the sample all the grid cells whose population density or land prices are in the top or bottom percentile (weighting by the number of households) of the sample. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.7: Robustness: alternative measures of distance to city center

	Log distance (1)	Population centroid (2)	Employment centroid (3)	Driving distance (4)	Driving time (5)
Sanitation infrastructure					
Log distance	-0.174***				
	(0.019)				
Distance to population centroid		-0.169***			
		(0.027)			
Distance to employment centroid			-0.183***		
			(0.024)		
Driving distance				-0.126***	
				(0.014)	
Driving time					-7.584***
					(1.405)
Observations	32604	32604	32604	32604	32604
Street infrastructure					
Log distance	-0.214***				
	(0.018)				
Distance to population centroid		-0.156***			
		(0.018)			
Distance to employment centroid			-0.182***		
			(0.016)		
Driving distance				-0.130***	
				(0.010)	
Driving time					-8.701***
					(0.842)
Observations	31187	31187	31187	31187	31187

Notes: The table display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. The variables measuring distance to the city center vary across columns; see main text for their description. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.8: Robustness to the use of tax assessments of land prices in São Paulo

	Data from CadUnico:		Data from IPTU:	
	sanit.	street	sanit.	street
	(1)	(2)	(3)	(4)
Distance to city center	-0.093***	-0.335***	-0.127***	-0.238***
	(0.010)	(0.032)	(0.012)	(0.025)
Observations	832	827	832	827

Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 3) or of the street infrastructure index (columns 2 and 4) on distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices. In columns 1 and 2, the proxy of land pries is calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$, as in baseline results. In columns 3 and 4, we use instead the logarithm of assessed land values per square meter, from the 2010 property tax (IPTU) assessment for the municipality of São Paulo. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample consists of all grid cells that are entirely within the municipality of São Paulo and for which there is IPTU data. Robust standard errors in parentheses. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.9: Robustness to the use of transaction prices in São Paulo

	Residential units:		Vacant land:	
	sanit.	street	sanit.	street
	(1)	(2)	(3)	(4)
Distance to city center	-0.088***	-0.262***	-0.103***	-0.293***
	(0.008)	(0.023)	(0.010)	(0.028)
Observations	769	768	634	634

Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 3) or of the street infrastructure index (columns 2 and 4) on distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices. In columns 1 and 2, the proxy of land prices is calculated as the average floor prices of residential unit transactions in the grid cell, to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. In columns 3 and 4, we use instead the average land price of vacant land transactions. Both variables are constructed based on the transaction values reported for the payment of the real estate transaction tax (ITBI) in 2010. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample consists of grid cells that are entirely within the municipality of São Paulo and for which the ITBI data contains at least one residential unit transaction (columns 1 or 2) or vacant land transaction (columns 3 and 4). Robust standard errors in parentheses. Statistical significance: * 10%, ** 5%, *** 1%.

Table A.10: Robustness: use CadUnico housing prices from 2013 or 2014

	Measures for 2013:		Measures for 2014:	
	sanit.	street	sanit.	street
	(1)	(2)	(3)	(4)
Distance to city center	-0.122***	-0.162***	-0.133***	-0.167***
	(0.023)	(0.020)	(0.022)	(0.020)
Observations	16399	16281	20878	20583

Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 3) or of the street infrastructure index (columns 2 and 4) on distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. In columns 1 and 2, the quality-adjusted log rent measure is constructed based only on CadUnico observations recorded in 2013, following from the estimation of equation A.2. In columns 3 and 4, we consider instead the observations recorded in 2014. The number of observations is lower than in our baseline results because, due to anonymity constraints, quality-adjusted log rent measures is available only for grid cells with at least 10 CadUnico families reporting rent expenditures in that year. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. Robust standard errors in parentheses. Statistical significance: * 10%, ** 5%, *** 1%.

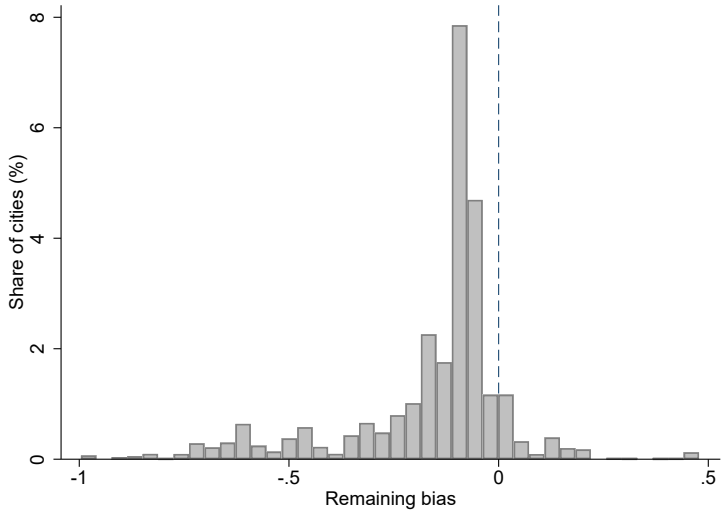
Table A.11: Robustness: within-municipality results

	sanitation infrastructure		street infrastructure	
	(1)	(2)	(3)	(4)
Distance to city center	-0.158**		-0.228***	
	(0.069)		(0.049)	
Distance to municipality center		-0.216***		-0.268***
		(0.059)		(0.051)
Observations	29207	29207	27966	27966

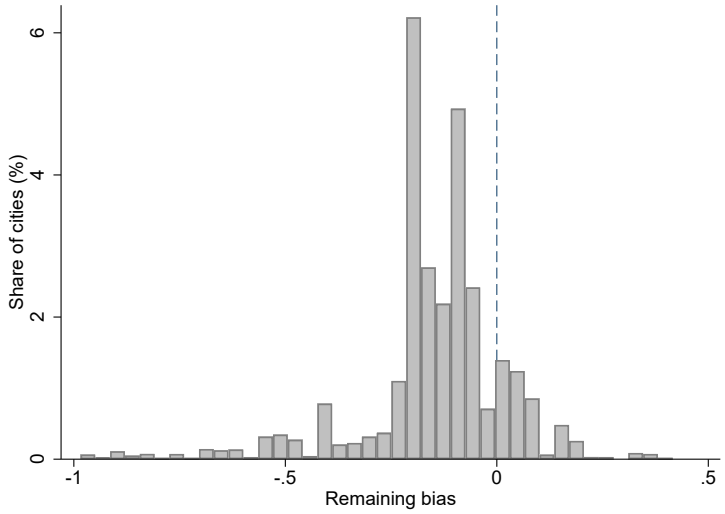
Notes: The table display estimates from a regression of the sanitation infrastructure index (columns 1 and 2) or of the street infrastructure index (columns 3 and 4) on municipality fixed effects, distance to the city center (columns 1 and 3) or to the municipality center (columns 2 and 4), in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The sample consists only of grid cells that are within of a single municipality. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

Figure A.1: Distribution of test results across cities

(a) Sanitation infrastructure



(b) Sreet infrastructure



Notes: The figures display histograms of city-specific estimates of a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. We display the point estimates of the coefficients on distance to the city center. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. In the histogram, each city is weighted by its population in 2010. To make the histograms readable, we exclude a few cities with coefficients below -1 or above 0.5. We use 40 equally spaced bins. The vertical dashed line indicates zero.

