Educational Segregation in a Dual Labor Market: Evidence from Bogotá

Edicson Luna*

August 2025

Abstract

This paper studies how labor-market duality between formal and informal employment contributes to educational residential segregation in large cities in developing countries, using data from the Encuesta Multipropósito for Bogotá. I argue that the central mechanism is the unequal ability to afford neighborhoods, rather than differences in the spatial incentives to access jobs. College-educated workers earn substantially more than non-college workers and are more likely to work in the formal sector, where wages are higher. As a result, they are better able to afford high-rent, high-amenity neighborhoods. The descriptive evidence shows that wage access varies across space within each education group, but the relative geography of formal and informal wage access is broadly similar across education groups. In other words, the locations that offer better labor-market access tend to be similar for college and non-college workers; what differs sharply is the level of earnings and the probability of obtaining formal employment. I then develop a quantitative spatial model with workers, firms, formal and informal sectors, commuting costs, neighborhood amenities, and exogenous housing prices. Counterfactual exercises show that a 20% subsidy to the minimum housing-cost component faced by low-educated households reduces the dissimilarity index from 0.437 to 0.310, while improving formal access for low-educated workers substantially raises formal employment but has little effect on residential segregation. Overall, labor-market duality shapes segregation mainly through the interaction between sectoral wage inequality and housing affordability.

*Ph.D. student in Economics at the University of Pennsylvania. I am grateful to my advisors Petra Todd and Mariaflavia Harari for their guidance and support. I also thank Shresth Garg and Chris Severen for valuable advice, and Juan Camilo Castillo, Benjamin Lester, Margaux Luflade, Jin Liu, and Holger Sieg for helpful comments and suggestions.

1 Introduction

Cities concentrate jobs, amenities, and housing unevenly, so individuals face very different access to opportunities depending on where they live. At the same time, urban space is not randomly organized, since people with similar levels of income and education tend to cluster in the same neighborhoods (Reardon & Bischoff, 2011; Friedman & Tom, 2023). Educational segregation is particularly important in developing countries, where college attainment remains limited and economic inequality is high. In these settings, residential location shapes not only housing and commuting conditions, but also everyday exposure to networks, information, employment opportunities, and neighborhood-level services financed by residents, such as maintenance, cleaning, and private security. Segregation is therefore more than a spatial pattern: it can reinforce social segmentation by reducing interaction across educational groups and limiting the diffusion of local advantages associated with more educated neighborhoods.

On the labor-demand side, large cities have a distinctive internal economic structure. Economic activity and employment are not randomly distributed across space; they tend to concentrate in central or highly accessible locations, where agglomeration forces, thicker labor markets, and proximity to other producers raise productivity (Lucas & Rossi-Hansberg, 2002; Ahlfeldt et al., 2015). In developing economies, these forces are likely to be especially relevant for formal employment. Formal jobs are more often tied to larger and more productive firms, while informal employment includes a large mass of self-employment, microenterprises, and small firms with lower productivity, limited scale, and weaker access to the agglomeration benefits associated with dense employment centers.[1] Labor-market duality can therefore translate into a spatially segmented urban labor market: higher-wage formal jobs with social benefits are disproportionately concentrated in central employment areas, while lower-wage informal work is more spatially dispersed and more closely tied to local demand.

In this paper, I study how labor-market duality between formal and informal employment contributes to residential segregation between low- and high-educated workers. I use Bogotá as my empirical setting. The central mechanism is income-based access to urban space. Low-educated workers earn substantially less than high-educated workers, and their earnings are even lower when they are employed informally. As a result, when neighborhoods differ in housing prices and local amenities, these wage differences translate into unequal residential choice sets: some high-amenity or high-price neighborhoods become financially inaccessible

[1] Informality may also be reinforced by institutional frictions. When labor regulations, payroll taxes, or compliance costs are high relative to firm productivity, some firms and workers may remain informal rather than absorb the costs of formality; see, for example, Levy (2008), La Porta & Shleifer (2014), and Ulyssea (2018).

or impose such a large housing burden that they are unattractive for low-wage workers. Importantly, this mechanism does not require high- and low-educated workers to face entirely different spatial distributions of labor-market opportunities. In the data, the geography of formal labor-market access is broadly similar across education groups: some areas offer better access than others for both groups. What differs sharply is the level of earnings attached to those opportunities, especially through the formal–informal divide. Thus, even when workers face similar relative spatial incentives, unequal wages can generate very different residential choices because the same neighborhood prices represent a much larger burden for low-educated and informal workers.

I document five stylized facts for Bogotá that give empirical content to this story. First, formal employment is spatially concentrated, whereas informal employment displays weaker evidence of spatial concentration. Second, highly educated workers earn substantially more and are much more likely to work in the formal sector, where wages are higher. Third, the geography of formal labor-market access is broadly similar across education groups, while informal work appears more local. Fourth, housing rents vary sharply across neighborhoods, so the same rent gradient imposes tighter affordability constraints on low-educated and informal workers. Finally, residential patterns reveal strong segregation by income and education. Together, these facts suggest that labor-market duality affects residential segregation less through sharply different job geographies across education groups and more through the way sectoral employment, wages, and housing prices jointly determine which neighborhoods different workers can afford.

To study these mechanisms, I develop a quantitative spatial model with a dual labor market in which workers jointly choose where to live, which sector to work in, and where to work. Workers differ by predetermined education type, indexed by whether they have completed college. In making residential and employment choices, households trade off wages, housing prices, commuting costs, and idiosyncratic Fréchet preference shocks whose distribution reflects, among other factors, neighborhood amenities. Access to formal employment is imperfect, capturing the fact that formal jobs are harder to obtain because of labor-market frictions, and it may differ by education to allow the model to capture potential heterogeneity. Housing enters preferences non-homothetically through a minimum housing requirement, so low-income households face tighter constraints when choosing high-rent neighborhoods.

On the production side, firms differ by sector and location. Formal and informal firms operate in each neighborhood under constant returns to scale and demand both low- and high-educated labor. Formal firms are more intensive in high-educated labor, offer social benefits, and benefit from productivity spillovers associated with the local concentration of educated workers. I do not endogenize the housing market: neighborhood housing prices are observed in the data and treated as given. The model therefore focuses on how workers sort across neighborhoods given the spatial distribution of rents and commuting costs, and on how labor-market duality, endogenous wages and amenities, and unequal housing affordability jointly contribute to residential segregation by education.

I discipline the model using a combination of direct measurement, external parameter values, gravity-based estimation, and internal calibration. Objects with direct empirical counterparts, including housing expenditure shares, housing prices, commuting distances, land area, and education-group population shares, are taken directly from the data. Preference and substitution parameters that are not central to the paper are set using values from the literature or reasonable benchmark values, while non-wage labor costs in the formal sector are set using Colombian labor regulations. I estimate sector-specific location-choice elasticities using gravity equations. The remaining parameters are calibrated internally because they govern the mechanisms emphasized in the paper: education-specific access to formal employment, non-homothetic housing demand, residential amenities, sector- and location-specific productivity, formal-sector skill intensity, and productivity spillovers. I choose these parameters so that the model matches empirical moments on formal employment by education, wage differences across sectors and education groups, the concentration of formal and informal employment in the central-east part of the city, residential shares across neighborhoods, and educational segregation.

I use the calibrated model to study counterfactual exercises around the recovered baseline equilibrium. The results show that residential segregation is most responsive to changes that directly relax the housing affordability constraint faced by low-educated households. A 20 percent reduction in the effective minimum housing cost for low-educated workers reduces the dissimilarity index from 0.437 to 0.310 and lowers the share of infeasible residence-work alternatives from 0.059 to 0.030. By contrast, increasing formal-sector access for low-educated workers raises their formal employment rate but leaves residential segregation essentially unchanged. Targeted commuting improvements from low-college neighborhoods produce a smaller decline in segregation, while a citywide reduction in commuting disutility slightly increases segregation. These exercises suggest that labor-market duality contributes to segregation mainly because unequal sectoral wages interact with neighborhood housing prices, making some parts of the city effectively inaccessible for low-educated and informal workers.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes Bogotá's context and the data. Section 4 presents the descriptive evidence. Section 5 develops the quantitative spatial model. Section 6 presents the estimation and calibration of parameters. Section 7 shows the counterfactual exercises. The final section concludes.

2 Literature review

This paper contributes to several strands of the literature.

First, this paper contributes to the literature on quantitative spatial models applied to cities in developing countries. The paper is most closely related to Zárate (2022), who study the interaction between formal and informal labor markets in an urban setting. It is also related to the heterogeneity analyses by skill level in Tsivanidis (2022), the spatial labor-market framework in Khanna et al. (2023), and Warnes (2020), who studies the effects of transport infrastructure improvements on segregation. A common feature of these studies is that spatial change is often driven by large-scale transportation improvements, particularly Bus Rapid Transit systems. In contrast, I do not study a transportation shock. I study how the existing geography of formal employment, unequal access to formal jobs, and neighborhood housing prices interact to generate residential segregation by education.

Second, this study contributes to the literature on educational segregation and housing-market segregation. A classic benchmark for thinking about segregation by education is Becker (1973), while Benabou (1993) shows how education-based segregation and local externalities can generate persistent spatial inequality. More recent work documents strong spatial segregation by skill across and within cities, driven by differences in labor-market opportunities, amenities, and urban consumption value (Diamond, 2016; Couture & Handbury, 2020). I connect this literature to recent work emphasizing how housing demand and housing prices shape spatial segregation. Gyourko et al. (2013) show how high housing prices can limit access to high-productivity locations; Couture et al. (2024) study how income growth affects urban spatial segregation and its distributional consequences; and Finlay & Williams (2025) show how non-homothetic housing demand can translate income inequality into spatial segregation. My contribution is to connect this affordability channel to labor-market duality: formal employment raises wages and is disproportionately accessed by high-educated workers, while informal employment is lower-paying and more common among low-educated workers. As a result, the same neighborhood rent schedule generates different feasible residential choice sets across education groups.

Third, the paper contributes to the literature on formal and informal labor markets in developing economies. It is most closely related to Bobba et al. (2021), who model a labor market with formal and informal jobs characterized by different productivity levels. A key distinction of my study is that I analyze formal and informal employment through a spatial lens and focus on how sectoral duality affects residential choices. In this sense, the paper is also related to Bobba et al. (2022), although I abstract from modeling the education decision itself. More broadly, the paper connects to the literature analyzing the behavior of the informal sector (Albrecht et al., 2009; Meghir et al., 2015; Ulyssea, 2018; Narita, 2020).

Finally, the paper relates to the literature on employment concentration, sorting, and agglomeration economies. Prior work shows that dense urban labor markets can improve matching, raise productivity, and amplify sorting across workers and firms (Wheeler, 2001; Gaubert, 2018). Relatedly, Ahlfeldt et al. (2015) provides a quantitative framework for studying the productivity and welfare effects of employment density within a city. I build on the idea that the spatial concentration of employment matters for urban outcomes, but shift the focus to the concentration of formal employment within a dual labor market. In this setting, concentrated formal employment affects residential segregation not only through the location of jobs, but also through the wage differences that determine which neighborhoods workers can afford.

3 Bogotá's Context and Data

Bogotá is Colombia's capital and largest city, with approximately 8 million residents and more than 12 million people living in its metropolitan area. The metropolitan area accounts for roughly one-fifth of the national population and close to 30% of Colombia's GDP.

The city has a strongly uneven spatial structure. A large share of high-income households, highly educated residents, and high-amenity neighborhoods is concentrated toward the north and northeast of the city. The eastern corridor, especially areas around Chapinero, Santa Fe, and La Candelaria, also contains a dense concentration of formal economic activity, including financial services, corporate offices, regulated commercial activity, national government institutions, and public-sector employment. This spatial organization makes Bogotá a useful setting to study how labor-market inequality and housing affordability interact with residential segregation.

The empirical analysis relies primarily on the Encuesta Multipropósito (EM) 2021, a large household survey conducted by Colombia's national statistical agency (DANE) in collaboration with Bogotá's city government. The relevant spatial unit is the *Unidad de Planeamiento Zonal* (UPZ), an urban planning unit used by Bogotá to organize the city below the level of *localidades*. The survey is representative at the UPZ level and provides detailed information on individuals' labor-market outcomes, wages, education, income, household characteristics, residential location, and workplace location. A key advantage of these data is that they combine features of a household survey and a mobility survey: I observe where workers live, where they work, and how much they earn, a combination that is rarely available in standard urban mobility data.

In the data, I observe 95 UPZ-level units. This number is lower than the official number of UPZs because some units appear grouped together in the survey. UPZs provide a spatial scale fine enough to study residential and workplace patterns while still being consistently observed in the data. I combine the EM 2021 with official geographic boundaries for Bogotá to compute distances between residential and workplace locations and to construct the spatial measures used throughout the analysis.

Workplace location is observed for roughly 75% of employed workers and missing for the remaining 25%. Since the descriptive analysis requires both residential and workplace location, the descriptive evidence is restricted to employed workers with observed workplace UPZ. Appendix Table B1 compares employed workers with observed and missing workplace UPZ. Workers with missing workplace location are less likely to be college-educated and formally employed, more likely to be informal or self-employed, and report longer commute times. These differences should be kept in mind when interpreting the descriptive patterns, which characterize the employed sample with observed workplace location.

4 Descriptive Evidence

This section documents five stylized facts that motivate the model. The central mechanism is that labor-market duality affects residential segregation through wage inequality and the housing-affordability constraints it creates.

First, formal employment is spatially concentrated, whereas informal employment displays weaker evidence of spatial concentration. Second, highly educated workers earn substantially more and are much more likely to work in the formal sector, where wages are higher. Third, the geography of formal wage access is broadly similar across education groups, while informal work appears more local. Fourth, housing rents vary sharply across neighborhoods, so the same rent gradient imposes much tighter affordability constraints on low-educated and informal workers. Finally, residential patterns reveal strong segregation by income and education.

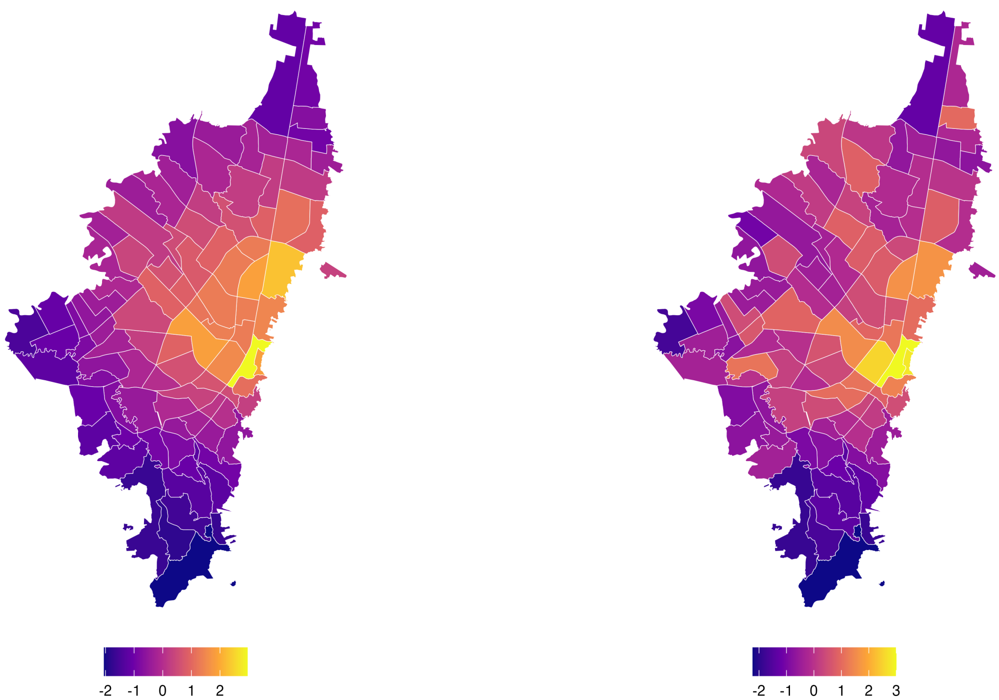
Formal employment is spatially concentrated. Formal employment is highly concentrated in the center-east of Bogotá and substantially less prevalent in peripheral areas. In contrast, informal employment displays a more dispersed spatial pattern, although informal employment is also relatively high in parts of the eastern side of the city.

