

# Skill allocation and urban amenities in the developing world

Jorge De la Roca\*<sup>‡</sup>

*University of Southern California*

Andrii Parkhomenko\*<sup>¶</sup>

*University of Southern California*

Daniel Velásquez-Cabrera\*<sup>§</sup>

*University of Michigan*

March 2023

**ABSTRACT:** We use individual geocoded data from Peru and document that the city-size wage premium is larger for low-skilled than for high-skilled workers, in contrast with most developed countries. We interpret this evidence using a model of location choice with private amenity goods and non-homothetic preferences. Skilled workers enjoy higher incomes and devote a higher expenditure share to amenity goods, such as private schools or upper-class neighborhoods. The supply of these amenities is subject to a fixed cost, and only sufficiently large cities have enough demand to offer them. Thus, skilled workers demand a higher wage premium to live in small cities, and the returns to working in a large city are smaller for them than for their unskilled counterparts. Our quantitative exercises indicate that the mechanism accounts for two-thirds of the gap in the city-size wage premium between high-skilled and low-skilled workers.

Key words: skill allocation, agglomeration, amenities

JEL classification: R10, R23, J31

\*We thank Seongmoon Cho and Evgeny Burinskiy for exceptional research assistance. We are grateful to Juan Pablo Chauvin and J.Vernon Henderson for valuable discussions and to Gilles Duranton, Jeff Lin, Bryan Stuart, and seminar and conference participants at Bocconi, Cal State Fullerton, George Washington, osus, the Philadelphia Federal Reserve, Uppsala, UBC, Wharton, and William & Mary for helpful comments and suggestions.

<sup>‡</sup>Sol Price School of Public Policy, University of Southern California, 650 Childs Way RGL 326, Los Angeles, CA 90089, USA (e-mail: [jdelaroc@usc.edu](mailto:jdelaroc@usc.edu); website: <http://jorgedelaroca.name>).

<sup>¶</sup>Marshall School of Business, Department of Finance and Business, University of Southern California, Hoffman Hall, 701 Exposition Blvd, Los Angeles, CA 90089, USA (e-mail: [parkhomenko.andrii@gmail.com](mailto:parkhomenko.andrii@gmail.com); website: <https://www.andrii-parkhomenko.com>).

<sup>§</sup>University of Michigan, Department of Economics, 238 Lorch Hall, 611 Tappan Ave., Ann Arbor, MI 48109, USA (e-mail: [danielvc@umich.edu](mailto:danielvc@umich.edu); website: <http://dvelasquezc.github.io>).

## 1. Introduction

Urbanization and skilled labor are primary ingredients of economic development. Developed countries exhibit high shares of skilled workers who concentrate in large urban areas and earn high nominal incomes. The positive relationship between city size and income—the city-size wage premium—is larger for the high-skilled and often used as evidence for the productive advantages of skill concentration (Moretti, 2004a; Glaeser and Resseger, 2010). In developing countries, the concentration of skills is even more striking, as most skilled workers live in a handful of large cities. The city-size wage premium, however, manifests itself differently for high- and low-skilled workers, compared to developed countries. Duranton (2016) finds that low-skilled informal sector workers have a higher city-size wage premium in Colombia than their high-skilled formal sector counterparts, while Quintero and Roberts (2018) show that in 12 out of 16 Latin American countries, the city-size wage premium is larger for workers in the informal sector.

Why are the returns to city size lower for high-skilled workers in developing countries? To answer this question, we use household geocoded survey data with rich information on labor market outcomes in Peru and rely on satellite data to delineate metropolitan areas. Using various skills measures, such as college education and labor market formality, we document that the city-size wage premium is much smaller for high-skilled workers than for low-skilled ones. We also show that the high living costs of large Peruvian metro areas more than offset any modest large-city wage premium for high-skilled workers. Moreover, we collect data on local amenities and find that some non-natural private amenities, such as private schools or upper-class neighborhoods, are severely undersupplied in small and medium-sized cities.

To formalize the relationship between the local supply of skills, amenities, and city-size wage premium, we build a spatial equilibrium model with low- and high-skilled workers. Workers choose their workplace location, considering wages, housing costs, amenities, and idiosyncratic preferences. There are two types of amenities—exogenous public amenities, such as weather or beautiful views, and endogenous private amenities, such as private schools or hospitals. We model preferences as non-homothetic, where housing is a necessity and private amenities are luxury goods. This utility specification follows from our survey data on household expenses, which indicate that high-skilled households spend a lower share of their income on housing and a much higher proportion on education and recreation.

The supply of private amenities is subject to fixed costs. This is a crucial and realistic assumption, given that providing services such as private education, local crime prevention, or park maintenance requires some initial lumpy capital investment. Given the fixed costs, only sufficiently large cities with high-income workers have enough demand to provide the amenities. Since high-skilled workers spend a higher share of income on amenities, they disproportionately choose to live in cities that supply them. This sorting of high-skilled workers into larger cities lowers their wages in such locations. Similarly, the undersupply of high-skilled workers in small cities increases local wages and leads to a flatter city-size wage premium.

To evaluate the importance of the underprovision of amenities in smaller cities, we build a quantitative version of the model with 66 metropolitan areas in Peru. We use school-level data on

national test scores in math and Spanish skills to identify high-quality private schools. We proxy the existence of amenities in a city if it houses at least one private school in the top 10 percentile of the national test-score distribution. In the data, 36 cities have such high-quality schools. In our benchmark economy, we set the fixed cost of amenity supply to ensure that the same number of cities supply the private amenity.

We then eliminate the fixed costs in our counterfactual experiment. In the absence of fixed costs, all cities supply private amenities. This leads to a reallocation of workers into smaller cities, but the effect is especially pronounced for the high-skilled who spend a higher share of their budget on these amenities. The re-sorting flattens the city-size wage premium for the low-skilled and makes it steeper for the high-skilled. The difference between the two premia is only one-third of the observed difference. In other words, the amenity channel accounts for about two-thirds of the difference in the city-size wage premium between high- and low-skilled workers.

This paper contributes to the literature that examines the location of workers by skill within a country (Berry and Glaeser, 2005; Moretti, 2012; Diamond, 2016; Davis, Mengus, and Michalski, 2021). This literature focuses on developed economies and emphasizes the growing divergence in the location choices of the high-skilled during the past decades as they have increasingly located in large, productive, and amenity-rich cities. We instead focus on the location choices of workers by skill within a developing country. Our paper is the first to note that the highly uneven provision of amenities lowers the city-size wage premium for the high-skilled, which contrasts with the larger city-size high-skilled premium in the developed world.

We also contribute to the long-standing literature that examines urbanization patterns in developing countries. Urban systems in the developing world suffer from the problem of the “missing middle,” whereby second and third-tier cities are unable to attract large numbers of high-skilled workers (Williamson, 1965; Zelinsky, 1971). Our findings on the underprovision of amenities in small and medium-sized cities help explain the limited mobility of the high-skilled and their excessive concentration in a handful of large cities. Lastly, we contribute to the growing literature that estimates the city-size wage premium for developing countries (Duranton, 2016; Combes, Démurger, and Li, 2015; Chauvin, Glaeser, Ma, and Tobio, 2017; Özgüzel, 2023). Our findings show that the city-size wage premium in Peru is slightly larger than available estimates for developed countries, confirming that agglomeration effects are more pronounced in developing countries (Combes and Gobillon, 2015).

The rest of the paper is organized as follows. Section 2 describes the data on workers, metropolitan areas, and local amenities. Section 3 presents our estimations on the city-size wage premium for the entire sample and the high- and low-skilled workers. We also show how urban costs and amenity provisions vary across cities of different sizes. Section 4 introduces the theoretical framework. Section 5 details the estimation and calibration of model parameters and builds a quantitative version of the model. Section 6 describes our counterfactual experiments that highlight the role of local amenities in explaining the difference in city-size wage premium between high-skilled and low-skilled workers.

## 2. Data

We begin by illustrating how we identify urban areas in Peru using satellite data. We then describe the household survey data and the selection criteria we follow to obtain our estimation samples. Next, we define our skills measures based on college attainment and formality status. Lastly, we present two amenity measures: the number of high-quality private schools and middle-class neighborhoods in an urban area.

### *Delineation of urban areas*

Urban economists conceive urban areas as integrated labor markets where workers and firms interact. However, delineating an urban or metro area as a locus of interaction is complex. Local authorities often designate a metro area as a set of administrative units (e.g., counties, municipalities) connected via commuting flows. If residents in two adjoining municipalities often cross boundaries, we expect them to be economically linked. Both municipalities would form a union, and then one can examine whether a third neighboring municipality is linked to such a union. Duranton (2015) follows this approach to delineate metropolitan areas in Colombia.<sup>1</sup>

Unfortunately, reliable commuting data are not available for the municipalities in Peru. We thus follow the cluster algorithm method, an alternative approach to defining metropolitan areas in developing countries (Dijkstra and Poelman, 2014). This approach uses a gridded population data set and identifies a spatially contiguous set of cells with high density, assuming those cells encompass an urban area.