Table A.12: Accounting for drivers of infrastructure provision

	Admin. decentr. (1)	Housing inform. (2)	Historical urban. (3)	Forward looking (4)	Emp. density (5)	Conges-tion (6)
Sanitation infrastructure						
Distance to city center	-0.110***	-0.156***	-0.096***	-0.136***	-0.135***	-0.110**
	(0.036)	(0.030)	(0.028)	(0.028)	(0.025)	(0.047)
Main municipality	0.163					
	(0.101)					
Share of *favelas*		-0.455***				
		(0.132)				
Share built by 1985			0.540***			
			(0.085)			
Yearly population growth				-4.139***		
				(0.443)		
Yearly rental price growth				0.358		
				(0.266)		
Employment density					0.001**	
					(0.000)	
Average speed to city center						
Observations	32604	32604	32604	32604	32604	32604
Street infrastructure						
Distance to city center	-0.134***	-0.175***	-0.097***	-0.149***	-0.109***	-0.125***
	(0.024)	(0.018)	(0.016)	(0.021)	(0.013)	(0.028)
Main municipality	0.105**					
	(0.050)					
Share of *favelas*		-0.875***				
		(0.058)				
Share built by 1985			0.648***			
			(0.048)			
Yearly population growth				-3.401***		
				(0.310)		
Yearly rental price growth				0.702**		
				(0.296)		
Employment density					0.002***	
					(0.000)	
Speed to city center						-0.004**
						(0.002)
Observations	31187	31187	31187	31187	31187	31187

Notes: The table display estimates from a regression of the sanitation infrastructure index (top panel) or of the street infrastructure index (bottom panel) on city fixed effects, distance to the city center, in units of 10 km, population density in 2010 and a proxy for land prices, calculated as the exponential of quality-adjusted log rent measures to the power of $\frac{1}{\alpha}$, where $\alpha = 0.35$. All observations are weighted by the number of households residing in the grid cell according to the 2010 Census. The equation also includes the additional controls reported in the table; see the main text for details. Robust standard errors in parentheses are clustered at the city level. Statistical significance: * 10%, ** 5%, *** 1%.

A.2 Data Appendix

A.2.1 Grid cells

Our units of observation are squared grid cells of size 0.01^{o}, roughly 1.1 km at the Equator. To prepare our sample, we first create a rectangular grid with these dimensions encompassing the Brazilian territory. We then use IBGE's shapefile of the Brazilian territory[18] to crop the grid so each cell is contained within Brazil. The resulting grid contains 7,118,659 cells. We calculate the area of each grid cell by reprojecting the grid onto the SAD 1969 Albers South America projection (ESRI:102033). Note that, as a result of differences in latitude and as some grid cells are cropped due to a location along the coast or international borders, the areas of the grid cells vary across the sample.

We build cross-walks between the grid cells and the census blocks (*setores censitários*) of the 2010 and 2022 Brazilian Censuses.[19] We do so by intersecting the census blocks and grid cells shapefiles and calculating the area of the intersecting polygons, according to the SAD 1969 Albers South America projection. We divide the intersecting polygon area by the full census block area to obtain the share of the census block that is within each grid cell that intersects it. Assuming that households and population are uniformly distributed through census blocks, we calculate the number of households and population in the intersecting polygons by multiplying this share with, respectively, the total number of household and population in the census block. We then construct estimates of total households and population in the grid cell by summing across these values among the intersecting polygons that form the cell. Finally, for each intersecting census block $\times$ grid cell pair, we define weights by dividing the intersecting polygon number of households by the number of households in the grid cell. These household (population) weights allows one to use weighted means to aggregate any census block-level variable at the grid cell level.

[18] Available at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html.

[19] The shapefiles of the 2010 and 2022 census blocs are available at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html?edicao=41826.

A.2.2 Census data

From the 2010 Census, we use data released at the census block level, which we aggregate into grid cells using the weights (discussed above). These data are publicly available at https://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/.Population and the number of households in each block is from the *Arquivo Básico* files. The sanitation infrastructure measures are from *Domicilio01* files. The street infrastructure measures are from the *Entorno02* files.

We also identify whether the census block is a *favela* in 2010 using the list of the blocks forming *aglomerados subnormais*. IBGE prepares this classification is prepared based on information about land titles and precarious infrastructure, and it is publicly available at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-favelas-e-comunidades-urbanas.html.

From the 2022 Census, we also use data released at the census block level and aggregate it into grid cells. In this case, we extract data on population and the number of households in each block. These data are publicly available at https://ftp.ibge.gov.br/Censos/Censo_Demografico_2022/Agregados_por_Setores_Censitarios/Agregados_por_Setor_csv/.

A.2.3 Housing price data

The main data source we use to construct housing price variables is the *Cadastro Único para Programas Sociais do Governo Federal* (CadÚnico), an administrative database compiled by the Brazilian Ministry of Social Development (*Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome*). CadÚnico is based on detailed questionnaires administered by municipal authorities across Brazil, primarily to identify and register low-income families eligible for various social assistance programs. This extensive dataset includes a wide array of demographic, income, housing, and expenditure information for millions of households. Due to its national coverage and standardized data collection procedures, CadÚnico offers a unique window into the living conditions and housing characteristics of economically vulnerable populations in urban and rural settings. Access to CadUnico is restricted and subject to a formal agreement with the ministry.

In addition to its rich sociodemographic content, the CadUnico database contains detailed address information for registered families, which allows for spatial analysis through georeferencing. To geolocate these addresses, we use the HERE Geocoding and Search API, processing over 14 million unique streets in the CadUnico databases from 2014 to 2020. The procedure involves filtering for active records with minimally complete address fields, georeferencing these addresses, and merging the coordinates to the CadUnico data.

We combine the CadUnico records databases for all these years and restrict the data so each observation refers to a family and update year interview. We consider only data based on interviews made between 2013 and 2020. There are over 104 million unique family-update year observations. Based on the geographical coordinates assigned to each observation, we identify in which grid cell the observation is located. We exclude some observations which are assigned to grid cells that do not belong to the municipalities each family was recorded in the CadUnico; this is plausibly due to georeferencing errors. We also exclude observations in grid cells that do not overlap with an urban census blocks in 2010 or 2022. The resulting database contains 86,825,601 household-level observations. Of these, 18.7% (16,249,745 households) report housing rent expenditures.