To describe these patterns more systematically, I construct a labor-based measure of sectoral market access. For each residential location r and sector $s \in \{F, I\}$, the index is defined as

$$\Phi_{rs}^{L} = \sum_{j=1}^{R} L_{sj}^{\theta_s} d_{rj}^{-\theta_s},$$

where L_{sj} denotes employment in workplace location j and sector s, d_{rj} is the distance between residential location r and workplace location j, and θ_s is the sector-specific commuting elasticity. I use $\theta_F = 1.64$ and $\theta_I = 4.09$ as estimated in Section 6. The index therefore assigns greater weight to nearby employment opportunities and allows the strength of distance decay to differ across sectors.

Figure 1 shows the spatial distribution of the standardized log labor-based market access index by sector. The formal-sector index is clearly concentrated in the center-east of Bogotá, while peripheral areas display substantially lower values. The informal-sector index is more spatially dispersed, although it also presents relatively high values in parts of the eastern side of the city. This descriptive pattern suggests that formal employment opportunities are more geographically clustered than informal employment opportunities.



(a) Formal sector

(b) Informal sector

Figure 1: Standardized logarithm of labor index by sector across Bogotá

To complement the maps, I implement a simple pairwise-distance localization test. The test is inspired by the logic of Duranton & Overman (2005), but adapted to the structure of the data, where employment is observed as sector-specific shares across workplace UPZs rather than as exact establishment locations. For each sector $s \in \{F, I\}$, let p_{sj} denote the share of total citywide employment in sector s located in workplace UPZ j. For each distance cutoff b, I compute

$$K_s(b) = \sum_{j \neq k} p_{sj} p_{sk} \mathbf{1}\{d_{jk} < b\},$$

where d_{jk} is the distance between workplace UPZs j and k. The statistic $K_s(b)$ measures the share of pairwise employment mass located within distance b, excluding pairs located in the same UPZ.[2]

I then compare the observed value of $K_s(b)$ with a counterfactual distribution obtained by randomly permuting sector-s employment shares across UPZs and recomputing the statistic many times. This preserves the overall distribution of sectoral employment shares but removes their observed spatial assignment across the city. If the observed curve lies above the 95 percent simulation envelope at short distances, I interpret this as evidence that employment in sector s is more spatially concentrated across nearby UPZs than would be expected under a random allocation of sectoral employment shares.

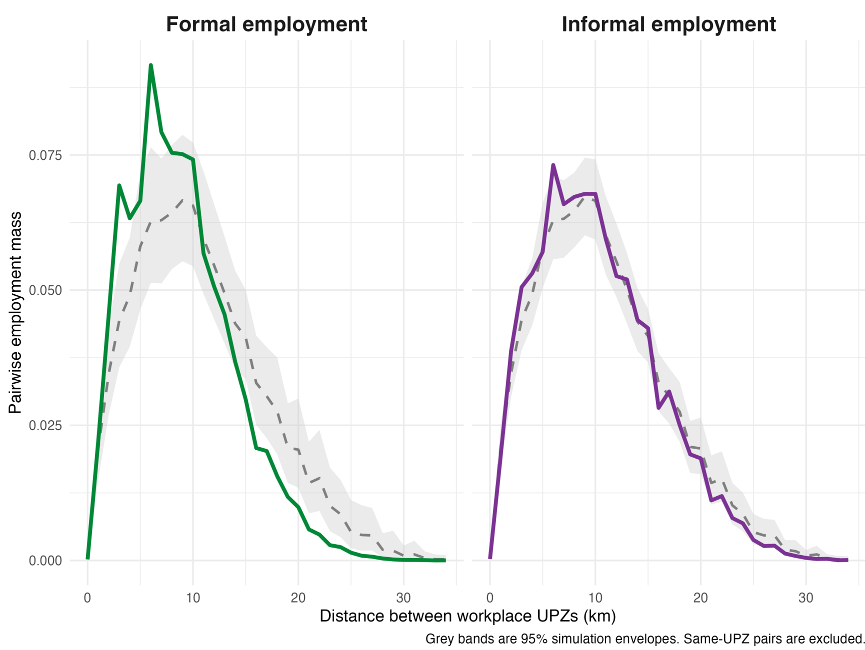


Figure 2: Localization Test

To provide complementary evidence using a wage-based commuter market access mea-

[2]This statistic captures concentration across neighborhoods, but not within neighborhoods. Since employment is observed at the workplace-UPZ level, the test cannot measure whether jobs are further clustered within specific parts of a UPZ, such as business corridors, commercial districts, or employment nodes. Moreover, because same-UPZ pairs are excluded, the statistic may understate the full extent of spatial concentration if employment is highly concentrated within particular UPZs.

sure. For each residential location r and sector $s \in \{F, I\}$, I define

$$\Phi_{rs} = \sum_{j=1}^{R} w_{sj}^{\theta_s} d_{rj}^{-\theta_s},$$

where w_{sj} is the average wage in workplace location j and sector s. The index is higher for residential locations that are close to high-wage workplace locations, with the strength of distance decay allowed to differ across sectors.

Using this measure, Appendix Figure A1 shows that formal wage access declines more steeply and more systematically with distance from *Las Nieves + Sagrado Corazón*, the part of the city with the highest index value for both sectors. In contrast, informal wage access displays a flatter and noisier relationship with distance. This reinforces the interpretation that formal labor-market opportunities are more strongly organized around the city's main employment center, while informal opportunities are more spatially dispersed.

Wages and formal access are substantially higher for highly educated workers. Wages differ sharply both across education groups and across employment sectors. Table 1 shows that college-educated workers earn substantially more than workers without college education. The average wage among college-educated workers is 3.86 million COP, compared with 1.28 million COP among non-college workers. Median wages display a similar gap. At the same time, formal employment pays substantially more than informal employment: the average formal wage is 2.80 million COP, compared with 1.32 million COP in the informal sector. These differences show that wage inequality is not only an education premium, but is also closely related to the sector in which workers are employed.

Table 1: Wages by Education and Employment Sector

Group	Mean wage (millions of COP)	Median wage
All workers	2.24	1.35
College	3.86	3.00
Non-college	1.28	1.07
Formal	2.80	1.80
Informal	1.32	0.91

These wage differences are reinforced by differences in access to formal employment. Table 2 reports the joint distribution of workers by education group and sector. College-educated workers represent 40.1 percent of employment, but they account for 32.3 percentage points of formal employment and only 7.8 percentage points of informal employment. In contrast, workers without college education account for 59.9 percent of employment and are almost evenly split between formal and informal jobs.

Table 2: Employment Shares by Education and Sector

	Informal	Formal	Total
No college	29.8%	30.0%	59.9%
College	7.8%	32.3%	40.1%
Total	37.6%	62.4%	100.0%

The implied formal employment rates differ substantially by education. Among college-educated workers, approximately 80.5 percent are employed in the formal sector, compared with 50.2 percent among workers without college education. Hence, highly educated workers have both higher wages conditional on employment and higher access to the sector that pays more. This composition amplifies wage differences across education groups: low-educated workers earn less on average, and they are also more exposed to informal employment, where wages are lower. Since formal employment is spatially concentrated, these differences make the geography of formal jobs especially relevant for understanding how labor-market duality translates into residential segregation.

Formal wage access is similar across education groups, while informal work is more local. The geography of formal wage opportunities is broadly similar for college and no-college workers. Residential locations with high formal wage access for college workers also tend to have high formal wage access for workers without college education. This suggests that educational segregation is unlikely to be driven primarily by the two education groups facing completely different maps of formal labor-market opportunities. The informal sector, however, displays a weaker common geography across education groups and appears more local, as reflected below in both wage-access patterns and commuting behavior.

To examine these patterns, I construct education-specific commuter market access (CMA) measures. For each residential UPZ r, education group e, and sector $s \in \{F, I\}$, I define wage-based CMA as

$$CMA_{rse} = \sum_{j} (w_{sje})^{\theta_s} d_{rj}^{-\theta_s},$$

where w_{sje} is the average wage of workers with education e in sector s and workplace UPZ j, d_{rj} is the distance between residential UPZ r and workplace UPZ j, and θ_s governs the distance elasticity for sector s. I then standardize the log of each CMA measure across residential UPZs.

Figure 3 compares standardized log formal wage access for college and no-college workers across residential UPZs. The points are tightly aligned around the 45-degree line, indicating that residential locations with high formal wage access for college workers also tend to have high formal wage access for no-college workers. Table 3 reports the corresponding correlations. The correlation in formal wage access across education groups is 0.896. This suggests that educational segregation is unlikely to be driven primarily by low- and high-educated workers facing completely different maps of formal labor-market opportunities.

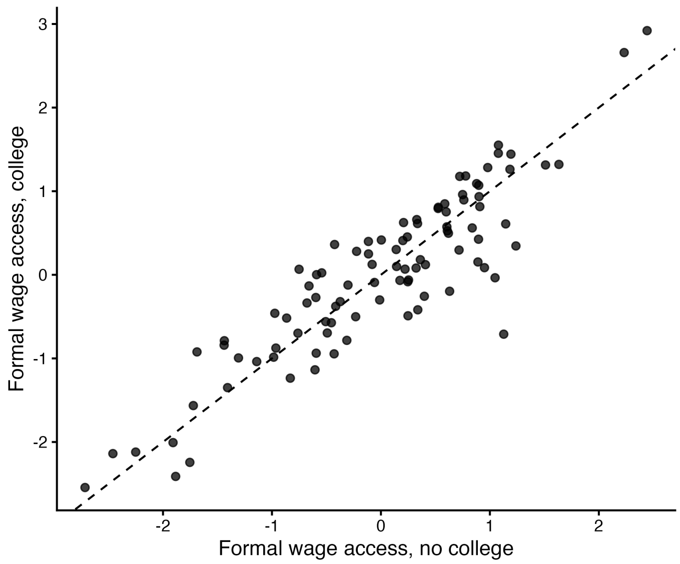


Figure 3: Formal wage access across education groups

Notes: The figure compares standardized log formal commuter market access (CMA) for college and no-college workers across residential UPZs. The dashed line is the 45-degree line. Each point is a residential UPZ.

Table 3: Similarity in Wage-Access Geography Across Education Groups

Statistic	Correlation
Formal wage access: college vs. no college	0.896
Informal wage access: college vs. no college	0.581

Notes: The table reports correlations across residential UPZs between standardized log CMA measures constructed separately by education group and sector.

The informal sector displays a weaker similarity across education groups. The correlation in informal wage access between college and no-college workers is 0.581, substantially below the corresponding correlation for formal wage access. This difference is consistent with informal employment being less organized around a common citywide employment structure and more closely tied to local labor-market conditions.

Commuting patterns provide additional evidence that informal employment is more local. Informal workers have shorter commutes than formal workers: the median commuting time is 30 minutes for informal workers and 40 minutes for formal workers, while the corresponding means are 35 and 48 minutes. The full commuting-time distributions, reported in Appendix Figure A2, show that informal workers are more likely to reach their jobs within shorter trips.

This distinction matters for the mechanism of the paper. Formal work provides access to a concentrated, citywide labor market, and the geography of formal wage access is broadly similar across education groups. The key difference across education groups therefore appears to be less spatial than economic. High- and low-educated workers do not seem to face sharply different spatial distributions of formal opportunities; rather, high-educated workers are more likely to obtain formal jobs and earn higher wages when they do. Low-educated workers, in contrast, are more exposed to informal employment and lower wages. Since informal employment is more local, their labor-market outcomes are more closely tied to nearby opportunities and local demand conditions.

The sectoral composition of informal employment reinforces this interpretation. Informality is especially prevalent in commerce, accommodation and food services, construction, household employment, arts, entertainment and recreation, and other services, as can be seen in Appendix Figure A3. Many of these activities are non-tradable and depend directly on nearby residential demand. This does not imply that all informal employment is local, but it suggests that an important component of informal work is concentrated in activities that serve nearby households. Complementary evidence in Appendix Figure A4 shows that informal CMA is more strongly associated with neighborhood income and the share of college-educated residents than formal CMA, consistent with informal wage opportunities being partly shaped by local demand in high-income and high-education areas.

Housing rents vary sharply across neighborhoods. Housing rents are considerably higher in some parts of Bogotá, especially in the north-eastern side of the city. This spatial variation is important because high rents can make some neighborhoods financially difficult to access for low-wage workers. In this sense, rents provide the link between the wage differences documented above and residential segregation: even if workers face similar broad maps of formal wage access, they may not be able to live in the same neighborhoods once housing costs are taken into account.

Appendix Figure A5 shows the spatial distribution of average rents across UPZs. Rents are highest in a small set of eastern and northern neighborhoods and substantially lower in much of the south and west of the city. This pattern suggests that the housing market creates a steep affordability gradient across space.

To quantify this affordability gradient, I compare average rents across UPZs with wages by education-sector group. I define a strict affordability measure: a UPZ is affordable for group g if the average rent in that UPZ is weakly below the wage of group g. This is not meant to be a conventional rent-burden threshold. Instead, it captures an extreme restriction: neighborhoods where average rent alone exceeds a worker's monthly wage are unlikely to be financially accessible to that group.

Table 4 shows large differences across education-sector groups. Using median wages, college-educated formal workers can afford 98.9 percent of UPZs, while no-college informal workers can afford only 55.8 percent. The gap remains large using mean wages: all UPZs are affordable for college formal workers, compared with 64.2 percent for no-college informal workers. These differences show how wage inequality and sectoral attachment translate into unequal feasible residential sets. High-rent neighborhoods are not equally accessible to all workers; they are much more restrictive for low-educated workers, especially those in informal employment.