Multiple data sources such as LandScan, Global Human Settlements, or WorldPop provide high-resolution geospatial data on population distribution at fine grids of 1 km<sup>2</sup> or lower scales. Researchers often define a high-density cluster as the union of contiguous grids with a density of 1,500 people per km<sup>2</sup> and an overall population of 50,000 people (Bosker, Park, and Roberts, 2021). Cells in low-density clusters have a lower density (500 people per km<sup>2</sup> or less) and a lower population threshold (20,000 or even as low as 5,000).

We follow the cluster algorithm approach proposed by Henderson, Nigmatulina, and Kriticos (2019)—an adaptation of Dijkstra and Poelman (2014)—to delineate Peruvian urban areas. We use data from WorldPop, a collection of open-access spatial demographic data sets focused on low and middle-income countries that combine census data and satellite imagery to map populations with high precision.<sup>2</sup> We utilize WorldPop’s 1km<sup>2</sup>-resolution annual population counts (using unconstrained top-down methods) for 2007–2019.

Following Henderson, Nigmatulina, and Kriticos (2019), we first smooth the spatial population data by assigning to each 1 km<sup>2</sup> cell the mean density of neighboring cells within 3 km. This smoothness circumvents geographic constraints to development, such as hills, rivers, landmarks, or big parks. For instance, in the case of Lima (the capital and largest urban area), multiple hills,

---

<sup>1</sup>The researcher must determine how to initiate the algorithm (i.e., select two administrative units) and choose the commuting threshold beyond which both units become interconnected. Duranton (2015) uses a threshold of 10 percent.

<sup>2</sup>See the technical details on the mapping methods at <https://www.worldpop.org/methods>. WorldPop has become a reliable and popular data source for measuring worldwide population distribution (Duranton and Puga, 2020).

three rivers, and several archaeological pre-Incan sites act as barriers to population expansion, yet population growth has eventually circumvented them.

We define high-density cells as those with over 1,500 people per km<sup>2</sup>. We group high-density cells into clusters or cores following a rook neighbor criterion, i.e., the reference cell should be adjacent to one of the four cells in the vertical or horizontal direction (excluding the diagonals). Subsequently, we define low-density cells as those with a density between 500 and 1,500 people per km<sup>2</sup>. These low-density cells constitute the metropolitan fringe, and the urban area ends when cells in the outskirts have low-density levels (below 500 people per km<sup>2</sup>).

The algorithm identifies 92 urban areas above 15,000 people using 2019 World Pop data. Urban areas cover 9,416 km<sup>2</sup> (around 0.76 percent of the Peruvian territory) and contain approximately 21.5 million people (67 percent of the population). Lima is the most populated metro area, with 11,069,922 inhabitants; Ferreñafe is the median metro area, with 43,807 inhabitants; and Calca is the smallest, with 15,234 inhabitants. While the number of urban areas above 15,000 people increased from 2007 to 2019, the vast majority remained well above that threshold throughout the analysis period.<sup>3</sup>

Figure 1 depicts the core and fringe of the Lima metropolitan area in 2019. Lima exhibits high-density levels across the board; thus, the fringe constitutes only a modest share of its territory. Most development has mainly formed in the south and southeast areas, expanding the commuting shed. The figure portrays an up-to-date map of main roads, revealing how our delineation method accurately captures Lima's surface.

### *Household data*

Our primary data source is the Peruvian National Household Survey (ENAHO), developed annually by the National Institute of Statistics and Informatics (INEI). We combine annual survey waves from 2007 to 2019—each annual wave is a growing sample from 26,527 households in 2007 to 43,868 households in 2019.

ENAHO collects rich data on individual income and demographics, with detailed information on primary and secondary activities, including wages, hours worked, fringe benefits, type of contract, economic sector, and occupation. It also gathers detailed data on household expenses, which allows us to calculate the share of household income spent on shelter, food, and other goods.

One crucial advantage is that households are geocoded at the cluster level, where each cluster contains around 140 households. This granular location identifier allows us to accurately assign households to urban areas and determine how far they locate from the city center. Household heads also provide the hypothetical value of their home if they were to rent it in the market. This information helps assess how property values vary within cities and by their degree of centrality.

---

<sup>3</sup>The number of urban areas above 15,000 people increased from XX in 2007 to 92 in 2019. However, the number remained stable after 2013 when 90 urban areas were above the 15,000 threshold. We drop urban areas with few household observations in our estimations, often those under 20,000 people.

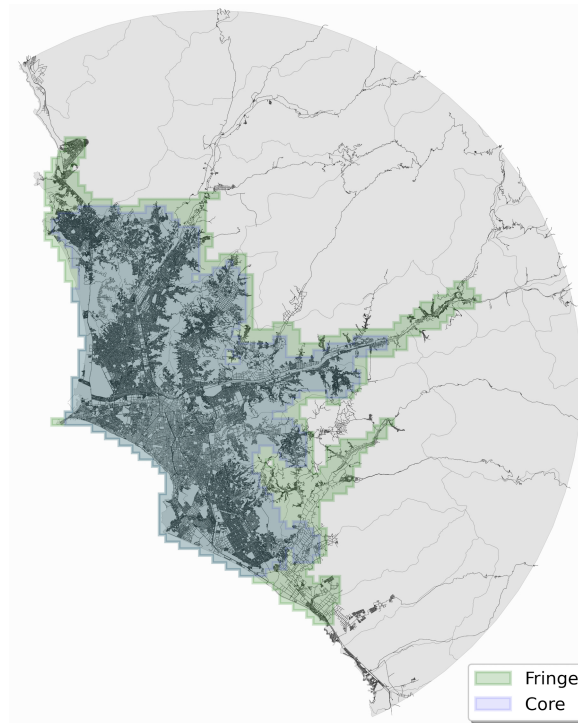


Figure 1: Lima metropolitan area, 2019

### *Sample selection*

The starting sample consists of a pooled cross-section of ENAHO annual waves from 2007 to 2019. There are 186,940 households and 735,507 individuals living in urban areas above 15,000 people. We restrict the sample to individuals between 16 and 70 years old and drop observations with missing values on age or gender. These restrictions reduce the sample to 180,024 households and 497,635 individuals.

Next, we keep only employed individuals, reducing the sample to 169,734 households and 348,212 individuals. We drop unpaid family and domestic workers since they do not report income or receive a sizable share of their income via in-kind transfers (e.g., food or shelter). We also eliminate employers since it is hard to disentangle their income from their business profits. Both restrictions reduce the sample to 161,338 households and 299,452 individuals.

We employ two samples for estimations. Our baseline sample excludes self-employed workers (115,786 observations) and workers in the military and the public sector (46,991 observations). The literature that estimates agglomeration effects focuses on workers employed in tradable services (e.g., manufacturing and business services), given that their nominal incomes largely reflect the marginal product of labor. Self-employed workers primarily earn higher incomes in larger cities because they benefit from higher prices (e.g., taxi drivers, street food sellers, or delivery workers). Similarly, we drop workers in the military and the public sector because their wages do not reflect agglomeration advantages; they are set at the national or regional level.

The final baseline sample includes 136,675 employees or manual workers in 95,778 households. We refer to the sample that includes self-employed workers and workers in the military and the

public sector as ‘extended sample.’ The following section will present workers’ socioeconomic characteristics in our baseline sample by skill. We turn first to describe our skill measures.

## *Skills*

Most studies that examine the distribution of skills across space in developed countries classify workers into high- and low-skilled based on college attainment (Moretti, 2004b; Berry and Glaeser, 2005; Diamond, 2016; Davis and Dingel, 2020). We first follow this standard skill classification (i.e., graduation from a 5-year university in the Peruvian case) and obtain that 18.9 percent of workers in the extended sample and 17.8 in the baseline sample are high-skilled. This is a first challenge for estimating how earnings vary by skill across cities since college graduates account for a smaller share of the labor force in developing countries. For instance, while the percentage of workers with college in US metropolitan areas was 32 percent in 2000, it was 15 percent in Brazil, 10 percent in China, and 21 percent in India (Dingel, Miscio, and Davis, 2021).<sup>4</sup> Thus, the researcher must estimate the college premium based on relatively few observations, especially in small cities.

We argue that using college attainment as a skill indicator in developing countries raises additional relevant concerns. First, college quality is highly uneven, particularly in the Peruvian context. Therefore, college attainment may be a highly imperfect measure of skills when universities in poorer and isolated regions lack infrastructure, have fewer resources and proficient professors, and fail to meet minimum quality standards.<sup>5</sup> Second, college availability is heavily biased toward larger cities. Thus many high-skilled young individuals from remote localities find it much harder to enroll in post-secondary institutions given financial constraints.<sup>6</sup> Third, household income and parental education may be more critical factors in the decision to attend college than high-school performance (Bacolod, De la Roca, and Ferreyra, 2021).<sup>7</sup> Altogether, given substantial spatial

---

<sup>4</sup>Dingel, Miscio, and Davis (2021) delineate metropolitan areas in the three developing countries using night lights and restrict the analysis to agglomerations above 100,000. The shares of college graduates in Brazilian, Chinese, and Indian cities correspond to 2010, 2000, and 2001, respectively. The Indian share may be slightly inflated since data on college attainment are not available for small constituencies of metropolitan areas.

