Urban costs around the world

Jordan Rosenthal-Kay*
Federal Reserve Bank of San Francisco

April 27, 2026

Abstract

Cities are engines of economic development, but the steep costs of urban scale have left large productivity gains unrealized. The rate at which urban costs rise with a city's size depends on its commuting costs and capacity to expand up and out. Using geospatial data, I partition polycentric agglomerations into 15,000 monocentric cities and use a structural model to measure these urban cost elasticities globally. Cities in the developing world grow by building out rather than up, even though residents face higher transportation costs, resulting in cost elasticities nearly twice as high as those in rich nations. Embedding these estimates in a quantitative spatial model of the world economy, I find that productivity growth since 1993 would have raised developing-world GDP per capita by an additional 6% had urban costs been at the global technological frontier. Road paving is one lever for lowering urban costs, but its returns depend sharply on targeting. Urban costs also shape nations' resilience to climate risk.

*Jordan.Rosenthal-Kay@sf.frb.org. I thank my advisors, Jonathan Dingel, Mikhail Golosov, Erik Hurst, and especially Esteban Rossi-Hansberg. This project greatly benefited from conversations with Rodrigo Adão, Milena Almagro, Elena Aguilar, Fernando Alvarez, Olivia Bordeu, Thomas Bourany, Juanma Castro-Vincenzi, Craig Chikis, Gilles Duranton, Ed Glaeser, Lu Han, Tom Hierons, Chang-Tai Hsieh, Greg Kaplan, Benny Kleinman, Hugo Lhuillier, John Mondragon, and Jeanne Sorin. I thank Ryden Iwamoto for excellent research assistance. All errors are my own.

1 Introduction

Cities around the world look different. Compared to cities in high-income nations, cities in the developing world have lower building heights, sprawl more, and are considerably denser, especially in their downtowns. In this paper, I argue that differences in urban form around the world reflect differences in *urban costs*: barriers to urban development. A city's urban costs depend on geophysical and regulatory constraints on development and the quality of its transportation infrastructure. Urban costs matter because, by limiting the size of cities, they limit an economy's ability to take advantage of the vast productivity benefits associated with urbanization. This paper asks, what barriers do cities around the world face to expanding, and what would the aggregate welfare gains be if these barriers were reduced? While substantial attention has been paid to quantifying the agglomeration benefits associated with urban life, this paper develops a framework to measure urban costs around the world and quantify their importance for national welfare.

Several facts suggest large economic gains from reallocating workers to the urban sector. First, there are large rural-urban wage gaps and substantial economic returns from rural-to-urban migration in the developing world (Gollin et al., 2014; Bryan et al., 2014).[1] Second, more populous and denser cities pay higher wages, controlling for selection on workers' unobserved abilities.[2] Urban costs constrain urbanization and therefore limit its benefits: if commuting costs or housing prices are high, or if the physical expansion of cities is burdensome, then it is difficult for people to agglomerate in space.

Understanding how urban costs affect economic development requires rich data on cities around the world. Traditional data sources on cities' wages, floorspace prices, and commuting patterns are not available with anything resembling worldwide coverage. To overcome the challenge of data availability, I combine theory and satellite data on cities' urban form – their physical characteristics and spatial organization – to understand differences in urban costs around the world. Data on urban form are available for all cities. Moreover, these data allow me to infer urban costs from cities' spatial patterns of development: whether they have accommodated demand for floorspace by building up, by building out, or by concentrating development in

[1] While these wage gaps may reflect the negative selection of low-productivity workers into agriculture to relieve a nation's 'food problem' (Lagakos and Waugh, 2013) or the positive selection of high-productivity individuals to skill-intensive cities (Young, 2013; Behrens et al., 2014; Herrendorf and Schoellman, 2018), the literature nonetheless documents sizable returns for urban migrants (Lagakos et al., 2020). Gollin et al. (2021) provides suggestive evidence that urban wage premia in sub-Saharan Africa do not reflect compensating differentials for poor urban amenities. See Lagakos (2020) for a review.

[2] See Combes and Gobillon (2015) for a review of the literature estimating agglomeration elasticities. There is suggestive evidence that these elasticities may be higher in developing nations (Chauvin et al., 2017; Ahlfeldt and Pietrostefani, 2019).

the urban core relative to the periphery.

I measure urban costs and quantify their importance in three steps. First, I develop a quantitative theory of cities that captures the key tradeoffs associated with urban scale and how they are reflected in urban form. Second, I use this theory to guide measurement of the parameters that determine the costs associated with scaling cities. Third, I embed my estimates in a general equilibrium model of an urban system which I calibrate to match the world economy. I use the calibrated model to understand how urban costs affect a nation's level of economic development, and study cost-effective policies to improve cities.

To inform my theory of cities, I first define in the data what I mean by a 'city.' My definition begins with the entirety of the Urban Centres Database from the Global Human Settlement Layer, which identifies urban agglomerations defined by convex sets of remotely sensed built cover associated with a population exceeding 50,000 and with a density over 1,500 persons per square kilometer. These urban agglomerations can encompass large sets of cities (e.g., the San Francisco Bay Area) or large polycentric cities like Tokyo. I partition these agglomerations into monocentric 'cities' by developing an urban-theory-informed estimator of the location and number of cities' central business districts, and apply this estimator using data on cities' built environment, as measured by the World Settlement Footprint 3D (WSF-3D). The WSF-3D is a satellite-observed (as opposed to modeled) dataset of built volume at a highly granular resolution (90m). This procedure results in a dataset of over 15,000 'cities,' breaking up, for example, Tokyo into Yokohama, Chiba, and downtown Tokyo (Chiyoda).

Using my dataset of cities, I document new facts on how urban form varies around the world. First, I find that polycentricity is predominantly a large-agglomeration, rich-world phenomenon: even conditional on city size, as measured by total agglomeration population, agglomerations in the rich world are more likely to be polycentric. Thus, even adjusting for city size, cities in the developing world look more monocentric than their rich-world counterparts.

I then examine aggregate measures of urban form by characterizing the average height and density of cities around the world. Adjusting for city size, there is considerably less built volume per person in developing-world cities. This is predominantly driven by the fact that developing-world cities are substantially shorter than those in the rich world, and only modestly denser, indicating that cities in the developing world achieve

their volume disproportionately with area rather than height: they build out, not up. These results suggest that the technology to build cities varies around the world, and that there are either barriers to vertical development in the developing world, or barriers to horizontal development in the rich world.

Second, the internal structure of cities varies around the world. In developing cities, there is more built volume in the urban core relative to the periphery. I show this by first nonparametrically estimating the 'skyline slope' of cities around the world, which is the gradient of average height with respect to distance from a city's downtown. Skyline slopes are monotonically flatter in cities as one moves up the development spectrum. Estimating the skyline slope of each of my fifteen thousand cities separately, I find that even after adjusting for city size, cities in the developing world have steeper skylines than rich-world cities. Through the lens of a simple urban model, this suggests either that commuting costs are greater in developing-world cities (concentrating demand at the urban core) or that floorspace is more elastically supplied in the developing world, reducing the need for sprawl.

I use these features of urban landscapes to measure urban costs. To do so, I develop a quantitative model of cities and an agricultural sector. In the model, differences in urban form reflect key parameters of the technologies that govern a city's capacity to expand. Cities crowd population at their core relative to their periphery if supply or demand for floorspace is greater in the city center. Supply is constrained by the capacity to build up, while demand depends on the transportation technology available in each city. Additionally, cities absorb population by building horizontally. All three of these aspects of cities' building and transportation technology determine each city's *urban cost elasticity*, which shapes how the costs of urban life scale with population size (Combes et al., 2019). My model incorporates key features from the urban and development literature. First, cities are monocentric and their internal structure depends on the transportation technology available in the city. Second, the economy is a dual-sector model, with a traditional (agricultural) sector and a modern (urban) sector featuring increasing returns and intermediate input linkages. In my model, workers sort between agriculture and the modern sector on the basis of comparative advantage, so that the marginal urban migrant has lower sector-specific human capital than the average urban worker, as in Lagakos and Waugh (2013). Finally, the model features a system of cities that vary in their technology, as in Henderson (1974) or Desmet and Rossi-Hansberg (2013). Unlike these frameworks, in my model, cities are linked through costly trade and costless migration, following standard assumptions in quantitative economic

geography (Allen and Arkolakis, 2014; Redding and Rossi-Hansberg, 2017).

The model generates several log-linear estimating equations that I use to recover three components of cities' technology: a city's ability to build up (the floorspace supply elasticity), to build out (the land supply elasticity), and the commuting cost elasticity. First, a city's skyline slope reveals the product of the commuting cost and floorspace supply elasticities. The empirical challenge is to disentangle the two. The relationship between the average height of a city and its average income per square kilometer is informative of a city's floorspace supply elasticity. To recover land supply elasticities, which govern how responsive a city's footprint is to changes in land rents at its periphery, I use time series variation in urban growth in terms of area and income. However, these relationships may reflect reverse causality: cities that are productive in building floorspace or expanding their footprint will have lower floorspace prices, drawing in labor, making these cities larger and richer in equilibrium. Consequently, I develop model-based instruments to estimate these causal parameters, and apply them in a generalized random forest (Athey et al., 2019) to learn spatial heterogeneity in cities' floorspace and land supply elasticities as a function of remotely sensed observables. Cities in the developing world have both higher commuting costs and less elastic floorspace supply. Land supply elasticities are on average lower in the rich world, and depend on land availability.

I combine the estimated parameters to construct each city's urban cost elasticity. In my framework, the urban cost elasticity is a land-supply-weighted average of congestion elasticities from rising floorspace prices and longer commutes. On average, cities in developing nations have urban cost elasticities that are nearly twice as large as those in the rich world. The largest driver of differences in urban cost elasticities across nations is differences in transportation technology, captured by the commuting cost elasticity.

I calibrate the model for nearly all countries to match their cities' population, output, building and transportation technology, and each country's intercity trade network and agricultural sector size and share of national income. To show that urban costs have constrained economic growth and structural transformation, I feed in observed productivity and population growth from 1993 to 2015. Had urban costs been at the global technological frontier – the level observed in East Asia – developing nations would have GDP per capita on average 6% larger today, while for rich nations, this number is only 0.14%. These gains are entirely driven by general equilibrium effects, as workers move out of agriculture and reallocate across cities. This exercise underscores the potential scope for lower urban costs to foster economic development.

To study one type of policy that might achieve these gains, I focus on road paving. Many developing-world cities are substantially under-paved, and the costs of road paving are well understood. For example, I find that only 50% of urban roads in Nigeria are paved. The benefits of transportation infrastructure are difficult to monetize, which may explain why developing countries have under-invested in these technologies. I use data on cities' transportation technology to estimate the effect of road paving on my estimates of cities' commuting cost elasticity and find that a 10pp increase in the share of roads paved lowers the commuting cost elasticity by over 5%. I show that the general equilibrium return on road paving depends on which cities are targeted by the policy: paving roads in productive cities can generate a sizable economic return over a long time horizon through spatial reallocation effects alone, but targeting policy at the least-paved cities diminishes these gains substantially. This illustrates how policy aimed at targeting intercity infrastructure disparities can be inadvertently misallocative.

Finally, I examine how resilient an urban system is to shocks that damage its cities by studying sea level rise. Cities in East and Southeast Asia are uniquely exposed to this climate risk. Feeding these climate shocks into the model, I find that nations with lower urban costs have greater capacity to adapt to these climate shocks.

Related literature This paper relates broadly to three bodies of work. First, it contributes to the literature examining global urban patterns by providing new facts on how cities' internal structure varies globally. Second, it advances the literature that examines city-level policy interventions using quantitative spatial models by developing a framework to understand the aggregate and distributional implications of such changes across countries. Third, by studying the determinants of the overall size of the urban sector and the population distribution across cities, this paper contributes to the literature on how the spatial (mis-)allocation of productive factors affects macroeconomic aggregates.

Several recent papers have studied various aspects of cities globally using geospatial data. Akbar et al. (2023a) and Akbar et al. (2023b) document that there are large differences in urban speed around the world, primarily driven by differences in freeflow speed. They find that cities in the rich world are 50% faster than those in developing nations. Harari (2020) documents that cities' spatial layout affects their growth and amenity value in India. I provide a quantitative framework that links a city's characteristics like speed and urban form to its size. The studies most closely related to the empirical component of this paper are those by Jedwab et al. (2021), Lall et al. (2021b), and Lall et al. (2021a), all of which study how cities'

physical characteristics vary with city and national income. Relative to these papers, I provide new facts on cities' internal structure and their polycentricity.[3] Moreover, I propose and quantify a theory of cities that explains variation in urban form as driven by differences in the building and transportation technology available around the world. I measure components of these technologies using model-consistent regressions, and I use my estimates to perform counterfactual analysis with a quantitative spatial model of the world.

A handful of recent papers use frameworks similar to mine to answer related questions. First, Ahlfeldt et al. (2023) use data on the construction of tall buildings in cities globally to quantify the welfare gains with changes in building technology since 1975. Coeurdacier et al. (2025) use data on changing agricultural and urban land-use patterns in France to understand how structural transformation has affected how cities grow. Deffebach et al. (2025) use data on income sorting within 121 to understand spatial inequality within cities across the development spectrum. Unlike these papers, I focus on three aspects of urban costs: the ability to build up, out, and transportation technology, and apply it to all cities around the world.

Second, this paper examines the aggregate effects of city-level policy interventions using tools from quantitative spatial economics. Past research has used case studies to understand the effects of changes to transportation technology within cities. For example, the effect of introducing Bus Rapid Transit systems on urban residents' welfare has been studied in Jakarta, Indonesia (Gaduh et al., 2022; Kreindler et al., 2023), Bogotá, Colombia (Tsivanidis, 2023) and Dar es Salaam, Tanzania (Balboni et al., 2020).[4] Unlike these papers, this work moves beyond case studies and studies how city-level policy interventions matter for the macroeconomy by accounting for equilibrium interactions across all cities within a country. My framework provides less detailed treatment of city structure in order to be comprehensive and draw comparisons around the world, assess the aggregate implications of high urban costs across all countries, and quantify the potential welfare gains associated with reducing such costs.

Third, this paper is about the role of cities in determining national welfare, and the urban costs that limit their size. Most similar in spirit to this work are Bryan and Morten (2019), who develop a model of frictional

[3] In a similar spirit, Mills and Tan (1980) collate population density gradient estimates for under 100 cities, and, similar to this paper's findings, finds steeper population density gradients in developing countries. Liotta et al. (2022) compute population density and rent gradients for 192 global cities to test predictions of the monocentric city model and find that the vast majority of cities globally have negative population density and rent gradients, consistent with theory. There is also a large remote-sensing literature on the spatial patterns of urban development around the world. For example, Frolking et al. (2024) analyze how urban land use patterns have changed over time using satellite data, and find that Asian cities have shifted from building horizontally to building vertically.

[4] See Bryan et al. (2020) for a review of this literature.

interregional migration, which they use to estimate the aggregate productivity effects of reducing migration barriers in Indonesia, and Desmet et al. (2017), who study the productivity effects of relaxing migration barriers in Asia. Lagakos et al. (2023) study the welfare gains of rural-to-urban migration in a dynamic model in which seasonal migration occurs as a reaction to uninsured negative shocks faced by agricultural households. Unlike these papers, which focus on the barriers faced by agricultural migrants, I instead focus on the urban costs that limit cities' size. This links work on rural-to-urban migration and aggregate productivity to an important macro-urban literature that seeks to understand the role the urban system plays in shaping national welfare. For example, Au and Henderson (2006) estimate agglomeration and congestion elasticities for cities in China, and find that Chinese cities are undersized relative to the social optimum. Desmet and Rossi-Hansberg (2013) study misallocation across cities and find that there would be much larger gains to eliminating frictions across cities in China than the United States. My framework allows me to study the macroeconomic impact of differences in cities' building and transportation technology for the entire world.

The rest of this paper is organized as follows. First, in Section 2, I develop a dataset of cities around the world, partitioning polycentric agglomerations into monocentric cities. Using these data, I show facts about cities globally. I then present a quantitative model of cities aimed at rationalizing these facts in Section 3. In Section 4, I use model-implied estimating equations to recover urban costs for all cities around the world. I embed these estimates in my quantitative spatial model and perform counterfactual analysis in Section 5. Section 6 concludes.

2 Urban form around the world

This section presents facts on how cities differ globally in terms of their urban form: both aggregate characteristics – their area, volume, and density – and their internal spatial structure.

Doing so first requires that I define what a city is in the data. I begin by creating a dataset of urban agglomerations within which I can measure the built environment. I then combine theory and machine learning tools to partition these agglomerations into a dataset of cities (e.g., splitting the San Francisco Bay Area into San Francisco, Oakland, and San Jose).

2.1 Building a dataset of cities

To define the location and spatial extent of a city, I use the shapefiles provided by the Urban Centre Database (GHS-UCDB, R2024a), via Copernicus, a program of the European Commission. The UCDB provides 11,422 polygons associated with urban developments of more than 50,000 people. These polygons are defined by combining a variety of modeled and remotely sensed data products compiled by Copernicus, and correspond to the intuitive notion of an urban agglomeration: a contiguous collection of cells with a population density exceeding 1,500 persons per square kilometer. I subset the sample to agglomerations with a 'high' quality flag, resulting in a dataset of 10,852 agglomerations. Importantly, this definition is remotely sensed, and does not make reference to political boundaries in its construction.

Within each agglomeration, I measure contemporary construction using the World Settlement Footprint 3D (WSF-3D, Esch et al., 2020; Esch et al., 2022), which combines observations from the Sentinel-1 and -2 satellites with a digital elevation model (TanDEM-X) to produce rasters of built volume at the 90m resolution globally. These data are superior to products like the built volume data in the Global Human Settlement Layer, which combines modeled height data with remotely sensed built surface data to infer volume.[5] Instead, the WSF-3D more closely resembles satellite-observed volume than a machine-learned product. Moreover, its measurement is globally consistent, guaranteeing measurement error is orthogonal to the level of development and the availability of validation data.