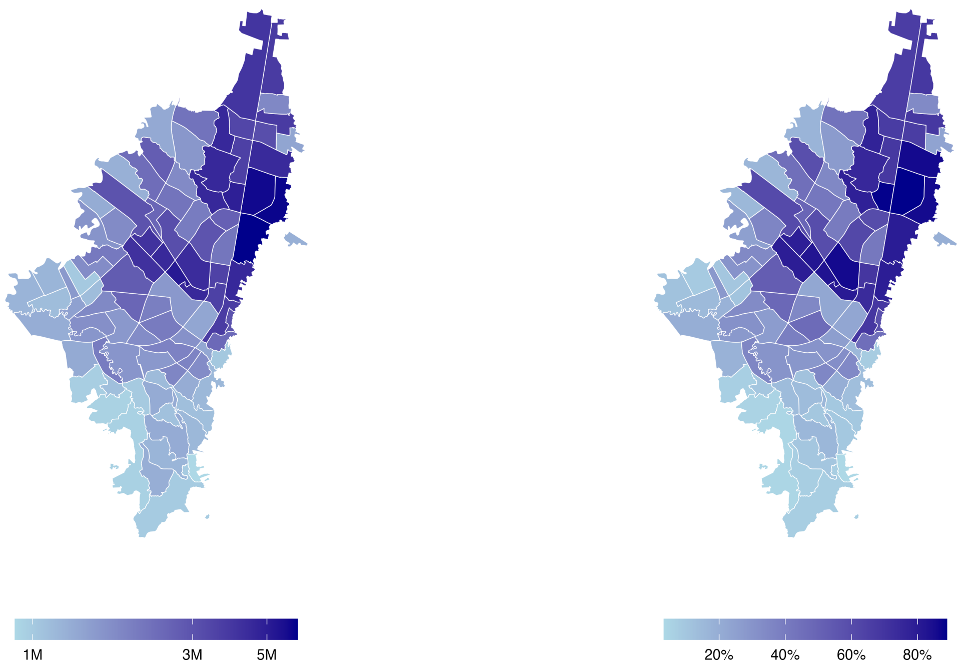
Table 4: Neighborhood Affordability by Education and Employment Sector

	Monthly wage		Share of UPZs affordable	
Group	Mean (millions of COP)	Median	Using mean wage (%)	Using median wage
College, formal	4.13	3.31	100.0	98.9
College, informal	2.73	1.90	97.9	90.5
No college, formal	1.56	1.25	83.2	74.7
No college, informal	1.00	0.90	64.2	55.8

Notes: The table reports the share of UPZs in which average rent is weakly below the wage of each education-sector group. Wages are monthly and expressed in millions of COP. The affordability measure is intentionally strict: a UPZ is classified as affordable for group g if its average rent is no greater than the mean or median wage of group g.

Residential segregation by income and education. Residential segregation by income and education is substantial across Bogotá. Figure 4 presents the spatial distribution of average income and the share of adult residents with a college degree across UPZs. Both variables are highly unevenly distributed across the city and are strongly positively correlated, with a correlation of 0.958. Higher-income and more educated residents are concentrated in the northeastern and central areas of the city, while many neighborhoods in the southern and southwestern peripheries display very low college shares and substantially lower average income.

These spatial patterns reveal a pronounced degree of residential segregation by socioeconomic status. In several peripheral neighborhoods, the share of college-educated residents is close to zero, whereas in some northeastern and central neighborhoods more than 80% of adults hold a college degree. Average income follows a similar geography. The close spatial alignment between income and education suggests that residential segregation in Bogotá is not only educational, but also reflects broader socioeconomic stratification.



(a) Income

(b) College

Figure 4: Average income and share with college by UPZ

To quantify this segregation, I use the dissimilarity, isolation, and exposure indices proposed by Massey & Denton (1987), adapted to the distribution of college and non-college residents across UPZs in Bogotá. The dissimilarity index captures unevenness in the spatial distribution of the two groups.[3] In this context, it measures the share of either college or non-college residents who would need to relocate across UPZs for both groups to have the same spatial distribution. The estimated dissimilarity index is 0.442. This implies that

[3] The dissimilarity index is defined as

$$D = \frac{1}{2} \sum_{r=1}^{R} \left| \frac{C_r}{C} - \frac{N_r}{N} \right|,$$

where C_r denotes the number of college-educated residents in UPZ r, N_r the number of non-college residents in UPZ r, and C and N denote the corresponding citywide totals.

44.2% of either college or non-college residents would need to relocate across UPZs for the two groups to be distributed identically across Bogotá.[4]

I then use the isolation and exposure indices of Massey & Denton (1987) to describe the neighborhood composition experienced by the average member of each group.[5] Table 5 reports the results. College-educated residents account for 40.14% of the citywide population, but the average college-educated resident lives in a UPZ where 55.75% of residents are also college-educated. Thus, the local college share experienced by the average college-educated resident is 15.61 percentage points above the citywide average. Conversely, the exposure of college-educated residents to non-college residents is 44.25%, well below the citywide non-college share of 59.86%.

Non-college residents display the mirror image of this pattern. Although they represent 59.86% of the citywide population, the average non-college resident lives in a UPZ where 70.32% of residents are also non-college and only 29.68% are college-educated. Taken together, these results indicate that college and non-college residents are not only unevenly distributed across Bogotá, but are also disproportionately exposed to members of their own group within neighborhoods. This provides further evidence that residential segregation by socioeconomic status is quantitatively meaningful across UPZs in the city.

Taken together, these five facts motivate a model in which labor-market duality affects residential segregation through wages, sectoral access, and housing affordability. Formal employment is spatially concentrated, but the relative geography of formal wage access is broadly similar across education groups: residential locations with better access to formal opportunities for high-educated workers also tend to offer better access for low-educated workers. The key difference is therefore not that formal opportunities are located in entirely different parts of the city for each group, but that high-educated workers are much more

[4] For reference, education-based dissimilarity indices of similar magnitude have been documented in other urban contexts. Friedman & Tom (2023) report an average index of 36.3% across U.S. metropolitan areas with populations above 500,000, while Chandler & Phillips (2020) estimate an education-based index of 44% for Washington, D.C. A closer comparison is Istanbul, where Pinarcioğlu & Işik (2009) report an index of 49.3%, although their comparison is between university graduates and individuals with only primary schooling rather than the broader group of residents without college education.

[5] The isolation index for college-educated residents is

$$P^*_{CC} = \sum_{r=1}^{R} \left(\frac{C_r}{C}\right)\left(\frac{C_r}{T_r}\right),$$

where T_r is the total population of UPZ r. The corresponding exposure of college-educated residents to non-college residents is

$$P^*_{CN} = \sum_{r=1}^{R} \left(\frac{C_r}{C}\right)\left(\frac{N_r}{T_r}\right).$$

Analogous expressions apply when the reference group is non-college residents.

Table 5: Group shares, isolation, and exposure across UPZs

	Share in city	College	Non-college
College	40.14%	55.75%	44.25%
Non-college	59.86%	29.68%	70.32%

Notes: The first column reports each group's share in the citywide population. The entries in the last two columns report the average neighborhood composition experienced by the representative member of each group. Considering only these last two columns, diagonal elements correspond to isolation indices, while off-diagonal elements correspond to exposure indices.

likely to obtain formal jobs and earn substantially higher wages. Low-educated workers, in contrast, are more exposed to informal employment, lower wages, and more local labor-market opportunities. When housing rents vary sharply across neighborhoods, these wage and sectoral differences translate into unequal feasible residential sets.

The next section develops a spatial equilibrium model that captures these mechanisms. Workers choose residence, sector, and workplace subject to wages, commuting costs, housing prices, and sector-specific access. Firms determine local labor demand and wages across formal and informal sectors. The model provides a framework to quantify how concentrated formal employment, unequal access to formal jobs, and neighborhood housing prices jointly contribute to residential segregation in Bogotá.

5 Model

In this section, I develop a spatial equilibrium model in the spirit of Ahlfeldt et al. (2015), building on Zárate (2022) to incorporate labor-market segmentation and on Tsivanidis (2022) to allow for worker heterogeneity. Households choose where to live, which sector to work in, and where in the city to work, taking housing costs as given. Firms operate in formal and informal sectors. Formal firms are more intensive in highly educated workers, pay social benefits, and benefit from productivity spillovers. Workers trade off wages, commuting costs, housing expenditures, and neighborhood amenities. Formal employment is subject to frictions that reduce the probability of obtaining a job and therefore lower the expected utility of formal work. In addition, neighborhood composition matters through an externality linked to the local share of college-educated residents.

5.1 Workers

There is a closed city with a discrete set of R neighborhoods. I use r to index residential locations and j to index workplace or production locations. The city is populated by a mass L of workers, normalized to 1. Workers belong to one of two predetermined education groups, low-educated (l) or high-educated (h), indexed by $e \in \{l, h\}$.

Workers make a joint choice over residence, firm type, and workplace. Let $s \in \{F, I\}$ denote firm type, where F refers to formal firms and I to informal firms. A worker of education type e who works in location j and is employed by a firm of type s earns wage w_{sje}.

Workers have non-homothetic preferences due to a minimum housing requirement. The indirect utility of a worker who resides in location r, works in location j, and is employed in sector s is given by

$$U_{rsj} = \left(\frac{C_r}{\alpha}\right)^{\alpha} \left(\frac{H_r - \bar{H}}{1-\alpha}\right)^{1-\alpha} d_{rj}^{-1} \varepsilon_{rsj}, \tag{1}$$

where C_r denotes consumption of the final composite good, H_r denotes housing consumption in residential location r, and $\bar{H}$ is the minimum housing requirement. The term d_{rj}^{-1} captures iceberg commuting costs between residence r and workplace j, while ε_{rsj} is an idiosyncratic taste shock over residence–sector–workplace alternatives. The parameter $\alpha \in (0, 1)$ governs the expenditure share on the final composite good, with $1 - \alpha$ corresponding to the housing share.

Given labor income w_{sje}, the worker faces the budget constraint

$$w_{sje} = P_r C_r + p_r^H H_r, \tag{2}$$

where P_r is the price of the final composite good in residential location r and p_r^H is the housing price in location r.

The final composite good combines goods produced by formal and informal firms:

$$C_r = \left[\omega C_{F,r}^{\frac{\sigma-1}{\sigma}} + (1-\omega) C_{I,r}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where $C_{F,r}$ and $C_{I,r}$ denote consumption of goods produced by formal and informal firms, respectively. The parameter $\omega \in (0, 1)$ governs the weight on formal goods, and $\sigma > 1$ is the elasticity of substitution between formal and informal goods consumption.

Taking the formal good as the numeraire, $p^F = 1$, the corresponding price index for the final composite good in residential location r is

$$P_r = \left[\omega^{\sigma} + (1-\omega)^{\sigma} (p_r^I)^{1-\sigma}\right]^{1/(1-\sigma)}, \tag{4}$$

where p_r^I is the price of the informal good in location r. See Appendix C.1 to see the derivation.

Formal jobs are not obtained with certainty, and access to them may differ across education groups. I capture this through an education-specific formal access term $\lambda_{Fe} \in (0, 1]$, which is common across workplace locations. This term summarizes reduced-form barriers to formal employment, including hiring requirements, information frictions, networks, minimum-wage constraints, or other institutional frictions that differ by education. I interpret λ_{se} as the probability that a worker of education group e can access a job in sector s, with $\lambda_{se} = \lambda_{Fe}$ if $s = F$ and $\lambda_{se} = 1$ if $s = I$.

If access is realized, the worker receives the payoff associated with alternative (r, s, j). If access is not realized, the worker receives zero labor income from that alternative, so the payoff associated with that job alternative is zero. The indirect utility is defined over the set of feasible alternatives satisfying $w_{sje} > p_r^H \bar{H}$. Alternatives that do not satisfy this condition are not available to the worker. Hence, the expected indirect utility associated with feasible alternative (r, s, j) is

$$\mathbb{E}_a \left[V_{rsje} \mid e, \varepsilon_{rsje} \right] = \lambda_{se} \frac{w_{sje} - p_r^H \bar{H}}{P_r^{\alpha} (p_r^H)^{1-\alpha}} d_{rj}^{-1} \varepsilon_{rsje}, \tag{5}$$

where informal access is set equal to one for both education groups. Thus, λ_{Fe} captures formal-sector access frictions relative to the informal sector.

Workers face a hierarchical location-employment choice. At the top level, they choose where to reside. Conditional on residence, they choose between formal and informal firms. Conditional on residence and firm type, they choose a workplace location.

To capture this structure, idiosyncratic preferences follow a three-level nested Fréchet distribution over residential location (r), sector (s), and workplace location (j). Residence is the top nest, sector is the intermediate nest, and workplace location is the bottom nest. This specification allows preference shocks to be correlated within nests, so the utility components (ε_{rsj}) are not independently distributed across all (r, s, j) alternatives. Instead, the nested joint distribution induces dependence among alternatives that share the same residence or sector. Let (B_r) denote the residence-specific amenity shifter and (B_s) the sector-specific amenity shifter. Following Zárate (2022), the joint distribution of (ε_{rsj}) is

$$G(\varepsilon) = \exp \left[- \sum_r B_r \left(\sum_s B_s \left(\sum_j \varepsilon_{rsj}^{-\theta_s} \right)^{\kappa/\theta_s} \right)^{\eta/\kappa} \right]. \tag{6}$$

Here, B_r denotes residential amenities in location r and B_s denotes firm-type-specific amenities. The parameter θ_s governs substitution across workplace locations within firm type s, κ governs substitution across firm types within a residential location, and η governs substitution across residential locations.

Motivated by the evidence in Diamond (2016) that amenities may endogenously improve in higher-skill locations, and following the specification in Tsivanidis (2022), I allow residential amenities to depend on both an exogenous component and the local college share:

$$B_{re} = \bar{B}_r \left(\frac{L_r^h}{L_r} \right)^{\psi_e}, \tag{7}$$

where L_r^h / L_r denotes the share of college-educated residents in location r, and ψ_e governs how strongly workers of education group e value amenities associated with the local concentration of college-educated residents.[6] The term $\bar{B}_r$ captures exogenous residential amenities, such as geography, baseline infrastructure, and other location characteristics that are not determined within the model. Allowing ψ_e to differ across education groups captures the possibility that college and non-college workers value these endogenous amenities differently.

5.2 Firms

In each workplace location j, there is a continuum of competitive firms. Firms operate as either formal or informal, indexed by $s \in \{F, I\}$. Both formal and informal firms hire college and non-college workers, indexed by $e \in \{h, l\}$, but they differ in their production technology, non-wage labor costs, and exposure to concentration spillovers.

Under constant returns to scale, aggregate production by firms of type s in location j can be represented by a representative firm. Let L_{sjh} and L_{sjl} denote college and non-college labor hired by firms of type s in location j. Output is given by

$$Q_j^s = A_j^s N_j^s, \tag{8}$$

where A_j^s is productivity and N_j^s is a composite labor input defined as

$$N_j^s = \left(\chi_s L_{sjh}^{\nu} + (1 - \chi_s) L_{sjl}^{\nu} \right)^{\frac{1}{\nu}}. \tag{9}$$

The parameter $\chi_s \in (0, 1)$ governs the importance of college labor in production for firm type s. I assume

$$\chi_F > \chi_I, \tag{10}$$

so that formal firms are relatively more intensive in college labor than informal firms.

[6] Tsivanidis (2022) provides a related motivation for Bogotá, where uneven public-good provision may lead higher-income or more educated residents to generate local amenities through private security, maintenance, and other neighborhood-level investments.

Formal employment is subject to non-wage labor costs, such as mandatory social benefits. Let $\tau_s \geq 0$ denote the proportional non-wage labor cost faced by firms of type s. I assume

$$\tau_F > 0, \qquad \tau_I = 0. \tag{11}$$

A representative firm of type s in location j solves

$$\Pi_j^s = p_j^s A_j^s N_j^s - (1 + \tau_s)\left(w_{sjh} L_{sjh} + w_{sjl} L_{sjl}\right), \tag{12}$$

where w_{sje} is the wage paid to workers of education type e employed by firms of type s in location j. The formal good is freely traded within the city and is taken as the numeraire, so $p^F = 1$. The informal good is non-traded and its price, p_j^I, is determined locally in equilibrium.