To obtain cross-sectional measures of housing prices, we pool observations from all years and estimate the following equation:

$$log(y_{igt}) = X_{igt}\theta_t + \lambda_t + \gamma_g^* + \epsilon_{igt}, \tag{A.1}$$

where y_{igt} is the rent expenditure of observation i in gridcell g and year t, X_{igt} is a vector of housing characteristics, λ_t are time fixed effects, and γ_g^* are grid cell fixed effects. The vector X_{igt} includes the number of rooms (as dummies for each value from 0 to 10), dummies for the housing materials (masonry, uncovered brick, wood, and other), and dummies indicating the floor materials (ceramic, concrete, dirt, and other). Our main measure of interest are the estimates for the grid cell fixed effects, $\hat{\gamma}_g^*$. These estimates measure the quality-adjusted log rents of a house in the grid cell g. We set 2020 as the reference year, four as the reference number of rooms, ceramic as the reference floor type, and masonry as the reference building material.

Besides estimating the grid cell fixed effects from equation A.1, we also restrict our sample to the interviews conducted in each of the sample years $t \in \{2013, ..., 2020\}$ and estimate the following equations:

$$log(y_{igt}) = X_{igt}\theta_t + \gamma_{gt} + \epsilon_{igt}. \tag{A.2}$$

By doing so, we obtain quality-adjusted housing rent measures $\hat{\gamma}_{gt}$ for each of the sample years. This allows us to examine not only the cross-sectional dimension of the rental price data, but also the evolution of rental prices over time, which we need for a secondary empirical result (see below).

As a restriction to guarantee the anonymity of CadUnico respondents required by the aggrement with the ministry, our final grid cell database contains only quality-adjusted rent measures based on rent information from at least 10 different families. Note that, by imposing this restriction, we sometimes have the measures based on the pooled sample, $\hat{\gamma}_g^*$, but not all of the yearly measures $\hat{\gamma}_{gt}$.

As a control in Table A.12, we use the yearly average rent growth in the grid cell. To calculate it, we fit a linear trend to the series of quality-adjusted log housing prices in each grid cell. We weight each yearly observation by the number of families reporting rent expenditures in that year. Note that, due to the anonymity restriction discussed above, this measure is only available for cells with 10 or more respondent families in at least two years.

A.2.4 City and city center definitions

To define cities, we use the definition of *arranjos populacionais*. This definition is provided by IBGE based on commuting flows between municipalities. The list of *arranjos populacionais* is available at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15782-arranjos-populacionais-e-concentracoes-urbanas-do-brasil.html.

We define the city center as the center of the largest municipality in the *arranjo populacional.* To identify the city center, we use the list of municipality seat locations compiled by IBGE in 2010 as part of the *Cadastro de Localidades Selecionadas*. These coordinates are available for download at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/27385-localidades.html. They refer to the centroid of the census block 0001 in the main district of each municipality, being usually located in the traditional city center of the municipality. We manually validate these locations for all cities with more than 1 million inhabitants. In a few cases, since the coordinates were at the edge of the city centers, we replace them by the coordinates of the main squares of the city center. We manually input city center coordinates for the following cities: São Paulo: Praça da Sé (-23.550, -46.633); Rio de Janeiro: Largo da Carioca (-22.907, -43.178); Belo Horizonte: Praça

da Liberdade (-19.932, -43.938); Recife: Praça do Marco Zero (-8.063, -34.871); Salvador: Praça Thomé de Souza (-12.975, -38.513); Belém: Forte do Presépio (-1.454, -48.504); Goiânia: Praça Cívica (-16.680, -49.255); São Luís: Praça Pedro II (-2.528, -44.305).

A.2.5 Other data

We also use the following auxiliary data sources in our analysis:

Relief: altitude data is from the Shuttle Radar Topography Mission (SRTM GL3) Global 90m dataset. We construct terrain ruggedness based on altitude. The data can be downloaded from OpenTopography at https://portal.opentopography.org/raster?opentopoID=OTSRTM.042013.4326.1.

Coastal cells: to identify coastal cells, we use a shapefile of the Brazilian coastline prepared by the National Water Agency (*Agência Nacional das Águas*, ANA), publicly available at https://metadados.snirh.gov.br/geonetwork/srv/api/records/0c698205-6b59-48dc-8b5e-a58a5dfcc989. We create a buffer of 0.01^{o} around the coastline, and identify a grid cell as coastal if it overlaps with this buffer. In this way, coastal cells are those with some areas that are roughly within 1.1 km of the coastline.

Soil: soil data is from the Harmonized World Soil Database (HWSD), v.2, publicly available at https://gaez.fao.org/pages/hwsd. This is a rich soil database with 1 km resolution. For each grid cell, we assign characteristics of the soil under its centroid. From this database, we use the soil category (unit symbol), including the phase, according to the WRB 2022; the soil bulk density (in grams per cubic centimeter); and the available water capacity (in milimeters per meter).

Urban areas: to measure the urban area of each grid cell in 1985 and 2020, we use MapBiomas land cover rasters (collection 8). Based on Landsat imagery, the MapBiomas project classifies the land cover of each 30 m $\times$ 30 m pixel into several natural and anthropic land covers, including whether the pixel is a built-up urban area. The ratio between the urban land areas in 1985 and 2010 proxies for how old is the built environment in the grid cell. The MapBiomas rasters are publicly available for download at https://brasil.mapbiomas.org/.

Employment locations: to construct employment density, we georeference establishment locations in the 2012 *Relação Anual de Informações Sociais* (RAIS), a matched employer-employee administrative

dataset covering the universe of formal employment in Brazil. The RAIS data is of restricted access, requiring approval from the Brazilian Ministry of Labor. We calculate employment density as total formal employment over grid cell area. We winsorize this variable to reduce the influence of outliers.

Driving time and distance: to measure the driving time and distance to the city center, we use queries from the HERE Geocoding and Search API. In these queries, we ask for the driving distances and times between the centroid of each cell and the city center for trips at 8 am and and 5 pm on a Wednesday. We then take the average of these two distances or times. The speed of a trip to the city center is also the average of the speeds in these two departure times.

Tax assessment data: as an alternative measure of land prices, we use property tax (*Imposto Predial e Territorial Urbano*, IPTU) assessment data for 2010 from the municipality of São Paulo. Based on property-level land price assessments and land area, we construct average land prices for each block of the city of São Paulo. We then construct a cross-walk between the grid cells and the blocks, assigning average land prices in each grid cell based on the areas of each block. We use these data only for grid cells that are fully within the municipality of São Paulo. The spreadsheet with IPTU assessments and the shapefile with the blocks of São Paulo are publicly available at https://geosampa.prefeitura.sp.gov.br.