Within each urban agglomeration, I extract the 2015 Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime luminosity measure (≈500m resolution) and population by way of WorldPop, which provides a lightly modeled population raster at the 100m level that does not use the built environment data as a modeling input.[6] Finally, I use a digital elevation model (Copernicus DEM GLO-30) combined with land cover data (classifications of land as soil, ice, and so on) to create a 10m-resolution raster indicating the potential for development to occur in that pixel, which I then aggregate to the level of the WSF-3D data. I classify land as developable if it has a slope with grade less than 15% and is not water, ice, wetland, or a mangrove.

[5] Previous versions of this project relied on the GHSL data.

[6] That is, I use the unconstrained data, which downscales administrative population estimates (census and projections) using covariates like land cover (water, ice, etc).

2.2 Identifying polycentricity: urban-theoretic partitions

The urban agglomerations in the data reflect an assemblage of what one might intuitively call 'cities.' For example, the San Francisco Bay Area polygon contains San Francisco, San Jose, and Oakland. Other agglomerations like Istanbul, Turkey, are famously polycentric and contain, at minimum, two 'cities' defined by the 'European' and 'Asian' sides, split by the Bosporus. The analysis throughout this paper defines cities as monocentric, convex clusters that may partition these larger urban units. In this section, I explain how I partition urban agglomerations into (potentially) smaller monocentric cities.

Theoretically, monocentricity means that there is a single central business district (CBD) to which agents commute for work, and the rest of urban land is used to house commuters to the CBD. I treat the number and location of these CBDs as unknown parameters to be estimated.

Suppose cities $k = 1, ..., K$ are circular with locations indexed in polar coordinates, (x, ϕ), where x denotes distance to a CBD, and ϕ is the polar angle.[7] Households in the city have a utility function, $U\left(C, A(x)h(x)\right)$ which they maximize by choosing C (consumption aggregator), $h(x)$ (floorspace) and x, subject to the budget constraint, $C+r(x)h(x) \leq w$. For now, I treat w as exogenous. Their location decision x trades off amenities $A(x)$ and floorspace prices $r(x)$. Amenities decay with respect to distance to the downtown, $A(x) \propto x^{-\tau}$, where τ captures commuting costs, which are paid for in time not spent at home enjoying $h(x)$.[8]

Developers supply floorspace with a constant returns technology in land (a fixed factor) and materials, resulting in upward sloping housing supply curves $h(x) = \bar{h}(x)r(x)^{\gamma}$, where $\bar{h}(x)$ is some exogenous shifter of the quality or quantity of land.

Then, in equilibrium, we have that the total amount of floorspace at location i in city k is

$$H_{ik} = \bar{H}_{ik}\left(x_{ik}\right)^{-\theta_k} \tag{1}$$

where $\theta_k = \tau_k\gamma_k$.

[7] That is, space is continuous. Have no fear, reader: I only consider symmetric allocations of economic activity. Hereon we can ignore ϕ; its introduction is meant to underscore that cities have some nonzero area.

[8] An alternative is that commuting costs are paid for in units of labor. While standard – this assumption has the unattractive property that purchasing power for housing declines across space, which generates variable elasticity price gradients that are sensitive to the income and price elasticity of floorspace demand. Mills (1972) failed to account for this fact, and this feature of the monocentric city model was not corrected until Brueckner (1982).

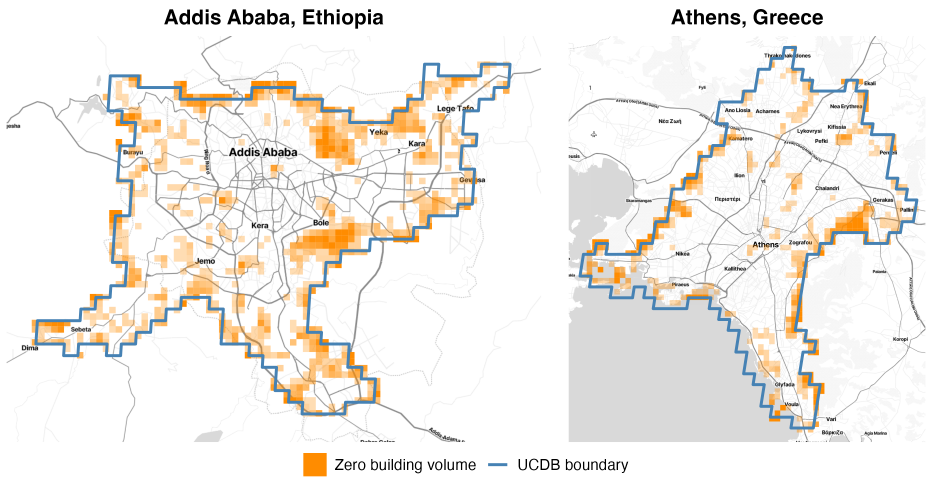

Figure 1: Structural zeros in Addis Ababa, Ethiopia, and Athens, Greece UCDB agglomerations. Addis is around 465km^2 with a population of 7.3M while Athens is 412km^2 with a population of around 3.2M. Colored pixels are aggregated to the 500m resolution and shaded by the share of zero-value WSF-3D pixels.

This model serves as the basis for CBD estimation and city partition. I observe building volume $H_i \geq 0$ on a grid of N spatial blocks within each urban agglomeration, as well as the fraction of land in each pixel that is developable, f_i. I rescale each urban agglomeration in the UCDB data onto the unit square $[0,1]^2$ and assign each pixel a location in the unit square corresponding to its centroid, μ_i.

I assume the statistical model,

$$H_i \sim \text{Poisson}\left(f_i \cdot \sum_{k=1}^{K} A_k (x_{ik})^{-\theta_k}\right) \tag{2}$$

where pixel i's distance to the CBD k, x_{ik}, is Euclidean. The key assumption here is additivity: that pixels aggregate over unobserved partitions of space in which agents are living in separate buildings. The rationale here is the superposition property of Poisson random variables, that if $H_{ik} \sim$Poisson$(A_k(x_{ik})^{-\theta_k})$, then $H_i = \sum_k H_{ik} \sim$Poisson$(\sum_{k=1}^{K} A_k(x_{ik})^{-\theta_k})$. As space is continuous in the model, the correct model-based analog of H_i is height, not volume. While pixels are of constant area, their developable area differs; f_i is the offset that corrects the scaling of volume data so that it is in constant proportion to height.

I use the Poisson model to rationalize zeros in the data that are plausibly structural: land with no development is more likely to occur on the fringe of a city, rather than near its CBD. Thus, accounting for these zeros will meaningfully affect estimates of θ_k. Upon inspection of the data, these structural zeros are far more common

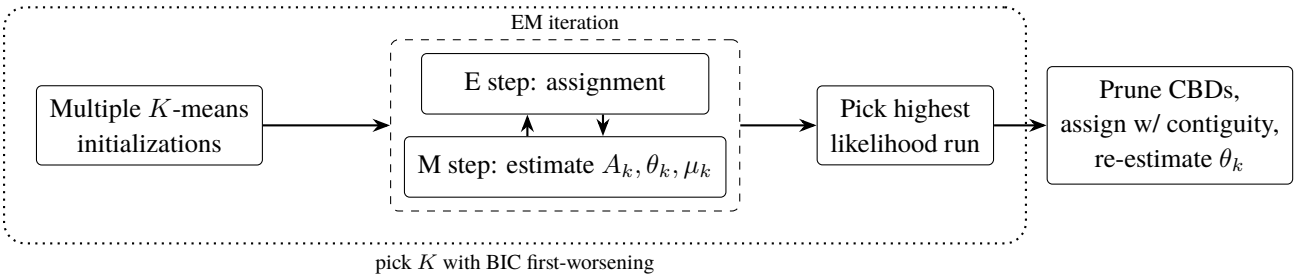

Figure 2: CBD detection algorithm for estimating the parameters Ω and K in (3).

in cities' peripheries than their downtowns, and more so in developing cities; see Figure 1 for an example of Athens, Greece and Addis Ababa, Ethiopia, which has about 4pp more empty pixels within the UCDB boundary.

The log likelihood for (2) is,

$$\ell(\Omega) = \sum_{i=1}^{N} H_i \ln \lambda_i - \lambda_i, \quad \lambda_i = f_i \sum_{k=1}^{K} A_k (x_{ik})^{-\theta_k}, \quad \Omega = \{\mu_k, \theta_k, A_k\}_{k=1}^{K}, \tag{3}$$

where $\mu_k \in [0,1]^2$ are the unknown CBDs. Solving for the unknown CBDs resembles the problem of K-means clustering and estimation is related, but it is not an off-the-shelf application for two reasons. First, the location of the CBDs must be jointly estimated with θ_k and A_k. Second, unlike K-means, in which space does not matter within a cluster assignment, here I care explicitly about spatial structure: I am looking for an isometric decay pattern centered at μ_k.

Estimating Ω by maximizing (3) is therefore nontrivial. I do so with an expectation-maximization (EM) algorithm. The EM algorithm consists of an iterative two-step process that alternates between an 'E' step and an 'M' step. The E step attributes mass in each cell i to each CBD k given the parameter estimates Ω. The Poisson model makes this attribution trivial: The expected contribution of H_i to CDB k is $\hat{H}_{ik} = a_{ik} H_i$ where $a_{ik} \propto A_k (x_{ik})^{-\theta_k}$. Given the assignment weights a_{ik}, I proceed to the M step in which A_k, θ_k and μ_k are updated. In particular, θ_k solves the weighted Poisson first-order condition and A_k can then be backed out in closed form.[9] To update μ_k, I take Newton steps by differentiating the likelihood with respect to μ_k at

[9] That is, θ_k solves,

$$\sum_i \left(\hat{H}_{ik} - f_i A_k (x_{ik})^{-\theta_k} \right) \ln(x_{ik}) = 0, \quad A_k = \frac{\sum_i \hat{H}_{ik}}{\sum_i f_i (x_{ik})^{-\theta_k}}.$$

the estimated parameter values and assignments. This is possible as μ_k enters through the distance function that produces $x_{ik} = \|\mu_i - \mu_k\|$.

I iterate between the E step and the M step until the likelihood function converges. In practice, I make a number of modifications to the algorithm to improve functionality: first, I aggregate data to the 500m resolution which massively reduces compute time, making it computationally feasible to estimate for every agglomeration in the data. I ensure that CBDs cannot be placed on undevelopable land with a guard on the Newton step, and place a soft penalty on the Newton step that prevents CBDs from colliding, ensuring separation of clusters as the algorithm explores the space. Additionally, I place a penalty on estimating $\theta_k < 0$ only during CBD detection to ensure the algorithm finds 'peaks' rather than 'troughs' in the spatial distribution of built volume.

I initialize guesses of μ_k using K-means, then use multiple restarts of the EM algorithm with different initializations and scale the number of restarts with city size. I choose $\hat{\Omega}$ from the highest-value likelihood among the set of restarts. I select K using Bayesian information criterion (BIC), a principled way of penalizing additional parameters that do not sufficiently increase the likelihood. I estimate CBDs for $K = 1, 2, ...$ with a first-worsening rule: if $\text{BIC}(K) > \text{BIC}(K-1)$ I set the city's number of centers $K^* = K - 1$. With K^* and $\hat{\Omega}$ in hand, I first prune CBDs with near-zero $\hat{A}_k$ or $\hat{\theta}_k \leq 0$, or for whom $\sum_k a_{ik}$ amounts to less than 2% of the total area.[10] I then assign pixels to a single CBD using a Potts model, which finds a hard assignment of pixels that maximizes the likelihood conditional on contiguity constraints. When assignments result in disconnected sets, I keep only the cluster that contains the CBD and reassign disconnected territories to the next-best CBD.

Once CBDs have been detected, I reestimate θ_k in a second step using the full, spatially granular WSF-3D data. In this second step, I form larger grids by binning up polar angles and radii, and weight observations inversely proportional to the number of observations in its radial bin. This weighting scheme is designed to undo the mechanical rise in observations one obtains from the simple geometry of circles. Second, I exclude a small ring of observations around the CBD from estimation, where the size of the ring is generated to be proportional to the overall area of the city. I fix the scale of this 'donut' so that in a city the size of Shanghai (3,100km^2) I remove a 2km radius hole. I do this because in practice, CBDs are not pointless masses but

[10] I always keep the largest A_k center in the case in which pruning implies I should remove all CBDs.

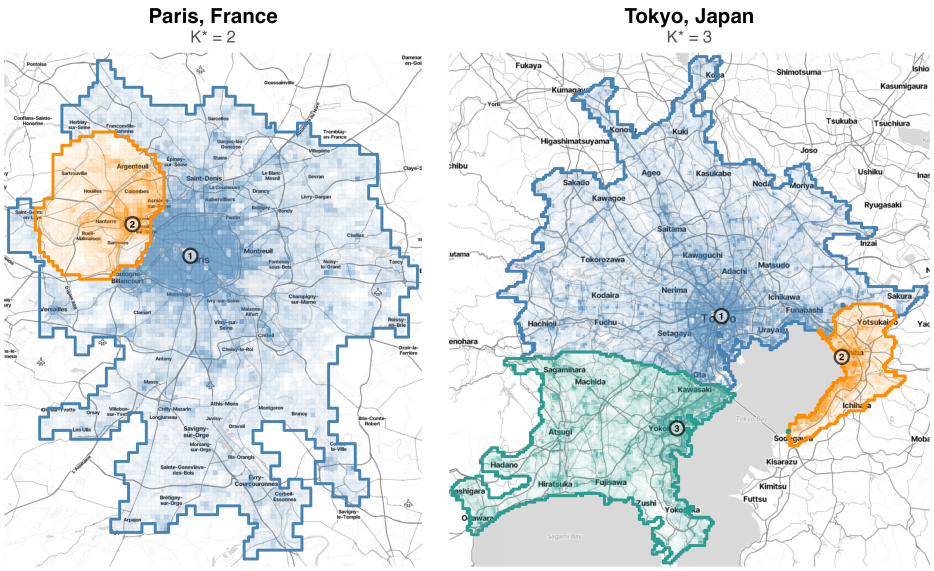

Figure 3: CBDs in Paris, France (left) and Tokyo, Japan (right). Pixels are aggregated to 500m and shaded by log built volume.

have area that scales with the size of the city. Within CBDs, the building volume gradient is quite flat, which biases down θ_k. A summary of the CBD detection algorithm and θ_k estimation is in Figure 2. All details are available in Appendix B.

2.3 Polycentricity around the world

The algorithm performs extremely well at finding well-known CBDs. For example, Figure 3 shows the CBDs detected in Paris, France ($K^* = 2$, one in central Paris and another surrounding La Défense), and Tokyo, Japan ($K^* = 3$, Chiyoda, Yokohama, and Chiba). By partitioning the UCDB agglomerations into potentially multiple cities, I expand the sample to 15,199 CBDs and their catchment areas, which I refer to as 'cities.'

In the left panel of Figure 4, I plot the average number of CBDs (K^*) in UCDB agglomerations within a country. Richer countries have more polycentric agglomerations: the average agglomeration in the US has about 1.75 central business districts, compared to 1.15 in India. This is not an artifact of simply the level of urbanization as this fact persists even when controlling for the size of the agglomeration. In the right panel, I plot the average K^* against log population of the agglomeration for both poor (World Bank 'Low' and 'Lower Middle' income countries) and rich countries. Cities in rich countries have more CBDs on average

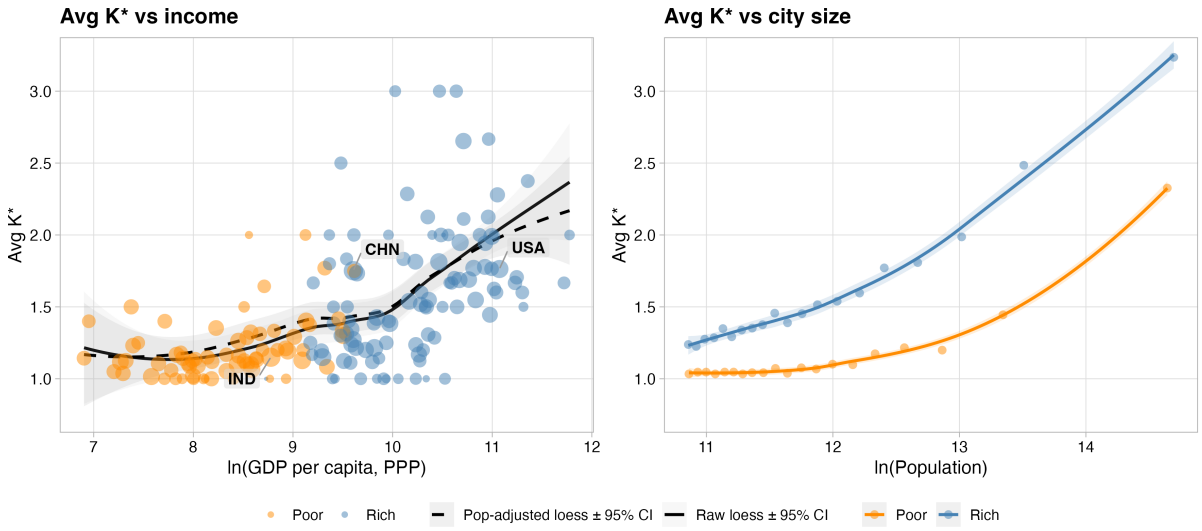

Figure 4: Average number of CBDs per agglomeration versus country-level log GDP per capita (left) and versus city size (right), split by the level of development.

conditional on agglomeration size across the entire distribution of agglomeration sizes. Polycentricity is predominantly a large- and rich-city phenomenon.

2.4 City height, area, and density

Using the over 15,000 cities constructed by partitioning the UCDB agglomerations, I now examine how cities differ around the world in terms of physical aggregates like height and area, as well as average population density.

A city's average height is defined as the total built volume in the city divided by the amount of developable land, so this definition includes developed areas of a city with no height, like roads and parks, and not just the height of structures. Area is measured as the total developable area in km^2. Mechanically, I can decompose built volume per person into average height and average area per person.