5.2.1 Concentration Spillovers

I allow productivity to differ across formal and informal firms. Formal-firm productivity depends on an exogenous component and on concentration spillovers generated by the local density of college employment. This specification is analogous to the productivity spillovers in Ahlfeldt et al. (2015), but I restrict spillovers to operate through the density of college employment within the same workplace location. Specifically,

$$A_j^F = \bar{A}_j^F \mathcal{A}_j^{\xi}, \tag{13}$$

where $\bar{A}_j^F$ is an exogenous productivity component and $\xi \geq 0$ measures the strength of local concentration spillovers. I define

$$\mathcal{A}_j \equiv \frac{L_j^h}{E_j}, \tag{14}$$

where L_j^h is total college employment in location j and E_j denotes land area in location j. Thus, formal productivity is higher in locations with greater density of college employment, holding fixed the exogenous productivity component $\bar{A}_j^F$.

Informal firms do not feature concentration spillovers. Their productivity is therefore given by

$$A_j^I = \bar{A}_j^I. \tag{15}$$

5.3 Equilibrium

Given the set of exogenous parameters

$$\Theta = \Big\{\alpha, \bar{H}, \omega, \sigma, \eta, \kappa, \nu, \xi, (d_{rj}^{-1})_{r,j\in R}, (\lambda_{se})_{s\in\{F,I\},e\in\{h,l\}}, (L_e, \psi_e)_{e\in\{h,l\}},$$

$$(B_s, \theta_s, \chi_s, \tau_s)_{s\in\{F,I\}}, (p_r^H, \bar{B}_r, E_r)_{r\in R}, (\bar{A}_j^s)_{s\in\{F,I\},j\in R}\Big\},$$

an equilibrium is a collection of endogenous objects

$$x = \Big\{\left(\pi_{re}, \left(\pi_{s|re}, \left(\pi_{j|rse}\right)_{j\in R}\right)_{s\in\{F,I\}}\right)_{r\in R, e\in\{h,l\}}, (L_{sje}, w_{sje})_{s\in\{F,I\},j\in R,e\in\{h,l\}},$$

$$\left(p_r^I, P_r, B_{re}\right)_{r\in R,e\in\{h,l\}}, \left(A_j^F\right)_{j\in R}\Big\}.$$

The choice probabilities determine the residential distribution by education, sectoral employment choices, and workplace choices. The objects (L_{sje}) and (w_{sje}) denote employment and wages by sector, workplace location, and education group. Together with the choice probabilities, sector-location employment (L_{sje}) can be aggregated or disaggregated to construct employment masses at any required level of the model, including residence-sector-workplace-education cells (r, s, j, e). The objects (p_r^I) and (P_r) denote, respectively, the informal-good price and the final composite price index in residential location (r). Residential amenities (B_{re}) are endogenous because they depend on the local college share, while formal productivity (A_j^F) is endogenous because it depends on the local density of college employment.

For computation, the equilibrium system can be reduced to a smaller set of unknowns as presented in the following subsections. Given a vector of wages, firm labor composites, and informal-good prices, all remaining endogenous objects can be recovered from household choice probabilities, labor aggregation, price-index equations, amenity equations, and productivity equations. I therefore solve directly for

$$\Big\{(w_{Fjh}, w_{Fjl}, w_{Ijh}, w_{Ijl})_{j\in R}, \left(N_j^F, N_j^I\right)_{j\in R}, \left(p_j^I\right)_{j\in R}\Big\}.$$

This gives a system with $7|R|$ unknowns: four wages, two sector-specific labor composites, and one informal-good price for each workplace location j.

Accordingly, the equilibrium system must provide $7|R|$ independent equations. In the baseline version of the model, these equations are given by labor-market clearing by sector, workplace location, and education group; zero-profit conditions for formal and informal firms;

and informal-good market clearing in each location.

5.3.1 Labor market clearing

Labor markets clear by workplace location, firm type, and education group. Since workers choose a residence r, a firm type $s \in \{F, I\}$, and a workplace location j, while firms hire college and non-college workers separately, the relevant labor market is indexed by (s, j, e). Thus, for each firm type s, workplace location j, and education group $e \in \{h, l\}$,

$$L^S_{sje} = L^D_{sje}. \tag{16}$$

These conditions pin down equilibrium wages by firm type, workplace location, and education group.

Let

$$y_{rsje} \equiv w_{sje} - p^H_r \bar{H} \tag{17}$$

denote the income available after paying for the minimum housing requirement in residential location r. The alternative (r, s, j, e) is feasible only if $y_{rsje} > 0$. In what follows, infeasible alternatives are assigned zero choice probability.

Using the worker optimization problem to derive labor supply, the mass of education-e workers supplied to firms of type s in workplace location j is

$$L^S_{sje} = L_e \sum_{r \in R} \pi_{re} \pi_{s|re} \pi_{j|rse}, \tag{18}$$

where L_e denotes the population share of education group e, with $\sum_{e \in \{h,l\}} L_e = 1$.

From the firm's problem, labor demand for high-educated workers is

$$L^D_{sjh} = N^s_j \left(\frac{p^s_j A^s_j \chi_s}{(1 + \tau_s) w_{sjh}} \right)^{\frac{1}{1-\nu}}, \tag{19}$$

while labor demand for low-educated workers is

$$L^D_{sjl} = N^s_j \left(\frac{p^s_j A^s_j (1 - \chi_s)}{(1 + \tau_s) w_{sjl}} \right)^{\frac{1}{1-\nu}}. \tag{20}$$

Therefore, for each workplace location j, labor market clearing gives four conditions. For formal firms, using $p_j^F = 1$, the conditions are

$$L_h \sum_{r \in R} \pi_{rh} \pi_{F|rh} \pi_{j|rFh} = N_j^F \left(\frac{A_j^F \chi_F}{(1+\tau_F) w_{Fjh}} \right)^{\frac{1}{1-\nu}}, \tag{21}$$

and

$$L_l \sum_{r \in R} \pi_{rl} \pi_{F|rl} \pi_{j|rFl} = N_j^F \left(\frac{A_j^F (1-\chi_F)}{(1+\tau_F) w_{Fjl}} \right)^{\frac{1}{1-\nu}}. \tag{22}$$

For informal firms, using $\tau_I = 0$, the conditions are

$$L_h \sum_{r \in R} \pi_{rh} \pi_{I|rh} \pi_{j|rIh} = N_j^I \left(\frac{p_j^I A_j^I \chi_I}{w_{Ijh}} \right)^{\frac{1}{1-\nu}}, \tag{23}$$

and

$$L_l \sum_{r \in R} \pi_{rl} \pi_{I|rl} \pi_{j|rIl} = N_j^I \left(\frac{p_j^I A_j^I (1-\chi_I)}{w_{Ijl}} \right)^{\frac{1}{1-\nu}}. \tag{24}$$

The probabilities are implied by the nested Fréchet structure. Conditional on residence r, firm type s, and education group e, the probability of choosing workplace location j is

$$\pi_{j|rse} = \frac{y_{rsje}^{\theta_s} d_{rj}^{-\theta_s}}{\sum_{k \in R: y_{rske} > 0} y_{rske}^{\theta_s} d_{rk}^{-\theta_s}}, \tag{25}$$

for all feasible alternatives satisfying $y_{rsje} > 0$, and $\pi_{j|rse} = 0$ otherwise. The residential price terms $P_r^\alpha (p_r^H)^{1-\alpha}$ drop from this expression because the probability is conditional on residence r. The formal-access term λ_{se} also drops from the conditional workplace choice because it is common across workplace locations within a given (r, s, e) nest.

Define the workplace inclusive value

$$\Phi_{rse} \equiv \lambda_{se}^{\theta_s} \sum_{k \in R: y_{rske} > 0} y_{rske}^{\theta_s} d_{rk}^{-\theta_s}. \tag{26}$$

The probability that a worker of education group e chooses firm type s, conditional on residence r, is then

$$\pi_{s|re} = \frac{B_s \Phi_{rse}^{\kappa/\theta_s}}{\sum_{s' \in \{F,I\}} B_{s'} \Phi_{rs'e}^{\kappa/\theta_{s'}}}. \tag{27}$$

Finally, the probability that a worker of education group e chooses residence r is

$$\pi_{re} = \frac{B_{re}\left[P_r^{\alpha}(p_r^H)^{1-\alpha}\right]^{-\eta}\left(\sum_{s'\in\{F,I\}} B_{s'}\Phi_{rs'e}^{\kappa/\theta_{s'}}\right)^{\eta/\kappa}}{\sum_{q\in R} B_{qe}\left[P_q^{\alpha}(p_q^H)^{1-\alpha}\right]^{-\eta}\left(\sum_{s'\in\{F,I\}} B_{s'}\Phi_{qs'e}^{\kappa/\theta_{s'}}\right)^{\eta/\kappa}}. \tag{28}$$

Together, these conditions provide $4|R|$ labor market clearing equations: one for each firm type, workplace location, and education group. The role of these conditions is to determine the equilibrium wage vector $\{w_{Fjh}, w_{Fjl}, w_{Ijh}, w_{Ijl}\}_{j\in R}$ such that, in each labor market (s, j, e), the mass of workers choosing to supply labor to that market equals the firm's demand for that type of labor. See Appendix C.2 for the full derivation.

5.3.2 Firm-Type Unit-Cost Conditions

Under perfect competition and constant returns to scale, firms earn zero profits in equilibrium. Combining zero profits with firms' conditional labor demands, Appendix C.3 derives the unit-cost conditions for formal and informal firms.

Since the formal good is the numeraire, $p_j^F = 1$ for all j. The formal-firm unit-cost condition is

$$1 = \frac{1}{A_j^F}\left[\chi_F^{\frac{1}{1-\nu}}\left((1+\tau_F)w_{Fjh}\right)^{-\frac{\nu}{1-\nu}} + (1-\chi_F)^{\frac{1}{1-\nu}}\left((1+\tau_F)w_{Fjl}\right)^{-\frac{\nu}{1-\nu}}\right]^{-\frac{1-\nu}{\nu}}. \tag{29}$$

The corresponding unit-cost condition for informal firms is

$$p_j^I = \frac{1}{A_j^I}\left[\chi_I^{\frac{1}{1-\nu}} w_{Ijh}^{-\frac{\nu}{1-\nu}} + (1-\chi_I)^{\frac{1}{1-\nu}} w_{Ijl}^{-\frac{\nu}{1-\nu}}\right]^{-\frac{1-\nu}{\nu}}, \tag{30}$$

where I use $\tau_I = 0$.

Together, these conditions provide $2|R|$ firm-type unit-cost equations: one for each sector and workplace location. The role of these conditions is to impose zero profits under perfect competition, so that the output price of each sector equals the minimum unit cost implied by wages, productivity, and non-wage labor costs. Combined with firms' conditional labor demand equations, they characterize the cost-minimizing mix of college and non-college labor used by formal and informal firms. These unit-cost conditions complement the labor market clearing conditions by jointly disciplining equilibrium wages and the sector-specific labor composites $\{N_j^F, N_j^I\}_{j\in R}$. See Appendix C.3 for the full derivation.

5.3.3 Informal Good Market Clearing

Let

$$y_{rsje} \equiv w_{sje} - p_r^H \bar{H} \tag{31}$$

denote income net of the minimum housing requirement for a worker who resides in r, works in location j, and is employed by a firm of type s. The alternative (r, s, j, e) is feasible only if $y_{rsje} > 0$.

For a worker of education group e, expected resources net of the minimum housing requirement, conditional on residing in location r, are

$$y_{re} \equiv \sum_{s\in\{F,I\}} \pi_{s|re} \sum_{j\in R} \pi_{j|rse} \lambda_{se} y_{rsje}. \tag{32}$$

The term λ_{se} appears because, if access to a job in sector s is not realized, labor income from that alternative is zero.

Aggregate net resources of residents in location r are therefore

$$Y_r \equiv \sum_{e\in\{l,h\}} L_e \pi_{re} y_{re}. \tag{33}$$

Because informal goods are non-traded, they must be produced and consumed in the same location. Therefore, equilibrium in the informal-good market requires

$$Q_r^I = D_r^I. \tag{34}$$

Using $Q_r^I = A_r^I N_r^I$ and the informal-good demand derived in Appendix C.4, this condition can be written as

$$A_r^I N_r^I = \frac{\alpha Y_r}{P_r \Xi_r^I}. \tag{35}$$

This condition links local informal production to local demand for informal goods. Locations with higher residential income generate greater demand for informal goods and therefore require greater informal production. The role of these conditions is to determine the vector of informal-good prices $\{p_r^I\}_{r\in R}$ such that, in each location, informal-good production equals informal-good demand. They provide one equation per location, for a total of $|R|$ equations.

Taken together, labor market clearing provides $4|R|$ equations, since markets clear separately by firm type, workplace location, and education group. The formal- and informal-firm unit-cost conditions provide $2|R|$ equations. Finally, informal-good market clearing provides $|R|$ equations. Hence, the equilibrium system delivers $7|R|$ equations to solve for the $7|R|$ endogenous variables: the four wages $\{w_{Fjh}, w_{Fjl}, w_{Ijh}, w_{Ijl}\}_{j\in R}$, the two composite labor inputs $\{N_j^F, N_j^I\}_{j\in R}$, and the informal-good prices $\{p_j^I\}_{j\in R}$.

5.3.4 Consistency Conditions

Equilibrium must satisfy consistency conditions linking endogenous amenities and formal productivity to the allocation they induce. First, residential composition by education is determined by workers' location choices:

$$L_r^e = L_e \pi_{re}, \qquad e \in \{l, h\}. \tag{36}$$

Total residence in location r is therefore

$$L_r = \sum_{e\in\{l,h\}} L_r^e. \tag{37}$$

Hence, the local college share is

$$\frac{L_r^h}{L_r}. \tag{38}$$

Residential amenities must then be consistent with the equilibrium residential composition:

$$B_{re} = \bar{B}_r \left(\frac{L_r^h}{L_r}\right)^{\psi_e}, \qquad e \in \{l, h\}. \tag{39}$$

Thus, amenities in location r depend on an exogenous component $\bar{B}_r$ and on the equilibrium share of college-educated residents. The parameter ψ_e allows the strength of this endogenous amenity channel to differ across education groups.

Second, formal-firm productivity must be consistent with the equilibrium density of college employment in each workplace location. Total college employment in location j is

$$L_j^h = L_{Fjh} + L_{Ijh}, \tag{40}$$

where L_{Fjh} and L_{Ijh} denote high-educated workers employed by formal and informal firms in workplace location j, respectively. Therefore, formal productivity satisfies

$$A_j^F = \bar{A}_j^F \left(\frac{L_j^h}{E_j}\right)^{\xi}. \tag{41}$$

This condition requires formal productivity to be consistent with the density of college employment generated by the equilibrium allocation.

Informal-firm productivity does not feature concentration spillovers and remains exogenous:

$$A_j^I = \bar{A}_j^I. \tag{42}$$

6 Estimation

This section describes how I bring the model to the data. I first construct the spatial objects that enter the model directly, including commuting costs, housing prices, land area, education-group population shares, and baseline amenity and productivity components. I then assign a subset of parameters using normalizations, values from the literature, or direct empirical counterparts. The remaining parameters are calibrated internally so that the model matches targeted moments related to formal employment access, wage differences across sectors and education groups, employment concentration, commuting patterns, and residential segregation by education.

6.1 Parameters

As mentioned above, the data come primarily from the Encuesta Multipropósito (EM), initially using the 2021 wave, with planned extensions to subsequent waves as they become available. The EM provides individual-level information on housing expenditures, education, labor-market status, formality, wages, and commuting behavior. I also use confidential workplace-location information to construct origin-destination commuting flows across UPZs. These data are used to measure commuting costs, housing prices, education-group population shares, and several empirical moments used in the calibration.