Real estate transactions data: to construct alternative measures of land prices, we use the records of property transactions subject to the *Imposto sobre Transmissão de Bens Imóveis* (ITBI), a municipal tax on real estate transactions, for the municipality of São Paulo. We consider two restrictions: sales of residential units, and sales of vacant land (*terrenos*). The ITBI data contains the city block in which the property is; we use the city blocks to assign the property to grid cells. In case the block overlaps with more than one grid cell, we partially account for the property sale according to the share of the block within each grid cell. We calculate average land prices for vacant land sales for 2010, and the average floor prices for residential sales for 2010 (for use in the robustness tests) and for 2013–2020 (for validation of our housing price data). The ITBI records we use are publicly available at https://prefeitura.sp.gov.br/web/fazenda/w/acesso_a_informacao/31501.

A.3 Validation of Housing Price Data

It is important for our empirical investigation that the quality-adjusted rental prices derived from the CadUnico indeed capture spatial variation in housing prices. Hence, we validate our data by comparing it with other measures of housing prices that, although not having the same spatial coverage or granularity, tend to be good proxies of housing prices. Our validation results show that, although our quality-adjusted housing prices estimates are not based on a representative sample of the population, it strongly predicts other measures that are.

Census. Our first comparison is with rental prices in the 2010 Census. Similarly to Cadastro Único, the extended questionnaire of the census includes household information on housing rents and charactersitics. Unlike the CadUnico, the census sample is representative of the population, but due to its sampling and the lack of spatial coordinates, the microdata from the census is representative only at the municipality level. To validate the cross-sectional performance of our housing price data, we estimate quality-adjusted rental housing price measures at the municipality level from both the census and CadUnico. We do so by estimating equation A.1 with municipality fixed effects as opposed to grid cell fixed effects. We then regress the housing price estimates from the representative census data on each one of the 9 estimates from CadUnico (one for each year from 2013 to 2020 and one aggregated). Table A.13 and Figure A.2 show the results.

Table A.13: Comparison with rent prices in the 2010 Census - regression estimates

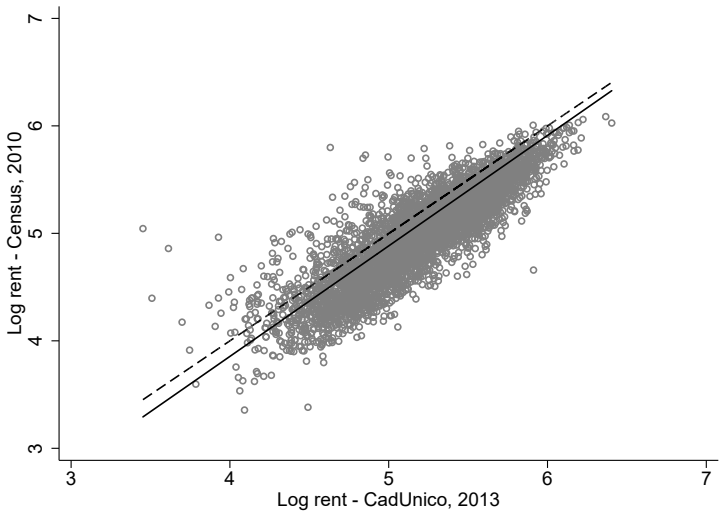
	2013 (1)	2014 (2)	2015 (3)	2016 (4)	2017 (5)	2018 (6)	2019 (7)	2020 (8)	all (9)
Index of log rents	1.028	0.987	0.959	0.941	0.971	0.981	0.988	1.005	1.015
	(0.047)	(0.041)	(0.041)	(0.041)	(0.041)	(0.037)	(0.037)	(0.037)	(0.037)
R-squared	0.798	0.804	0.783	0.770	0.757	0.759	0.765	0.776	0.803
Observations	4707	5083	5108	5156	5253	5259	5282	5073	5522

Notes: The table displays ordinary least squares estimates of a regression of quality-adjusted log rents estimated for municipalities using the 2010 Census data against quality-adjusted log rents estimated using the CadUnico data. The latter variable is constructed for the year indicated at the top of the column. In column 9, we use the measure based on all years. Note that the number of observations varies across columns, as some small municipalities may lack a measure for some years. Observations are weighted by the total urban population in the 2010 Census. Standard errors in parentheses are robust to heteroskedasticity.

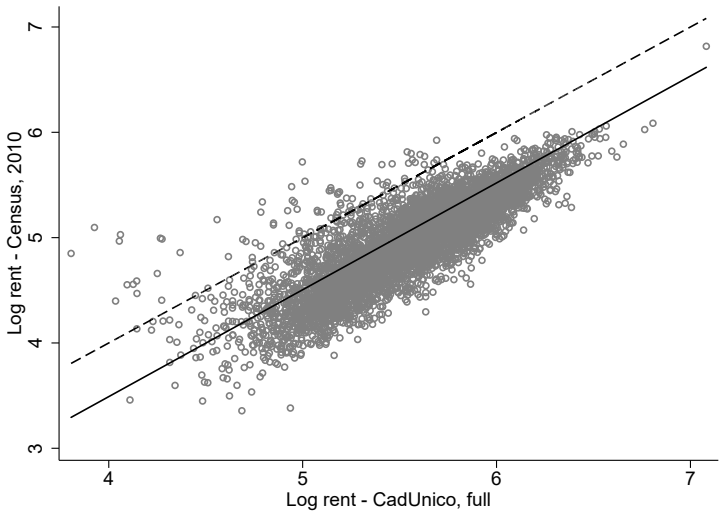
Those correlation between the census and CadUnico housing prices are high and close to one. Moreover, the R-squared is above 0.75 in all cases, indicating a high predictive power. When analyzing the plot of the

Figure A.2: Comparison with rent prices in the 2010 Census - scatterplots

(a) Against 2013



(b) Rent - all years combined



Note: The figure displays scatterplots of quality-adjusted log rents estimated for municipalities using the 2010 Census data (in the vertical axis) against quality-adjusted log rents estimated using the CadUnico data (in the horizontal axis). In Panel (a), the horizontal axis uses the measures estimated for 2013. In Panel (b), the horizontal axis uses the measures estimated using data for all years combined. The solid line indicates the linear regression fit, with each municipality weighted by its 2010 urban population. The dashed line indicates the 45-degree line.

census-based measure against those from the CadUnico, we see the estimates clustering around a fitted line that is parallel to the 45-degree line and a bit to the right due to inflation since 2010.

Real estate transactions. To validate our data with respect to the within-city dimension, we use data from the *Imposto sobre Transmissão de Bens Imóveis* (ITBI), a municipal tax on real estate transactions. São Paulo, the most populous municipality in Brazil, publishes detailed ITBI information, recording details and the city block of all taxed transactions. This allows us to compute the average floor prices of all residential units sold in each grid cell within São Paulo. For comparability with the CadUnico data, we use the ITBI records for 2013–2020.