In Figure 5, I loosely follow the approach of Jedwab et al. (2021) and ask how the characteristics of a city depend on its nation's level of development. Each scatterplot shows the simple average characteristic of cities within each nation versus the country's level of development as measured by (log) GDP/capita. Each plot shows both a lowess smoother of this relationship (solid line), and the relationship conditional on a city's

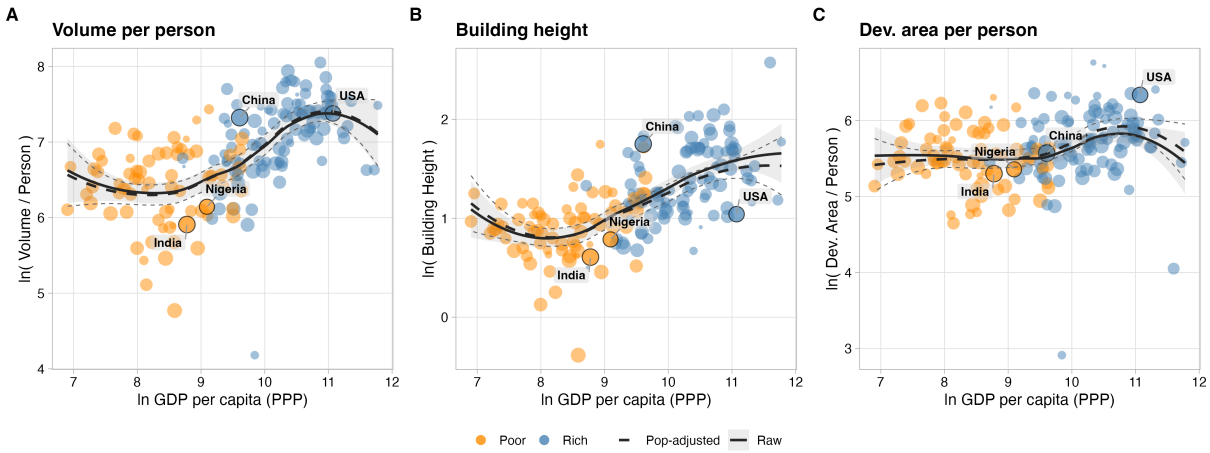

Figure 5: Average city characteristics versus a country's level of development. Left: built volume per person. Middle: average height per person. Right: developable area per person. $N = 15,199$ cities in 10,852 UCDB agglomerations.

size (dashed line).[11]

Cities in the developing world have substantially less built volume per person, on the order of one hundred log points or more.[12] This difference is primarily driven by height: rich world cities are considerably taller than their counterparts in poor nations, and are only modestly less dense.

In Appendix Table A1, I regress the difference between log city height and log developable area per person on country GDP/capita (with and without controlling for city size and primacy). When this coefficient is zero, it means the composition of built volume in terms of height or area is not systematically related to income. Instead I find that a 10% increase in GDP per capita shifts the composition of built volume per person towards vertical instead of horizontal construction by 1.1% ($p < 0.01$) indicating that cities in the developing world achieve their volume disproportionately with area rather than height: they build out, not up. These results suggest that the technology to build cities varies around the world, and that there are either barriers to vertical development in the developing world, or horizontal barriers to development in the rich world.

[11] That is, residualized against an estimated polynomial in city population and a dummy for whether a city is the largest in its country. I include this control because primate cities are often the focal site of government-led development, especially in nations with weak states (Ades and Glaeser, 1995).

[12] A one hundred log point difference here means that cities in the developing world have roughly 63% less built volume per capita than cities of a similar size in the rich world.

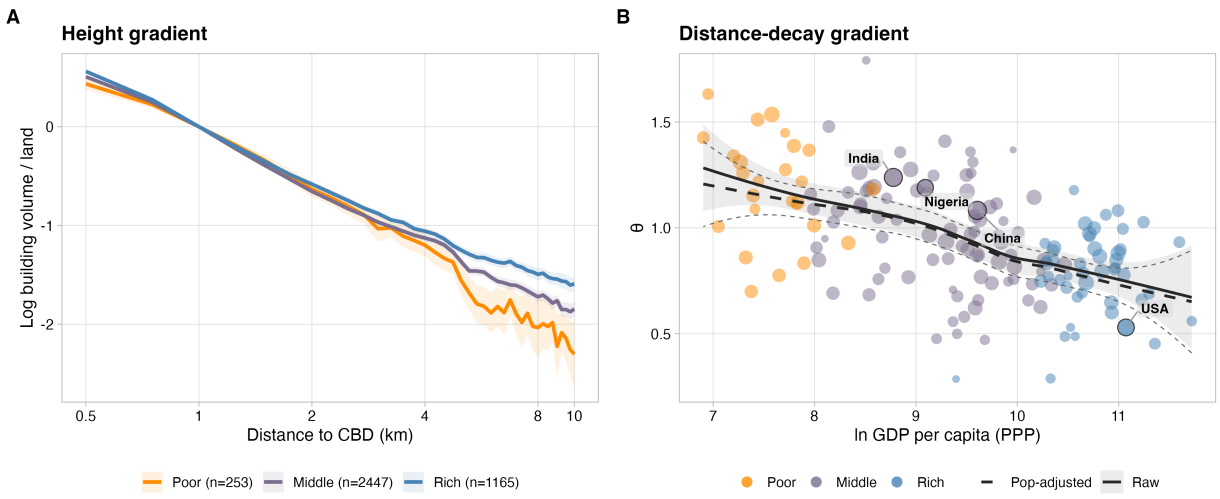

Figure 6: The left panel plots the average log building volume per unit land against distance bins of distance to the CBD. Distance bins are 250m wide and are presented on a log scale; the curve is normalized to zero at 1km. The sample excludes India and China, who have so many cities they dominate the curve. Panel B plots the distance-decay parameter θ_k (estimated at 100m resolution in an area-scaled donut around the CBD) against country-level log GDP per capita (PPP). Each point is a country average across non-primate cities; point size reflects the number of cities. The solid line is a loess smoother on raw country means; the dashed line adjusts for city population (quadratic polynomial residualization at the city level). Poor = Low + Lower Middle income; Rich = Upper Middle + High income (World Bank classification).

2.5 Cities' internal structure

To better understand differences in cities' internal structure around the world, I examine estimates of θ_k for cities around the world.

In the left panel of Figure 6, I plot an estimate of cities' 'skyline slopes,' by which I mean the gradient of built volume per unit land with respect to log distance to the downtown. This provides a nonparametric estimate of how θ_k varies by the level of development. I average height in 250-meter concentric rings around each city's estimated CBD, and then I normalize height at a 1km distance from a city's estimated CBD (to remove level differences across cities made prominent in Figure 5). Averaging within distance bins removes zeros, making a log-transform of the y-axis possible without issue. I then average across cities by the World Bank income group definition, omitting India and China, and plot 95% confidence bands for the estimate of the mean. The first thing to note is that log-linearity provides a fairly accurate description of cities' skyline slopes. Second, there appears to be a clear monotonic relationship between a country's level of development and the skyline slope of its cities. This pattern captures not only the height of structures in the urban core

relative to the periphery, but also the thinning out of cities: relatively more urban land is undeveloped in the periphery. Building heights both decay more rapidly and cities thin out more quickly as you leave the downtown in developing cities compared to those in the rich world.

In the right panel of Figure 6, I plot estimates of the average θ_k across cities within a country against a country's level of development as measured by its log GDP per capita. I overlay a lowess smoother, both for the simple mean, and one in which I regression-adjust θ_k for differences in city size. The nonparametric relationship in the left panel persists with the parametric estimates of skyline slopes across the development spectrum: cities are flatter in the rich world. This pattern is also clearly visible in the nightlights data; see Appendix Figure A1.

Greater built volume density in the urban core relative to the periphery can reflect either a greater demand or supply for floorspace in the urban core, reflected by the fact that $\theta = \gamma \cdot \tau$: Demand for the urban core decreases with the quality of a city's transportation infrastructure. When commuting is costly, agents desire to live closer to the CBD, bidding up floorspace prices, incentivizing development, and ultimately resulting in a steeper skyline.

In summary, compared to cities in high-income nations, cities in the developing world accommodate demand for floorspace by building out, not up, yet pack in built mass relatively more in their urban cores, resulting in steeper skylines.

3 Urban costs in general equilibrium

The theory in the previous section showed that in a monocentric city, a city's skyline reflects the product of the floorspace supply elasticity and the commuting cost elasticity; $\theta_i = \gamma_i \tau_i$. That is to say, cities have steep skylines if commuting costs are high – incentivizing households to cluster in the urban core – or if it is easy to build vertically. The empirical challenge is to disentangle these two forces that shape cities' skylines. Doing so will allow me to determine whether the ability to scale cities is more constrained in the developing world relative to the rich world.

That developing cities tend to build out rather than up suggests that it is unlikely that floorspace is elastically

supplied in the developing world. However, we need a formal theory of cities to both define a city's urban cost elasticity and to demonstrate what moments in the data identify the key elasticities.

In this section, I deepen the monocentric city model to define a city's urban cost elasticity, which measures the welfare cost of scaling cities. I then embed this model in a quantitative system-of-cities model with endogenous rural-to-urban migration. This general equilibrium model accounts for important features of the spatial and development literature like frictional inter-regional trade, and Roy-like sorting in the labor market. More importantly, this general equilibrium model reveals how different cross-sectional and time series moments identify the key parameters of each city's urban cost elasticity. It is the structure of this model that disciplines estimation in Section 4 and is quantified for counterfactual analysis in Section 5.

3.1 Defining the urban cost elasitcity

Defining the urban cost elasticity requires making parametric assumptions on the technology and market structure to build cities up and out. In each city i, there is an infinite pool of identical potential developers. Developers choose whether to enter, and upon entering, operate a technology at a single site to build floorspace vertically after paying a fixed cost to incorporate marginal land into the city. The technology available to builders for building floorspace H_i at distance x is,

$$H_i(x) = \tilde{Z}_i^H (M)^{\frac{\gamma_i}{1+\gamma_i}} (T)^{\frac{1}{1+\gamma_i}}, \tag{4}$$

where M is materials, measured in units of the numeraire, and T is land, and Z_i^H is vertical construction TFP. At each location, available urban land is in fixed supply $T_i(x, \phi) = 1$. To incorporate marginal land into the city, developers must pay a fixed cost $F_i(x)$ that is rising in x, $F_i(x) = \tilde{Z}_i^X (x)^{2/\rho_i}$. Developers are competitive and take prices as given. A developer constructing floorspace at x solves, $\max_M \; q_i(x)H_i(x) - M - F_i(x)$, subject to the production technology (4). Developers only enter and construct floorspace so long as land rents, $r_i(x) \geq F_i(x)$, where, $r_i(x) = \max_M q_i(x)H_i(x) - M$.

This structure results in two supply curves,

$$H_i(x) = Z_i^H q_i(x)^{\gamma_i}, \quad \pi X_i^2 = Z_i^X r(X_i)^{\rho_i}.$$

On the demand side, there are L_i households who maximize $U(c, A_i(x)h)$ subject to a budget constraint. It is convenient to define effective housing, $\check{h} = A_i(x)h$ whose price is $\check{q}_i(x) = q_i(x)/A_i(x)$ so that the budget constraint is $c + \check{q}_i\check{h} \leq w_i$, and expenditure is $\mathcal{E}(\check{q}_i, w_i)$. Marshallian housing demand is $h(x) = \mathcal{E}_i/q_i(x)$.

Spatial equilibrium implies that (indirect) utility $V(\check{q}_i(x), w_i) = \bar{v}_i$ on $[0, X_i]$, which here means that $\check{q}_i(x) = \check{q}_i$; i.e., effective prices must everywhere be the same, so that $\mathcal{E}_i$ is spatially invariant.[13] The effective price $\check{q}_i$ is a city-level endogenous object that encodes all urban costs in this model: both commuting costs and floorspace prices.

Invoking market clearing assumptions within the city allows us to aggregate across urban space and define the urban cost elasticity. Clearing the floorspace market at each x gives $H_i(x) = h_i(x)L_i(x)$ which $L_i(x) \propto (\check{q}_i^{1+\gamma_i}/\mathcal{E}_i)x^{-\tau_i(1+\gamma_i)}$. Labor market clearing means that total city population $L_i = 2\pi \int_0^{X_i} xL_i(x)dx$, which implies $L_i \propto (\check{q}_i^{1+\gamma_i}/\mathcal{E}_i)X_i^{2-\tau_i(1+\gamma_i)}$.[14] Land rent is a fixed fraction of floorspace revenue, so $r_i(x) \propto q_i(x)^{1+\gamma_i}$. Free entry of developers gives $r_i(X_i) = F_i(X_i)$ at the periphery. In logs we have two equations in two unknowns,

$$\ln \mathcal{E}_i + \ln L_i = (1+\gamma_i)\ln \check{q} + \big(2 - \tau_i(1+\gamma_i)\big)\log X_i + c_0 \qquad \text{(floorspace + labor market clearing)}$$

$$\ln \check{q}_i = \left(\tau_i + \frac{2}{\rho_i(1+\gamma_i)}\right)\ln X_i + c_1 \qquad \text{(free entry)}$$

where c_0 and c_1 are exogenous constants.

Using free entry to substitute for X_i, we have, that,

$$\ln \check{q}_i = \underbrace{\left(\frac{1}{1+\rho_i}\frac{1}{1+\gamma_i} + \frac{\rho_i}{1+\rho_i}\frac{\tau_i}{2}\right)}_{\equiv \kappa_i}(\ln L_i + \ln \mathcal{E}_i) + \tilde{c}_0. \tag{5}$$

Here, κ_i is the *urban cost elasticity*, and it measures how urban costs shift with population or expenditure per household on floorspace. Totally differentiating (5) reveals how urban costs rise with respect to different

[13] Effective prices encode space through the 'Muth condition,' that $\frac{d\log q_i(x)}{d\log x} = -\tau_i$. In a model with commuting costs paid for in dollars, not utils, $\mathcal{E}_i$ varies across space. Ignoring this caused Mills (1972) to illegitimately conclude that his monocentric city model with Cobb-Douglas preferences generated exponential rent gradients. As pointed out by Brueckner (1982), the rent gradient in a model with real commuting costs is neither isoelastic nor exponential.

[14] Note this requires $\tau_i(1+\gamma_i) < 2$ for the integral to converge.

shocks,

$$d\ln \check{q}_i = \kappa_i \left(d\ln L_i + (1+\epsilon_{\check{h},\check{q}_i}) d\ln \check{q}_i + \epsilon_{\mathcal{E},w_i} d\ln w_i \right) \tag{6}$$

where $\epsilon_{\check{h},\check{q}_i}$ is (Marshallian) own-price elasticity of effective floorspace demand (under Cobb-Douglas preferences, this is -1) and $\epsilon_{\mathcal{E},w_i}$ is the income elasticity of housing demand (under homothetic preferences, this is 1).[15]

To translate this into welfare requires that we define $d\ln V / d\ln \check{q}_i = -\beta$ so that $-\beta\kappa_i$ is the elasticity of city consumption-equivalent welfare with respect to city population in city k, holding wages and goods prices fixed.[16] That is, it answers the question, if a city grew 1% in population, what percent increase in consumption would be required to keep residents as well off as they were before, holding wages and goods prices fixed? It is important to hold wages and goods prices fixed, because changes in these prices may reflect the benefits of urbanization, through, for example, agglomeration effects, or spillovers through the trade network that engender increased market access. The urban cost elasticity contains all components of the urban technology: how well cities can build up, build out, and how costly residents find it to spread out. To gain intuition on how it operates, consider two edge cases,

$$\rho_i = 0 \implies \kappa_i = \frac{1}{1+\gamma_i}, \quad \text{and} \quad \rho_i = \infty \implies \kappa_i = \frac{\tau_i}{2}.$$

When land supply is completely inelastic ($\rho_i = 0$) as a city grows in population, land is increasingly scarce. The ability for a city to accommodate people is determined entirely by its ability to build 'up' (γ_i) and so all congestion forces operate through the floorspace market. When land supply is completely elastic ($\rho_i = \infty$), any upwards pressure in land prices from a population increase incentivizes developers to expand the city horizontally so as to completely offset congestion through the housing market. The consequence of a larger city is that commutes are on average longer, lowering resident welfare.

[15] Note that this means when $(1+\epsilon_{\check{h},\check{q}_i})$ is not -1 (e.g., under CES preferences),

$$\left.\frac{d\ln \check{q}_i}{d\ln L_i}\right|_{d\ln w_i = 0} = \frac{\kappa_i}{1-(1+\epsilon_{\check{h},\check{q}_i})\kappa_i}.$$

When effective floorspace is a gross complement to non-floorspace production, urban costs are amplified: households cannot substitute for high urban costs by trading urban consumption $A(x)h(x)$ for tradeables while holding their location fixed. Equation (6) can easily be written in Slutsky form, $d\ln \check{q}_i = \kappa_i \left(d\ln L_i + (1+\epsilon^{\text{Hicksian}}_{\check{h},\check{q}_i}) d\ln \check{q}_i + \epsilon_{\mathcal{E},w_i} \left(d\ln w_i - s_{\check{h}} d\ln \check{q}_k \right) \right)$ where $s_{\check{h}}$ is the share of income devoted to floorspace.

[16] This is the definition of the urban cost elasticity in Combes et al. (2019).

3.2 Intercity equilibrium

I now broaden the framework to include many cities and a rural sector that are linked through trade and migration. Doing so makes endogenous city sizes L_i and wages w_i. This model serves as both the backbone to estimation, clarifying what cross-sectional variation can identify the elements of κ, the urban cost elasticity, and provides a general equilibrium framework within which I can do counterfactual analysis.

In the model, I treat each country as a closed economy with $i = 1, ..., N$ cities and a rural sector, denoted by R. There is a mass of $\bar{L}$ workers. Cities are disks with endogenous area πX_i^2, as specified in Section 3.1.