Table 6 summarizes the parameters used in the quantitative version of the model. I divide them into four groups. The first group is measured directly from the data. This includes the housing expenditure share, the distribution of workers by education group, housing prices by residential location, commuting costs between residential and workplace locations, and land area by location. The second group is taken from the literature. This includes preference and substitution elasticities governing residential choice, sector choice, workplace choice, and substitution between formal and informal goods. The third group is fixed by normalization or external institutional information. For example, I normalize the sector amenity terms B_F and B_I to one, normalize informal-sector access to one, and set τ_F using Colombian non-wage labor cost regulations while imposing $\tau_I = 0$. The fourth group is calibrated internally to match targeted moments in the data.

I set the housing expenditure share, $1-\alpha$, equal to its sample average in the EM. The estimated share is 0.337, implying $\alpha = 0.663$. I also take the education-group population shares directly from the EM, with $L_h = 0.4014$ for college workers and $L_l = 0.5986$ for non-college workers. Housing prices, commuting distances, land area, and the location characteristics used to construct E_r are also measured directly from the data and Bogotá maps.

I take the residential choice elasticity from Tsivanidis (2022). Since the baseline abstracts from education-specific residential elasticities, I impose a common value of η, equal to the education-share-weighted average of the two estimates reported there. This yields $\eta = 2.142$. I also take the education-specific amenity spillover elasticities from Tsivanidis (2022), setting $\psi_l = 0.730$ for non-college workers and $\psi_h = 1.002$ for college workers. I set the elasticity of substitution between formal and informal goods to $\sigma = 2$, set $\omega = 0.5$, corresponding to equal weights on the two components of the final composite good as in Autor & Dorn (2009), and set $\nu = 0.2$. The workplace-choice elasticity θ_s is estimated separately by sector, while the sector-choice elasticity κ is set to 2.8 following Zárate (2022). Finally, I set non-wage labor costs to $\tau_F = 0.45$ in the formal sector and $\tau_I = 0$ in the informal sector, using Colombian labor regulations.

The remaining parameters are calibrated internally because they govern the mechanisms emphasized in the paper. The formal access parameters λ_{Fe} capture education-specific barriers to formal employment, while informal access is normalized to one for both education groups. The minimum housing requirement $\bar{H}$ governs non-homothetic housing demand, the baseline residential amenity terms $\bar{B}_r$ discipline residential segregation across neighborhoods, the formal skill-intensity parameter χ_F governs the relative demand for college labor in the formal sector, and the productivity spillover elasticity ξ determines how local college employment affects formal productivity. I set $\chi_I = 0.25$ for the informal sector. Finally, the baseline productivity terms $\bar{A}_j^s$ are only partially calibrated: rather than estimating a separate productivity parameter for every sector-location pair, I calibrate a formal-sector productivity scale and an additional productivity shifter for the central-east employment cluster, defined as the group of UPZs Pardo Rubio, Chapinero, Las Nieves, Sagrado Corazón, La Macarena, La Candelaria, Chicó Lago, and El Refugio. These internally calibrated parameters are chosen so that the model matches empirical moments on formal employment by education, wage differences across sectors and education groups, the concentration of formal and informal employment in the central-east part of the city, residential shares across neighborhoods, and educational segregation.

Table 6: Model parameters

Parameter	Value / Source
α	0.663; computed from the Encuesta Multipropósito
$\bar{H}$	Calibrated
ω	0.5; equal weights as in Autor & Dorn (2009)
σ	2
η	2.142; weighted average from Tsivanidis (2022)
κ	2.8; from Zárate (2022)
ν	0.2
ξ	Calibrated
$(d_{rj}^{-1})_{r,j\in R}$	Constructed from UPZ-to-UPZ distances using Bogotá maps and estimated elasticities θ_s
$(\lambda_{se})_{s\in\{F,I\},e\in\{h,l\}}$	Calibrated for $s = F$ and normalized to one for $s = I$
$(L_e)_{e\in\{h,l\}}$	$L_h = 0.4014$ and $L_l = 0.5986$; computed from the Encuesta Multipropósito
$(\psi_e)_{e\in\{h,l\}}$	$\psi_l = 0.730$ and $\psi_h = 1.002$; from Tsivanidis (2022)
$(B_s)_{s\in\{F,I\}}$	Normalized $B_F = 1$ and $B_I = 1$
$(\theta_s)_{s\in\{F,I\}}$	Estimated
$(\chi_s)_{s\in\{F,I\}}$	χ_F calibrated and $\chi_I = 0.25$
$(\tau_s)_{s\in\{F,I\}}$	$\tau_F = 0.45$ and $\tau_I = 0$; from Colombian regulations
$(p_r^H)_{r\in R}$	Computed from the Encuesta Multipropósito
$(\bar{B}_r)_{r\in R}$	Calibrated
$(E_r)_{r\in R}$	Computed from Bogotá maps
$(\bar{A}_j^s)_{s\in\{F,I\},j\in R}$	Partially calibrated: formal-sector scale and central-east productivity shifter

6.2 Internally Estimated Parameters

Estimation of θ_I and θ_F. To estimate the workplace-choice elasticities, I follow the approach in Zárate (2022). I parameterize iceberg commuting costs as

$$d_{rj} = \exp(\delta\, \mathrm{time}_{rj}), \tag{43}$$

where time_{rj} denotes commuting time between residential location r and workplace location j. The parameter δ governs the semi-elasticity of iceberg commuting costs with respect to travel time.

I estimate δ using a multinomial logit model of transportation-mode choice, in which workers choose among bus, private car, and walking. For each origin-destination pair, I construct predicted travel times by transportation mode. Identification comes from within-origin-destination variation across modes: conditional on the same origin and destination, workers are more likely to choose modes with shorter travel times. Restricting the sample to individuals who do not work from home, I obtain $\delta = 0.027$, somewhat larger than the estimate reported by Zárate (2022).

I then recover the workplace-choice elasticities separately by employment status using observed origin-destination-sector flows. In the model, workers may differ in their access to formal employment by education group. However, these education-specific access frictions are not separately identified in this reduced-form step. Instead, the estimating equation is based on aggregate commuting flows by sector, so education-specific barriers to formal access are absorbed by sector-destination attractiveness and fixed effects.

Appendix C.5 shows that combining the workplace-choice probability in equation (25) with the commuting-cost parameterization in equation (43) yields the estimating equation

$$\ln \pi_{j|rse} = \beta_s \,\text{time}_{rj} + \gamma_{jse} + \gamma_{rse} + \varepsilon_{rjse}, \tag{44}$$

where the coefficient on commuting time satisfies

$$\beta_s = -\theta_s \delta. \tag{45}$$

Thus, δ captures the disutility of commuting time at the mode-choice level, while β_s captures the reduced-form sensitivity of workplace choices to commuting time within each employment status. Given an estimate of δ, the structural workplace-choice elasticity is recovered as

$$\theta_s = -\frac{\beta_s}{\delta}. \tag{46}$$

Estimating equation (44) separately by employment status and restricting the sample to commuters, I obtain $\beta_I = -0.110$ and $\beta_F = -0.044$. These estimates imply workplace-choice elasticities of $\theta_I = 4.09$ and $\theta_F = 1.64$. The larger elasticity for informal workers indicates that informal employment choices are more sensitive to commuting time, consistent with more localized informal labor markets. In contrast, the lower elasticity for formal workers implies that formal workers are willing to search over a broader set of workplace locations. Compared with Zárate (2022), my estimates are similar for informal workers but smaller for formal workers. This difference may reflect differences in sample composition, the use of post-pandemic commuting patterns in the 2021 EM, or changes in workplace flexibility after COVID-19, which weakened the relationship between formal employment choices and observed commuting time. It may also reflect unobserved barriers to formal employment that are absorbed by the fixed effects.

Calibration of $\bar{B}_r$, $\bar{H}$, and labor-market parameters. The remaining parameters are calibrated using a sequential procedure that separates the residential block of the model from the labor-market block, while still allowing the two to interact through equilibrium choices. The calibrated objects are the baseline residential amenities $\{\bar{B}_r\}_{r\in R}$, the minimum housing requirement $\bar{H}$, the education-specific formal-access parameters λ_{Fl} and λ_{Fh}, the formal skill-intensity parameter χ_F, and the formal productivity spillover elasticity ξ. I also calibrate two parsimonious productivity shifters for the formal sector: a common formal-sector productivity scale, μ_A^F, which shifts formal productivity in all locations, and a central-east productivity shifter, μ_A^{CE}, which raises formal productivity in the main employment cluster. These two shifters allow the model to match both the overall formal wage premium and the spatial concentration of formal employment without estimating a separate productivity parameter for every location.

The calibration proceeds in three steps. First, I solve a simplified version of the model with $\bar{H} = 0$ and common residential amenities to obtain plausible starting values for the labor-market parameters. Second, holding these initial labor-market parameters fixed, I calibrate the residential block. For each candidate value of $\bar{H}$, I choose $\{\bar{B}_r\}_{r\in R}$ so that the model matches the empirical aggregate residential distribution across UPZs. I then select $\bar{H}$ to discipline the observed degree of residential segregation by education, measured by the dissimilarity index. Third, holding fixed the calibrated residential block, I recalibrate the labor-market parameters

$$x = \left(\lambda_{Fl}, \lambda_{Fh}, \mu_A^F, \mu_A^{CE}, \chi_F, \xi\right).$$

The vector x is chosen to match moments that discipline the central mechanisms of the model: formal employment shares by education, wage ratios across sectors and education groups, and the spatial concentration of formal and informal employment in the central-east cluster. For each candidate vector, I solve the full equilibrium system and compute the model-implied moments. The objective function is a weighted sum of squared relative deviations between model and data moments, minimized using the Nelder–Mead algorithm. I impose economically motivated restrictions requiring positive formal access, weakly higher formal access for college workers than for non-college workers, $\chi_F \in (0, 1)$, and $\xi \geq 0$. Appendix C.6 provides the full calibration algorithm, the objective function, and the numerical implementation details.

Table 7 reports a positive minimum housing requirement, $\bar{H} = 0.414$, which gives the model a non-homothetic affordability channel: housing costs impose a tighter constraint on low-wage workers and therefore help translate wage inequality into residential segregation. The formal-access parameters are low for both education groups, with $\lambda_{Fl} = 0.164$ and $\lambda_{Fh} = 0.163$, implying that formal employment is difficult to access relative to informal employment. In this calibration, differences in formal-sector attachment are therefore driven less by large exogenous differences in access parameters and more by equilibrium wages, sectoral productivity, skill intensity, and residential affordability. The calibrated value of $\chi_F = 0.789$ indicates that formal firms are strongly intensive in college labor, while the formal productivity scale $\mu_A^F = 4.975$, the central-east shifter $\mu_A^{CE} = 1.495$, and the spillover elasticity $\xi = 0.069$ allow the model to generate both a formal wage premium and a spatial concentration of formal employment.

Table 7: Calibrated parameter values

Parameter	Calibrated value
$\bar{H}$	0.414
λ_{Fl}	0.164
λ_{Fh}	0.163
μ_A^F	4.975
μ_A^{CE}	1.495
χ_F	0.789
ξ	0.069

Table 8 shows that the model matches several targeted moments closely. The model reproduces the overall degree of educational residential segregation, with a dissimilarity index of 0.437 compared with 0.442 in the data. It also matches the formal employment share among non-college workers and the formal employment share in the central-east cluster almost exactly. The model, however, overpredicts the formal employment share of college workers and somewhat overpredicts both the formal-informal wage ratio and the college-non-college wage ratio. These discrepancies indicate that the calibration generates a slightly stronger concentration of high-educated workers in the formal sector and somewhat larger equilibrium wage gaps than observed in the data. The largest mismatch among the targeted moments is the informal employment share in the central-east cluster: the model predicts 0.049, compared with 0.087 in the data. This suggests that the model makes informal employment too dispersed, or alternatively that it does not yet capture all forces that attract informal activity to high-income central and eastern neighborhoods, such as local service demand or commercial density.

Table 8: Targeted and model-implied moments

Moment	Empirical value	Model value
Dissimilarity index	0.442	0.437
Formal share, college workers	0.808	0.909
Formal share, non-college workers	0.502	0.505
Formal/informal wage ratio	2.120	2.324
College/non-college wage ratio	3.000	3.280
Formal employment share in central-east cluster	0.147	0.150
Informal employment share in central-east cluster	0.087	0.049

Additional validation moments are reported in Appendix Table B2. The model closely reproduces the spatial pattern of educational segregation, with a correlation of 0.935 between empirical and model-implied college shares by UPZ. It also matches the formal CMA gradient, although it overpredicts average commuting distances.

7 Counterfactual Analysis

Table 9 reports a set of counterfactual exercises around the recovered baseline equilibrium. The baseline dissimilarity index is $D = 0.437$, with a formal employment share of 0.909 among high-educated workers and 0.505 among low-educated workers. The baseline also implies a formal-informal wage ratio of 2.324, a high- to low-education wage ratio of 3.280, and a share of infeasible residence-work alternatives equal to 0.059. Each counterfactual changes one component of the calibrated economy while keeping the remaining structural environment fixed.

Counterfactual 1. Housing subsidies for low-educated workers. The first counterfactual reduces the minimum housing-cost component for low-educated households by 20%. Formally, for $e = \ell$, the housing-cost term changes from $\bar{H}p_r^H$ to $(1 - 0.20)\bar{H}p_r^H$, while it remains unchanged for high-educated households. This intervention produces the largest change in residential segregation. The dissimilarity index falls from 0.437 to 0.310, a reduction of about 29%. The share of infeasible alternatives also declines sharply, from 0.059 to 0.030. By contrast, labor-market outcomes change only modestly. The formal employment share among low-educated workers falls slightly, from 0.505 to 0.495, and the formal-informal wage ratio declines from 2.324 to 2.314. These results suggest that the subsidy reduces segregation mainly by relaxing the residential affordability constraint, rather than by substantially changing sectoral employment or equilibrium wages.

Counterfactual 2. Higher formal access for low-educated workers. The second counterfactual increases formal-sector access for low-educated workers. Formally, I set the formal-access parameter for low-educated workers to $\lambda_{F\ell} = 0.30$ (previously 0.164), while leaving all other access parameters unchanged. This intervention substantially changes sectoral allocation but has almost no effect on residential segregation. The formal employment share among low-educated workers rises from 0.505 to 0.600, while the formal-informal wage ratio falls from 2.324 to 1.934. The decline in the wage ratio reflects both the expansion of formal employment among lower-wage workers and the equilibrium adjustment of wages as effective access to formal jobs increases. However, the dissimilarity index remains essentially unchanged, moving from 0.437 to 0.437. This result separates labor-market access from residential affordability: improving access to formal jobs raises formal employment among low-educated workers, but in this calibration it does not relax the housing affordability constraint enough to change where households live.

Counterfactual 3. Improved commuting access from low-education neighborhoods. The third counterfactual improves commuting access from low-college residential locations. Formally, for origins r classified as low-college neighborhoods, I replace travel times time_{rj} with 0.75time_{rj} for every workplace destination j, while leaving travel times from all other origins unchanged. This intervention produces a moderate decline in residential segregation: the dissimilarity index falls from 0.437 to 0.417, a reduction of about 4.5%. Formal employment shares and wage ratios remain close to their baseline values. These results suggest that targeted improvements in commuting conditions can reduce segregation, but their effect is smaller than the effect of directly relaxing the housing affordability constraint, because the cost of living in high-rent neighborhoods remains unchanged.