<sup>5</sup>In 2014, the Peruvian government established SUNEDU, the National Superintendency of Higher University Education, to grant operating licenses to universities that met minimum quality standards. The licensing process revealed substantial disparities in quality across the 145 universities nationwide. Fifty universities did not meet minimum standards, most of them being private and poorly funded institutions. A ranking based on scientific output and research impact as inputs could only rate 45 universities with academic publications. Further, only 31 of those institutions had an H5 index above 4 in the Web of Science database, i.e., four publications were cited at least four times in other indexed articles between 2014 and 2018 (Superintendencia Nacional de Educación Superior Universitaria, 2020). Such institutions were largely concentrated in Lima (14) and a handful of cities—Arequipa and Trujillo with three institutions; Huancayo, Cusco, Piura, Cajamarca, Iquitos, Puno, Huaraz, Huancavelica, and Chachapoyas with only one. This spatial disparity in quality indicates that young individuals in most Peruvian cities cannot attend a public or private university that provides an education with adequate standards.

<sup>6</sup>Public universities are tuition-free, yet the admissions rates are substantially lower than for private institutions. While for-profit private universities admitted 80 percent of applicants in 2019, public universities admitted only 16 percent (Superintendencia Nacional de Educación Superior Universitaria, 2021). To get accepted into public universities, students often take an aptitude and knowledge test offered twice a year.

<sup>7</sup>Bacolod, De la Roca, and Ferreyra (2021) find that household income and parental education are much more important determinants for college mobility in Colombia than individual ability, proxied through standardized scores in a national aptitude test. Similarly, household income is the most important determinant of university enrollment and completion in Peru, followed by parental education. Students from the lowest consumption quintile were 28 percent less likely to attend college in 2018 than those from the top quintile (Superintendencia Nacional de Educación Superior Universitaria, 2020).

disparities in quality and accessibility, many talented young individuals cannot attend college. In contrast, others manage to enroll in one but study at a college of poor quality.

We propose formality status in the job market as an alternative indicator of skills. Hiring a formal worker is more expensive for employers, as they must pay social security and fringe benefits, so they incur in an additional cost only if the worker is sufficiently productive or skilled to compensate for the cost. This labor-market driven skill classification allows to classify some workers as high-skilled despite being high school graduates or as low-skilled even when they have acquired some post-secondary education.

Studies that use household survey data generally follow two criteria to classify workers as formal or informal based on their job or socioeconomic characteristics. The first one is the productive perspective which classifies informal workers as those with low productivity, often associated with unskilled self-employed workers, salaried workers in small firms (usually below five workers), or unpaid family workers. The second one is the legalistic or social protection perspective, which classifies informal workers based on the mandated rights and benefits that employers provide in their jobs.

We follow the legalistic perspective and consider three definitions for formality status that vary in their intensity. The *lax* formality definition classifies workers as formal if they are hired under a contract. The *moderate* definition labels workers as formal if they meet the lax criteria and receive social security benefits. Lastly, the *strict* definition categorizes a worker as formal if they meet the moderate definition, earn bonus payments, and get paid vacations. Therefore, the pool of high-skilled workers drops with the number of formal status requirements. In section 3, we present summary statistics for both skill definitions.

### *Urban amenities*

We consider a restricted set of amenities, namely those that are rival, excludable, and not free to consume. We argue that such amenities, conceived as local private goods that charge access fees, are a critical factor influencing skilled workers' location in developing countries. Given the absence of a welfare state that provides adequate standards of education, health, sanitation, safety, and other services, individuals must secure the private provision of several services.

One salient example is private education. The majority of high-skilled workers in Peru send their offspring to private schools since there is a widespread belief that they are of higher quality. To measure school quality, we use data from the Evaluación Censal de Estudiantes (ECE), an annual standardized math and verbal skills test administered by the Ministry of Education. The ECE measures the academic performance of most second-grade students, targeting all public and private schools that meet two criteria, having an enrollment of five second-grade students in a test year and using Spanish as the primary language of instruction. We pool student-level annual normalized scores from 2007 to 2016 and characterize high-quality schools as those in the top 10 percentile of the national test-score distribution.

The data reveal that private schools are overrepresented in the higher percentiles of the



distribution, validating the beliefs of high-skilled parents.<sup>8</sup> Crucially for our interests, the ECE geocodes all schools so that we can assign them to metropolitan areas. We find that 36 cities have at least one high-quality private school, as illustrated in panel (a) of figure 5. Appendix A describes the distribution of test scores and examines how the number of schools varies across cities of different sizes under different thresholds.

Another pertinent example of a private amenity is an upper-class neighborhood. Such neighborhoods are amenities desired by high-skilled workers since they provide parks and recreation, offer better sanitation services (e.g., garbage collection), and have lower crime levels.<sup>9</sup> We use data on census zones to identify the number of upper-class neighborhoods in each metropolitan area. Specifically, we define an upper-class neighborhood as a zone with a share of residents with college above 30.2 percent, the 90<sup>th</sup> percentile of the distribution across zones within urban areas.<sup>10</sup> We find that 35 cities have at least one upper-class neighborhood and only 13 of 63 cities with a population under 100,000 contain one. Panel (b) of figure 5 describes how the number of upper-class neighborhoods varies across cities, and appendix A provides further details.

### 3. Empirics

We begin describing our baseline sample of individuals by skill level, i.e., individuals with and without a college degree and individuals working in the formal or informal sector. Next, we present four stylized facts on the wage premium by skill level, rent city-size premium, and local private goods that will motivate our spatial equilibrium model.

#### *Summary statistics*

We present summary statistics by skill level in table 1. The first two columns classify workers by formality status in the extended sample, following a lax definition of formality. The last two columns classify workers by formality status in the baseline sample, according to a moderate definition of formality. Formal workers account for 39 percent in the extended sample and 37 percent in the baseline sample; hence, even after excluding workers in rural areas and domestic and unpaid family workers, informality as a share of the Peruvian labor force is considerable.

We notice an imperfect mapping between education and formality status. While around half of the formal sector workers in our baseline sample have attained a post-secondary degree, 37 percent are high-school graduates or dropouts. Similarly, whereas informal sector workers have, on average,

---

<sup>8</sup>However, public schools have on average higher test scores. Many private schools have very low scores and are thus overrepresented in the lower deciles of the distribution.

<sup>9</sup>Many Peruvian municipalities have recruited civilian personnel to surveil the streets and collaborate with the national police. Residents fund these surveillance teams (“Serenazgos”) via property taxes. In 2019, 55 percent of municipalities offered a Serenazgo service, though with marked differences in surveillance services and equipment that varied with income. For example, in some wealthy municipalities of Lima, such as San Isidro and Miraflores, the number of security residents per guard was 59 and 118, respectively. In contrast, the ratio varied between 2,900 and 5,600 residents in lower-income municipalities (Instituto Nacional de Estadística e Informática, 2019).

<sup>10</sup>Census zones cover the entire Peruvian territory. The 2017 Census lays out 6,399 zones in our set of urban areas—86 cities with a population greater than 15,000 people in 2019. Zones in urban areas have an average of 1,906 individuals. We exclude zones with fewer than 100 individuals and drop individuals under 25 to calculate the share of residents with a college degree.

Table 1: Summary statistics

	Extended sample		Baseline sample	
	Informal	Formal	Informal	Formal
Share	60.6%	39.4%	63.0%	37.0%
<u>Education</u>				
Primary	23.1%	3.8%	14.1%	3.9%
Incomplete secondary	17.2%	4.6%	15.8%	5.4%
Secondary	33.4%	21.4%	35.5%	27.5%
Incomplete tertiary	12.8%	15.1%	16.9%	16.7%
Technical degree	8.5%	20.5%	10.2%	20.0%
University degree	5.0%	34.6%	7.4%	26.6%
Years of education	9.8	13.6	10.8	13.0
<u>Individual characteristics</u>				
Female	46.6%	42.2%	37.4%	33.9%
Age	39.7	39.3	33.2	37.6
Married	56.4%	56.4%	42.4%	56.4%
<u>Household characteristics</u>				
Household size	3.93	4.09	4.13	4.21
Poverty status	17.7%	4.7%	17.9%	4.5%
Monthly rent (S/)	275.8	491.6	277.3	525.7
Rent / income	35.6%	25.4%	27.7%	25.3%
<u>Job characteristics</u>				
Hourly wage (S/)	25.3	51.5	26.2	50.0
Monthly income (S/)	827	2,000	984	2,191
Hours	41.9	44.8	43.7	49.4
Years of tenure	5.8	7.5	2.5	5.5
<u>Firm size</u>				
0 - 20 workers	95.3%	19.4%	78.9%	19.1%
21 - 50 workers	2.2%	8.3%	8.2%	13.1%
51 - 100 workers	0.8%	6.2%	3.8%	10.7%
101 - 500 workers	0.9%	11.3%	4.4%	21.5%
500+ workers	0.7%	54.8%	4.7%	35.7%
Median firm size	2.0	70.0	5.0	150.0
<u>Expenditures</u>				
Education, recreation, culture	5.5%	8.0%	5.7%	7.9%
Educ, recreat, culture + health	12.4%	15.9%	12.4%	15.9%

*Notes:* The first two columns classify workers in the extended sample following a lax definition of formality. The last two columns classify them in the baseline sample using a moderate formality definition. See the text for details on the formality definitions.