Household preferences, and location choice Households have Cobb-Douglas preferences over a final good (taken as the numeraire) and effective floorspace $A_i(x)h$ on which they spend a fraction β of their income. Each household draws a skill s which is distributed iid Fréchet with shape parameter ε. over the rural and urban sector. Over cities, households also draw a vector of preference shocks distributed iid Fréchet across cities with shape parameter ψ. Indirect urban utilities are, $V_i = \bar{A}_i s w_i (\check{q}_i)^{-\beta}$, and in the rural sector $\mathcal{V}_R$.

Production Each city produces a unique traded variety indexed by i, and the final good Y, is produced by a representative firm with technology,

$$Y = \left(\sum_{i=1}^{N} (x_i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

Each city's good is produced combining other urban output with human capital subject to local agglomeration forces,

$$y_i = Z_i (S_i)^{1-\eta} (M_i)^{\eta}, \quad M_i = \left(\sum_{i=1}^{N} (m_{ji})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad Z_i = \bar{Z}_i (L_i)^{\zeta}. \tag{7}$$

Intercity trade costs are iceberg, so that $\delta_{ji} \geq 1$ reflects the amount of output shipped from j to i for one unit to arrive in i.

Equilibrium A full definition of the model and its equilibrium is given in Appendix C. Given a city partial equilibrium (in which w_i, L_i are treated as exogenous, pinning down $\check{q}_i \propto (w_i L_i)^{\kappa_i}$), an intercity general equilibrium is, given, $\{\bar{A}_i, \bar{Z}_i, \kappa_i\}$, a set of $\{w_i, L_i, p_i\}$ over cities and an overall size of the urban sector L_U

such that: (i) Migration probabilities are given by,

$$\frac{L_U}{\bar{L}} = \frac{(\mathcal{V}_U)^{\varepsilon}}{(\mathcal{V}_U)^{\varepsilon} + (\mathcal{V}_R)^{\varepsilon}}, \tag{8}$$

$$\frac{L_i}{L_U} = \frac{(V_i)^{\psi}}{\sum_{j=1}^{N}(V_j)^{\psi}} \tag{9}$$

where $\mathcal{V}_U = \left(\sum_{i=1}^{N}(V_i)^{\psi}\right)^{1/\psi}$; (ii) Prices are marginal costs gross iceberg costs,

$$p_{ji} \propto \delta_{ji}\frac{w_j^{1-\eta}P_j^{\eta}}{Z_j}, \quad P_j = \left(\sum_k p_{kj}^{1-\sigma}\right)^{1-\sigma};$$

(iii) the goods market clears for each good,

$$p_i y_i = p_i^{1-\sigma}\sum_i w_i L_i + \eta\sum_j \left(\frac{\delta_{ij}p_i}{P_j}\right)^{1-\sigma} p_j y_j,$$

and $\sum_i p_i^{1-\sigma} = 1$ by definition of the numeraire.

Discussion This model captures several important mechanisms emphasized in the spatial and development literature. First, there is an upward sloping supply of labor to the urban sector generated by workers sorting on comparative advantage, as in Lagakos and Waugh (2013) and Young (2013), or Bryan and Morten (2019). The extent of this sorting is governed by ε, which controls skill dispersion. Thus, rural-urban wage gaps reflect selection and the marginal migrant does not experience wage gains equivalent to the wage gap. In calibrations I set $\mathcal{V}_R = A_R w_R$ where w_R is labor productivity in agriculture, and does not adjust in equilibrium, consistent with workers earning their average product, as in Lewis (1954).[17] A_R captures forces keeping workers in agriculture, such as informal insurance networks (Munshi and Rosenzweig, 2016) or wedges from migration costs.[18]

Trade costs δ_{ij} provide the connective tissue by which cities are linked, making the geography of each country's urban system matter in shaping aggregate productivity. Unlike models where trade costs affect the local price index (e.g., Allen and Arkolakis, 2014), each city's position in the intercity trade network affects the

[17] If farmers own the land they till and consume land directly, then this is the indirect utility function, see Appendix C.3.

[18] When migration costs are not bilateral, but instead depend on an origin and destination component, they are not separable from amenities, as in Desmet et al. (2018).

wage: cities that are connected to productive places enjoy a lower intermediates price index and consequently a higher wage. As urban goods are differentiated, expanding a city's size moves down its output demand curve; in effect, urban labor demand slopes down. This congestion force is attenuated due to agglomeration forces, which increase city productivity as a function of city size ($\zeta > 0$). Including intermediate inputs and roundabout production function creates a multiplier effect of local productivity increases that has been shown to be important for both aggregate development accounting (Krugman and Venables, 1995; Jones, 2011) and the spatial transmission of shocks within nations (Caliendo et al., 2018).

4 Estimating the urban technology

To understand how urban form expresses urban costs, I use the geospatial data on cities (introduced in Section 2) and the structure of the model of Section 3 to identify the parameters γ_i,ρ_i, and τ_i and construct urban cost elasticities κ_i globally.

4.1 Estimating the floorspace supply elasticity, γ_i

Integrating the equilibrium amount of floorsapce $H_i(x, \phi)$ across locations (x, ϕ) within each city to compute a city's average height $\bar{H}_i$ reveals that,[19]

$$\ln \bar{H}_i = \frac{\gamma_i}{1+\gamma_i}\left(\ln w_i + \ln \frac{L_i}{\pi X_i^2}\right) + e_i. \tag{10}$$

There are two issues with using (10) to identify γ_i. First is an endogeneity issue: the structural error e_i contains contains Z_i^H.[20] Cities that have high productivity in floorspace construction will on average have lower floorspace prices, and in equilibrium, such cities will be larger in terms of L_i and πX_i^2, which can influence wages through price and agglomeration effects. Consequently, the regressors of interest are endogenous.

[19] Defining,

$$\bar{H}_i = \frac{2\pi \int_0^{X_i} x H_i(x) dx}{\pi X_i^2}.$$

[20] In particular,

$$e_i = \frac{1}{1+\gamma_i} \ln Z_i^H + \ln\left(\frac{2}{2-\tau_i\gamma_i} \cdot (\beta(1-\tau_i(1+\gamma_i)/2))^{\frac{\gamma_i}{1+\gamma_i}}\right).$$

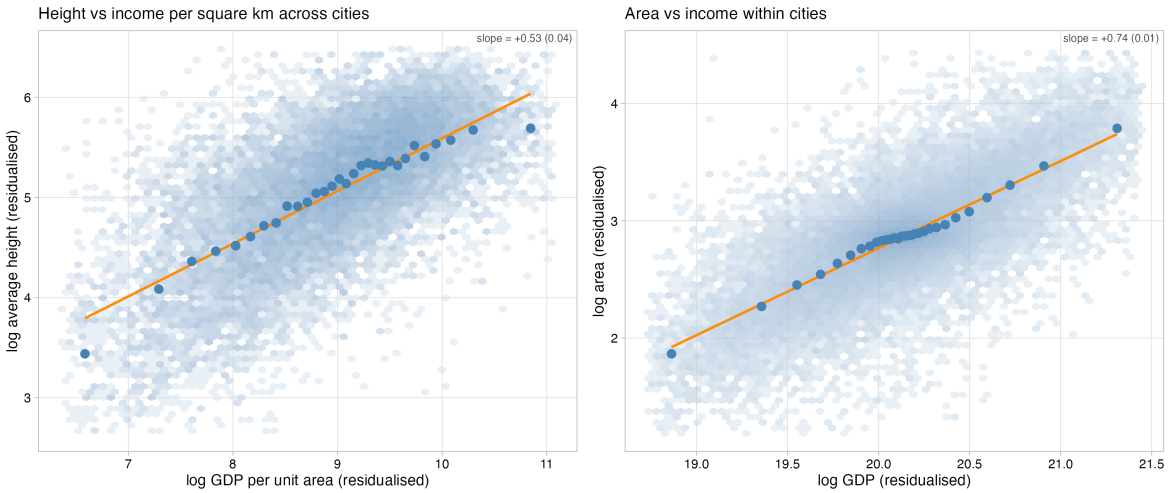

Figure 7: Binsregs of the relationships' between cities average height and income per square kilometer across cities, and area against income within cities. Residualized observation density is shown as a 'hexbin' cloud beneath the 'binsreg' (the solid blue circles, Cattaneo et al., 2024) and the OLS fit (orange). The relationship on the left is net of country fixed effects in the 2015 cross-section; the data in the right panel is net of agglomeration and region-year fixed effects leveraging the 1993-2015 time series.

Second, I only observe built volume in the cross-section, and consequently cannot estimate a separate γ_i for each city using within-city variation. The left panel of Figure 7 shows the average height and income per square kilometer relationship in the data. While this is not causal, and does not reflect heterogeneity, the relationship is clearly approximately log-linear, consistent with the model providing an accurate description of the data.

To handle endogeneity, I use instrumental variables combined with controls. I instrument for $\log w_i$, which is measured in the data as city GDP/capita, using $\log \bar{Z}_i^y$, which I recover by inverting the calibrated model (see Section 5). This is possible because $\bar{Z}_i^y$ is identified without knowledge of γ_i, so long as we know w_i, L_i and δ_{ij}, alongside η and σ. I control for population density, geophysical covariates G_i, their interaction, and country fixed effects. Using this instrument isolates the component of wages driven by city productivity from variation in wages driven by goods prices and agglomeration effects. I include country fixed effects (ξ_n) to difference out country-specific determinants of height and productivity, so the orthogonality assumption is that $\log \bar{Z}_i^y \perp e_i \mid L_i/\pi X_i^2, \xi_n, G_i$. Put simply, the assumption is that tradeable goods productivity is uncorrelated with floorspace construction productivity within a country, conditional on population density and geophysical controls. For example, Los Angeles is productive because its fair weather and diverse landscape made it ideal for film production, but skyscraper construction is stymied by seismic activity and varied

bedrock depth across the Los Angeles basin.

To capture heterogeneity, I model $\frac{\gamma_i}{1+\gamma_i} = \Gamma(G_i)$, a function of geophysical controls. To do so, I rely on the 'instrumental forest' machinery introduced in Athey et al. (2019). They extend the results of Wager and Athey (2018) for the estimation of heterogeneous treatment effects using random forests to generalized moment equations, including the standard instrumental variables orthogonality condition.[21] I include in G_i the share of developable land within 10km of a city's CBD, a city's slope, ruggedness, elevation, distance to the coast, and the clay, sand, and water content of its soil, as well as estimates of its soil density. I handle fixed effects by demeaning regressors and controls within countries, but I do not demean the predictive covariates that enter $\Gamma()$ so that the learner can leverage cross-country variation in geophysical characteristics to predict γ_i heterogeneity.

4.2 Estimating the land supply elasticity, ρ_i

In equilibrium,

$$\ln \pi X_i^2 = \frac{\rho_i}{1+\rho_i} \ln(w_i L_i) + f_i. \tag{11}$$

The land supply elasticity ρ_i therefore governs how much cities physically expand as their income grows. Identification here faces the same challenge as in estimating γ_i: the structural error f_i contains Z_i^X.[22] When productivity in urban land production, Z_i^X, is high, floorspace is less dear, and the city is more attractive. As a result, a naive regression of area on GDP would fail to identify ρ_i due to endogeneity bias. To estimate ρ_i, I leverage panel variation in urban area and GDP, which is absent for built volume data. I see city areas using 1993, 2000, and 2015 observations with the urban land classification from the WSF and GDP data from Kummu et al. (2025) at UCDB agglomeration level.[23] The panel variation here is compelling in that it

[21] Random forests (Breiman, 2001) are a machine learning tool used to nonparametrically estimate unknown functions in high-dimensional spaces. A substantial statistical literature has studied their properties; see, e.g., Hastie et al. (2009) for a review. Take the model $y_i = \beta(x)w_i + u_i$ where $w_i \not\perp u_i$ but z_i is a valid instrument. $\beta(x)$ is identified from the moment condition $\mathbb{E}[z_i(y_i - \beta(x)w_i)] = 0$. A generalized random forest learns $\beta(x)$ so that this moment equation holds by finding weights $\omega_i(x)$ so that $\hat{\beta}(x)$ solves $\sum_i \omega_i(x) z_i(y_i - \beta w_i) = 0$.

[22] The structural error is,

$$f_i = \ln \pi + \frac{\rho_i}{1+\rho_i} \ln\left(\frac{\beta(1-\tau_i(1+\gamma_i)/2)}{\pi(1+\gamma_i)\tilde{Z}_i^X}\right).$$

[23] Here, I aggregate cities to their UCDB areas to avoid performing an ad-hoc downscaling of GDP data to the monocentric city level when there is polycentricity in the UCDB agglomeration, and then treat all cities within an agglomeration as facing the same land supply elasticity.

uses how cities actually grow, rather than the cross-sectional relationship (which represents a long-run supply elasticity) and allows me to control for city and region-year fixed effects. A city fixed effect differences out all time-invariant components of urban land production productivity, which may be driven by fixed, geological factors. Urban land production productivity may be varying over time as technological advances in, e.g., electrical grid, paving, or sewage technology, have allowed cities to expand horizontally.

To model heterogeneity in land supply elasticities, I again treat the land supply as a function of geophyiscal observables G_i, $\frac{\rho_i}{1+\rho_i} = \varrho(G_i)$. The city income data is a population-weighted downscaling of subnational GDP measures and is measured with substantial noise. I rely on this as high-resolution nightlights are not available. I assume this measurement error is classical, and use a low-resolution harmonized nightlights series (Li et al., 2020) to instrument for city GDP to correct for this measurement error. To jointly capture this heterogeneity and run instrumental variables regressions, I again rely on an instrumental forest.

4.3 Spatial heterogeneity in urban cost elasticity estimates

Floorspace supply elasticity results Shutting down heterogeneity but flexibly controlling for geophysical characteristics (via 'double machine learning' in a partially-linear instrumental variables regression; Chernozhukov et al., 2018) I estimate a pooled $\hat{\gamma} = 0.44$ (s.e., 0.05; first-stage F-statistic 95.3). Allowing for city-level heterogeneity through G_i, the global is median $\hat{\gamma}_i = 0.52$ (pop-weighted mean is 0.64), and the 10-90 percentile range is $[0.24, 0.91]$; see Appendix Table A9. The model requires that $\hat{\Gamma}(G_i) = \widehat{\frac{\gamma_i}{1+\gamma_i}}$ is in $(0, 1)$ and, reassuringly, 96% of observations fall into that range. For the model calibration, I use Bayesian posterior mean estimates of $\hat{\gamma}_i$ under a Uniform(0,1) prior, which has no effect over most of the support of γ_i but shrinks estimates in the tails towards the mean so that all posterior means fall within (0,1). The most 'important' geophysical features for predicting floorspace supply elasticities are the soil characteristics (water content, clay content, and bulk density, and their interactions).[24] In the left panel of Appendix Figure A6, I order $\hat{\Gamma}(G_i)$ estimates and display 95% CIs produced by the instrumental forest. I can exclude 0 and 1 fromo the confidence interval for most estimates (except those in the tails). In the left panel of Appendix Figure A8, I compare my estimates of $\hat{\gamma}_i$ to those of Baum-Snow and Han (2024) for U.S. cities. Our estimates coincide in magnitude, and are weakly correlated (regression coefficient 0.23, s.e. 0.09). Estimation

[24] Features are important if the random forest regularly uses it to form splits of the sample.

details and results are available in Appendix D.

Land supply elasticity results Shutting down heterogeneity, the pooled land supply elasticity is $\hat{\rho} = 0.91$ (s.e., 0.11; first-stage F-statistics 75.7) This average may seem surprisingly low given how much undeveloped land there is in most parts of the world. The majority of habitable land is neither inhabited nor used for agriculture, and urban land represents a tiny fraction of overall habitable land (Ritchie and Roser, 2019). This might suggest that land is in effectively free supply. These results suggest that there are increasing marginal costs associated with weaving peripheral land into the urban fabric, possibly due to increased complexity of laying down urban infrastructure (electricity and plumbing), deteriorating land quality (Los Angeles must expand into the Hollywood hills) or a need to reclaim land from the sea (as in Singapore). Allowing for heterogeneity across agglomerations with the instrumental forest, the median land supply elasticity is 1.1 (pop-weighted mean 1.5), and the 10-90 percentile range is [0.5,2.9]. While 99% of estimates of $\hat{\varrho}(G_i)$ are theory-consistent, falling within (0,1), I nonetheless impose a uniform prior to shrink theory-inconsistent estimates onto (0,1); see Appendix Table A9. The most important predictive geophysical features are the share of developable land with 10km, the clay and sand content of soil, ruggedness, and coastal distance. The right panel of Appendix Figure A6 shows the sorted-$\hat{\rho}_j$ estimates and their 95% confidence interval, allowing me to reject 0 and 1 over much of the support of $\hat{\varrho}(G_i)$. In the right panel of Appendix Figure A8, I compare my estimates to those of Saiz (2010) for U.S. cities. On average, my estimates of ρ_i are around half as large, but the correlation in estimates is strong (regression slope = 0.47, s.e. 0.12)

Commuting cost elasticity estimation The final parameter I require is τ_i, which I estimate by diving the estimates $\hat{\theta}_i$ from Section 2 by my estimated $\hat{\gamma}_i = \hat{\Gamma}(G_i)/(1 - \hat{\Gamma}(G_i))$. The median city has $\hat{\tau}_i$=1.91 (pop-weighted mean 2.15); 10-90 percentile range [0.7,4.5]. 95% of $\hat{\tau}_i$ are theory-consistent ($\hat{\theta}_i > 0$ and $\hat{\gamma}_i^{\text{raw}} > 0$). To impose the constraints of the theory, I use an empirical Bayes procedure in which I posit truncated normal prior on $[0, \infty)$ for $\hat{\theta}_i$, estimate the parameters of the prior country-by-country, and recover posterior means, which I then divide by the 'shrunken' $\hat{\gamma}_i$ estimates so $\hat{\tau}_i > 0$. See Appendix D.2 for details. To validate these estimates, I correlate $\ln \hat{\tau}_i$ with within-city driving-speed measures taken from Akbar et al. (2023b) for over one thousand global cities (Appendix Table A10). Though noisily estimated, measures of freeflow speed are negatively correlated with with τ_i, the expected sign.