Counterfactual 4. Citywide reduction in commuting frictions. The fourth counterfactual reduces the commuting-friction parameter citywide by 30%. Formally, I replace δ_{time} with $0.70\delta_{\text{time}}$ for all origin-destination pairs, while leaving the travel-time matrix time_{rj} unchanged. Unlike the targeted commuting intervention, this broad reduction slightly increases residential segregation: the dissimilarity index rises from 0.437 to 0.445. Although the increase is small, it is informative. A citywide reduction in commuting frictions benefits all workers, rather than specifically relaxing the constraints faced by low-educated households or low-college neighborhoods. Since housing prices and wage differences remain unchanged, the intervention does not directly expand access to high-rent neighborhoods for low-educated workers. Generalized mobility improvements therefore need not reduce educational segregation unless they disproportionately improve access for disadvantaged groups or are combined with policies that relax housing affordability constraints.

Counterfactual 5. Combined intervention. The final counterfactual combines all previous targeted and citywide interventions. This scenario generates the lowest dissimilarity index: D falls from 0.437 to 0.307, a reduction of about 30%. The share of infeasible alternatives declines from 0.059 to 0.027, while the formal employment share among low-educated workers rises from 0.505 to 0.601. The combined intervention therefore affects both residential affordability and labor-market allocation. However, the reduction in segregation is only slightly larger than under the housing-subsidy counterfactual alone. This indicates that, in the current calibration, relaxing housing affordability is the dominant force behind the reduction in residential segregation, while improved formal access mainly changes the sectoral composition of employment.

Table 9: Counterfactual results from recovered baseline calibration

Scenario	D	$s_{F\|h}$	$s_{F\|\ell}$	w_F/w_I	w_h/w_ℓ	q_{inf}
Baseline	0.437	0.909	0.505	2.324	3.280	0.059
(C1) Housing subsidy	0.310	0.906	0.495	2.314	3.235	0.030
(C2) Low-ed formal access	0.437	0.913	0.600	1.934	3.346	0.055
(C3) Low-college commute	0.417	0.908	0.504	2.312	3.267	0.059
(C4) Citywide commute	0.445	0.910	0.508	2.308	3.279	0.060
(C5) Combined	0.307	0.910	0.601	1.876	3.302	0.027

Notes: D is the residential dissimilarity index.$s_{F|h}$ is the formal employment share among high-educated workers.$s_{F|\ell}$ is the formal employment share among low-educated workers.w_F/w_I is the formal-informal wage ratio.w_h/w_ℓ is the high- to low-education wage ratio.q_{inf} is the share of infeasible alternatives.

Overall, the counterfactuals indicate that educational segregation in the model is most responsive to policies that relax the residential affordability constraint faced by low-educated households. Labor-market access matters for formal employment, and targeted commuting improvements can modestly reduce segregation, but neither mechanism generates a decline in segregation comparable to the direct housing-cost intervention. These results support the interpretation that sectoral wage inequality and housing prices interact to make parts of the city effectively inaccessible for low-educated workers, especially when they remain in lower-paying informal employment.

8 Conclusion

This paper studies how labor-market duality between formal and informal employment contributes to residential segregation by education within a city. Using Bogotá as the empirical setting, I show that formal employment is spatially concentrated, informal work is more dispersed and local, and workers without college education are much more exposed to informal employment and lower wages. The main mechanism is not that low- and high-educated workers face entirely different maps of formal wage access. Instead, the key difference is the level of earnings attached to labor-market opportunities. When housing rents vary sharply across neighborhoods, lower wages restrict the residential choice sets of low-educated workers, especially when they remain in informal employment.

To quantify this mechanism, I develop a spatial equilibrium model in which workers choose residence, sector, and workplace. Workers differ by education, face formal-sector access frictions, and make choices subject to wages, housing prices, commuting costs, and amenities. Housing demand is non-homothetic through a minimum housing requirement, so low-income workers are more constrained in high-rent neighborhoods. The calibrated model reproduces the overall degree of educational residential segregation in Bogotá and captures key features of the formal-informal wage structure and spatial employment patterns.

The counterfactual exercises indicate that the housing affordability channel is central. Reducing the effective minimum housing cost faced by low-educated households lowers the dissimilarity index from 0.437 to 0.310, while substantially reducing the share of infeasible alternatives. Increasing formal-sector access for low-educated workers raises their formal employment rate, but has little effect on residential segregation. Commuting interventions have smaller and less uniform effects: targeted improvements from low-college neighborhoods reduce segregation modestly, while a citywide reduction in commuting disutility slightly increases segregation.

The results should be interpreted as model-based comparative statics rather than as evaluations of specific policies. The current framework holds housing prices fixed and abstracts from endogenous housing supply, migration, and fiscal costs. Future work should incorporate richer local-demand forces for informal services, endogenous housing prices, and additional sources of neighborhood amenities. Nevertheless, the analysis shows that formal-informal labor-market segmentation can have important spatial consequences. Sectoral wage inequality changes who can afford to live in different parts of the city, making labor-market duality a relevant mechanism behind educational residential segregation.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, *83*(6), 2127–2189.

Albrecht, J., Navarro, L., & Vroman, S. (2009). The effects of labour market policies in an economy with an informal sector. *The Economic Journal*, *119*(539), 1105–1129.

Autor, D. H., & Dorn, D. (2009). *Inequality and specialization: the growth of low-skill service jobs in the united states.* National Bureau of Economic Research.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political economy*, *81*(4), 813–846.

Benabou, R. (1993). Workings of a city: location, education, and production. *The Quarterly Journal of Economics*, *108*(3), 619–652.

Bobba, M., Flabbi, L., & Levy, S. (2022). Labor market search, informality, and schooling investments. *International Economic Review*, *63*(1), 211–259.

Bobba, M., Flabbi, L., Levy, S., & Tejada, M. (2021). Labor market search, informality, and on-the-job human capital accumulation. *Journal of Econometrics*, *223*(2), 433–453.

Chandler, J. P., & Phillips, J. (2020). Racial, education, and income segregation in the district of columbia. *District of Columbia, Office of Planning. Retrieved May*, *22*, 2024.

Couture, V., Gaubert, C., Handbury, J., & Hurst, E. (2024). Income growth and the distributional effects of urban spatial sorting. *Review of Economic Studies*, *91*(2), 858–898.

Couture, V., & Handbury, J. (2020). Urban revival in america. *Journal of Urban Economics*, *119*, 103267.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *American economic review*, *106*(3), 479–524.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077–1106.

Finlay, J., & Williams, T. C. (2025). Housing demand, inequality, and spatial sorting. *Journal of International Economics*, 104171.

Friedman, S., & Tom, T. (2023). Residential segregation by education in the US, 2016–2020. In *Population and development in the 21st century: Between the anthropocene and anthropocentrism.* IntechOpen.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, *108*(11), 3117–3153.

Gyourko, J., Mayer, C., & Sinai, T. (2013). Superstar cities. *American Economic Journal: Economic Policy*, *5*(4), 167–199.

Khanna, G., Nyshadham, A., Ramos-Menchelli, D., Tamayo, J. A., & Tiew, A. (2023). *Spatial mobility, economic opportunity, and crime*. Harvard Business School.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of economic perspectives*, *28*(3), 109–126.

Levy, S. (2008). Good intentions. *Bad Outcomes (Washington, DC: Brookings Institution, 2008)*.

Lucas, R. E., & Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, *70*(4), 1445–1476.

Massey, D. S., & Denton, N. A. (1987). Trends in the residential segregation of blacks, hispanics, and asians: 1970-1980. *American sociological review*, 802–825.

Meghir, C., Narita, R., & Robin, J.-M. (2015). Wages and informality in developing countries. *American Economic Review*, *105*(4), 1509–1546.

Narita, R. (2020). Self-employment in developing countries: A search-equilibrium approach. *Review of Economic Dynamics*, *35*, 1–34.

Pinarcioğlu, M. M., & Işik, O. (2009). Segregation in istanbul: patterns and processes. *Tijdschrift voor economische en sociale geografie*, *100*(4), 469–484.

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. *American journal of sociology*, *116*(4), 1092–1153.

Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. *Unpublished manuscript*, *18*.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, *108*(8), 2015–2047.

Warnes, P. E. (2020). Transport infrastructure improvements and spatial sorting: Evidence from buenos aires.

Wheeler, C. H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics*, *19*(4), 879–899.

Zárate, R. D. (2022). Spatial misallocation, informality, and transit improvements. *Development Research*.

A Figures

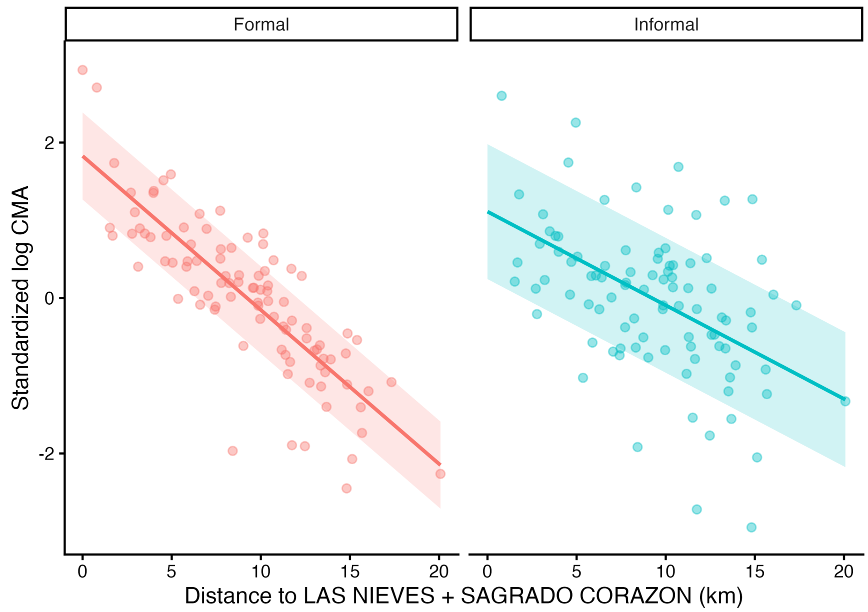


Figure A1: CMA decline from the center

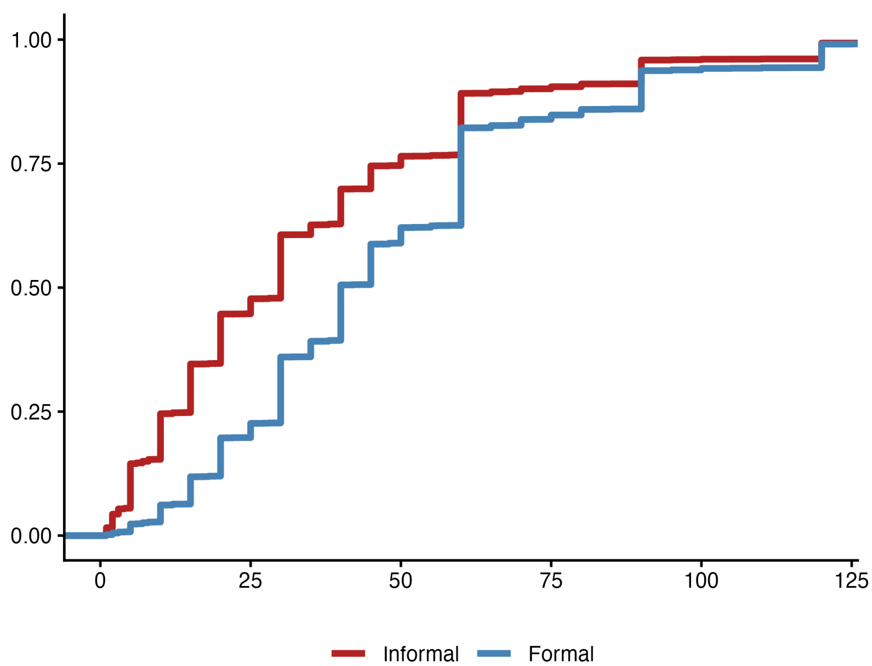


Figure A2: Commuting time cumulative distribution by sector (minutes)