2.2 fewer years of education (10.8 vs. 13), around a third have attended some post-secondary studies and yet remain informal. We also note substantial income disparities since hourly wages for formal sector workers are twice as high (50 vs. 26.2 soles). The gap in monthly incomes amplifies as formal sector workers work six extra hours a week, have accumulated more job tenure, and work in

much larger firms. Lastly, higher incomes of formal sector workers result in lower shares of income spent on rent (27.7 vs. 25.3 percent) and a larger share spent on local non-tradable goods, such as education, recreation, culture, and health (15.9 vs. 12.4 percent).

***Fact 1: the city-size wage premium in Peru is slightly larger than in developed countries***

The city-size wage premium summarizes how wages vary across cities of different sizes. Extant studies for developed countries have estimated an elasticity of wages with respect to city size between 0.02 and 0.05, indicating pervasive agglomeration economies.<sup>11</sup> We follow Combes and Gobillon (2015) and estimate such elasticity in two stages.

First, let us assume that the log monthly wage of worker  $i$  in city  $c$  at time  $t$ ,  $w_{ict}$ , is given by:

$$w_{ict} = \sigma_c + \mathbf{x}_{it}'\beta + \varepsilon_{ict}, \quad (1)$$

where  $\sigma_c$  is a city fixed-effect,  $\mathbf{x}_{it}$  is a vector of individual and job characteristics, and  $\varepsilon_{ict}$  is an error term. Our set of controls includes gender, marital status, a non-Spanish language indicator, indicators for years of education, potential experience, tenure, indicators for 2-digit sectors and occupations, firm-size categories, and years. In the second stage, we regress:

$$\hat{\sigma}_c = \gamma \ln(\text{city size}) + \varepsilon_c, \quad (2)$$

where  $\gamma$  is the city-size wage premium or the elasticity of wages with respect to city size (a measure of population or experienced density).

Figure 2 plots the estimated city fixed effects against city size, proxied using population. We find marked spatial differences in wages even for observationally-equivalent workers. For instance, a worker in Lima earns around 39 percent more than one in Huaraz (106,000 people), and we note the largest wage differential of 60 percent between Lima and Huancavelica (32,000 people). We estimate an elasticity of wages with respect to city size of 0.0508. Therefore, doubling city size is associated with an approximate increase of 3.6 percent in wages ( $2^{0.0508} - 1 \approx 3.6$ ) after controlling for differences attributable to education, experience, tenure, occupation, or sector. The estimated elasticity for the extended sample is almost identical at 0.0516, indicating that self-employed and public sector workers do not alter the results.

Our elasticity is slightly above the range for developed countries and similar to the one Duranton (2016) estimates for Colombia (0.051) using household survey data. Other recent studies that use survey data for developing countries obtain elasticities in that range. Quintero and Roberts (2018) estimate elasticities between 0.025 and 0.090 across Latin American countries with a population above 15 million people and Özgüzel (2023) finds an elasticity of 0.057 for Turkey.<sup>12</sup> We have conducted several robustness estimations that resulted in the city-size wage premium in Peru

<sup>11</sup>See Combes and Gobillon (2015) and Ahlfeldt and Pietrostefani (2019) for reviews of available estimates.

<sup>12</sup>Quintero and Roberts (2018) estimate city-size wage premia of 0.046 for Argentina, 0.071 for Brazil, 0.028 for Chile, 0.025 for Colombia, 0.037 for Ecuador, 0.090 for Guatemala, 0.051 for Mexico, and 0.051 for Peru. Other researchers have estimated much larger elasticities for large developing countries, such as 0.077 for India (Chauvin, Glaeser, Ma, and Tobio, 2017), 0.099–0.192 for China (Combes, Démurger, and Li, 2015; Chauvin, Glaeser, Ma, and Tobio, 2017), and 0.202–0.295 for Java-Bali, Indonesia (Bosker, Park, and Roberts, 2021).

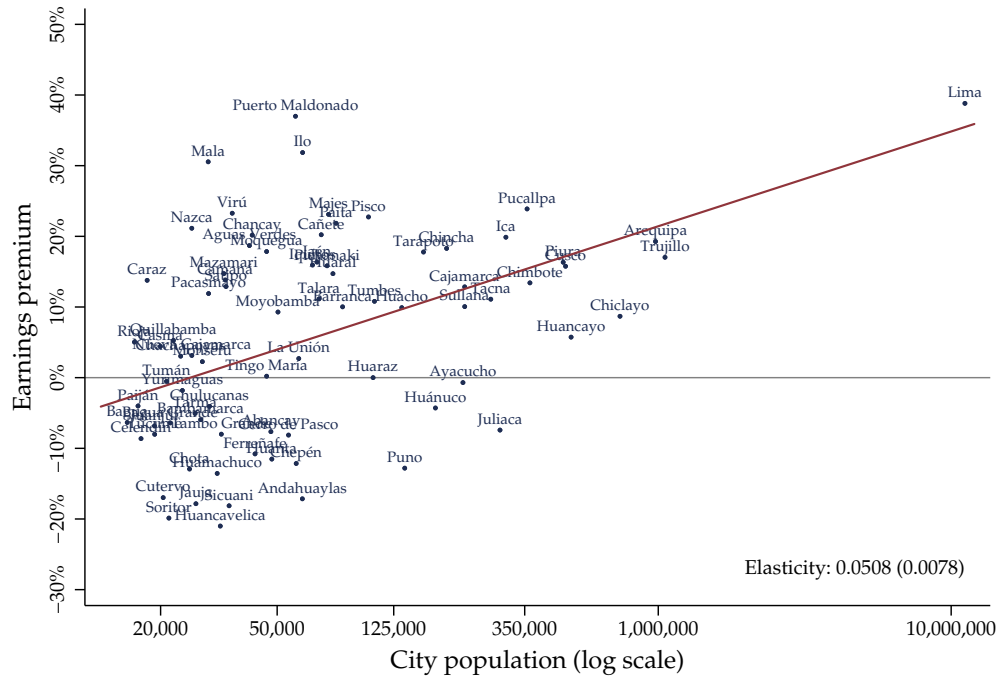


Figure 2: City-size wage premia, baseline sample

varying from 0.0437 to 0.0516 depending on the income definition, sample selection, and city size measure.<sup>13</sup>

The literature has emphasized several concerns regarding the estimation of the city-size wage premium. First, individuals with higher skills may sort into larger cities (Combes, Duranton, and Gobillon, 2008). The usual solution is introducing worker fixed effects in equation (1) to control for time-invariant worker characteristics. However, we cannot follow that strategy since we lack a large panel that follows workers. Instead, we control for a large set of individual and job characteristics to partially alleviate sorting concerns. Second, larger cities may provide benefits that accumulate over time as individuals acquire more valuable experience (De la Roca and Puga, 2017). We rely on a short ENAHO panel to test for such dynamic effects and do not find evidence of steeper wage profiles for workers in Lima or the next three largest cities.<sup>14</sup> Lastly, unobserved city characteristics (e.g., favorable business climate) correlated with size may drive the positive association we estimate in

<sup>13</sup>The city-size wage premium drops to 0.0437 when we use log hourly income as the outcome variable. The elasticity of hours worked with respect to city size is 0.0072 and 0.0121 in the baseline and extended samples, respectively, yet only statistically significant at the ten percent level in the latter. Thus we find suggestive evidence that individuals work more hours in larger cities. Further, we use wages for the primary job (including the value of in-kind payments) in our estimation; however, the elasticity remains unaltered when considering total income (primary and secondary activities).

<sup>14</sup>The ENAHO panel interviews respondents living in the same dwelling up to five consecutive years. Unfortunately, we are unable to follow workers when they move dwellings. However, we can include worker fixed effects in equation (1) and estimate how wages vary over time for individuals working across cities of different sizes. We do not find evidence of dynamic effects since the value of the experience acquired in Lima, or in the next three largest cities (Arequipa, Trujillo, and Chiclayo), is not statistically significantly different from the value of overall experience.

equation (2). We instrument city size using city population in 1940 and obtain a similar elasticity.<sup>15</sup>

***Fact 2: the city-size wage premium is lower for high-skilled workers***

We obtain the city-size wage premium by skill level estimating the following first and second stage regressions:

$$\begin{aligned}
 w_{ict} &= \sigma_c + \sigma_c \times \mathbb{1}(\text{high skills}) + \mathbb{1}(\text{high skills}) + \mathbf{x}'_{it}\beta + \varepsilon_{ict}, & (3) \\
 \hat{\sigma}_c &= \gamma_{\text{low skills}} \ln(\text{city size}) + \varepsilon_c, \\
 \hat{\sigma}_c + \hat{\sigma}_c \times \mathbb{1}(\text{high skills}) &= \gamma_{\text{high skills}} \ln(\text{city size}) + \varepsilon_c,
 \end{aligned}$$

where  $\mathbb{1}$  is an indicator function.