Within countries, larger cities have on average lower $\hat{\tau}_i$, consistent with high commuting costs acting as

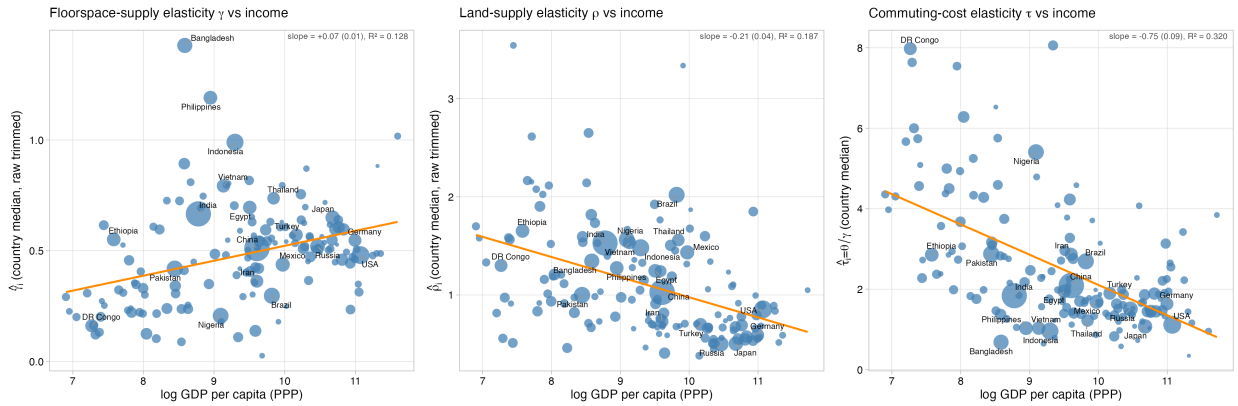

Figure 8: Country-median estimates of γ_i (floorspace supply elasticity), ρ_i (land supply elasticity) and τ_i (commuting cost elasticity). Each point is a country and each point's size reflects a country's population. The top right of each figure shows regression statistics: floorspace supply increases in log GDP/cap (slope 0.07, s.e. 0.01), land supply falls in log GDP/cap (slope -0.21, s.e., 0.04), as do commuting costs (slope -0.75, s.e. 0.09).

a negative amenity: cities with worse transportation infrastructure are smaller. A 10% increase in $\hat{\tau}_i$ is associated with a 6% decrease in population (p <0.01; see Appendix Table A2). This could of course reflect endogenous congestion or political-economy investment incentives — accounting for these is an active area (Allen and Arkolakis, 2022; Bordeu, 2023). Average population density strongly predicts $\hat{\tau}_i$: more dense cities have higher τ_i, consistent with the theory. Polycentric cities have lower commuting cost elasticities, consistent with high commuting costs creating the incentives for a second CBD to emerge, as in Fujita and Ogawa (1982), so as to endogenously produce lower experienced commuting costs for urban residents.

East Asian cities have some of the lowest estimated commuting cost elasticities — the population-weighted average τ_i is 0.85 in Indonesia, 0.88 in Malaysia, 0.91 in Vietnam, and 1.11 in Korea. OECD economies range from $\tau = 1.22$ in the United States and 1.34 in France to 2.32 in Chile. The commuting cost elasticity is on average much higher in the developing world: across sub-Saharan Africa the pop-weighted mean is 4.47 (median 4.54). The sub-continent is heterogeneous — in Uganda the country average is 2.98, while Nigeria (4.41) and South Africa (4.60) sit near the regional average.

Cross-country comparisons of urban cost elasticities Figure 8 shows how the median estimates of γ_i, ρ_i, and τ_i vary across countries. Floorspace is more elastically supplied in the rich world while land is less elastically supplied. Commuting costs fall dramatically with development.

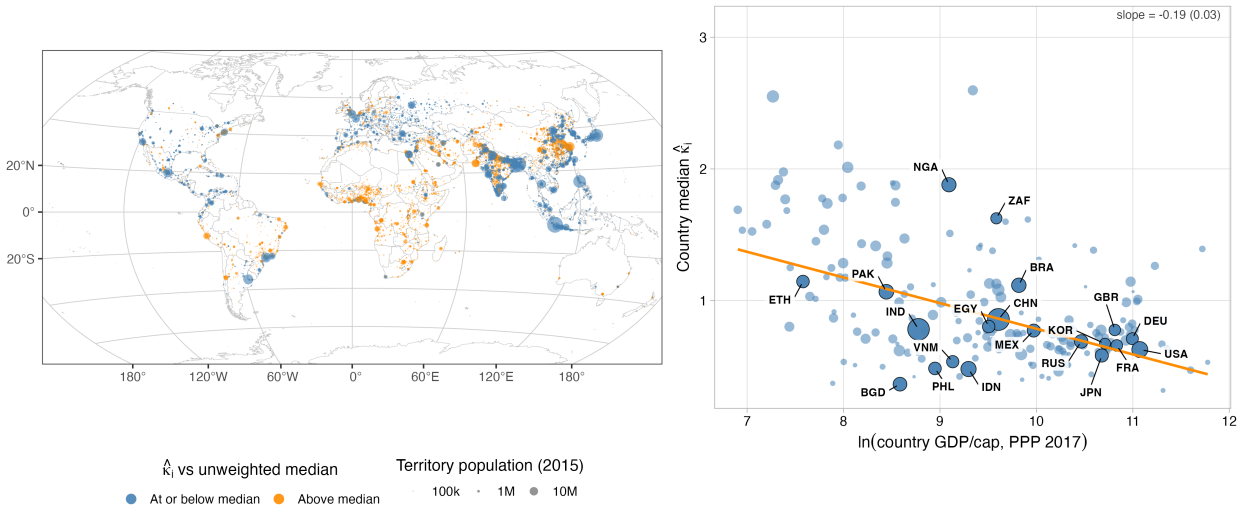

Figure 9: Global distribution of the urban cost elasticity $\hat{\kappa}_i$. Left: a world map; Right: scatter of country-level median κ_i against log GDP/cap in 2015. Point size is country 2015 population; the orange line is a population-weighted least-squares fit with slope -0.19 (s.e. 0.03). Not pictured: Eswatini with median κ_i=3.64.

.

Summarizing the estimates so far, cities in poor nations grow by building *out*, not *up*, as, on the margin, land is more elastically supplied than vertical floorspace. While this greater ability to expand horizontally ameliorates congestion forces brought about by competition for floorspace on extant urban land, households in developing world cities must on average travel farther to commute to work. As cities' commuting costs are higher in the developing world than those in the developed world, this suggests that urban costs may be higher in poor nation cities.

Armed with $\hat{\gamma}_i$, $\hat{\rho}_i$, and $\hat{\tau}_i$ I construct $\hat{\kappa}_i$ for each city in the data with a plug-in estimator. The median city has a κ_i equal to 0.81 and the 10-90 percentiles range is [0.46,1.65]. Urban costs are strongly associated with development: Figure 9 shows the spatial distribution of urban cost elasticities (colored by above/below median) and country level median urban cost elasticities vs income per capita. Most cities above-median κ_i are in sub-Saharan Africa, parts of Latin America, and some parts of South Asia and central China. Median urban cost elasticities fall sharply with development and plateau in the rich world.

This relationship with development is robust to controlling for city size and therefore does not reflect compositional differences in the urban systems of rich and poor countries. Within countries, a 10% increase in the urban cost elasticity is associated with a 5% ($p < 0.01$) decrease in population, as is expected in the

theory. Cities with higher urban costs are on average denser in terms of population, and polycentric cities have lower urban costs on average; see Appendix Table A2. In Appendix Table A3 decomposes global variance in urban cost elasticities into a within-country, between-country and intensive margin ($\frac{1}{1+\rho_i}\frac{1}{1+\gamma_i}$) and extensive margin ($\frac{\rho_i}{1+\rho_i}\frac{\tau_i}{2}$) components, as well as their covariance. About 40% of the variance in urban cost elasticities is cross-country, and both the cross-country and within country components are almost entirely by the extensive margin (34% is between-country extensive margin, while 58% is within-country extensive margin) and the covariance intensive-extensive margin covariance plays a small role between countries (4%) and almost no role within countries.

How do different cities and regions achieve different urban cost elasticities? Appendix Table A4 breaks down the distribution of urban costs elasticities and their components by region. The global technological frontier is in East and Southeast Asia. Countries like Japan, Singapore, and Malaysia have cities with high γ (median 0.8) and modest τ (median 1.14), in contrast to cities in North America, which face steeper floorspace supply curves despite having similarly low commuting costs. Floorspace supply is even less elastically supplied in sub-Saharan Africa (median γ 0.2), but there cities can more easily grow horizontally (median ρ 1.4) despite facing commuting cost elasticities nearly four times as large as those in North America (median τ 4.7).

5 How important are urban costs?

Urban cost elasticities are reduced-form objects that capture both unmodeled geophysical constraints and policy variables. They are therefore subject to the Lucas critique, and may change endogenously as cities invest in transportation infrastructure or reduce regulations that inhibit both vertical and horizontal urban growth. Nonetheless, my framework allows me to assess how these growth barriers have affected economic development and will continue to do so if left unchanged. In this section, I ask three related questions about how urban costs shape each nation's macroeconomy.

First, I assess what GDP per capita growth would have been from 1993 to 2015 had urban costs been lower than observed. Next, I consider what policies reduce urban costs, and whether such policies are cost effective. To do so, I study urban road paving. I estimate the effect of paving roads on commuting cost elasticities, and study counterfactuals in which I pave roads under a fixed budget, subject to different targeting rules. My goal

is to speak directly to policies that lower urban costs.

Finally, I examine how urban costs determine a nation's ability to adapt to shocks to its cities. Cities in the developing world, particularly those in East and Southeast Asia, are threatened by sea level rise (Hsiao, 2025). Which cities are threatened, and to what extent a nation's urban system can adapt to sea level rise, depends on its geography and the ability of its cities to scale and absorb climate migrants.

5.1 Model calibration

To perform this counterfactual analysis, I need to quantify the rest of the model. Quantification follows three steps. First, in addition to the elasticities estimated in Section 4, I select values for the remaining model elasticities from the literature. I then compute trade costs δ_{ij} across all cities globally. Finally, I invert the model by recovering fundamentals (e.g., A_i) that exactly rationalize the data in each year (1993, 2000, and 2015) as a model equilibrium, exactly matching national accounts data from the World Development Indicators for every country. I use the index n to index nations. Calibration details are in Appendix E.

Externally calibrated parameters Table 1 reports all externally calibrated parameter values. I set the elasticity of substitution across urban varieties $\sigma = 5$, which is near the median value across estimates of this elasticity (across countries) in Bajzik et al. (2020) and is used in similar studies. For the elasticity of urban TFP with respect to population, I set $\zeta = 0.04$, which is the average wage elasticity with respect to density in the meta-analysis of Ahlfeldt and Pietrostefani (2019) and is the elasticity of urban TFP with respect to density in Combes et al. (2010). There is some evidence that agglomeration economies may be stronger in the developing world (Chauvin et al., 2017), but the evidence is not conclusive. Estimating agglomeration economies is notoriously difficult (Combes et al., 2011) and not the focus of this study, so I choose the same parameter for all cities. I set $\eta = 0.5$, which is standard for the literature, consistent with the intermediate share of gross output (Jones, 2011).

I set the intercity migration elasticity to 3.3, and across U.S. counties in Monte et al. (2018). The long-run rural-to-urban migration elasticity ε is a key parameter for governing how model simulations behave: it both controls the amount of counterfactual urbanization and how negatively selected on skills the marginal rural-to-urban migrant is. I choose $\varepsilon = 2$, which is between a range of estimates in the literature. On the low

Parameter	Value	Description	Source
σ	5	CES elasticity across city varieties; trade elasticity $\sigma - 1 = 4$	Standard; see e.g., Simonovska and Waugh (2014) and Bajzik et al. (2020)
η	0.5	Firm cost share on intermediate bundle (vs. labor)	(Jones, 2011)
ζ	0.04	Agglomeration elasticity	Ahlfeldt and Pietrostefani (2019)
ψ	3.3	Cross-city migration elasticity	Monte et al. (2018)
ε	2.0	Rural→urban migration elasticity	Middle of 1.2 (Sahai and Bailey, 2022) to 3.3 (Bryan and Morten, 2019)
β_n	–	Share of income spent on floorspace	Calibrated for each country using ICP data.

Table 1: Externally calibrated parameters.

end and in the short-run, Sahai and Bailey (2022) estimate a migration elasticity in India (which has a low urbanization level) of around 1.17 using data from Facebook. On the other hand, Lagakos and Waugh (2013) estimate a much less dispersed distribution of skills in the United States, though Eckert and Peters (2022) estimate $\varepsilon = 1.7$ in the U.S. during the 19th century. However, accounting for selection and the correlation of skills across locations, Bryan and Morten (2019) estimate a migration elasticity of over 3 in Indonesia, and Tombe and Zhu (2019) estimate a migration elasticity, accounting for selection, of 2.5 in China. My choice of 2 coincides with the choice made for skill dispersion in the United States in Hsieh et al. (2019). I set β separately for each country using consumption expenditure data from the World Bank's 2017 International Comparison Program (ICP), and extrapolate using a regression of β on log GDP/capita for countries not covered by the survey. This captures non-homotheticities in the floorspace expenditure function. While income elasticities with respect to floorspace tend to be estimated at less than 1, the share of income spent on floorspace *rises* in GDP per capita, suggesting that food nonhomotheticities strongly dominate housing.[25] This means that poorer countries are *less* sensitive to κ_i than those in the rich world.

Trade costs In the model, each country is in autarky, but within countries, cities trade intermediates with each other. To get trade costs across cities, I first estimate the average drive time between cities in every country using the Open Street Maps routing data, d_{ij}. To convert this to iceberg trade costs, I assume $\delta_{ij} =$

[25]Estimates of the income elasticity of housing demand range from unit elastic to mildly inelastic. In the U.S., both Davis and Ortalo-Magné (2011) and Comin et al. (2021) estimate a number near 1, as implied by Cobb-Douglas, while Albouy et al. (2016) and Finlay and Williams (2022) estimate an elasticity closer to 0.7. In China, Murray and Sun (2017) estimate the same elasticity. Malpezzi (1999) surveys estimates of housing demand price and income elasticities in developing economies from earlier literature that broadly accord with these patterns.

$(1 + d_{ij})^{-\bar{\delta}}$ where d_{ij} is measured in hours.[26] Trade costs only enter the model raised to the power $1 - \sigma$, so I seek to estimate $-\bar{\delta}(\sigma - 1)$. To that end, I estimate a gravity regression with a pseudo-Poisson maximum likelihood estimator (Silva and Tenreyro, 2006) by projecting the value of road-based shipments across cities in the United States onto log travel duration, controlling for origin and destination fixed effects, using the 2017 Commodity Flow Survey (CFS). Appendix Table A11 reports the results of this regression; I obtain a coefficient $\widehat{-\bar{\delta}(\sigma - 1)} = 1.27$. Full details are available in Appendix E.

Inversion Armed with the externally calibrated parameters, trade costs, city-level output and population estimates, and country-level GDP (in PPP-adjusted constant 2015 USD), the share of value added in agriculture, and employment measures in agriculture and nonagriculture from the World Development Indicators, I am able to invert the model for each country to find values of $\bar{Z}_i^y$, A_i, Z_i^H, and Z_i^X which exactly rationalize these as an equilibrium of the model.

5.2 Growth under lower urban costs

To understand the importance of urban costs for economic development, and to illustrate mechanisms, I study how global growth would have unfolded from 1993 to 2015 were urban costs lower. To do so, I feed in observed productivity, amenity, and population growth that exactly move the 1993 economy to its 2015 values, and then ask how this growth would have unfolded were urban costs lower.[27] This is the right kind of counterfactual to study when changing technological primitives that appear as *elasticities*. As Greaney (2026) notes, elasticities govern how the economy reacts to shocks; naively changing them without feeding in new shocks can rotate supply curves in a way that 'unanchors' them from the observed equilibrium. By how much and in what direction these curves shift depends on the scale of the data, thus rendering the counterfactual predictions unidentified. By feeding in growth shocks to an economy 'anchored' at its 1993 value, the simulations I perform are identified.

I focus on GDP, rather than welfare, because the elasticities represent technological constraints that directly

[26] I prefer travel duration, opposed to distance, as routing sometimes requires complex mode changes, e.g., via ferries in Indonesia, see Appendix Figure A9.

[27] I use the city-level GDP data from Kummu et al. (2025) to fit the cross-city distribution of income within countries, and scale it up to hit WDI aggregates in this counterfactual. I do so instead of relying on nightlights as the DMSP-OLS nightlights are very noisy for poorly-lit cities in the developing world in 1993 and subsequently predict unreasonable productivity growth over this period.

affect welfare (through housing and commuting costs). Lowering urban costs would therefore mechanically increases welfare, and more so in the developing world, where urban costs are higher. Urban costs, however, only indirectly affect GDP through how they shape the spatial distribution of economic activity. In other words, I am isolating only the general equilibrium effects of lowering urban costs, rather than their direct effects. The changes to GDP per capita I measure only stem from how changes to the size of the urban sector and the distribution of people across cities. Thus, these counterfactuals abstract from the gains associated with how lower urban costs affect city-level productivity by improving the spatial allocation of labor *within* cities (see, e.g., Zárate, 2022; Severen, 2023; Tsivanidis, 2023; Chen et al., 2024; Sorin, 2025). This separation between outcome and primitive mirrors the approach taken in Bryan and Morten (2019), who study how lower migration barriers affect aggregate productivity and not welfare.

In particular, I rescale every city's κ_i by a country-specific factor so that the country's median urban cost elasticity aligns with the median level observed in East and Southeast Asia (0.56, roughly the global technological frontier). This preserves within-country heterogeneity and the correlation between κ_i and city-level characteristics like productivity. I then feed in the productivity, amenity, and population growth shocks that would in absence of a change to κ_i, move the 1993 economy to the 2015 exactly. I run this counterfactual for every country in the data.