Table A.14: Comparison with sales prices in São Paulo's ITBI records

	2013	2014	2015	2016	2017	2018	2019	2020
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Index of log rents	0.664	0.810	0.649	0.761	1.001	1.127	1.125	0.853
	(0.143)	(0.104)	(0.099)	(0.148)	(0.117)	(0.116)	(0.103)	(0.116)
R-squared	0.132	0.183	0.107	0.135	0.193	0.230	0.262	0.159
Observations	648	704	674	696	701	724	718	702

Notes: The table displays ordinary least squares estimates of a regression of log sales prices per floor area, measured from the São Paulo ITBI data, against quality-adjusted log rents estimated using the CadUnico data. Both variables are constructed for the year indicated at the top of the column. Note that the number of observations varies across columns, as some grid cells may lack either ITBI or CadUnico records for some years. Observations are weighted by the total number of households in the 2010 Census. Standard errors in parentheses are robust to heteroskedasticity.

For each year, we regress the log average sale floor price in the grid cell on our quality-adjusted rent prices. The estimated coefficients, reported in Table A.14, show that the correlation is high in every year. For 6 of the 8 years (2014, 2016–2020), the coefficient is statistically indistinguishable from one. The R-squared for the within-city is lower than for the across-municipality validation, which is plausibly be due to sale prices being distinct than rent prices, sales being less frequent than rental prices and potentially more selected, and the fact that sale prices are not adjusted for quality, as the ITBI data lacks house characteristics.

A.4 Theory Details and Extensions

A.4.1 Proof of Proposition 1

Consider first the comparative statics with respect to r. From the indifference condition 1, take the total derivative with respect to r, holding g fixed. This yields $q_r = -\frac{v_y}{v_q} y'(r)$. By Roy's Identity, $-\frac{v_y}{v_q} = h^m(q,y) = h(r,g)$, implying $q_r = \frac{y'(r)}{h(r,g)} < 0$, since $y'(r) < 0$. Given this decline in prices, we use the Hicksian demand to obtain $h_r = h_q^h \cdot q_r > 0$, as $h_q^h < 0$. The derivatives of housing supply and land prices follow from chain's rule: $s_r(r,g) = s'(q) q_r(r,g) < 0$ and $p_r(r,g) = p'(q) q_r(r,g) < 0$. Finally, the derivative on population density satisfies $d_r = s_r - h_r$, yielding the expression $d_r(r,g) = \left[s'(q) - h_q^h\left(q, \frac{\omega}{a(g)}\right)\right] q_r(r,g) < 0$.

Consider now the comparative statics with respect to g. From the indifference condition 1, take derivatives of g, holding r fixed. This yields $q_g = -\frac{a'(g)}{a(g)} \cdot \frac{v}{v_q} > 0$, as the indirect utility is decreasing in prices. Given this decline in prices, we use the Hicksian demand to obtain $h_g = h_q^m \cdot q_g$, which is negative because $h_q^m < 0$ since housing is a normal good. The rest of the proof is identical to the paragraph above. Hence, $s_g(r,g) = s'(q) q_g(r,g) > 0$, $p_g(r,g) = p'(q) q_g(r,g) > 0$ and $d_g(r,g) = \left[s'(q) - h_q^m\left(q, y(r)\right)\right] q_g(r,g) > 0$.

A.4.2 Extension to QSMs

In this appendix, we show how our results apply to quantitative spatial models. We focus on spatial equilibrium in the residential market, and thus we shut down any general equilibrium labor productivity effects (for instance, due to agglomeration economies). In any other aspect, the structure of the model follows standard QSM functional form assumptions, including Cobb-Douglas preferences and locational preference shocks distributed according to a Fréchet distribution. This demonstrates that our results are not specific to monocentric cities.

The economy consists of N urban neighborhoods, indexed by r, and one rural location, denoted by $r = 0$. Let L_0 be the total population in the economy. The utility of living in the rural location is exogenous and given by w_0, and commuting from the rural location to the city is unfeasible. The utility of urban residents, in contrast, depends on their consumption of housing, of a private tradable good, of urban infrastructure, and of exogenous neighborhood amenities. For simplicity, we assume that neighborhoods are equal-sized, with

an area of one. The neighborhood density d_r is thus equal to its total population. Each individual draws a vector of independent preference shocks b_{rn} over pairs of residential location r and workplace location n, distributed according to a Fréchet distribution with shape parameter β. Let y_n be the wages of workers in n, a_r be a common amenity shifter for those living in r, and τ_{rn} be iceberg transport costs for workers commuting from r to n; we assume that all these parameters are exogenous. Each resident supplies one unit of labor. As in our main model, we denote housing prices by q_r, per capita housing consumption by h_r, and housing supply (floor space built per unit area) by s_r. The price of the tradable private good is equal to one. Preferences are isoelastic, and the share of housing is μ. These preferences imply that the Marshallian demand for housing is $h = \mu y q^{-1}$. Neighborhood infrastructure is a composite good g_r, which enters multiplicatively as g_r^{γ} in the utility as an amenity shifter. Given these preferences, the indirect utility of a worker living in neighborhood r and working in neighborhood n is a function of the wages, commuting costs, infrastructure, amenities, housing prices and Fréchet shocks according to:

$$v_{rn} = a_r g_r^{\gamma} q_r^{-\mu} \frac{y_n}{\tau_{rn}} b_{rn} \equiv \omega_{rn} b_{rn}$$

Note that ω_{rn} is the utility of living in r and commuting to n, net of the Fréchet shock. This term is common across all individuals.

Housing supply. Profit maximizing developers use land and capital to build housing according to a constant returns to scale, Cobb-Douglas technology. Let α denote the land share. Developers choose housing supply in neighborhood r in order to maximize profits

$$\max_{m} \quad (1-\alpha)^{\alpha-1} q_r \varphi_r m^{1-\alpha} - m - p_r,$$

where m denotes the quantity of materials per unit of land, φ_r is an exogenous housing productivity shifter for neighborhood r, and p_r denotes land prices. Real estate development is competitive, so land prices adjust until developers make zero profits in equilibrium. These conditions imply that housing supply is $s_r = \varphi_r^{1/\alpha} q_r^{\frac{1-\alpha}{\alpha}}$ and land prices are $p_r = \alpha \varphi_r^{\frac{1}{\alpha}} q_r^{\frac{1}{\alpha}}$.

Equilibrium commuting flows. Since b_{rn} is a Fréchet shock, standard properties of the Fréchet distribution allow us to characterize commuting flows, average utility, and average income. These results are standard in the QSM literature; see, for instance, Heblich, Redding, and Sturm (2020) for details. The total

flow of commutes between residence r and workplace n is given by

$$F_{rn} = \frac{\omega_{rn}^{\beta}}{\sum_{m=1}^{N} \omega_{rm}^{\beta}} d_r = \frac{y_n^{\beta} \tau_{rn}^{-\beta}}{\sum_{m=1}^{N} y_m^{\beta} \tau_{rm}^{-\beta}} d_r,$$

The Fréchet distribution properties also imply that the expected utility of residents of location r is:

$$\omega_r = \Gamma\left(\frac{\beta-1}{\beta}\right)\left(\sum_{n=1}^{N} \omega_{rn}^{\beta}\right)^{\frac{1}{\beta}} = a_r g_r^{\gamma} q_r^{-\mu} \overline{y}_r \tag{A.3}$$

where $\overline{y}_r = \Gamma\left(\frac{\beta-1}{\beta}\right)\left(\sum_{n=1}^{N} y_n^{\beta} \tau_{rn}^{-\beta}\right)^{1/\beta}$ is the average wage of residents from r and $\Gamma()$ is the gamma function.