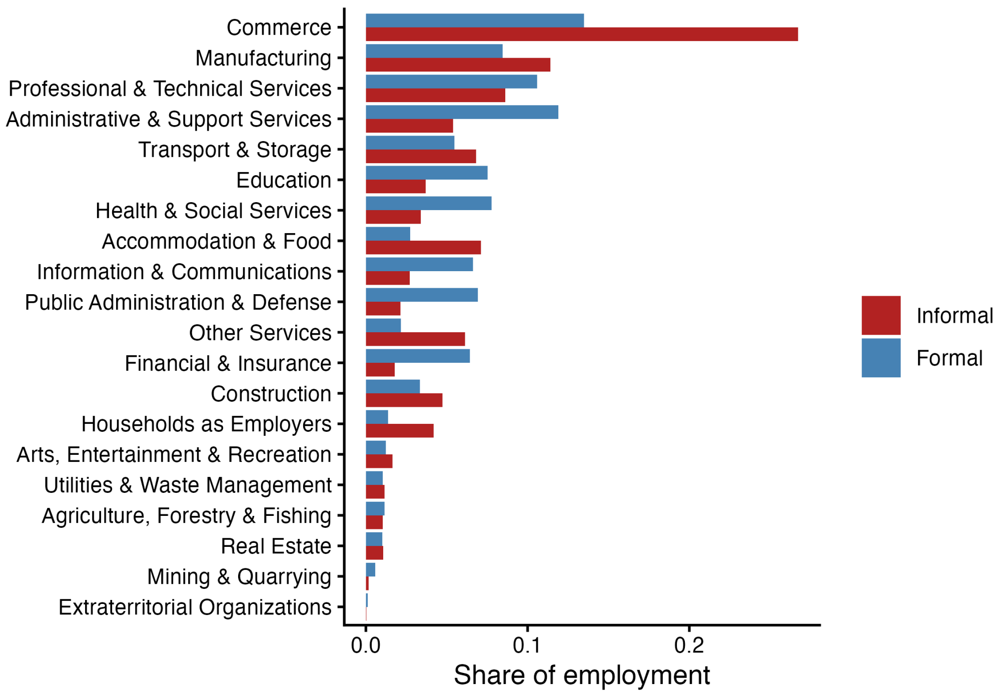


Figure A3: Share of total employment by industry

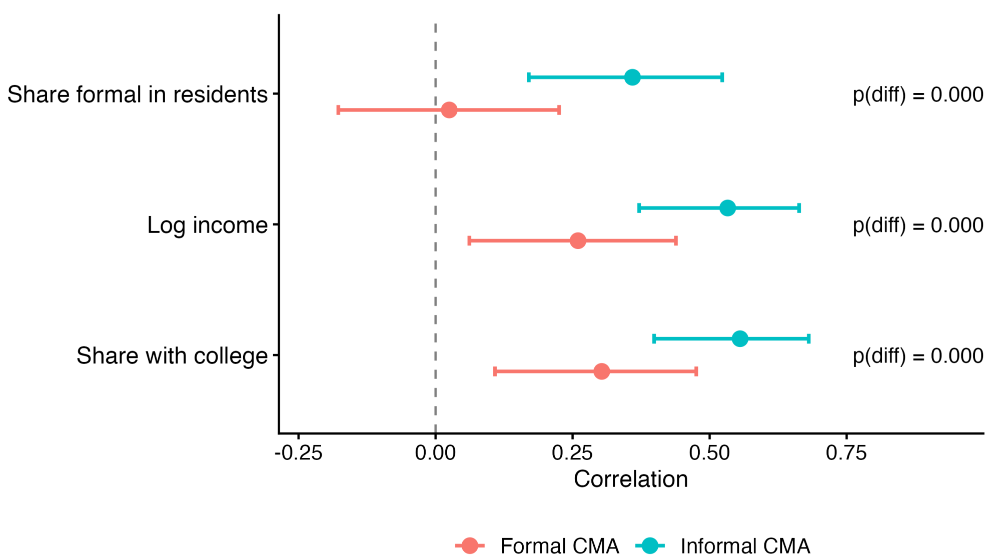
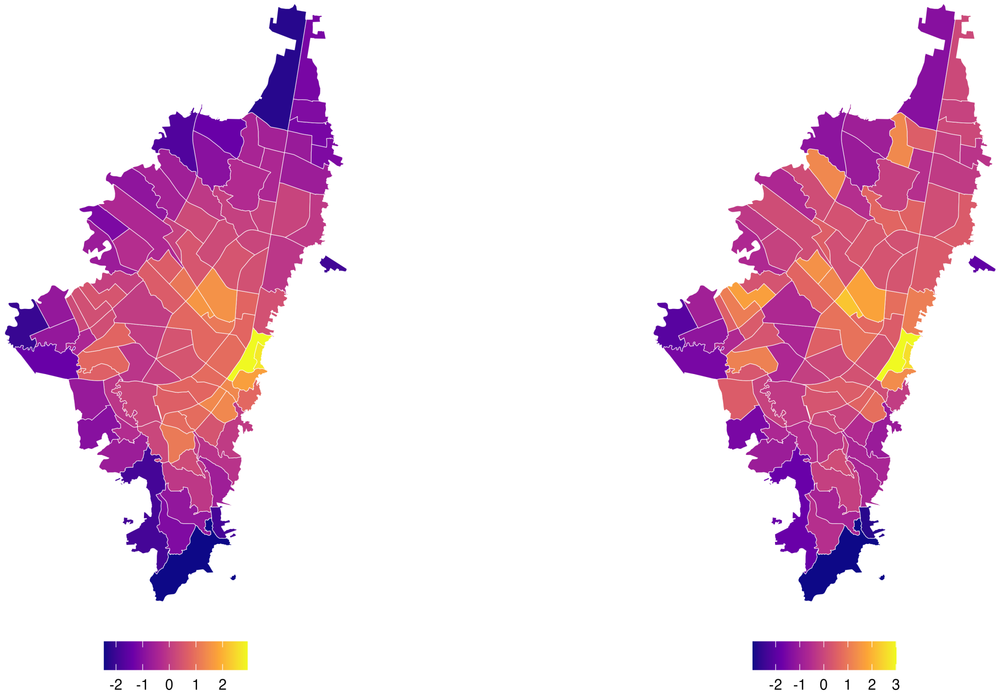


Figure A4: UPZ characteristics and commuter market access by sector

Notes: The figure reports Pearson correlations between UPZ-level characteristics and standardized log commuter market access (CMA) in the formal and informal sectors. Horizontal bars denote 95% confidence intervals. Reported p-values correspond to tests of equality between the correlation with formal CMA and the correlation with informal CMA for the same UPZ-level characteristic, accounting for the fact that both CMA measures are observed for the same UPZs.



(a) Formal sector

(b) Informal sector

Figure A5: Standardized logarithm of CMA by sector across Bogotá

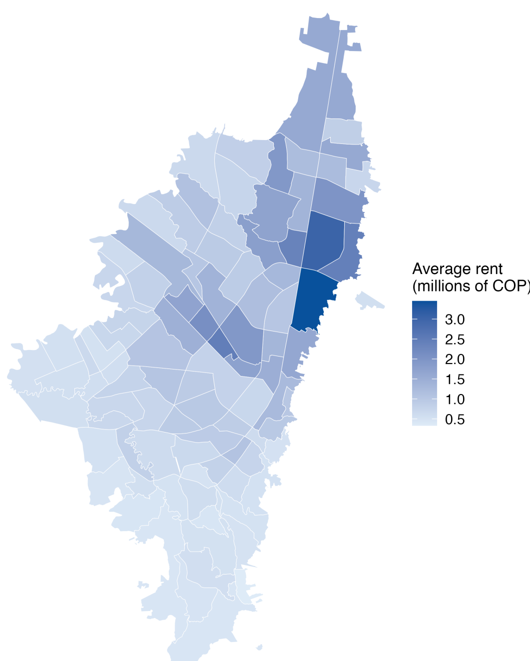


Figure A6: Rent in Bogotá by UPZ

B Tables

Table B1: Attrition in Workplace Location

Variable	Workplace UPZ Observed	Workplace UPZ Missing	Difference	p-value
Male	0.533	0.566	0.033	0.000
Age	38.327	37.667	-0.660	0.000
College	0.423	0.227	-0.196	0.000
Formal employee	0.543	0.455	-0.089	0.000
Independent professional / business owner	0.104	0.051	-0.052	0.000
Informal employee	0.114	0.176	0.062	0.000
Self-employed	0.239	0.318	0.080	0.000
Commute time	24.914	45.291	20.377	0.000

Notes: The sample is restricted to employed workers. The table compares workers with observed and missing workplace UPZ. The difference is defined as the mean for workers with missing workplace UPZ minus the mean for workers with observed workplace UPZ. Employment-group variables are indicators, so their means correspond to group shares. Number of workers with observed workplace UPZ: 71476; number with missing workplace UPZ: 24193.

Table B2: Additional model validation moments

Panel A. Spatial labor-market validation		
Moment	Empirical	Model
CMA slope, formal sector	-0.198	-0.193
CMA slope, informal sector	-0.121	-0.147
Mean commute distance, formal sector	5.020	9.875
Mean commute distance, informal sector	3.280	8.225

Panel B. Residential segregation validation	
Statistic	Value
Correlation between empirical and model college share by UPZ	0.935

Notes: Panel A compares empirical and model-implied spatial labor-market moments. CMA slopes are obtained from regressions of standardized log commuter market access on distance to the main wage center. Mean commute distances are measured in kilometers. Panel B reports the correlation across UPZs between the empirical and model-implied college share.

C Quantification

C.1 Derivation of the CES Price Index

This appendix derives the price index associated with the CES aggregator

$$C_r = \left[\omega C_{F,ri}^{\frac{\sigma-1}{\sigma}} + (1-\omega) C_{I,ri}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}. \tag{47}$$

Taking the formal good as the numeraire, let $p^F = 1$, and denote by $p \equiv p_r^I$ the price of the informal good in residential location r. For a target level of the composite good $\bar{C}$, the household solves the expenditure minimization problem

$$\min_{C_F, C_I} \; C_F + pC_I \tag{48}$$

subject to

$$\left[\omega C_F^{\rho} + (1-\omega) C_I^{\rho}\right]^{1/\rho} = \bar{C}. \tag{49}$$

where $\rho \equiv \frac{\sigma-1}{\sigma}$ and restriction satisfies equality because it binds.

The Lagrangian is

$$\mathcal{L} = C_F + pC_I + \lambda \left\{\bar{C} - \left[\omega C_F^{\rho} + (1-\omega) C_I^{\rho}\right]^{1/\rho}\right\}. \tag{50}$$

The first-order conditions are

$$1 = \lambda \left[\omega C_F^{\rho} + (1-\omega) C_I^{\rho}\right]^{1/\rho - 1} \omega C_F^{\rho-1}, \text{and} \tag{51}$$

$$p = \lambda \left[\omega C_F^{\rho} + (1-\omega) C_I^{\rho}\right]^{1/\rho - 1} (1-\omega) C_I^{\rho-1}. \tag{52}$$

Since $\rho - 1 = \frac{\sigma-1}{\sigma} - 1 = -\frac{1}{\sigma}$, and defining $A \equiv \lambda \left[\omega C_F^{\rho} + (1-\omega) C_I^{\rho}\right]^{1/\rho - 1}$, thus they become

$$1 = A\omega C_F^{-1/\sigma}, \text{ and} \tag{53}$$

$$p = A(1-\omega) C_I^{-1/\sigma}. \tag{54}$$

Solving for C_F and C_I gives

$$C_F = (A\omega)^{\sigma}, \text{ and} \tag{55}$$

$$C_I = \left(\frac{A(1-\omega)}{p}\right)^{\sigma}. \tag{56}$$

Now substitute these expressions into the binding CES constraint.

First, because $\sigma\rho = \sigma \cdot \frac{\sigma-1}{\sigma} = \sigma - 1$, hence

$$C_F^\rho = (A\omega)^{\sigma-1}, \text{ and} \tag{57}$$

$$C_I^\rho = \left(\frac{A(1-\omega)}{p}\right)^{\sigma-1}. \tag{58}$$

Therefore,

$$\omega C_F^\rho + (1-\omega)C_I^\rho = \omega(A\omega)^{\sigma-1} + (1-\omega)\left(\frac{A(1-\omega)}{p}\right)^{\sigma-1} = A^{\sigma-1}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]. \tag{59}$$

Substituting into the CES constraint,

$$\bar{C} = \left\{A^{\sigma-1}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]\right\}^{\sigma/(\sigma-1)} = A^\sigma\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]^{\sigma/(\sigma-1)}. \tag{60}$$

Minimum expenditure is $E = C_F + pC_I$, and taking (55) and (56),

$$E = (A\omega)^\sigma + p\left(\frac{A(1-\omega)}{p}\right)^\sigma = A^\sigma\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]. \tag{61}$$

From (60),

$$A^\sigma = \bar{C}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]^{-\sigma/(\sigma-1)}, \tag{62}$$

and substituting into (61),

$$E = \bar{C}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]^{-\sigma/(\sigma-1)}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right] = \bar{C}\left[\omega^\sigma + (1-\omega)^\sigma p^{1-\sigma}\right]^{1/(1-\sigma)}. \tag{63}$$

By definition, the CES price index P_r satisfies $E = P_r\bar{C}$, and therefore,

$$P_r = \left[\omega^\sigma + (1-\omega)^\sigma (p_r^I)^{1-\sigma}\right]^{1/(1-\sigma)}. \tag{64}$$

C.2 Labor Market Clearing Conditions Derivation

From the supply side, since total population is normalized to one, L_{sje}^S can be interpreted as the share of the city population with education e supplied to firms of type s in workplace location j. These objects are determined by workers' optimization.

Let

$$y_{rsje} \equiv w_{sje} - p_r^H \bar{H} \tag{65}$$

denote the income available after paying for the minimum housing requirement in residential location r. The alternative (r, s, j, e) is feasible only if $y_{rsje} > 0$. Infeasible alternatives are assigned zero choice probability.

Based on equation (5) and the properties of the nested Fréchet distribution given in equation (6), the probability that a worker of education group e living in r chooses workplace location j, conditional on firm type s, is

$$\pi_{j|rse} = \frac{y_{rsje}^{\theta_s} d_{rj}^{-\theta_s}}{\sum_{k \in R: y_{rske} > 0} y_{rske}^{\theta_s} d_{rk}^{-\theta_s}}, \tag{66}$$

for all feasible alternatives satisfying $y_{rsje} > 0$, and $\pi_{j|rse} = 0$ otherwise. The residential price terms $P_r^{\alpha}(p_r^H)^{1-\alpha}$ drop from this expression because the probability is conditional on residence r. The access term λ_{se} also drops from the conditional workplace choice probability because it is common across workplace locations within a given (r, s, e) nest.

Define the workplace inclusive value

$$\Phi_{rse} \equiv \lambda_{se}^{\theta_s} \sum_{k \in R: y_{rske} > 0} y_{rske}^{\theta_s} d_{rk}^{-\theta_s}. \tag{67}$$

The term λ_{se} enters the inclusive value because it affects the expected payoff from accessing sector s for workers of education group e. Informal access is set equal to one for both education groups, while formal access may differ by education.

Similarly, the probability that a worker of education group e chooses firm type s, conditional on residence in r, is

$$\pi_{s|re} = \frac{B_s \Phi_{rse}^{\kappa/\theta_s}}{\sum_{s' \in \{F,I\}} B_{s'} \Phi_{rs'e}^{\kappa/\theta_{s'}}}. \tag{68}$$

The probability that a worker of education group e chooses residential location r is

$$\pi_{re} = \frac{B_{re} \left[P_r^{\alpha}(p_r^H)^{1-\alpha}\right]^{-\eta} \left(\sum_{s' \in \{F,I\}} B_{s'} \Phi_{rs'e}^{\kappa/\theta_{s'}}\right)^{\eta/\kappa}}{\sum_{q \in R} B_{qe} \left[P_q^{\alpha}(p_q^H)^{1-\alpha}\right]^{-\eta} \left(\sum_{s' \in \{F,I\}} B_{s'} \Phi_{qs'e}^{\kappa/\theta_{s'}}\right)^{\eta/\kappa}}. \tag{69}$$

The term B_{re} denotes the residential amenity faced by education group e in location r. It may differ across education groups because endogenous amenities depend on the local college share through the education-specific elasticity ψ_e.

Given these probabilities, the mass of workers of education group e who choose residence r, firm type s, and workplace location j is

$$L_{rsje}^S = L_e \pi_{re} \pi_{s|re} \pi_{j|rse}, \tag{70}$$

where L_e denotes the population share of education group $e \in \{l, h\}$, with $\sum_{e\in\{l,h\}} L_e = 1$.

Summing over residential locations yields the labor supply from education group e to firms of type s in workplace location j:

$$L^S_{sje} = \sum_{r\in R} L^S_{rsje} = L_e \sum_{r\in R} \pi_{re}\pi_{s|re}\pi_{j|rse}. \tag{71}$$

From the demand side, the firm's first-order conditions imply

$$p^s_j A^s_j \chi_s (N^s_j)^{1-\nu} L^{\nu-1}_{sjh} = (1+\tau_s) w_{sjh}, \tag{72}$$

$$p^s_j A^s_j (1-\chi_s)(N^s_j)^{1-\nu} L^{\nu-1}_{sjl} = (1+\tau_s) w_{sjl}. \tag{73}$$

Solving for factor demands yields

$$L^D_{sjh} = N^s_j \left(\frac{p^s_j A^s_j \chi_s}{(1+\tau_s) w_{sjh}} \right)^{\frac{1}{1-\nu}}, \tag{74}$$

and

$$L^D_{sjl} = N^s_j \left(\frac{p^s_j A^s_j (1-\chi_s)}{(1+\tau_s) w_{sjl}} \right)^{\frac{1}{1-\nu}}. \tag{75}$$

Labor market clearing holds separately by firm type, workplace location, and education group. Thus, for each location j, there are four labor market clearing conditions.