Using the moderate definition of formality, we first categorize skills by formality status (i.e., a worker is formal if they have a job contract and receives social security benefits). Figure 3 shows the joint estimation of the city-size wage premium and displays the second stage of equation (3), again plotting the estimated city fixed effects for both skill types on the city population. While formal sector workers earn more than informal sector workers across the board, the formal workers' city-size wage premium is much flatter. For instance, relative to an observationally-equivalent informal sector worker in Huancavelica, informal sector workers in Huaraz and Lima earn 20 and 70 percent more, respectively. In turn, formal sector workers in Huancavelica, Huaraz, and Lima earn 72, 87, and 108 percent more, reflecting a substantial wage gap with informal sector workers that, however, shrinks with city size.

We estimate a low elasticity of wages with respect to city size of 0.0161 for formal sector workers, much smaller than for their informal sector counterparts (0.0552). The difference in elasticities using other formality definitions are pretty similar. When we follow the lax formality definition to classify workers, we obtain an elasticity for formal sector workers of 0.0239 and one of 0.0584 for informal sector workers. Further, when using the strict formality definition, the informal sector elasticity is higher by a factor of 3.7 (0.0149 vs. 0.055).<sup>16</sup>

We confirm our findings on the skill disparities in the city-size wage premia when using college attainment to measure skills. We obtain an elasticity of 0.036 for college graduates and a larger elasticity of 0.054 for non-college graduates.<sup>17</sup> We also note that the overall wage gap (adjusted for workers' characteristics) between workers with and without college is less salient, suggesting that college is a more noisy signal to proxy for workers' productivity. Our much larger elasticities

<sup>15</sup>Reverse causality may also drive the positive association between wages and city size. More productive or prosperous places may attract residents and, in turn, increase city size. Typical instruments use long-lagged population measures or soil characteristics to predict current city size (Ciccone and Hall, 1996; Combes, Duranton, Gobillon, and Roux, 2010). Cities that were large in the past due to favorable agriculture or historical relevance are likely to remain large today; however, the productivity drivers that led them to grow back in time are less fundamental today.

<sup>16</sup>We observe relatively few formal workers in small cities under a strict formality definition. While the estimated elasticity is low at 0.0149 and statistically significantly different from zero at the ten percent level, we take this result cautiously since formal sector workers make up only 29 percent of the sample.

<sup>17</sup>The results hold in the extended sample. The city-size wage premium for formal workers is much lower at 0.0164 than for informal sector workers at 0.0628. Similarly, the elasticity for workers with college is lower than for non-college graduates (0.026 vs. 0.054).

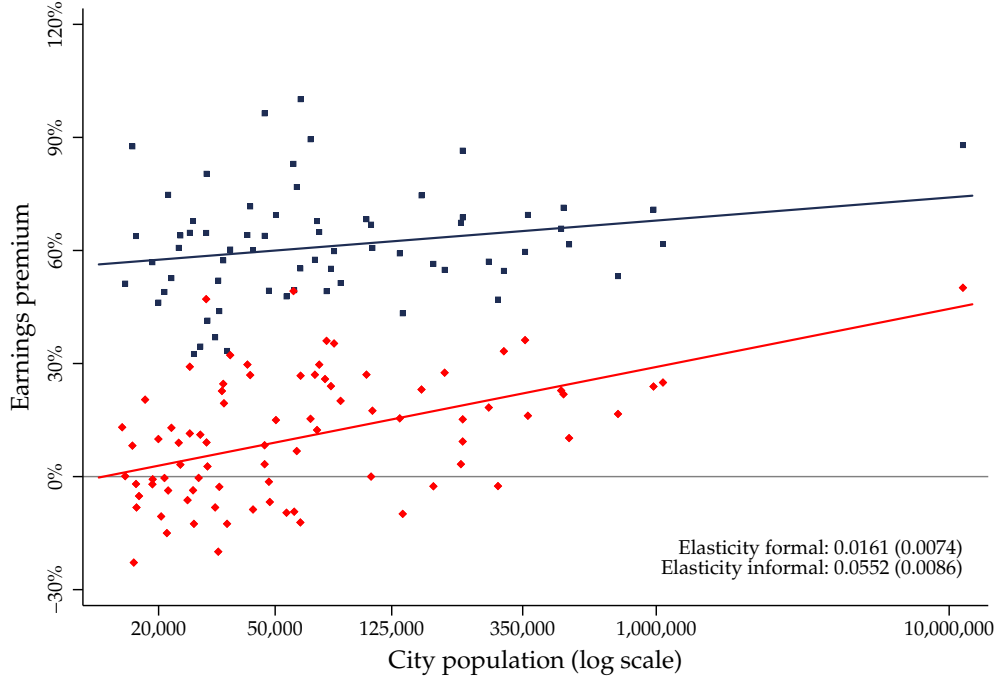


Figure 3: City-size wage premia, formal and informal workers

for informal sector workers are consistent with the findings of Duranton (2016) in Colombia and Quintero and Roberts (2018) for most big Latin American countries.

**Fact 3: urban costs increase with city size for high- and low-skilled workers**

Housing and transportation represent the core of urban costs. Urban intra-city location models characterize a tradeoff between land values and transportation costs. Higher accessibility near the central business district (CBD) results in higher land values. Accordingly, to assess urban costs across cities of different sizes, we compare land values precisely at the CBD, where households' transportation costs are minimized (Combes, Duranton, and Gobillon, 2019).

Let us assume that the log of the monthly rent of household  $h$  in city  $c$  at time  $t$  is given by:

$$r_{hct} = \eta_c - \mathbf{f}(\mathbf{d}_{hct})' \delta + \mathbf{x}'_{ht} \theta + \varepsilon_{ict} \quad (4)$$

where  $r_{hct}$  are self-reported values of hypothetical market rents and  $\mathbf{f}(\mathbf{d}_{hct})$  is a polynomial of distance to the CBD.<sup>18</sup> We designate the "Plaza de Armas" (Main Square) as the CBD following the Spanish foundation of most cities. Our set of controls includes the number of rooms, indicator variables on whether the house has property title, access to a water pipe or drainpipe, and indicators for types of materials on external walls, floors, and roofs.

Since our goal is to estimate rental values at the CBD, we estimate equation (4) and predict rents using sample means for all covariates and a distance value of zero, i.e., we obtain the (hypothetical)

<sup>18</sup>Households report the hypothetical value of their home if they were to rent on the market. Unfortunately, we lack information on land values and must rely on self-reported rents. We restrict the analysis to single-family homes.







city  $j$  is given by  $N_{sj}$ .

Workers live one period, and consume a traded consumption good  $c_g$  (numeraire) and housing  $c_h$ . Workers also receive utility from local skill-specific amenities  $X_{sj}$ . These amenities represent either a non-rival and/or non-excludable public good (e.g. public schools, roads, public parks, etc.) or natural amenities, such as climate, ocean views, etc. The amenities are assumed to be exogenous.

The worker's utility is represented by  $u(c_g, c_h)X_{sj}$ , where  $u$  is an increasing function in both arguments. A household who has chosen to reside in location  $j$  chooses  $c_g$  and  $c_h$  to maximize  $u(c_g, c_h)$  subject to the budget constraint  $c_g + rc_h \leq w$ , where  $r$  is the housing rent. The indirect utility function is then given by  $v(w, r)$ .

### *Production of the numeraire*

In each city, there is a representative firm which combines the two types of labor in a CES production function in order to produce the numeraire consumption good,

$$Y_j = \tilde{N}_j,$$

where

$$\tilde{N}_j \equiv \left[ (A_{Lj}N_{Lj})^\theta + (A_{Hj}N_{Hj})^\theta \right]^{\frac{1}{\theta}}$$

is local labor supply in efficiency units. The elasticity of substitution between low and high skills is  $1/(1 - \theta)$ . Wages are equal to the marginal product of each type of labor,

$$w_{sj} = A_{sj} \left( 1 + \left( \frac{A_{kj}N_{kj}}{A_{sj}N_{sj}} \right)^\theta \right)^{\frac{1-\theta}{\theta}}, \quad (5)$$

where  $k = H$  if  $s = L$  and  $k = L$  if  $s = H$ . The equilibrium wage of type  $s$  depends on local productivity of each type of labor as well as on the relative supply of skill  $s$ .

### *Production of housing*

Housing is produced using land ( $L_j$ ) and non-land inputs ( $K_{hj}$ ) with a Cobb-Douglas production function, following Combes, Duranton, and Gobillon (2019):

$$H_j = \chi_j L_j^\eta K_{hj}^{1-\eta}.$$

We assume that the numeraire good can be converted into the non-land input at no cost. As a result, the price of the non-land input is equal to 1 in all locations. The total exogenous supply of land in the city is given by  $\Lambda_j$  and is owned by absentee landlords. There is no alternative use of land, therefore landlords will sell it at any positive price. As a result, the optimal land input is  $L_j = \Lambda_j$ . The equilibrium rent in location  $j$  is equal to

$$r_j = \frac{1}{\chi_j} \left( \frac{\Lambda_j}{\eta} \right)^\eta \left( \frac{1}{1-\eta} \right)^{1-\eta}, \quad (6)$$

where  $l_j$  is the local price of land equal to

$$l_j = \frac{\eta r_j}{\Lambda_j} \sum_{s \in \{L, H\}} c_h(w_{sj}, r_j) N_{sj}. \quad (7)$$

### Consumption and location choice

For individual  $n$  of skill  $s$ , the indirect utility of residing in location  $j$  is

$$V_{sjn}(w_{sj}, r_j, X_j) = v(w_{sj}, r_j) X_j + \sigma \varepsilon_{sjn},$$

where  $\varepsilon_{sjn}$  is the location preference shock which follows the standard Extreme Value Type I distribution and  $\sigma > 0$  is the scale parameter.