Population distribution across neighborhoods. Let L be the total population of the city. Note that $L = \sum_{n=1}^{N} d_n$. Using again properties of the Fréchet distribution and combining with the expressions for ω_r, the share of the city population residing in neighborhood r is given by

$$\theta_r = \frac{d_r}{L} = \frac{\omega_r^{\beta}}{\sum_{t=1}^{N} \omega_t^{\beta}} = \frac{a_r^{\beta} g_r^{\beta\gamma} q_r^{-\beta\mu} \left(\sum_{n=1}^{N} y_n^{\beta} \tau_{rn}^{-\beta}\right)}{\sum_{t=1}^{N} a_t^{\beta} g_t^{\beta\gamma} q_t^{-\beta\mu} \left(\sum_{n=1}^{N} y_n^{\beta} \tau_{tn}^{-\beta}\right)}, \tag{A.4}$$

while the expected utility of city residents is given by:

$$\omega = \Gamma\left(\frac{\beta-1}{\beta}\right)\left(\sum_{r=1}^{N} \omega_r^{\beta}\right)^{\frac{1}{\beta}} = \Gamma\left(\frac{\beta-1}{\beta}\right)\left[\sum_{r=1}^{N} a_r^{\beta} g_r^{\beta\gamma} q_r^{-\beta\mu} \left(\sum_{n=1}^{N} y_n^{\beta} \tau_{rn}^{-\beta}\right)\right]^{\frac{1}{\beta}}$$

Supply of city residents. Considering now the choice between living in the city or in the rural location, the Fréchet distribution implies that the supply of city residents is upward-sloping in ω according to equation:

$$L = \frac{\omega^{\beta}}{\omega^{\beta} + \omega_0^{\beta}} L_0 \tag{A.5}$$

Moreover, the expected utility of residents in the economy, W, is also increasing in ω:

$$W = \Gamma\left(\frac{\beta-1}{\beta}\right)\left(\omega^{\beta} + \omega_0^{\beta}\right)^{\frac{1}{\beta}}$$

Note that, since the supply of city residents is upward sloping in ω and not affected by the distribution of

neighborhood infrastructure, the problem of a social planner who wants to maximize W consists of choosing $(g_r)_{r=1}^N$ in order to maximize the total demand for city residents, as in the monocentric city model case. We now derive this demand curve.

Demand for city residents. The total demand for city residents follows from the identity $L = \sum_{r=1}^N d_r$. There are two ways we can write density as a function of housing first. First, we write population density as housing supply per block divided by individual demand for residential floor space using the expressions for the housing supply s_r and the Marhsallian demand h_r. This yields the following expression for density at location r:

$$d_r = \frac{s_r}{h_r} = \mu^{-1} \varphi_r^{1/\alpha} q_r^{1/\alpha} \overline{y}_r^{-1}$$

We obtain a second expression for density as $d_r = \theta_r L$ using the neighborhood choice condition A.4. We use the definitions of ω and $\overline{y}_r$ to simplify the expression for θ_r. This results in:

$$d_r = a_r^{\beta} g_r^{\beta\gamma} q_r^{-\mu\beta} \overline{y}_r^{\beta} \omega^{-\beta} L$$

Comparing both expressions for density allows us to isolate housing prices q_r as

$$q_r = \mu^{\frac{\alpha}{1+\mu\beta\alpha}} \varphi_r^{-\frac{1}{1+\mu\beta\alpha}} a_r^{\frac{\beta\alpha}{1+\mu\beta\alpha}} g_r^{\frac{\alpha\beta\gamma}{1+\beta\alpha\mu}} \overline{y}_r^{\frac{\alpha+\beta\alpha}{1+\mu\beta\alpha}} \omega^{-\frac{\beta\alpha}{1+\mu\beta\alpha}} L^{\frac{\alpha}{1+\mu\beta\alpha}}.$$

Plugging the prices back into any of the expressions for density equation results in

$$d_r = \xi_r g_r^{\frac{\beta\gamma}{1+\beta\alpha\mu}} \omega^{-\frac{\beta}{1+\mu\beta\alpha}} L^{\frac{1}{1+\mu\beta\alpha}},$$

where $\xi_r = \mu^{-\frac{\beta\alpha\mu}{1+\beta\alpha\mu}} a_r^{\frac{\beta}{1+\mu\beta\alpha}} \overline{y}_r^{\frac{\beta(1-\mu\alpha)}{1+\mu\beta\alpha}}$ is a local density shifter consisting of exogenous parameters. It includes exogenous amenities, housing construction productivity, and commuter market access. Density d_r is also determined by the city level equilibrium variables: it is increasing in city size L and decreasing in the utility of city residents ω. It is also affected by the provision of neighborhood infrastructure g_r, with a constant elasticity $\frac{\beta\gamma}{1+\beta\alpha\mu}$.[20]

Finally, by summing the density across all locations $r = 1, ..., N$ and rearranging, we obtain a downward-

[20] This elasticity differs from the elasticity in the monocentric city model of Section 6 by the presence of the Fréchet parameter β. When $\beta \to \infty$, this elasticity converges to $\frac{\gamma}{\alpha\mu}$.

sloping demand curve for city residents:

$$L = \left(\sum_{r=1}^{N} \xi_r g_r^{\frac{\beta\gamma}{1+\beta\alpha\mu}} \right)^{\frac{1+\beta\alpha\mu}{\beta\mu\alpha}} \omega^{\frac{-\beta}{\beta\alpha\mu}} \tag{A.6}$$

Optimal provision rule. Equations A.6 and A.5 form a supply and demand system that determines the equilibrium values of L and ω. As in the monocentric city case, the provision of infrastructure only affects the demand curve, not the supply. Hence, in order to maximize ω (and thus also W), a social planner chooses a sequence $(g_r)_{r=1}^{N}$ to maximize the demand for city residents. More specifically, consider the maximization of $\sum \xi_r g_r^{\frac{\beta\gamma}{1+\beta\alpha\mu}}$, which is the only component of the demand depending on g_r. We thus write the following maximization problem:

$$\max_{g_r} \sum_{r=1}^{N} \xi_r g_r^{\frac{\beta\gamma}{1+\beta\alpha\mu}} \quad \text{s.t} \quad \sum_{r=1}^{N} c_r g_r = C_0,$$

where c_r is the unit cost of providing infrastructure in neighborhood r and C_0 is the total expenditure in public goods.