For high-educated workers employed by formal firms, using $p^F_j = 1$:

$$L_h \sum_{r\in R} \pi_{rh}\pi_{F|rh}\pi_{j|rFh} = N^F_j \left(\frac{A^F_j \chi_F}{(1+\tau_F) w_{Fjh}} \right)^{\frac{1}{1-\nu}}. \tag{76}$$

For low-educated workers employed by formal firms, using $p^F_j = 1$:

$$L_l \sum_{r\in R} \pi_{rl}\pi_{F|rl}\pi_{j|rFl} = N^F_j \left(\frac{A^F_j (1-\chi_F)}{(1+\tau_F) w_{Fjl}} \right)^{\frac{1}{1-\nu}}. \tag{77}$$

For high-educated workers employed by informal firms, using $\tau_I = 0$:

$$L_h \sum_{r\in R} \pi_{rh}\pi_{I|rh}\pi_{j|rIh} = N^I_j \left(\frac{p^I_j A^I_j \chi_I}{w_{Ijh}} \right)^{\frac{1}{1-\nu}}. \tag{78}$$

For low-educated workers employed by informal firms, using $\tau_I = 0$:

$$L_l \sum_{r \in R} \pi_{rl} \pi_{I|rl} \pi_{j|rIl} = N_j^I \left(\frac{p_j^I A_j^I (1 - \chi_I)}{w_{Ijl}} \right)^{\frac{1}{1-\nu}} . \tag{79}$$

C.3 Firm-Type Unit-Cost Conditions Derivation

Under perfect competition and constant returns to scale, firms earn zero profits in equilibrium whenever production is positive. For any firm type $s \in \{F, I\}$ and workplace location j, the representative firm solves equation (12). Taking productivity A_j^s as given by the individual firm, zero profits imply

$$p_j^s A_j^s N_j^s = (1 + \tau_s) \left(w_{sjh} L_{sjh} + w_{sjl} L_{sjl} \right) . \tag{80}$$

From the firm's first-order conditions, conditional factor demands are

$$L_{sjh} = N_j^s \left(\frac{p_j^s A_j^s \chi_s}{(1 + \tau_s) w_{sjh}} \right)^{\frac{1}{1-\nu}} , \tag{81}$$

and

$$L_{sjl} = N_j^s \left(\frac{p_j^s A_j^s (1 - \chi_s)}{(1 + \tau_s) w_{sjl}} \right)^{\frac{1}{1-\nu}} . \tag{82}$$

Substituting (81) and (82) into (80) yields

$$\begin{aligned} p_j^s A_j^s N_j^s = {} & (1 + \tau_s) w_{sjh} N_j^s \left(\frac{p_j^s A_j^s \chi_s}{(1 + \tau_s) w_{sjh}} \right)^{\frac{1}{1-\nu}} \\ & + (1 + \tau_s) w_{sjl} N_j^s \left(\frac{p_j^s A_j^s (1 - \chi_s)}{(1 + \tau_s) w_{sjl}} \right)^{\frac{1}{1-\nu}} . \end{aligned} \tag{83}$$

Dividing by N_j^s and factoring out $(p_j^s A_j^s)^{1/(1-\nu)}$ from the right-hand side gives

$$p_j^s A_j^s = (p_j^s A_j^s)^{\frac{1}{1-\nu}} \left[\chi_s^{\frac{1}{1-\nu}} \big((1 + \tau_s) w_{sjh} \big)^{-\frac{\nu}{1-\nu}} + (1 - \chi_s)^{\frac{1}{1-\nu}} \big((1 + \tau_s) w_{sjl} \big)^{-\frac{\nu}{1-\nu}} \right] . \tag{84}$$

Therefore, the unit-cost condition for firms of type s in location j is

$$p_j^s = \frac{1}{A_j^s} \left[\chi_s^{\frac{1}{1-\nu}} \big((1 + \tau_s) w_{sjh} \big)^{-\frac{\nu}{1-\nu}} + (1 - \chi_s)^{\frac{1}{1-\nu}} \big((1 + \tau_s) w_{sjl} \big)^{-\frac{\nu}{1-\nu}} \right]^{-\frac{1-\nu}{\nu}} . \tag{85}$$

Equation (85) gives one unit-cost condition per firm type and workplace location. The formal and informal conditions are obtained as special cases.

Since the formal good is freely traded within the city and is taken as the numeraire, $p_j^F = 1$ for all j. Hence, the formal-firm unit-cost condition is

$$1 = \frac{1}{A_j^F}\left[\chi_F^{\frac{1}{1-\nu}}\big((1+\tau_F)w_{Fjh}\big)^{-\frac{\nu}{1-\nu}} + (1-\chi_F)^{\frac{1}{1-\nu}}\big((1+\tau_F)w_{Fjl}\big)^{-\frac{\nu}{1-\nu}}\right]^{-\frac{1-\nu}{\nu}}. \tag{86}$$

For informal firms, $\tau_I = 0$, and the informal-good price is determined in equilibrium. Therefore, the informal-firm unit-cost condition is

$$p_j^I = \frac{1}{A_j^I}\left[\chi_I^{\frac{1}{1-\nu}} w_{Ijh}^{-\frac{\nu}{1-\nu}} + (1-\chi_I)^{\frac{1}{1-\nu}} w_{Ijl}^{-\frac{\nu}{1-\nu}}\right]^{-\frac{1-\nu}{\nu}}. \tag{87}$$

C.4 Informal-Good Demand and Market Clearing

Let p_r^I denote the price of the informal good in residential location r. The final composite good combines the formal good and the informal good through a CES aggregator, and the corresponding price index is P_r, derived in Appendix C.1. Total expenditure on the final composite good for household i living in location r is

$$E_{ri} = P_r C_{ri}. \tag{88}$$

To derive demand for informal goods, note that conditional cost minimization implies that the marginal rate of substitution between informal and formal goods must equal the relative price. Since the formal good is the numeraire, $p^F = 1$, this condition can be written as

$$\frac{1-\omega}{\omega}\left(\frac{C_{F,ri}}{C_{I,ri}}\right)^{\frac{1}{\sigma}} = p_r^I. \tag{89}$$

Rearranging yields

$$C_{F,ri} = C_{I,ri}\left(\frac{p_r^I \omega}{1-\omega}\right)^{\sigma}. \tag{90}$$

Substituting this expression into the CES composite and then into the expenditure function gives

$$E_{ri} = P_r C_{I,ri}\, \Xi_r^I, \tag{91}$$

where

$$\Xi_r^I \equiv \left[\omega^{\sigma}(p_r^I)^{\sigma-1}(1-\omega)^{1-\sigma} + (1-\omega)\right]^{\frac{\sigma}{\sigma-1}}. \tag{92}$$

Solving for $C_{I,ri}$ yields individual demand for the informal good:

$$C_{I,ri} = \frac{E_{ri}}{P_r \Xi_r^I}. \tag{93}$$

Under preferences over the final composite good and housing with a minimum housing requirement, households allocate a share α of resources net of the minimum housing requirement to the final composite good. Let y_{ri} denote household i's realized resources net of the minimum housing requirement in residential location r. Then

$$E_{ri} = \alpha y_{ri}. \tag{94}$$

Therefore, individual demand for informal goods can be written as

$$C_{I,ri} = \frac{\alpha y_{ri}}{P_r \Xi_r^I}. \tag{95}$$

Since all households living in residential location r face the same informal-good price p_r^I and the same composite-good price index P_r, the term $P_r \Xi_r^I$ is common across residents. Aggregate demand for informal goods in location r is therefore obtained by summing individual demand across all households living there:

$$D_r^I = \sum_{i \in r} C_{I,ri} = \frac{\alpha \sum_{i \in r} y_{ri}}{P_r \Xi_r^I}. \tag{96}$$

Define aggregate net resources of residents in location r as

$$Y_r \equiv \sum_{i \in r} y_{ri}. \tag{97}$$

Then aggregate demand for informal goods can be written as

$$D_r^I = \frac{\alpha Y_r}{P_r \Xi_r^I}. \tag{98}$$

Using the worker choice probabilities, aggregate net resources in location r are

$$Y_r = \sum_{e \in \{h,l\}} L_e \pi_{re} \sum_{s \in \{F,I\}} \pi_{s|re} \sum_{j \in R} \pi_{j|rse} \lambda_{se} \left(w_{sje} - p_r^H \bar{H} \right). \tag{99}$$

The term λ_{se} captures the probability that access to the job is realized. If access is not realized, labor income from that alternative is zero.

Since the informal good is non-traded, the informal-good market clears locally. Therefore, for each location r,

$$Q_r^I = D_r^I. \tag{100}$$

Using $Q_r^I = A_r^I N_r^I$, this condition becomes

$$A_r^I N_r^I = \frac{\alpha Y_r}{P_r \Xi_r^I}. \tag{101}$$

Thus, aggregate demand for informal goods in residential location r is proportional to aggregate resident resources net of the minimum housing requirement, adjusted by the local composite-good price index and the relative-price term Ξ_r^I.

C.5 Derivation of the Gravity Equation

This appendix derives the estimating equation used to recover the workplace-choice elasticities θ_s. For this auxiliary estimation, I abstract from the Stone–Geary housing term by setting $\bar{H} = 0$. Under this restriction, the conditional workplace-choice probability has a standard gravity structure in wages and commuting costs.

From equation (25), for a worker of education group e, the conditional probability of choosing workplace location j given residential location r and employment status s is

$$\pi_{j|rse} = \frac{(\lambda_{se} w_{js})^{\theta_s} d_{rj}^{-\theta_s}}{\Phi_{rse}}, \tag{102}$$

where

$$\Phi_{rse} \equiv \lambda_{se}^{\theta_s} \sum_{k \in R} (w_{ks})^{\theta_s} d_{rk}^{-\theta_s}. \tag{103}$$

Substituting the commuting-cost parameterization from equation (43),

$$d_{rj} = \exp(\delta \, \mathrm{time}_{rj}),$$

into equation (102) yields

$$\pi_{j|rse} = \frac{\left(\lambda_{je}^{s} w_{js}\right)^{\theta_s} \exp(-\theta_s \delta \, \mathrm{time}_{rj})}{\Phi_{rse}}. \tag{104}$$

Taking logs gives

$$\ln \pi_{j|rse} = \theta_s \ln w_{js} + \theta_s \ln \lambda_{je}^{s} - \theta_s \delta \, \mathrm{time}_{rj} - \ln \Phi_{rse}. \tag{105}$$

Equation (105) has a convenient fixed-effects structure. The wage and access terms vary by workplace location, employment status, and education group, but not by residential origin. I therefore absorb them into a destination–status–education fixed effect:

$$\gamma_{jse} \equiv \theta_s \ln w_{js} + \theta_s \ln \lambda^s_{je}. \tag{106}$$

Similarly, the denominator term varies by residential origin, employment status, and education group, but not by workplace destination. I absorb it into an origin–status–education fixed effect:

$$\gamma_{rse} \equiv -\ln \Phi_{rse}. \tag{107}$$

Using these definitions, equation (105) can be written as

$$\ln \pi_{j|rse} = \beta_s \,\mathrm{time}_{rj} + \gamma_{jse} + \gamma_{rse} + \varepsilon_{rjse}, \tag{108}$$

where

$$\beta_s = -\theta_s \delta. \tag{109}$$

The error term ε_{rjse} captures measurement error and specification error in the empirical implementation. Given an estimate of δ, the workplace-choice elasticity is recovered as

$$\theta_s = -\frac{\beta_s}{\delta}. \tag{110}$$

C.6 Calibration Details

The calibrated objects are the baseline residential amenities $\{\bar{B}_r\}_{r\in R}$, the minimum housing requirement $\bar{H}$, the education-specific formal-access parameters λ_{Fl} and λ_{Fh}, the formal skill-intensity parameter χ_F, and the formal productivity spillover elasticity ξ. I also calibrate two parsimonious productivity shifters for the formal sector: a common formal-sector productivity scale, μ_A^F, which shifts formal productivity in all locations, and a central-east productivity shifter, μ_A^{CE}, which raises formal productivity in the main employment cluster. These two shifters allow the model to match both the overall formal wage premium and the spatial concentration of formal employment without estimating a separate productivity parameter for every location.

Informal access is normalized to one for both education groups, and the informal-sector skill-intensity parameter is fixed at $\chi_I = 0.25$. The central-east employment cluster is defined as the group of UPZs Pardo Rubio, Chapinero, Las Nieves, Sagrado Corazón, La Macarena, La Candelaria, Chicó Lago, and El Refugio.

The calibration proceeds in three steps. The first step solves a simplified version of the model with no minimum housing requirement, $\bar{H} = 0$, and common residential amenities across locations. This version abstracts from the endogenous residential block and is used only to obtain plausible starting values for the labor-market parameters. In particular, it provides initial values for formal access, productivity, and skill-intensity parameters that generate reasonable formal-sector shares, wage ratios, and spatial employment patterns.

The second step calibrates the residential block. Holding the initial labor-market parameters fixed, I choose the vector of baseline residential amenities $\{\bar{B}_r\}_{r\in R}$ so that the model matches the empirical aggregate residential distribution across UPZs. For a given candidate value of $\bar{H}$, the model implies residence probabilities π_{re} for each location r and education group $e \in \{l, h\}$. The model-implied aggregate residential share in location r is

$$\hat{\pi}_r = \sum_{e\in\{l,h\}} L_e \pi_{re},$$

where L_e is the population share of education group e. The amenity vector $\{\bar{B}_r\}_{r\in R}$ is updated iteratively until $\hat{\pi}_r$ matches the empirical residential share in each UPZ. I then choose $\bar{H}$ to discipline the empirical level of residential segregation by education, measured by the dissimilarity index. For each candidate $\bar{H}$, I compute the model-implied dissimilarity index $D^{\text{model}}(\bar{H})$ and compare it with the empirical target $D^{\text{data}} = 0.442$.

The third step recalibrates the labor-market block holding the residential block fixed. Given the calibrated $\{\bar{B}_r\}_{r\in R}$ and $\bar{H}$, I choose the vector

$$x = \left(\lambda_{Fl}, \lambda_{Fh}, \mu_A^F, \mu_A^{CE}, \chi_F, \xi\right),$$

where λ_{Fl} and λ_{Fh} are formal-access parameters for non-college and college workers, respectively; μ_A^F is the common formal-sector productivity scale; μ_A^{CE} is the central-east formal productivity shifter; χ_F is the formal-sector high-skill intensity; and ξ governs productivity spillovers from local college employment density.

C.6.1 Targeted moments and objective function

The labor-market calibration chooses x to match a vector of empirical moments. The targeted moments are the formal employment share among college workers, the formal employment share among non-college workers, the formal-to-informal wage ratio, the college-to-non-college wage ratio, the share of formal employment in the central-east cluster, and the share of informal employment in the central-east cluster.

Let $m_k(x)$ denote the model-implied value of moment k, and let m_k^{data} denote its empirical counterpart. The calibration minimizes a weighted sum of squared relative deviations:

$$\mathcal{L}(x) = \sum_k \omega_k \left(\frac{m_k(x) - m_k^{\text{data}}}{m_k^{\text{data}}} \right)^2,$$

where ω_k is the weight assigned to moment k. In the baseline calibration, I place relatively higher weight on matching formal employment shares by education and the informal employment share in the central-east cluster, since these moments are central for disciplining formal access and the degree of spatial concentration across sectors.

C.6.2 Numerical implementation

For each candidate parameter vector x, I solve the full equilibrium system using a nonlinear solver. The equilibrium unknowns are expressed in logs for numerical stability. These include four wage vectors, two sector-specific labor-composite vectors, and the informal-good price vector. Given the equilibrium solution, I recover the model-implied moments and evaluate the loss function.

The outer minimization of $\mathcal{L}(x)$ is performed using the Nelder–Mead simplex algorithm. During the search, I impose economically motivated parameter restrictions: formal access must be positive; high-educated workers cannot have lower formal access than non-college workers; the common formal productivity scale and the central-east productivity shifter are restricted to reasonable ranges; $\chi_F \in (0, 1)$; and $\xi \geq 0$. Candidate vectors that violate these restrictions or fail to produce a converged equilibrium receive a large penalty in the objective function.

This sequential strategy should be interpreted as a preliminary quantification approach. It first uses a simplified model to obtain plausible labor-market starting values, then uses the residential distribution and segregation moments to discipline $\{\bar{B}_r\}_{r \in R}$ and $\bar{H}$, and finally recalibrates the labor-market parameters holding the residential block fixed. This reduces the dimensionality of the calibration problem while preserving the main equilibrium links between residential segregation, formal access, wages, and the spatial concentration of employment.