Table 2 shows the results of this counterfacutal. Each row reports population-weighted statistics by World Bank income group classification over this period. The first two columns report factual urbanization and GDP/capita growth over this 22 year period. The developing world urbanize substantially and GDP/capita growth modestly exceeded that of the rich world. Cross-country income inequality fell modestly (σ-convergence, the standard deviation of log GDP/cap across countries, fell by -4.55 log points).

Columns 3-5 report the change in urban output per urban worker (i.e., urban productivity) under three different counterfactuals. In column (3), I fix productivity and rural-to-urban migration and record the aggregate productivity effects of allowing workers to reallocate across the urban system given observed productivity growth. The spatial allocation of labor modestly improves in high income countries but actually *falls* in low income countries. This is because κ_i is negatively correlated with city productivity in the developing world but there is almost no correlation in the rich world (see Appendix Table A5). Relaxing urban costs allows low-productivity cities to scale, attenuating the growth of the highest-productivity cities. The urban

	Factual, 1993–2015		Δ vs factual				
Income group	Δurb (pp)	GDP/cap (%)	Fixed urb (%)	+R→U (%)	+agglom (%)	Δurb (pp)	GDP/cap (%)
High	+3.8	+52.0	+0.08	−0.03	+0.00	+0.20	+0.17
Upper middle	+9.8	+70.1	−0.05	−0.58	−0.51	+0.93	+0.34
China	+28.1	+549.7	−0.12	−1.95	−1.72	+2.80	+7.34
Lower middle	+12.8	+98.6	+0.03	−1.29	−1.12	+1.53	+1.11
India	+18.4	+181.2	+0.67	−0.46	−0.20	+1.36	+3.88
Low	+9.5	+75.2	−0.08	−5.33	−4.71	+4.55	+6.40
σ-convergence	−2.79	−4.55	+0.07	+1.08	+0.97	−0.82	−0.50

Table 2: Country-level statistics for factual urbanization and income growth 1993-2015 versus counterfactual growth under lower urban costs. Columns 1 and 2 report factual gains, and columns 3-7 report changes from that baseline in the counterfactual. Each country's urban cost elasticity distribution is dilated (or widened) so that its median is 0.56, the median for East/Southeast Asian cities. All columns report population-weighted means within World Bank income group, with China and India shown separately. Urbanization is measured as the share of employment in the urban sector. Columns 3-5 show the percent change in urban output per worker, i.e., urban productivity. Column 3 fixes the size of the urban sector (only allowing for productivity/amenity changes). Column 4 allows for population growth and rural to urban migration. Column 5 allows for agglomeration effects. Columns 6 and 7 report country level changes in urbanization and total GDP per capita (measured in constant 2015 dollars. Total GDP adjusts for the agricultural sector by holding agricultural productivity fixed. The final row measures σ-convergence: the standard deviation of the percent of employment in the urban sector or 100$\times$ the log of GDP/capita.

system becomes less biased towards urban 'giants' (Ades and Glaeser, 1995) in the developing world, but the cost of this is reduced output. Allowing for rural-to-urban migration exacerbates this as the marginal rural-to-urban migrant has lower skills than the average urban worker, so urban productivity falls (Column 4). However, allowing for agglomeration effects attenuates these losses (Column 5). Nonetheless, GDP per capita growth is quite high in the developing world: GDP per capita growth would have been 6.4% points higher, compared to only 0.17% in the rich world.[28] This is because rural-urban wage gaps are tremendous in the developing world, and the share of employment in the urban sector would have grown by 4.55 percentage points. Consequently, cross-country GDP per capita dispersion would have declined modestly more increasing σ-convergence by 11% of the factual 1993-2015 decline. Appendix Figure A3 maps the changes in urbanization (left panel) and the changes to GDP (right panel) in this counterfactual.

To better understand these mechanisms, I zoom in on India in Figure 10. In the left panel, I show how the distribution of population across Indian cities would have evolved under factual 1993-2015 productivity growth, holding fixed the size of the population and rural to urban migration. Broadly, second-tier cities grow more

[28] These small gains in the rich world are consistent with the limited impact of housing supply regulations on aggregate productivity through the spatial reallocation channel, as Greaney (2026) finds when correcting the analysis done in Hsieh and Moretti (2019).

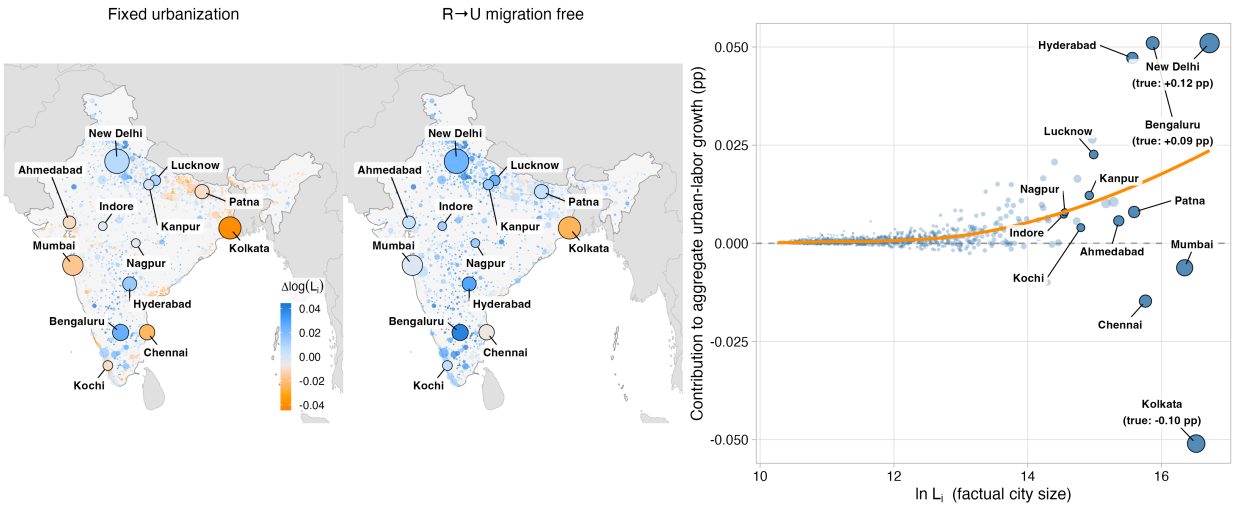

Figure 10: 1993-2015 urbanization in India with lower urban cost elasticities. Each panel shows city-level changes in log employment 1993-2015 between factual growth and counterfactual growth with the κ_i distribution so its median is 0.56. Polycentric cities are aggregated to the UCDB agglomeration level. Left panel: the change in the spatial distribution of urban population without population change or rural-to-urban migration. Middle: city-level population changes allowing for population growth and rural-to-urban migration. Right: each city's contribution to the total growth of the urban sector versus its 2015 factual population. A LOESS smoother is shown in orange.

than India's megacities. Population declines in regional anchors like Kolkata and Patna spill over to satellite cities by reducing their productivity through the intercity trade network. By contrast, Delhi, Hyderabad, and Bengaluru and their surrounding cities grow. In the middle panel, I show the population change allowing for the full population change, including endogenous rural-to-urban migration and agglomeration effects. Most cities grow, but the distribution of growth roughly follows the spatial pattern under fixed urbanization. In the right scatter, I show the growth contribution of each city to overall urbanization: there is substantial reallocation across the major cities, while most second-tier cities grow.

5.3 The price of urbanization

A convenient way to summarize how urban costs interact with a nation's urban system – its cities' productivity, amenities, and spatial position – is what I call *the price of urbanization*. I ask how big a uniform wage subsidy s would have to be to induce a 1% increase in the size of the urban sector, in excess of that implied by a given migration elasticity. That is, without urban costs, a standard location choice model would imply that to inducing a 1% change in urbanization requires a $s_{\text{logit}} = 0.01/\varepsilon(1 - L_U/\bar{L})$ increase in urban wages

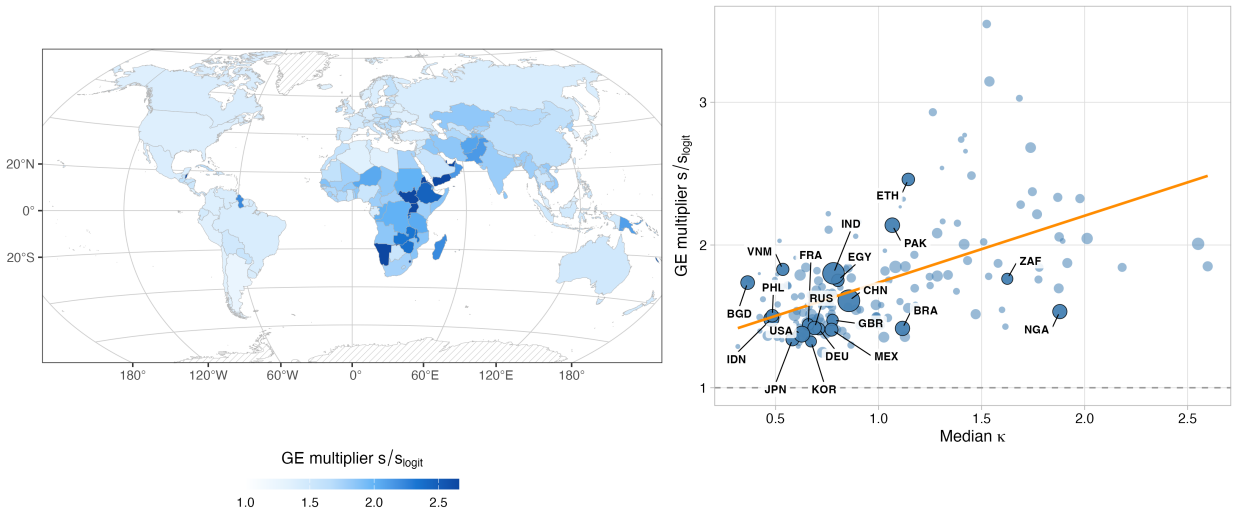

Figure 11: The price of urbanization: the country-level general equilibrium multiplier s/s_{logit}, where s is the uniform urban wage subsidy required to increase the urban population by 1% in the quantified model. s_{logit} is the wage subsidy implied by a simple location choice model absent urban costs and spatial heterogeneity within the urban sector. Left: map of s/s_{logit}. Right: s/s_{logit} versus a country's median urban cost elasticity, with the OLS fit in orange. The scatter omits 2 outliers, Qatar and Eswantini.

to increase the urban sector's size by 1%; I define the price of urbanization as s/s_{logit}.

I compute the price of urbanization for every country in the 2015 calibration of the model. Figure 11 displays a map of this 'price' (left panel) . The price of urbanization is about 1.5 times larger than implied by the migration elasticity and urbanization share in much of the rich world, but this multiplier is closer to two in much of the developing world. This general equilibrium multiplier is strongly correlated with a country's median urban cost elasticity (right panel). Countries with higher urban cost elasticities require a disproportionately larger 'urban pull' (or, conversely, agricultural 'push') to move its population into its cities.

5.4 Road-paving and the value of targeting

The analysis of Section 5.2 revealed that while lower urban costs can foster economic development allowing cities to scale and absorb rural-to-urban migrants, *which* cities grow determines the aggregate productivity effects. While urban cost elasticities are on average much higher in the developing world, within developing countries, the most productive cities already face lower urban costs.

Policy aimed at lowering urban costs must account for this fact: targeting cities with high urban costs a priori

may induce misallocation by allowing the least productive cities in a nation's urban system to grow at the expense of its most productive. In this section, I study urban road paving as one type of policy aimed at lowering cities' urban costs.

Some attention in the development literature has focused on the impacts of road paving.[29] McIntosh et al. (2018) studies the impact of road paving in urban Mexico, finding that the return on investment on road paving, as measured by its capitalization in real estate values was around 100%, while Gonzalez-Navarro and Quintana-Domeque (2016) find a more modest 9% return on investment when studying road paving in a single city in Mexico.[30] I focus on the general equilibrium effects of urban road paving on a long time scale. Increased road paving can lower commuting costs and therefore urban costs, allowing cities to better scale, thus influence the spatial distribution of economic activity and the overall size of the urban sector.

Urban roadways in the developing world are considerably under-paved. To see this, I extract road data for every city in the world using OpenStreetMaps (OSM, via Boeing, 2024). Using these data, I compute the share of urban roadways (in km) that are unpaved.[31] The left panel of Figure 12 show this data globally: nearly all urban roadways are paved in the rich world but only 50% of urban roadways are paved in Nigeria. In the right panel, I plot the relationship between the log of τ_i and the share of a city's roads that are paved net. Cities with more paved roads tend to have lower commuting cost elasticities on average, consistent with the model.

The relationship in the right panel of Figure 12 may not be causal. For example cities with low τ_i may be bigger in equilibrium and have larger municipal budgets to invest in road paving, reflecting reverse causality. To better estimate a causal relationship between paving roads and the commuting cost elasticity τ_i, I take a selection-on-observables approach to identification. I regress $\ln \tau_i$ on the share of roads paved, controlling for additional characteristics of a city's road network (its size, intersection density, lane count, and road width,

[29] The majority of this literature focuses on road paving to foster market access between cities and agricultural regions. For example, Storeygard (2016) studies how road surface affects transportation costs and the income of cities in sub-Saharan Africa, while Aggarwal (2018) studies the effect of paved roads on market integration in India villages. My framework allows me to assess the role of these intercity transportation costs on national income, but as the focus of this paper is on urban costs, I omit this analysis.

[30] Additionally, both Olken (2007) and Wong et al. (2013) study the efficacy of local governments in providing paved roads in Indonesia and China, respectively, but these studies focus on the role of corruption and local governance in effectively implementing this policy.

[31] In practice, I compute the share of roadways with a surface tag for which the tag is dirt, sand, ground, gravel, earth, grass, mud, compacted, rock, wood, clay, soil, unpaved, fine gravel, terra, and all other tags I count as paved. I do this because the share of roads that have any tag at all is much lower in developing world cities compared to rich world cities. Likely, my calculation overstates how many roadways are paved if paved roads are more likely to have a surface tag.

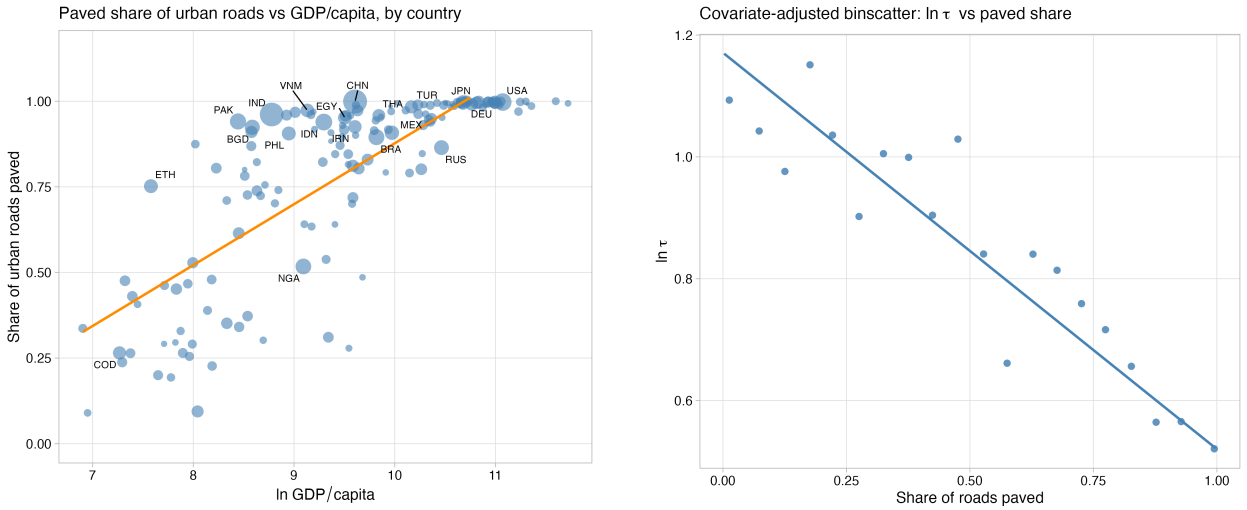

Figure 12: Urban road paving and commuting costs around the world. Left: the share of urban roads paved (OSM data extracted within UCDB shapefiles) against log GDP per capita. Each point is a country sized by population and the orange line reflects the OLS fit. Right: binsreg of $\ln \tau_i$ against the paved road share.

connectivity, and the share of its roads with a surface tag) as well as its city-level characteristics (population density, population, nightlights per capita, and in indicator for whether it is the largest city in the country).[32] Conditioning on these covariates, I find that a 10pp increase in the share of roads paved would lower τ_i by 5.6%; see Appendix Table A6.

I use this relationship to map simulated road paving around the world to changes in the urban cost elasticity in a variety of counterfactuals. To pin down the cost of urban road paving, I use cost estimates from the World Bank's Road Costs Knowledge System (ROCKS) database, which estimates that upgrading 1km of road to a bituminous two land road in 2015 USD ranges from \$92K to \$1.35M. In Table 3, I compare three scenarios, asking how economic development would have unfolded from 1993 to 2015 with more road paving. First, in Panel A, I estimate cost of paving every urban roadway globally as a share of 1993 GDP, and then assess the return on investment of this policy of this by measuring the counterfactual change in GDP induced by paving these roads. For lower-middle and low-income countries, the median cost estimate across countries (using the 'central' price of road paving from ROCKS) for paving all urban roadways is between 8-36% of baseline GDP, but would have only produced a return of \$0.02 to \$0.04 in general-equilibrium gains for every dollar invested.[33] These low gains are because most of the urban roadways targeted are in unproductive cities with

[32] Ideally, I would control for country fixed effects, Netting out country fixed effects eliminates most variation in the share of roads paved, and the within-country relationship is noisy and sometimes flat or wrong-signed in rich world countries.