Let η be the Lagrange multiplier. The first order condition results in

$$\frac{\beta\gamma}{1+\beta\alpha\mu} \xi_r g_r^{\frac{\gamma\beta}{1+\mu\beta\alpha}-1} = c_r \eta.$$

Rearranging, we obtain the same optimal provision rule as in the monocentric city model (equation 5)

$$g_r = \frac{\varepsilon_{d,g}}{c_r \lambda} d_r,$$

where $\lambda = \eta \omega^{-\frac{\beta}{1+\mu\beta\alpha}} L^{\frac{1}{1+\mu\beta\alpha}}$ is a rescaling of η and $\varepsilon_{d,g} = \frac{\beta\gamma}{1+\mu\beta\alpha}$ is the elasticity of population density with respect to infrastructure. It is straightforward that an extension where the planner also accounts for land prices will again result in the same condition as equation 6.

A.5 Details and Extensions of the Quantitative Results

A.5.1 Proof of Proposition 4

Since residents are indifferent between living in any location within the city, our proof considers the welfare of households residing in the city center. Denote by $d(0)$ and $g(0)$, respectively, the population density and infrastructure provision at the city center. Since infrastructure is log-linear, the budget constraint is a function of $g(0)$ and the gradient of infrastructure provision Δ_g, $C_0 = \int_0^\infty g(0)e^{-\Delta_g r}2\pi r dr = \frac{2\pi g(0)}{\Delta_g^2}$. Hence, $log\ g(0) = log\ C_0 - log\ 2\pi + 2log\ \Delta_g$. Similarly, the integral over population density yields an expression for the log city size as a function of density at the city center and the density gradient, $log\ L = log\ d(0) + log\ 2\pi - 2log\ \Delta_d$.

To solve for equilibrium city size L and urban welfare ω, we proceed in four steps. First, we use the equation for log density 10, evaluated at $r = 0$, to obtain an expression for $log\ d(0)$ as a function of $log\ g(0)$. Second, we use the budget constraint to replace $log\ g(0)$ by a function of the total funds C_0 and the gradient Δ_g. These two steps imply that population density at the city center is

$$log\ d(0) = c_d + \frac{1-\alpha\mu}{\alpha\mu}\log\ y_0 - \frac{1}{\alpha\mu}log\ \omega + \frac{\gamma}{\alpha\mu}log\ \frac{C_0}{2\pi} + 2\frac{\gamma}{\alpha\mu}log\ \Delta_g,$$

This expression makes it clear that, the greater the gradient of infrastructure, the higher is population density at the city center, as residents move to it due to the concentrated neighborhood infrastructure.

In the third step, we use the linear relationship between $log\ d(0)$ and $log\ L$ to obtain a downward-sloping isoelastic demand for city residents. Combining with the supply for city residents derived from the distribution of urban taste shocks (equation 8), we have the following supply and demand system:

$$\begin{aligned} log\ L &= c_L^s + \beta log\ \omega \\ log\ L &= c_L^d + 2\big(\frac{\gamma}{\alpha\mu}log\ \Delta_g - log\ \Delta_d\big) - \frac{1}{\alpha\mu}log\ \omega, \end{aligned}$$

where $c_L^d = c_d + \frac{1-\alpha\mu}{\alpha\mu}log\ y_0 + \frac{\gamma}{\alpha\mu}log\ \frac{C_0}{2\pi} + log\ 2\pi$.

In the fourth step, we take log deviations with respect to the optimal equilibrium, which satisfies $\kappa = 0$. The supply of city residents becomes $log\ \frac{L}{L^*} = \beta log\ \frac{\omega}{\omega^*}$ and the demand for city residents becomes

$log\ \frac{L}{L^*} = 2\big(\frac{\gamma}{\alpha\mu} log\ \frac{\Delta_g}{\Delta_g^*} - log\ \frac{\Delta_d}{\Delta_d^*}\big) - \frac{1}{\alpha\mu} log\ \frac{\omega}{\omega^*}$. The expressions in the proposition are simply the solution to this system.

A.5.2 Robustness of Quantitative Results

Below, we consider additional quantitative exercises that examine the robustness of the results from Section 6. Table A.15 reports the result of these exercises.

Table A.15: Robustness of quantitative results

	Aggregate consequences on: urban welfare	city sizes	Equivalent variation in infrastructure budget:	Add. steepness due to bias in infrastructure:
Add aggl. economies	-0.12%	-0.53%	- 3.71%	27.27%
Lower $\beta = 1.65$	-0.12%	-0.19%	- 3.71%	27.27%
Closed city ($\beta \to 0$)	-0.13%	–	- 3.71%	27.27%
Open city ($\beta \to \infty$)	–	-1.92%	- 3.71%	27.27%
Lower $\alpha = 0.25$	-0.04%	-0.18%	- 2.71%	14.09%
Higher $\alpha = 0.45$	-0.17%	-0.77%	- 4.69%	41.00%
Lower $\kappa = 0.046$	-0.03%	-0.14%	- 1.07%	15.17%
Higher $\kappa = 0.131$	-0.18%	-0.80%	- 7.27%	36.89%
Lower $\gamma = 0.029$	-0.08%	-0.38%	- 3.79%	20.90%
Higher $\gamma = 0.039$	-0.17%	-0.76%	- 5.61%	45.24%
Bounded city:				
$\overline{r} = 50$ km	-0.29%	-1.30%	-10.91%	27.27%
$\overline{r} = 40$ km	-0.30%	-1.36%	-11.40%	27.27%
$\overline{r} = 30$ km	-0.31%	-1.40%	-11.79%	27.27%
$\overline{r} = 20$ km	-0.32%	-1.44%	-12.06%	27.27%
$\overline{r} = 10$ km	-0.32%	-1.46%	-12.22%	27.27%

Agglomeration economies. We add agglomeration economies, allowing the wages y_0 to depend on city population L according to $log\ y_0 = log\ \overline{y_0} + \xi \log\ L$, where $\overline{y_0}$ is an exogenous component and ξ is the elasticity of wages to city size. This simple model extension does not change the expressions for the density and infrastructure gradients Δ_d and Δ_g, but it slightly changes the relationship between the gradient

differences and welfare or city sizes to:

$$log\ \frac{\omega}{\omega^*} = -\frac{2}{1+\beta[\alpha\mu-\xi(1-\alpha\mu)]}\left(\alpha\mu\ log\ \frac{\Delta_d}{\Delta_d^*} - \gamma\ log\ \frac{\Delta_g}{\Delta_g^*}\right)$$
$$log\ \frac{L}{L^*} = -\frac{2\beta}{1+\beta[\alpha\mu-\xi(1-\alpha\mu)]}\left(\alpha\mu\ log\ \frac{\Delta_d}{\Delta_d^*} - \gamma\ log\ \frac{\Delta_g}{\Delta_g^*}\right)$$

We calibrate an agglomeration elasticity of $\xi = 0.052$, as estimated by Chauvin et al. (2017) for Brazil. The aggregate effects remain similar, of -0.12% on welfare and -0.53% on city sizes.