Workers choose location at the beginning of a period, taking into account wages, prices, amenities, as well as an idiosyncratic location preference shock. As a result, the equilibrium supply of type- $s$  labor in location  $j$  is

$$N_{sj} = \frac{\exp [v(w_{sj}, r_j) X_{sj}]^{\frac{1}{\sigma}}}{\sum_{k \in \mathcal{J}} \exp [v(w_{sk}, r_k) X_{sk}]^{\frac{1}{\sigma}}} \times N_s, \quad (8)$$

where  $N_s$  is the exogenous total supply of workers of type  $f$  in the economy.

### Spatial equilibrium and city-size wage premia

The following definition characterizes a spatial equilibrium in this model.

**Definition 1 (spatial equilibrium)** *A spatial equilibrium consists of local wages  $w_{sj}$ , rents  $r_j$ , land prices  $l_j$ , and labor supply of each type  $N_{sj}$ , such that equations (5), (6), (7), and (8) are satisfied.*

What is the relationship between city size and wages for low- and high-skilled workers? We first make simplifying assumptions on the relationship between city size and skill-specific labor productivity and amenities. We will relax these assumptions in our quantitative model. In particular, we assume that  $A_{sj}$  is a monotonic function of city size and that

$$\frac{dA_{Lj}}{dN_j} > \kappa_A \frac{dA_{Hj}}{dN_j}, \quad (9)$$

where  $\kappa_A > 0$  is a constant. We also assume that  $X_{sj}$  is a monotonic function of city size and that

$$\frac{dX_{Hj}}{dN_j} > \kappa_X \frac{dX_{Lj}}{dN_j}, \quad (10)$$

where  $\kappa_X > 0$  is a constant. Then in proposition 1, we highlight necessary and sufficient conditions that result in a steeper relationship between wages and city size for low-skilled workers, as observed in the data and documented in section 3.

**Proposition 1 (city-size wage premia)** *Let  $\kappa_A > \Omega_A$  or  $\kappa_X > \Omega_X$ .<sup>19</sup> Then the city-size wage premium is larger for low-skilled workers, i.e.,*

$$\frac{dw_{Lj}}{dN_j} > \frac{dw_{Hj}}{dN_j}.$$

**Proof** See appendix section B. □

<sup>19</sup>Please see the proof for explicit expressions for  $\Omega_A$  and  $\Omega_X$ .

### *Model with endogenous supply of private amenities*

So far we have assumed that local amenities  $X_j$  are an exogenous attribute of a city freely available to all local residents. Some amenities, such as weather or beautiful views, are indeed exogenous. In practice, however, many goods and services that we call “amenities,” such as schools, grocery stores, or crime, depend on local population size and its composition (Diamond, 2016; Almagro and Domínguez-Iino, 2021). Moreover, not all amenities are non-rival public goods. Restaurants, private schools, and middle-class neighborhoods are characterized by limited supply and are not free to consume.

In what follows, we extend the model along three dimensions. First, to account for endogeneity as well as rival and excludable nature of some amenities, we add a local amenity good  $c_a$  to the consumer’s choice set and introduce firms that produce the good. Second, to account for the fact that the consumption share of local amenity goods differs by income (see table ??), we allow preferences to be non-homothetic. Third, we allow labor productivity to depend on the city size. We keep using  $X_{sj}$  to represent *exogenous* local amenities.

#### *Private amenity goods and non-homothetic preferences*

Now, on top of consuming the tradable good and housing, a worker consumes a local amenity good  $c_a$ . The utility function is

$$u(c_g, c_h, c_a) = c_g^{\omega_g} (c_h - \bar{c}_h)^{\omega_h} (c_a + \bar{c}_a)^{\omega_a}, \quad (11)$$

where  $\bar{c}_h > 0$ , that is a household must consume a minimum positive amount of housing, and  $\bar{c}_a > 0$ , that is a household would only choose positive consumption  $c_a$  if  $c_g$  and  $c_h$  are high enough.<sup>20</sup> That is, we model the private amenity good as a luxury good and housing as a necessity good. We assume homotheticity of degree 1, i.e.,  $\omega_g + \omega_h + \omega_a = 1$ .

Since private amenities are a luxury good, the functional form of the indirect utility depends on whether household income is high enough to afford it. This gives rise to two cases. If household income is sufficiently high to ensure  $c_a > 0$ , then the indirect utility is

$$v(w, p, r) = \omega_g^{\omega_g} \omega_h^{\omega_h} \omega_a^{\omega_a} (w + p\bar{c}_a - r\bar{c}_h) r^{-\omega_h} p^{-\omega_a}.$$

If income is low, it is optimal not to consume the amenities, i.e.,  $c_a = 0$ , and the indirect utility is

$$v(w, p, r) = \frac{\omega_g^{\omega_g} \omega_h^{\omega_h}}{(\omega_g + \omega_h)^{\omega_g + \omega_h}} \left( (w - r\bar{c}_h)^{\omega_g + \omega_h} \bar{c}_a^{\omega_a} \right) r^{-\omega_h}.$$

#### *Supply of private amenity goods*

In each city, there is a representative firm that produces private amenities. The production function is linear in input  $K_a$ , which is produced with a one-to-one technology using the traded consumption good, similarly to  $K_h$  in the production function for housing. Therefore, since the traded good is a

<sup>20</sup>This functional form is common in the literature on structural transformation. See, among others, Herrendorf, Rogerson, and Valentinyi (2013).

numeraire, the price of  $K_a$  is 1 in all locations. The production function is

$$Q_j = B_j K_{aj}, \quad (12)$$

where  $B_j$  is the productivity term. The production of private amenities is subject to fixed cost  $\phi \geq 0$ . Hence, the profit is

$$p_j Q_j - K_{aj} - \phi,$$

where  $p_j$  is the equilibrium price of private amenities in city  $j$ .

#### *Demand for private amenity goods*

Let  $\Psi(p_j, r_j)$  denote the minimal level of income such that a household optimally chooses to consume a positive quantity of private amenities. It is straightforward to show that this threshold is

$$\Psi(p_j, r_j) = \frac{1 - \omega_a}{\omega_a} p_j \bar{c}_a + r_j \bar{c}_h.$$

Local demand for private amenities depends on the extensive margin (whether income of a given group is high enough to buy the amenities) and the intensive margin (the level of income, in case it is above the threshold). The demand can be represented as a function that encompasses three cases:

$$Q_j^d = \begin{cases} \omega_a (w_{Lj} N_{Lj} + w_{Hj} N_{Hj}) - (\omega_a r_j \bar{c}_h + (1 - \omega_a) p_j \bar{c}_a) N_j & \text{if } w_{Lj} > \Psi(p_j, r_j), w_{Hj} > \Psi(p_j, r_j) \\ \omega_a w_{sj} N_{sj} - (\omega_a r_j \bar{c}_h + (1 - \omega_a) p_j \bar{c}_a) N_{sj} & \text{if } w_{sj} > \Psi(p_j, r_j), w_{kj} \leq \Psi(p_j, r_j) \\ 0 & \text{if } w_{Lj} \leq \Psi(p_j, r_j), w_{Hj} \leq \Psi(p_j, r_j) \end{cases} \quad (13)$$

In the first case, both high- and low-skilled workers have enough income to purchase the amenity goods. In the second case, only workers of skill  $s$  do. In the third case, neither high-skilled nor low-skilled workers have sufficient income.

#### *Market clearing*

Recall that the production of local private amenities requires paying the fixed cost  $\phi$ . The fixed cost represents the idea that many types of amenities are provided by relatively large establishments, e.g., schools, theaters, or shopping malls. If local demand for amenities is low, the fixed cost may be too high and local firms may not be able to supply them without incurring losses. In this case, the supply of local amenities is zero and the market does not clear. The cities where the amenities are supplied are the cities with sufficiently large populations that have sufficiently high incomes. Moreover, all else equal, higher demand for local amenities will reduce its prices. The equilibrium price is given by a decreasing function,

$$p_j(Q_j^d), \quad (14)$$

defined explicitly in appendix section B.