[33] This does not account for the welfare gain of paving roads that would be captured by a WTP measure nor does it account for how paving cities roads improves the productivity of the city itself. It merely reflects the additional gains from urban road paving

		Cost as % of 1993 GDP			ROI (2015 GDP gain / cost)		
Income group	N	Low	Central	High	Low	Central	High
Panel A: Close the paving gap (pave every paveable territory, no cap)							
High	60	0.00	0.02	0.06	1.55	0.33	0.11
Upper middle	48	0.37	1.75	5.46	0.19	0.04	0.01
Lower middle	47	1.60	7.46	23.2	0.20	0.04	0.01
Low	21	7.75	36.3	113	0.08	0.02	0.01
Panel B: Y/L-targeted ($\geq$100K pop), 10% of close-gap cost, 10pp walk to 75%							
High	60	0.00	0.00	0.00	-0.46	-0.10	-0.03
Upper middle	48	0.00	0.00	0.02	0.11	0.02	0.01
Lower middle	47	0.03	0.14	0.45	0.39	0.08	0.03
Low	21	0.54	2.51	7.81	0.24	0.05	0.02
Panel C: Least-paved-targeted ($\geq$100K pop), 10% of close-gap cost, 10pp walk to 75%							
High	60	0.00	0.00	0.00	-0.64	-0.14	-0.04
Upper middle	48	0.00	0.00	0.02	-0.55	-0.12	-0.04
Lower middle	47	0.03	0.14	0.45	0.04	0.01	0.00
Low	21	0.54	2.51	7.81	0.05	0.01	0.00

Table 3: Counterfactual analysis of paving cities roads. Each panel shows a different simulation globally. Panel A shows the costs (as a percent of 1993 GDP) and ROI estimates based on paving all urban roads globally and then feeding in 1993 to 2015 productivity and amenity shocks as well as population growth. Each row displays the country-level median for the World Bank income group. Panel B targets the policy by paving roads in the highest productivity cities (with over 100K persons) until 75% of its roads are paved, working down the productivity distribution until 10% of the paving gap is closed. Panel C repeats this exercise, but targets the least paved cities.

low shares of paved roads at baseline.

In panel B, I implement a targeted policy in which I pave roads in the most productive cities first (measured by 1993 GDP/cap) paving roadways until 75% of urban roadways are paved, and working down the productivity distribution (for cities over 100K persons) until 10% of the paving gap is closed. For low- and lower-middle income countries this only costs 2.5% and 0.1% of baseline GDP, respectively, and doubles the ROI. In the rich world, targeting is still ineffective (the ROI is negative) because the almost all of the most productive cities are entirely paved, and so even the most productive city with incomplete paving is relatively unproductive. In Panel C, I show the results from a policy 'mistake' of targeting the least-paved cities. Directing resources to the least paved cities incentivizes workers to stay in the least productive cities, making the policy in the developing world less than 1/5th as effective as productivity targeting. Thus, a policy of 'fixing where it

due to altering the distribution of labor across cities and agriculture.

worse' or 'levelling up' aimed at reducing geographic disparities can be massively misallocative.

5.5 Urban costs inhibit resilience to sea-level rise

Climate change threatens cities around the world through increased pressure from rural-to-urban migration, increased flood risk (Kocornik-Mina et al., 2020; Bearpark et al., 2024), and through sea level rise (Hsiao, 2025).[34] Sea level rise destroys both developable land and built infrastructure in coastal cities. I examine how urban costs determine the efficacy of inter-city migration as an adaptation mechanism to the this threat. Thus, this analysis sets aside the costs and benefits associated with cities adapting *in situ* (Gandhi et al., 2022; Hsiao, 2023; Hsiao, 2024).

To perform this analysis, I use the Sea Level Rise and Urban Infrastructure Data Set (SUDS, Hsiao, 2025), which reports the amount of urban land, and the number of facilities (schools, hospitals, clinics, and seven other OSM-tagged categories), and roadway length inundated at every 0.1 m of sea level rise up to 5m of sea level rise. To simulate sea level rise damages in the model, I regress the log of urban amenities A_i on the log number of facilities and roadway length, controlling for city size, density, income per capita, all net of country fixed effects (Appendix Table A7). Treating this association as causal allows me to map loss in facilities and roadways from flooding into amenity shocks in the model. In addition, I compute the effect of lost land on floorspace prices, similar to the approach taken in Desmet et al. (2021).[35] Importantly, the shocks I feed into the model have no direct effect on productivity. This likely renders the GDP loss estimates to be very conservative, as it isolates only the general equilibrium effect of shifting the urban population distribution across cities and the agricultural sector in reaction to pure amenity and urban cost shocks.

In the left panel of Figure 13 I show the global GDP loss under sea level rise of up to 2 meters.[36] East and Southeast Asian nations are most exposed to this shock. In the right panel I show which cities experience coastal flooding in Southeast Asia, and how the population reallocates across the urban system. I display these changes under 1.5m of sea level rise, as the Mekong Delta in Vietnam, Manila Bay, Phillipines, and

[34] There is some evidence that climate change is driving urbanization in sub-Saharan Africa (Henderson et al., 2017) and Mexico (Nawrotzki et al., 2017).

[35] Through the lens of the model decreasing the amount of land by 1% increases floorspace prices by $1/(1+\gamma_i)$.

[36] I exclude the Netherlands from this calculation as it becomes so inundated in this simulation that it entirely de-urbanizes. The Netherlands alone drives $\sim 40\%$ of the global GDP loss at 2m of sea level rise.

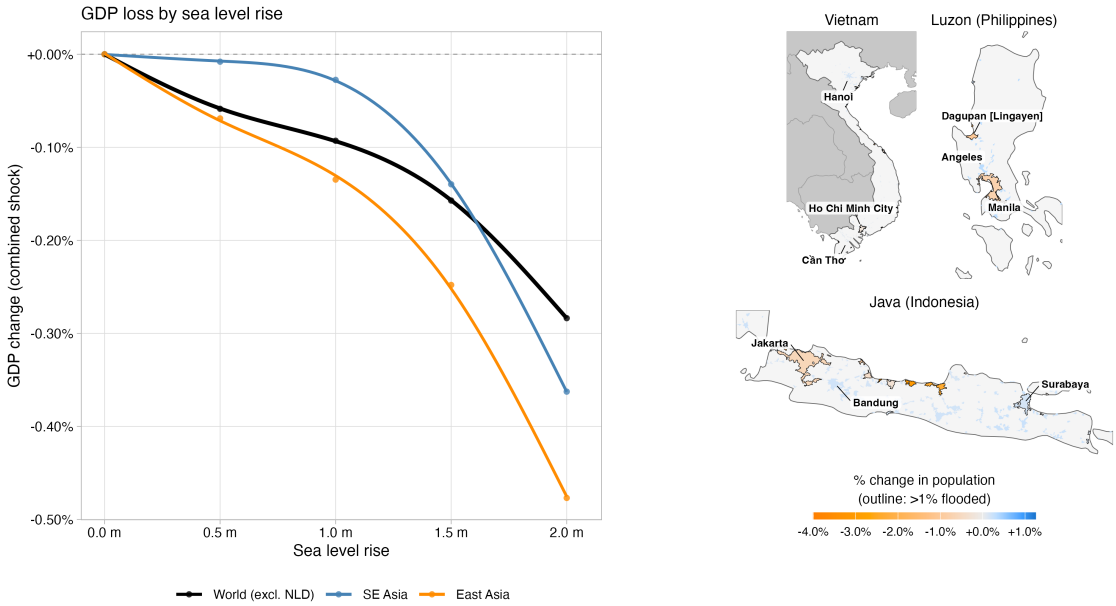

Figure 13: The effects of sea level rise on GDP. Left: Global GDP loss from sea level rise (excluding the Netherlands) and in Southeast (SE) Asia and East Asia. Right: city-level population changes in several key Southeast Asian regions.

coastal Java, Indonesia, all begin to flood in concert. Several major coastal cities are threatened, and the population generally moves away from the coast. On Java, Surabaya faces some flooding but in fact grows in terms of population, absorbing climate refugees from Jakarta and elsewhere.

To isolate the role urban costs play in governing urban systems' resilience to this shock, I project the simulated GDP loss across countries at 2m of sea level rise onto exposure (% of urban land flooded) and its interactions with the level of urbanization and a country's median urban cost elasticity (Appendix Table A8). At the median κ_i, the elasticity of GDP loss with respect to exposure is 0.1, implying that for every 10pp increase in urban land flooded is associated with a 1% GDP loss, and this nearly doubles moving from the median κ_i to the 75th percentile of κ_i across countries. In other words, countries with higher urban cost elasticities are less able to adapt to sea level rise, as their displaced workers cannot be absorbed cheaply into surviving cities.

6 Conclusion

This paper argues that developing countries have failed to reap the full benefits of urban agglomeration because the costs of urban scale are too high. This paper measures those costs and quantifies their macroeconomic burden.

To quantify differences in urban technology across space, and how they matter for economic development, I first develop an econometric approach to partitioning polycentric urban agglomerations into monocentric 'cities' and identifying their urban cores. I find that, save for a few megacities in the developing world, polycentricity is primarily a rich-world phenomenon. Using this new database of cities, I document basic facts about cities which suggest that urban costs are high in low-income nations, relative to the rich world. Cities in developing countries have accommodated demand for floorspace by building out, rather than up, suggesting that there are technological or regulatory barriers to vertical expansion in developing nations. Yet, developing world cities crowd physical mass in their urban cores: the gradient of built height with respect to the distance to a city's downtown monotonically flattens in national income; city skylines are steeper in the developing world. Urban theory suggests that a greater incentive to build in the core, rather than the periphery, is driven by high commuting costs.

To discipline what urban costs are, and how I can use geospatial data to measure the key structural parameters that govern cities' ability to scale, I develop a general equilibrium model of cities and an agricultural sector linked through trade and migration. In the model, cities differ in terms of their urban technology: their ability to build vertically, horizontally, and the costs of commuting within them. All three components of a city's urban technology govern its urban cost elasticity. The urban cost elasticity is a land-supply elasticity weighted average of congestion forces that operate through the floorspace market and through transportation costs. The model implies several estimating equations to recover components of the urban technology.

To take these equations to the data, I combine my city-partitioning estimator with generalized random forests that learn how urban cost parameters vary with remotely sensed observables. My estimates are consistent with the patterns described above: floorspace supply is less elastic in low-income nations, but land is more elastically supplied in low-income nations. However, commuting costs are much higher in low-income nations. Consequently, urban cost elasticities are nearly twice as high in low-income nations' cities compared

to those in high-income nations.

I then quantify my model by calibrating it to nearly every country in the world. I first ask how growth from 1993–2015 would have unfolded had urban cost elasticities been at the global technological frontier. Feeding in productivity shocks that exactly rationalize the observed path of growth, GDP per capita would have grown by an additional 6.4% over this period relative to less than 1% in the rich world, modestly reducing international income inequality. These gains accrue only through the general equilibrium impact of urban costs on income, which operate by changing the spatial distribution of population across cities and the overall size of the urban sector. I develop a simple statistic to understand how the rich spatial heterogeneity and general equilibrium interactions in the model shape a nation's ability to urbanize: the 'price of urbanization:' how much larger a uniform urban wage subsidy would have to be to induce a 1% increase in urbanization, relative to a partial-equilibrium benchmark using the rural-to-urban migration elasticity. This multiplier is around 2 in developing nations, compared to only 1.5 in rich nations, suggesting that high urban costs have slowed the pace of structural transformation in the developing world.

To understand what available policies can reduce urban costs, I explore the efficacy of urban road paving around the world. The efficacy of this intervention depends on how it is targeted: many high-urban cost cities are unproductive, and so urban policy aimed at reducing spatial inequality can be inherently misallocative, redirecting resources to the least productive systems and holding back aggregate economic development. Instead, well-targeted policy can more than double the return on investment.

Finally, I study how a nation's urban costs shape its resilience to the climate risks that threaten its cities. I do so by examining sea level rise, to which East and Southeast Asian cities are uniquely exposed. Simulating damages from sea level rise in the model, I find that nations with lower urban costs can better adapt to sea level rise, all else equal.

The framework and estimates I have provided could additionally be leveraged to think about several questions in macroeconomics and economic development. First, the model could be augmented to include international trade, and used to study how a nation's openness to trade affects the urbanization process. Second, the framework could be used to better understand urbanization via structural transformation by incorporating richer heterogeneity with the agricultural sector, internal bilateral migration costs, and nonhomothetic preferences.

Overall, this study highlights the importance of city-level characteristics in explaining differences in macroeconomic aggregates. Nations with better urban technology have higher aggregate productivity because they can better reap the benefits of cities' productivity advantages. This underscores the importance of research that focuses on how to improve cities around the world. Cities are rapidly changing, but several important features of cities have been absent in this study and demand further attention. First, modes of transit vary across cities, with people in the developing world increasingly using informal transit options and motorcycles. Policy that allows cities to better adapt to these changes in transportation technology may be an effective way of reducing transportation costs. Second, many developing cities have slums, which governments routinely clear. Whether these slums decrease urban productivity or provide access to cities for people who otherwise would not live in them remains an open question (Monge-Naranjo et al., 2022). As cities are replete with externalities, understanding the costs and benefits of urban policy interventions remains an important agenda. This study has shown that differences in urban costs across cities and nations have large macroeconomic consequences: in short, when it comes to improving the quality of cities, the stakes are large.

References

Ades, Alberto F and Edward L Glaeser (1995). "Trade and circuses: explaining urban giants". In: *The Quarterly Journal of Economics* 110.1, pp. 195–227.

Aggarwal, Shilpa (2018). "Do rural roads create pathways out of poverty? Evidence from India". In: *Journal of Development Economics* 133, pp. 375–395.

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

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani (2019). "The economic effects of density: A synthesis". In: *Journal of Urban Economics* 111, pp. 93–107.

Akbar, Prottoy, Victor Couture, Gilles Duranton, and Adam Storeygard (2023a). "Mobility and congestion in urban India". In: *American Economic Review* 113.4, pp. 1083–1111.

— (2023b). *The fast, the slow, and the congested: Urban transportation in rich and poor countries*. Tech. rep. National Bureau of Economic Research.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu (2016). *Housing demand, cost-of-living inequality, and the affordability crisis*. Tech. rep. National Bureau of Economic Research.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *The Quarterly Journal of Economics* 129.3, pp. 1085–1140.

— (2022). "The welfare effects of transportation infrastructure improvements". In: *The Review of Economic Studies* 89.6, pp. 2911–2957.

Allen, Treb, Costas Arkolakis, and Xiangliang Li (forthcoming). "On the Equilibrium Properties of Spatial Models". In: *American Economic Review: Insights*. Forthcoming.

Athey, Susan, Julie Tibshirani, and Stefan Wager (Apr. 2019). "Generalized Random Forests". In: *The Annals of Statistics* 47.2, pp. 1148–1178. DOI: 10.1214/18-AOS1709.

Au, Chun-Chung and J Vernon Henderson (2006). "Are Chinese cities too small?" In: *The Review of Economic Studies* 73.3, pp. 549–576.

Bajzik, Josef, Tomas Havranek, Zuzana Irsova, and Jiri Schwarz (2020). "Estimating the Armington elasticity: The importance of study design and publication bias". In: *Journal of International Economics* 127, p. 103383.

Balboni, Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi (2020). "Transportation, gentrification, and urban mobility: The inequality effects of place-based policies". In: *Preliminary Draft* 3.

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

Bearpark, Tom, Archana Patankar, and Ashwin Rode (2024). "Rainfall and death in a developing megacity". In: *Working paper*.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud (2014). "Productive cities: Sorting, selection, and agglomeration". In: *Journal of Political Economy* 122.3, pp. 507–553.

Boeing, Geoff (2024). "Modeling and Analyzing Urban Networks and Amenities with OSMnx". In: *Working paper*.

Bordeu, Olivia (2023). "Commuting infrastructure in fragmented cities". In: *Job Market Paper, University of Chicago Booth School of Business* 6.

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Brueckner, Jan K (1982). "A note on sufficient conditions for negative exponential population densities". In: *Journal of Regional Science* 22.3, pp. 353–359.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2014). "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh". In: *Econometrica* 82.5, pp. 1671–1748.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis (2020). "Cities in the developing world". In: *Annual Review of Economics* 12, pp. 273–297.

Bryan, Gharad and Melanie Morten (2019). "The aggregate productivity effects of internal migration: Evidence from Indonesia". In: *Journal of Political Economy* 127.5, pp. 2229–2268.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte (2018). "The impact of regional and sectoral productivity changes on the US economy". In: *The Review of economic studies* 85.4, pp. 2042–2096.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng (2024). "On binscatter". In: *American Economic Review* 114.5, pp. 1488–1514.

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

Chen, Liming, Rana Hasan, Yi Jiang, and Andrii Parkhomenko (2024). "Faster, taller, better: Transit improvements and land use policies". In: *Journal of Development Economics* 171, p. 103322.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney K. Newey, and James Robins (2018). "Double/Debiased Machine Learning for Treatment and Structural Parameters". In: *The Econometrics Journal* 21.1, pp. C1–C68. DOI: 10.1111/ectj.12097.

Coeurdacier, Nicolas, Florian Oswald, and Marc Teignier (2025). "Structural change, land use and urban expansion". In: *Review of Economic Studies*, rdaf091.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2011). "The identification of agglomeration economies". In: *Journal of economic geography* 11.2, pp. 253–266.

— (2019). "The costs of agglomeration: House and land prices in French cities". In: *The Review of Economic Studies* 86.4, pp. 1556–1589.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2010). "Estimating agglomeration economies with history, geology, and worker effects". In: *Agglomeration economics*. University of Chicago Press, pp. 15–66.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "The empirics of agglomeration economies". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 247–348.

Comin, Diego, Danial Lashkari, and Martí Mestieri (2021). "Structural change with long-run income and price effects". In: *Econometrica* 89.1, pp. 311–374.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin (2020). "Fast Poisson estimation with high-dimensional fixed effects". In: *The Stata Journal* 20.1, pp. 95–115.