Migration elasticity. We also examine the robustness of our results with respect to the migration elasticty β. Our calibration of $\beta = 4.5$ uses an estimate in the higher end of the literature. In the robustness test, we use the lower value of $\beta = 1.65$ from Ahlfeldt, Baum-Snow, and Jedwab (2023). The welfare effects are only a little larger at -0.11%, while the effects on city sizes fall even further to -0.19%. We can also consider the two canonical cases of the monocentric city model: the closed city, in which population L is fixed, and the open city, in which welfare ω is fixed. These canonical cases correspond, respectively, to our economy when $\beta \to 0$ or $\beta \to \infty$. The welfare losses in the closed city case are 0.13%, and the city size losses in the open city economy are of 1.92%. Note that these values give upper bounds on the effects on welfare, which are always small, and on city sizes, which may be meaningful in the long-run as city populations adjusts.

Land share. Note that we calibrate the share of land in housing production externally $\alpha = 0.35$, based on the literature (Combes, Duranton, and Gobillon 2021). Since this number was not estimated for our setting, we assess the robustness of our results to alternative values of α. We consider $\alpha = 0.25$ and $\alpha = 0.45$. Note that, to keep mattching the density gradient $\Delta_d = 0.450$, these new values for α also change the calibrated values of γ, which are now, respectively, 0.018 and 0.051. The results, in Table A.15, show that our findings are sensitive to this parameter choice. The effects of the misallocation are increasing in α, as it increases the costs of density. Welfare impacts vary from 0.04% o 0.17%, while the additional steepness to the population density gradient varies from 14% to 41%.

Bias of infrastructure. We also reconsider our calculations using alternative values for the infrastructure bias. First, we consider a lower value, $\kappa = 0.046$, which is the median value of the semi-elasticities derived from Table 4. In our baseline quantitative exercise, we use the mean instead of the median. The choice of κ also affects the calibration of γ, which is now 0.037. The equilibrium consequences become smaller, impacting welfare in -0.03% and city sizes in -0.14%, but still adding 15.7% to the density gradient. Second,

we consider a higher value, $\kappa = 0.131$, which is the mean value of the semi-elasticities derived from Table A.2. In contrast to Table 4, the results in Table A.2 control only for population density, not for land prices. It is a stricter test of optimality, focusing on the restricted problem of maximizing only resident's welfare, not a convex combination with landowners' welfare. The choice of κ again affects the calibration of γ, which is now 0.032. The equilibrium consequences become higher, impacting welfare in -0.18% and city sizes in -0.80%, while 36.9% to the density gradient.

Amenity value of infrastructure. A key parameter in our model is γ, which determines how much urban infrastructure impacts local amenities. Note that, since we calibrate $\gamma = 0.034$ in order to match the density gradient given the other parameters, an assessment of the robustness of our results to γ requires changes in some other parameter so the gradient remains matched. In Table A.15, we vary γ to 0.029 and 0.039 and thus recalibrate τ to keep $\Delta_d = 0.450$. The welfare losses are increasing with γ, varying from 0.08% to 0.17%. The additional steepness of the density gradient due to the bias in infrastructure also increases with γ, from 21% when $\gamma = 0.29$ to 45% when $\gamma = 0.39$.

Bounded city. For our baseline results, we set $\overline{r} = \infty$, effectively assuming that the city extends over an unbounded landscape. This assumption simplifies our results, but in reality city areas are finite. By setting a finite $\overline{r}$, our results on the gradients, expressed in Proposition 1, remain unchanged. However, the expressions for the welfare and city sizes change slightly. This occurs because the solution for integrals of the form $\int_0^{\overline{r}} e^{-\Delta r} 2\pi r dr$ is $2\pi\Delta^{-2}\left[1 - e^{-\Delta\overline{r}}\left(1 + \Delta\overline{r}\right)\right]$. The term under brackets is new to the finite $\overline{r}$ case, and it is carried over to the solution, yielding:

$$log\ \frac{\omega}{\omega^*} = -\frac{2}{1+\beta\alpha\mu}\left\{\alpha\mu\left[log\ \frac{\Delta_d}{\Delta_d^*} - \frac{1}{2}log\ \frac{1-e^{-\Delta_d\overline{r}}\left(1+\Delta_d\overline{r}\right)}{1-e^{-\Delta_d^*\overline{r}}\left(1+\Delta_d^*\overline{r}\right)}\right] - \gamma\left[log\ \frac{\Delta_g}{\Delta_g^*} - \frac{1}{2}log\ \frac{1-e^{-\Delta_g\overline{r}}\left(1+\Delta_g\overline{r}\right)}{1-e^{-\Delta_g^*\overline{r}}\left(1+\Delta_g^*\overline{r}\right)}\right]\right\}$$

$$log\ \frac{L}{L^*} = -\frac{2\beta}{1+\beta\alpha\mu}\left\{\alpha\mu\left[log\ \frac{\Delta_d}{\Delta_d^*} - \frac{1}{2}log\ \frac{1-e^{-\Delta_d\overline{r}}\left(1+\Delta_d\overline{r}\right)}{1-e^{-\Delta_d^*\overline{r}}\left(1+\Delta_d^*\overline{r}\right)}\right] - \gamma\left[log\ \frac{\Delta_g}{\Delta_g^*} - \frac{1}{2}log\ \frac{1-e^{-\Delta_g\overline{r}}\left(1+\Delta_g\overline{r}\right)}{1-e^{-\Delta_g^*\overline{r}}\left(1+\Delta_g^*\overline{r}\right)}\right]\right\}$$

The results of these expressions depend on the city boundary $\overline{r}$. In Table A.15, we vary the boundary from 50 km to 10 km. The welfare effects are decreasing in the cutoff, and up to three times larger than if assuming an infinite landscape. The intuition for these larger effects is that, with a smaller urban area, spatial equilibrium is less effective in reducing the effects of the sub-optimal provision of infrastructure. These results show that, by assuming an infinite landscape, we are being conservative on the welfare and city size effects.

References

Ahlfeldt, Gabriel M, Nathaniel Baum-Snow, and Remi Jedwab (2023). "The skyscraper revolution: Global economic development and land savings".

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio (2017). "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States". *Journal of Urban Economics* 98, pp. 17–49.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2021). "The Production Function for Housing: Evidence from France". *Journal of Political Economy* 129.10, pp. 2766–2816. DOI: 10.1086/715076.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm (2020). "The making of the modern metropolis: evidence from London". *The Quarterly Journal of Economics* 135.4, pp. 2059–2133.