Table 2: Parameter values

Parameter	Value	Source of target	Moment	
			Model	Data
<b>Internally calibrated parameters</b>				
Exogenous city-skill taste shifter	$X_{sj}$	Labor supply by city-skill	dist < 0.1%	
Exogenous city-skill productivity	$\bar{A}_{sj}$	Wage by city-skill	dist < 0.1%	
Construction sector TFP	$\chi_j$	Rents by city	dist < 0.1%	
Housing cons. share parameter	$\omega_h = 0.143$	Housing cons. share	0.267	0.267
(2) non-homotheticity parameter	$\bar{c}_h = -0.095$	Share difference, H-L	-0.024	-0.024
Amenity cons. share parameter	$\omega_a = 0.258$	Amenity cons. share	0.138	0.138
(2) non-homotheticity parameter	$\bar{c}_a = 0.135$	Share difference, H-L	0.035	0.035
Fixed cost of amenities	$\phi = 1.74 \times 10^{-4}$	# cities with amenities	36 of 66	36 of 66
<b>Externally calibrated parameters</b>				
Elast. of substitution, $1/(1 - \theta)$	$\theta = 0.6$	Card (2009)		
Elasticity of labor supply, $1/\sigma$	$\sigma = 0.35$	Hornbeck and Moretti (2018)		
Agglomeration externality	$\rho = 0.05$	standard value		
Land share	$\eta = 0.2$	Combes, Duranton, and Gobillon (2019)		
TFP of amenity production	$B_j = 1$ for all $j$			

### Agglomeration externalities

We assume that productivity of labor of type  $s$  depends on an exogenous city-type term  $\bar{A}_{sj}$  as well as total labor supply,

$$A_{sj} = \bar{A}_{sj} N_j^\rho,$$

where  $\rho$  measures the strength of the production agglomeration externality.

### Equilibrium and its properties

The following definition characterizes a spatial equilibrium in the extended version of the model.

**Definition 2 (spatial equilibrium)** *A spatial equilibrium consists of local wages  $w_{sj}$ , rents  $r_j$ , land prices  $l_j$ , labor supply of each type  $N_{sj}$ , as well as private amenity prices  $p_j$ , such that equations (5), (6), (7), and (8) are satisfied, and the supply of private amenities (12) is either equal to the demand (13) or is zero.*

## 5. Quantification

Next we build a quantitative version of the model. The model economy consists of 66 metropolitan areas in Peru. Table 2 summarizes parameter values used to construct the quantitative version of the model.

There are several vectors of location-specific parameters that we calibrate to match location-specific moments. In particular, we set the values for the exogenous city-type amenity shifter  $X_{sj}$  by ensuring that employment of each skill type in each metro area is the same in the model and in the data. We calibrate the exogenous city-skill labor productivity  $\bar{A}_{sj}$  to match the observed wages by

location and skill. Finally, we calibrate the TFP in the construction sector  $\chi_j$  to match the observed housing rents in each city.

There are also several economy-wide parameters that we calibrate internally. First, there are four parameters in the utility function that determine consumption shares of each type of good. Recall that the utility function is non-homothetic (equation 11), and therefore expenditure shares will depend on income. This specification allows us to calibrate  $\omega_h$  to match the aggregate housing consumption share and  $\bar{c}_h$  to match the difference in the housing consumption share between formal and informal workers. Similarly, we calibrate  $\omega_a$  and  $\bar{c}_a$  to match the aggregate amenity consumption share and the share difference between low- and high-skilled workers.

We calibrate  $\phi$ , the fixed cost of providing consumption amenities, to the number of cities that supply the private amenity good. As a proxy, we use the number of private schools that are top-10 nationally in math and language. If a given city has at least one top-10 school in math and one top-10 school in language, then we say that the city supplies the private amenity good. Otherwise it does not. We find that, according to this definition, 36 out of 66 cities supply the private amenity good.

Finally, we take several parameters from the literature. We use the elasticity of substitution between high- and low-skilled workers of 2.5, as estimated by Card (2009). This implies that  $\theta = 0.6$ . The scale parameter of the Extreme Value Type I distribution determines the (inverse) elasticity of local labor supply. We use  $\sigma = 0.35$ , following Hornbeck and Moretti (2018). The agglomeration externality is set to  $\rho = 0.05$ , a standard value in the literature (Combes and Gobillon, 2015). The land share in housing construction is 0.2, following the estimates in Combes, Duranton, and Gobillon (2019). Finally, since we do not observe prices of local amenity goods, we assume that the TFP of amenity-producing firms is the same in all locations and, without loss of generality, we normalize this productivity to 1.

## 6. Counterfactual experiments

One possible explanation why low-skilled workers enjoy a higher city-size wage premium than high-skilled workers is that amenities make up a larger share of high-skilled workers' consumption basket. Since small cities do not provide amenities, high-skilled workers desire to live in medium and large cities exceeds that of low-skilled workers. As the relative supply of high-skilled workers in large cities goes up, their wages go down (see equation 5) and the city-size wage premium for high-skilled workers falls. To put it differently, high-skilled workers are willing to accept somewhat lower wages in large cities in exchange for the possibility of consuming amenities.

To evaluate the role of amenity provision in generating the difference between city-size wage premia for high- and low-skilled workers, we run a counterfactual experiment in which we eliminate the fixed cost by setting  $\phi = 0$ . In this scenario, amenities are supplied in all cities and high-skilled workers do not have to avoid small cities in order to be able to consume amenities. As a result, all workers reallocate from large to small cities. However, since high-skilled workers spend a larger share of their budget on the amenities, their migration responds more in the counterfactual scenario. Figure 6 demonstrates this result.

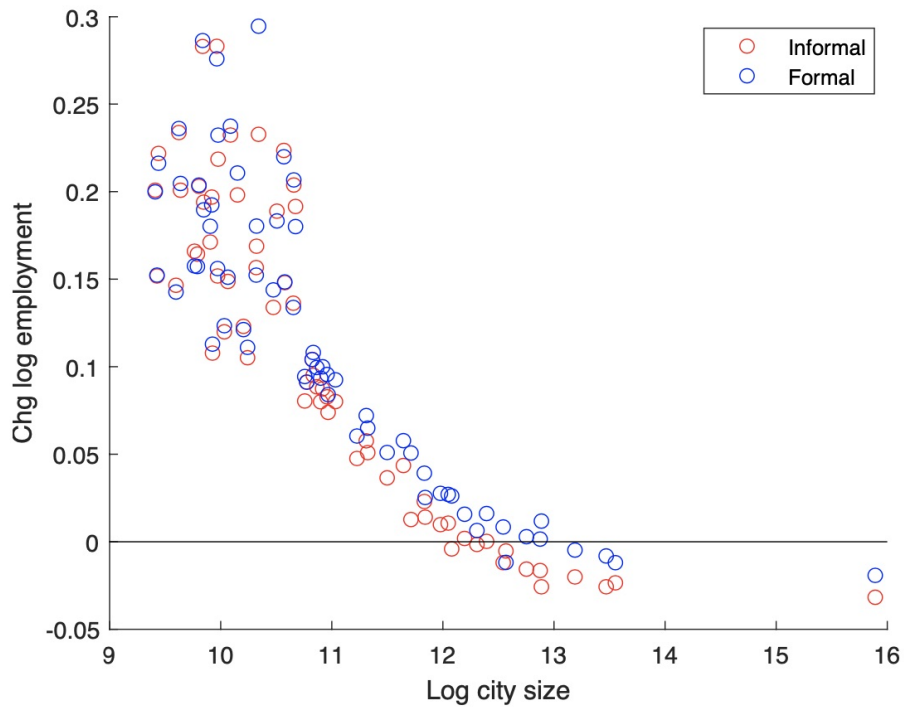
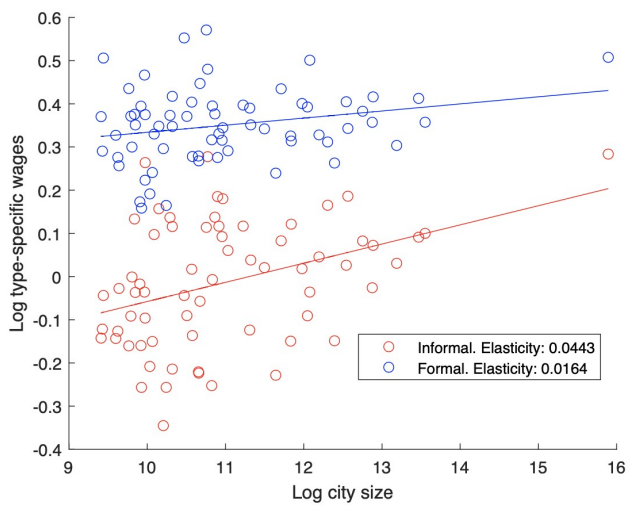


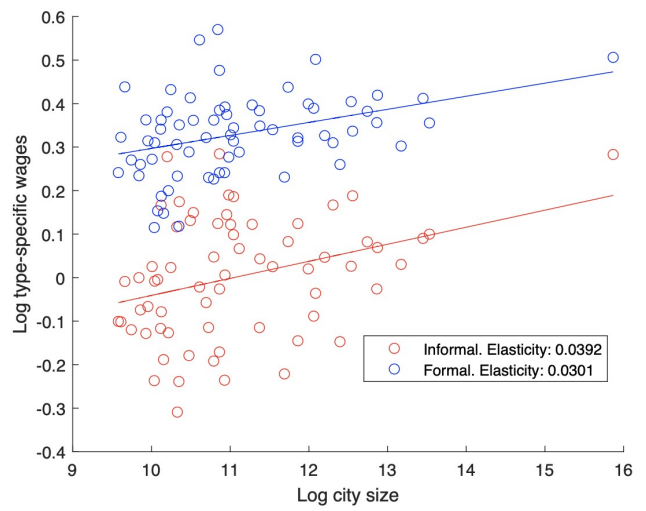
Figure 6: Local labor supply, benchmark vs. counterfactual

Consequently, the relative supply of high-skilled workers in large cities falls and their wages increase there. At the same time, the relative supply of high-skilled workers in small cities goes up which lowers their wages there. As a result, as Figure 7 shows, the gap in city-size wage premia between low- and high-skilled workers shrinks by 67%. In other words, our model accounts for 2/3 of the observed gap in city-size wage premia.<sup>21</sup>

<sup>21</sup>The observed gap between the elasticities is  $0.0443 - 0.0164 = 0.0279$  percentage points. In the counterfactual scenario, the gap shrinks to  $0.0392 - 0.0301 = 0.0091$  which is equal to  $0.0091/0.0279 = 33\%$  of the observed gap. Therefore, the model accounts for 67% of the observed gap in city-size wage premia between low- and high-skilled workers.



Panel (a): benchmark



Panel (b): counterfactual

Figure 7: Wages, benchmark vs. counterfactual



## Appendix A. Urban amenities

To be completed.

## Appendix B. Proofs

### *Proof of Proposition 1*

TBC.

### *Equilibrium Prices of Private Amenities*

Therefore, combining profit maximization of local-good firms and the market-clearing condition, we obtain the following price of the local good:

$$p_j(Q_j^d) = \begin{cases} \frac{-p_{2j} - \sqrt{p_{2j}^2 - 4p_{1j}p_{3j}}}{2p_{1j}}, & \text{if } p_j Q_j^d > \phi \\ \notin \mathbb{R}, & \text{if } p_j Q_j^d \leq \phi. \end{cases}$$

In other words, if the demand for the local good is high enough so that the local-good producer can cover the fixed cost and still earn non-negative profit, then the price is given by the first expression above. Otherwise, if the demand is too low to cover the fixed cost, amenity producers do not operate and the market does not clear.

The expressions for  $p_{1j}$ ,  $p_{2j}$ , and  $p_{3j}$  depend on the levels of  $w_{Ij}$  and  $w_{Fj}$ :

1. If  $w_{Ij} > \Psi(p_j, r_j)$ ,  $w_{Fj} > \Psi(p_j, r_j)$ , then

$$\begin{aligned} p_{1j} &\equiv -(1 - \omega_a)\bar{c}_a N_j \\ p_{2j} &\equiv \omega_a (w_{Ij} - r_j \bar{c}_h) N_{Ij} + \omega_a (w_{Fj} - r_j \bar{c}_h) N_{Fj} + (1 - \omega_a)\bar{c}_a \frac{N_j}{B_j} \\ p_{3j} &\equiv -\frac{\omega_a (w_{Ij} - r_j \bar{c}_h) N_{Ij} + \omega_a (w_{Fj} - r_j \bar{c}_h) N_{Fj}}{B_j} - \phi \end{aligned}$$

2. If  $w_{Ij} \leq \Psi(p_j, r_j)$ ,  $w_{Fj} > \Psi(p_j, r_j)$ , then

$$\begin{aligned} p_{1j} &\equiv -(1 - \omega_a)\bar{c}_a N_{Fj} \\ p_{2j} &\equiv \omega_a (w_{Fj} - r_j \bar{c}_h) N_{Fj} + (1 - \omega_a)\bar{c}_a \frac{N_{Fj}}{B_j} \\ p_{3j} &\equiv -\frac{\omega_a (w_{Fj} - r_j \bar{c}_h) N_{Fj}}{B_j} - \phi \end{aligned}$$

Note that if  $w_{Ij} \leq \Psi(p_j, r_j)$ ,  $w_{Fj} \leq \Psi(p_j, r_j)$ , then  $Q_j^d = 0$  and a market-clearing price does not exist.

## References

- Ahlfeldt, Gabriel M. and Elisabetta Pietrostefani. 2019. The economic effects of density: A synthesis. *Journal of Urban Economics* 111: 93–107.
- Almagro, Milena and Tomás Domínguez-Iino. 2021. Location sorting and endogenous amenities: Evidence from Amsterdam. Preprint, University of Chicago.
- Bacolod, Marigee, Jorge De la Roca, and María Marta Ferreyra. 2021. In search of better opportunities: Sorting and agglomeration effects among young college graduates in Colombia. *Regional Science and Urban Economics* 87: 103656.
- Berry, Christopher R. and Edward L. Glaeser. 2005. The divergence of human capital levels across cities. *Papers in Regional Science* 84(3): 407–444.
- Bosker, Maarten, Jane Park, and Mark Roberts. 2021. Definition matters. Metropolitan areas and agglomeration economies in a large-developing country. *Journal of Urban Economics* 125: 103275.
- Card, David. 2009. Immigration and inequality. *American Economic Review* 99(2): 1–21.
- Chauvin, Juan Pablo, Edward L. Glaeser, Yueran Ma, and Kristina Tobio. 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics* 98: 17–49.
- Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1): 54–70.
- Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li. 2015. Migration externalities in Chinese cities. *European Economic Review* 76: 152–167.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2): 723–742.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2019. The costs of agglomeration: House and land prices in French cities. *Review of Economic Studies* 86(4): 1556–1589.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *Agglomeration Economics*. Chicago, IL: Chicago University Press, 15–65.
- Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In Gilles Duranton, J. Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5. Amsterdam: Elsevier, 247–348.
- Davis, Donald R. and Jonathan I. Dingel. 2020. The comparative advantage of cities. *Journal of International Economics* 123: 103291.
- Davis, Donald R., Eric Mengus, and Tomasz Michalski. 2021. Labor market polarization and the great urban divergence. Working Paper 26955, National Bureau of Economic Research.
- De la Roca, Jorge and Diego Puga. 2017. Learning by working in big cities. *Review of Economic Studies* 84(1): 106–142.
- Diamond, Rebecca. 2016. The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *American Economic Review* 106(3): 479–524.

- Dijkstra, Lewis and Hugo Poelman. 2014. A harmonised definition of cities and rural areas: the new degree of urbanisation. Regional Working Paper 1, European Commission Directorate-General for Regional and Urban Policy.
- Dingel, Jonathan I., Antonio Miscio, and Donald R. Davis. 2021. Cities, lights, and skills in developing economies. *Journal of Urban Economics* 125: 103174.
- Duranton, Gilles. 2015. A proposal to delineate metropolitan areas in Colombia. *Revista Desarrollo y Sociedad* 75: 223–264.
- Duranton, Gilles. 2016. Agglomeration effects in Colombia. *Journal of Regional Science* 56(2): 210–238.
- Duranton, Gilles and Diego Puga. 2020. The economics of urban density. *Journal of Economic Perspectives* 34(3): 3–26.
- Glaeser, Edward L. and Matthew G. Resseger. 2010. The complementarity between cities and skills. *Journal of Regional Science* 50(1): 221–244.
- Henderson, J. Vernon, Dzhamilya Nigmatulina, and Sebastian Kriticos. 2019. Measuring urban economic density. *Journal of Urban Economics* : 103188.
- Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2013. Two perspectives on preferences and structural transformation. *American Economic Review* 103(7): 2752–89.
- Hornbeck, Richard and Enrico Moretti. 2018. Who benefits from productivity growth? Direct and indirect effects of local tfp growth on wages, rents, and inequality. Working Paper 24661, National Bureau of Economic Research.
- Instituto Nacional de Estadística e Informática. 2019. *Perú: Indicadores de Gestión Municipal 2019*. Lima, Peru: Instituto Nacional de Estadística e Informática.
- Moretti, Enrico. 2004a. Human capital externalities in cities. In J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: Elsevier, 2243–2291.
- Moretti, Enrico. 2004b. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review* 94(3): 656–690.
- Moretti, Enrico. 2012. *The New Geography of Jobs*. New York, NY: Houghton Mifflin Harcourt.
- Özgüzel, Cem. 2023. Agglomeration effects in a developing economy: evidence from turkey. *Journal of Economic Geography* XX(XX): XX–XX.
- Quintero, Luis E. and Mark Roberts. 2018. Explaining spatial variations in productivity. Evidence from Latin America and the Caribbean. Policy Research Working Paper 8560, World Bank.
- Superintendencia Nacional de Educación Superior Universitaria. 2020. *II Informe bienal sobre la realidad universitaria en el Perú*. Lima, Peru: Superintendencia Nacional de Educación Superior Universitaria.
- Superintendencia Nacional de Educación Superior Universitaria. 2021. *III Informe bienal sobre la realidad universitaria en el Perú*. Lima, Peru: Superintendencia Nacional de Educación Superior Universitaria.
- Williamson, Jeffrey G. 1965. Regional inequality and the process of national development: A description of the patterns. *Economic Development and Cultural Change* 13(4, Part 2): 1–84.
- Zelinsky, Wilbur. 1971. The hypothesis of the mobility transition. *Geographical Review* 61(2): 219–249.