Davis, Morris A and François Ortalo-Magné (2011). "Household expenditures, wages, rents". In: *Review of Economic Dynamics* 14.2, pp. 248–261.

Deffebach, Peter, David Lagakos, Yuhei Miyauchi, and Eiji Yamada (2025). *The spatial distribution of income in cities: New global evidence and theory*. Tech. rep. National Bureau of Economic Research.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss (2021). "Evaluating the Economic Cost of Coastal Flooding". In: *American Economic Journal: Macroeconomics* 13.2, pp. 444–486. DOI: 10.1257/mac.20180336.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg (2017). "Asia's Geographic Development". In: *Asian Development Review* 34.2, pp. 1–24.

— (2018). "The geography of development". In: *Journal of Political Economy* 126.3, pp. 903–983.

Desmet, Klaus and Esteban Rossi-Hansberg (2013). "Urban accounting and welfare". In: *American Economic Review* 103.6, pp. 2296–2327.

Eckert, Fabian and Michael Peters (2022). *Spatial structural change*. Tech. rep. National Bureau of Economic Research.

Esch, Thomas, Elisabeth Brzoska, Stefan Dech, Benjamin Leutner, Daniela Palacios-Lopez, Annekatrin Metz-Marconcini, Mattia Marconcini, Achim Roth, and Julian Zeidler (2022). "World Settlement Footprint 3D-A first three-dimensional survey of the global building stock". In: *Remote sensing of environment* 270, p. 112877.

Esch, Thomas, Julian Zeidler, Daniela Palacios-Lopez, Mattia Marconcini, Achim Roth, Milena Mönks, Benjamin Leutner, Elisabeth Brzoska, Annekatrin Metz-Marconcini, Felix Bachofer, et al. (2020). "Towards a large-scale 3D modeling of the built environment—joint analysis of TanDEM-X, Sentinel-2 and open street map data". In: *Remote Sensing* 12.15, p. 2391.

Finlay, John and Trevor C. Williams (2022). "Housing Demand, Inequality, and Spatial Sorting". In: Working Paper.

Frolking, Steve, Richa Mahtta, Tom Milliman, Thomas Esch, and Karen C Seto (2024). "Global urban structural growth shows a profound shift from spreading out to building up". In: *Nature Cities*, pp. 1–12.

Fujita, Masahisa and Hideaki Ogawa (1982). "Multiple equilibria and structural transition of non-monocentric urban configurations". In: *Regional science and urban economics* 12.2, pp. 161–196.

Gaduh, Arya, Tadeja Gračner, and Alexander D Rothenberg (2022). "Life in the slow lane: Unintended consequences of public transit in Jakarta". In: *Journal of urban economics* 128, p. 103411.

Gandhi, Sahil, Matthew E Kahn, Rajat Kochhar, Somik Lall, and Vaidehi Tandel (2022). *Adapting to flood risk: Evidence from a panel of global cities*. Tech. rep. National Bureau of Economic Research.

Gollin, Douglas, Martina Kirchberger, and David Lagakos (2021). "Do urban wage premia reflect lower amenities? Evidence from Africa". In: *Journal of Urban Economics* 121, p. 103301.

Gollin, Douglas, David Lagakos, and Michael E Waugh (2014). "The agricultural productivity gap". In: *The Quarterly Journal of Economics* 129.2, pp. 939–993.

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

Greaney, Brian (2026). "Housing constraints and spatial misallocation: Comment". In: *American Economic Journal: Macroeconomics* 18.2, pp. 409–428.

Harari, Mariaflavia (2020). "Cities in bad shape: Urban geometry in India". In: *American Economic Review* 110.8, pp. 2377–2421.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). "Random forests". In: *The elements of statistical learning: Data mining, inference, and prediction*. Springer, pp. 587–604.

Henderson, J Vernon (1974). "The sizes and types of cities". In: *The American Economic Review* 64.4, pp. 640–656.

Henderson, J Vernon, Adam Storeygard, and Uwe Deichmann (2017). "Has climate change driven urbanization in Africa?" In: *Journal of development economics* 124, pp. 60–82.

Herrendorf, Berthold and Todd Schoellman (2018). "Wages, human capital, and barriers to structural transformation". In: *American Economic Journal: Macroeconomics* 10.2, pp. 1–23.

Hsiao, Allan (2023). "Sea level rise and urban adaptation in Jakarta". In: *Working paper*.

— (2024). "Sea level rise and urban inequality". In: *AEA Papers and Proceedings*. Vol. 114. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 47–51.

— (2025). "Sea Level Rise and Urban Infrastructure". In: *AEA Papers and Proceedings*. Vol. 115, pp. 557–562.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow (2019). "The allocation of talent and us economic growth". In: *Econometrica* 87.5, pp. 1439–1474.

Hsieh, Chang-Tai and Enrico Moretti (2019). "Housing constraints and spatial misallocation". In: *American economic journal: macroeconomics* 11.2, pp. 1–39.

Jedwab, Remi, Prakash Loungani, and Anthony Yezer (2021). "Comparing cities in developed and developing countries: Population, land area, building height and crowding". In: *Regional Science and Urban Economics* 86, p. 103609.

Jones, Charles I (2011). "Intermediate goods and weak links in the theory of economic development". In: *American Economic Journal: Macroeconomics* 3.2, pp. 1–28.

Kocornik-Mina, Adriana, Thomas KJ McDermott, Guy Michaels, and Ferdinand Rauch (2020). "Flooded cities". In: *American Economic Journal: Applied Economics* 12.2, pp. 35–66.

Kreindler, Gabriel, Arya Gaduh, Tilman Graff, Rema Hanna, and Benjamin A Olken (2023). *Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta*. Tech. rep. National Bureau of Economic Research.

Krugman, Paul and Anthony J Venables (1995). "Globalization and the Inequality of Nations". In: *The quarterly journal of economics* 110.4, pp. 857–880.

Kummu, Matti, Maria Kosonen, and Sina Masoumzadeh Sayyar (2025). "Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022". In: *Scientific Data* 12.1, p. 178.

Lagakos, David (2020). "Urban-rural gaps in the developing world: Does internal migration offer opportunities?" In: *Journal of Economic perspectives* 34.3, pp. 174–192.

Lagakos, David, Samuel Marshall, Ahmed Mushfiq Mobarak, Corey Vernot, and Michael E Waugh (2020). "Migration costs and observational returns to migration in the developing world". In: *Journal of Monetary Economics* 113, pp. 138–154.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh (2023). "The welfare effects of encouraging rural–urban migration". In: *Econometrica* 91.3, pp. 803–837.

Lagakos, David and Michael E Waugh (2013). "Selection, agriculture, and cross-country productivity differences". In: *American Economic Review* 103.2, pp. 948–980.

Lall, Somik, Mathilde Lebrand, Hogeun Park, Daniel Sturm, and Anthony Venables (2021a). *Pancakes to Pyramids: City form to promote sustainable growth*. World Bank.

Lall, Somik, Mathilde Lebrand, and Maria Edisa Soppelsa (2021b). "The Evolution of City Form". In.

Lewis, W. Arthur (May 1954). "Economic Development with Unlimited Supplies of Labour". In: *The Manchester School* 22.2, pp. 139–191.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao (2020). "A harmonized global nighttime light dataset 1992–2018". In: *Scientific data* 7.1, p. 168.

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

Malpezzi, Stephen (1999). "Economic analysis of housing markets in developing and transition economies". In: *Handbook of regional and urban economics* 3, pp. 1791–1864.

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

Mills, Edwin S. (1972). *Studies in the Structure of the Urban Economy*. Baltimore, MD: Johns Hopkins Press.

Mills, Edwin S. and Jee Peng Tan (1980). "A comparison of urban population density functions in developed and developing countries". In: *Urban studies* 17.3, pp. 313–321.

Monge-Naranjo, Alexander, Pedro Cavalcanti Ferreira, and Luciene Torres de Mello Pereira (2022). "Of Cities and Slums". In: *Working paper*.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg (Dec. 2018). "Commuting, Migration, and Local Employment Elasticities". In: *American Economic Review* 108.12, pp. 3855–90.

Morris, Carl N (1983). "Parametric empirical Bayes inference: theory and applications". In: *Journal of the American statistical Association* 78.381, pp. 47–55.

Munshi, Kaivan and Mark Rosenzweig (2016). "Networks and misallocation: Insurance, migration, and the rural-urban wage gap". In: *American Economic Review* 106.01, pp. 46–98.

Murray, Michael P and Guoqing Sun (2017). "The demand for space in China". In: *Journal of Urban Economics* 98, pp. 214–222.

Nawrotzki, Raphael J, Jack DeWaard, Maryia Bakhtsiyarava, and Jasmine Trang Ha (2017). "Climate shocks and rural-urban migration in Mexico: exploring nonlinearities and thresholds". In: *Climatic change* 140, pp. 243–258.

Olken, Benjamin A (2007). "Monitoring corruption: evidence from a field experiment in Indonesia". In: *Journal of political Economy* 115.2, pp. 200–249.

Redding, Stephen J and Esteban Rossi-Hansberg (2017). "Quantitative spatial economics". In: *Annual Review of Economics* 9, pp. 21–58.
Ritchie, Hannah and Max Roser (2019). "Half of the world's habitable land is used for agriculture". In: *Our World in Data*. https://ourworldindata.org/global-land-for-agriculture.
Sahai, Harshil and Michael Bailey (2022). "Social Networks and Internal Migration: Evidence from Facebook in India". In: *Working Paper*.
Saiz, Albert (2010). "The geographic determinants of housing supply". In: *The Quarterly Journal of Economics* 125.3, pp. 1253–1296.
Severen, Christopher (2023). "Commuting, labor, and housing market effects of mass transportation: Welfare and identification". In: *Review of Economics and Statistics* 105.5, pp. 1073–1091.
Silva, J. M. C. Santos and Silvana Tenreyro (2006). "The Log of Gravity". In: *The Review of Economics and Statistics* 88.4, pp. 641–658.
Simonovska, Ina and Michael E Waugh (2014). "The elasticity of trade: Estimates and evidence". In: *Journal of international Economics* 92.1, pp. 34–50.
Sorin, Jeanne (2025). "Public roads on private lands: Land costs and optimal road improvements in urban uganda". In: IGC.
Storeygard, Adam (2016). "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa". In: *The Review of economic studies* 83.3, pp. 1263–1295.
Tombe, Trevor and Xiaodong Zhu (2019). "Trade, migration, and productivity: A quantitative analysis of China". In: *American Economic Review* 109.5, pp. 1843–1872.
Tsivanidis, Nick (2023). "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio". In: *Working Paper*.
Wager, Stefan and Susan Athey (2018). "Estimation and inference of heterogeneous treatment effects using random forests". In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242.
Wong, Ho Lun, Renfu Luo, Linxiu Zhang, and Scott Rozelle (2013). "Providing quality infrastructure in rural villages: The case of rural roads in China". In: *Journal of Development Economics* 103, pp. 262–274.
Young, Alwyn (2013). "Inequality, the urban-rural gap, and migration". In: *The Quarterly Journal of Economics* 128.4, pp. 1727–1785.
Zárate, Román D (2022). "Spatial misallocation, informality, and transit improvements: Evidence from mexico city". In: *Policy Research Working Papers* 9990.

A Additional tables and figures

A.1 Additional figures

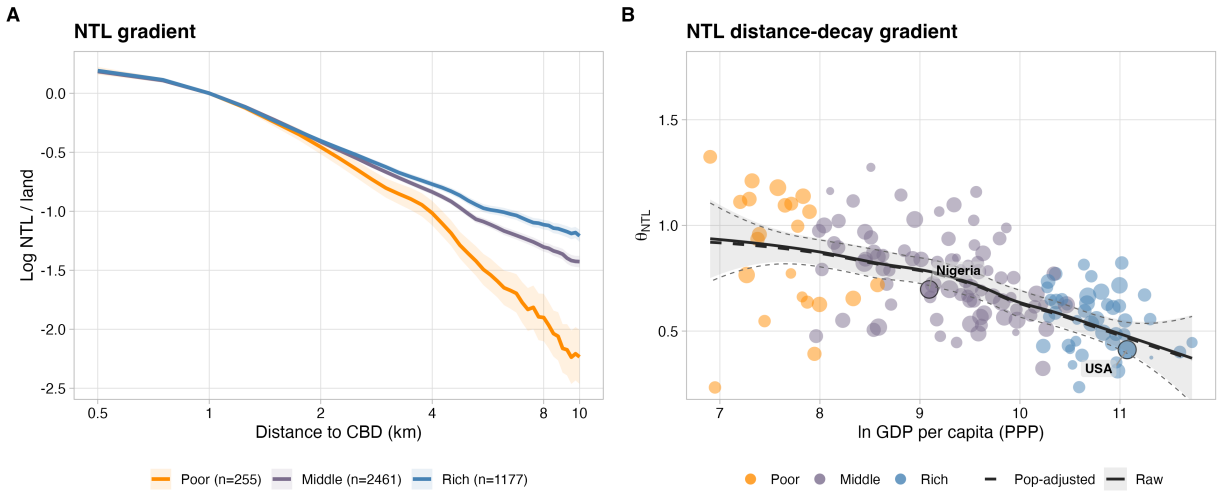

Figure A1: Internal structure of cities using nighttime lights in place of building volume. The left panel plots the average log nighttime luminosity per unit land against distance bins to the CBD, estimated by Poisson regression with city-cluster fixed effects and city-clustered standard errors; the curve is normalized to zero at 1km and the sample excludes India and China. The right panel plots country-average estimates of the NTL distance-decay gradient θ_k^{NTL} against country-level log GDP per capita (PPP), with points colored by World Bank income group (Poor = Low income, Middle = Lower + Upper Middle, Rich = High income). I use the Black Marble VIIRs as nightlight data. Point size reflects the number of cities; solid line is a loess smoother on raw country means, dashed line adjusts for city population.

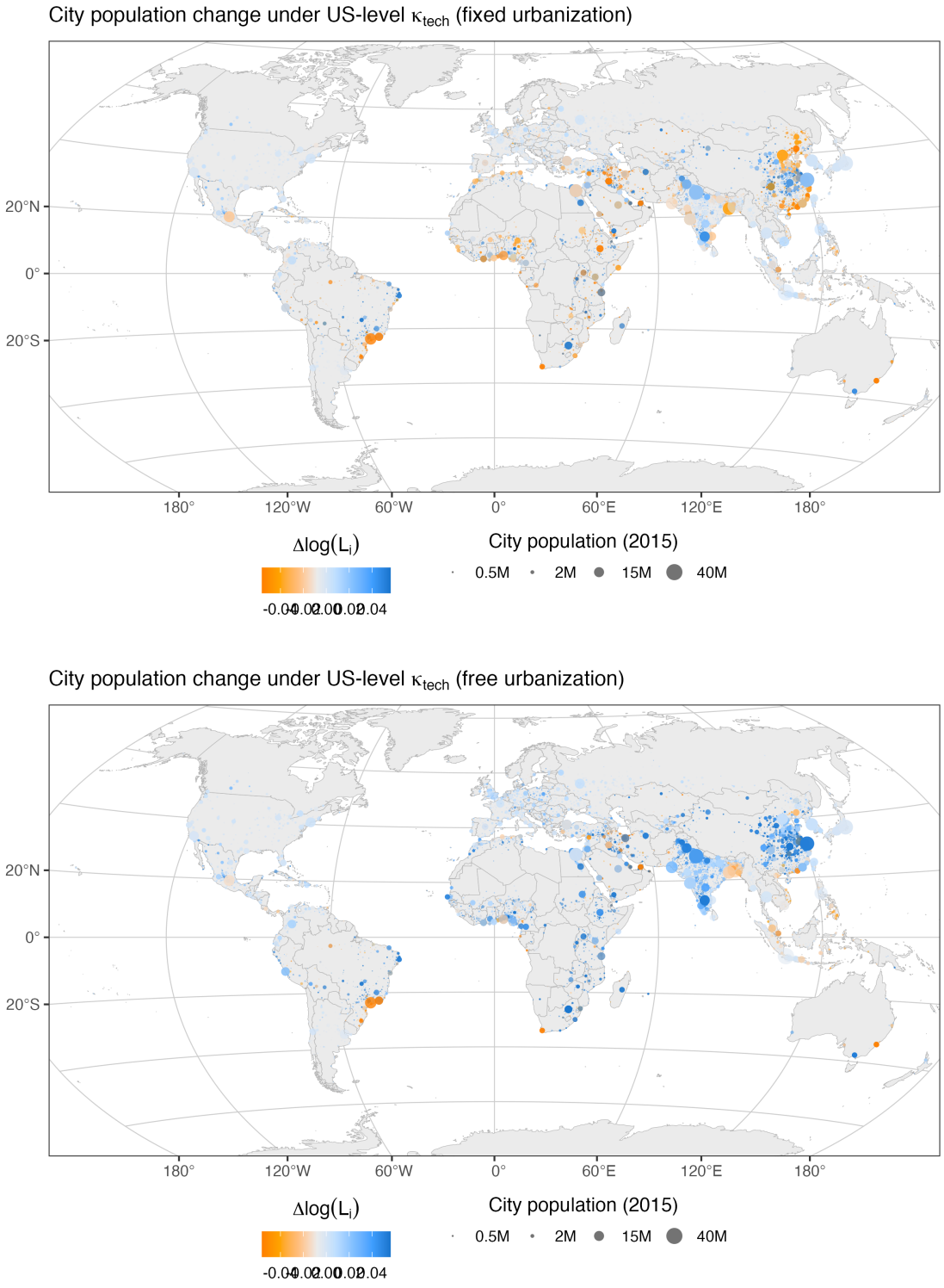

Figure A2: Maps of city-level population changes with and without urbanization, feeding in 1993-2015 productivity growth, amenity changes, and population growth, with κ_i shrunk to the technological frontier (distributions dilated so the within-country median is 0.56). Top: without population growth or rural to urban migration; bottom with population growth. Each point is a city, sized by its population in 2015. Orange cities shrank in size while blue ones grew.

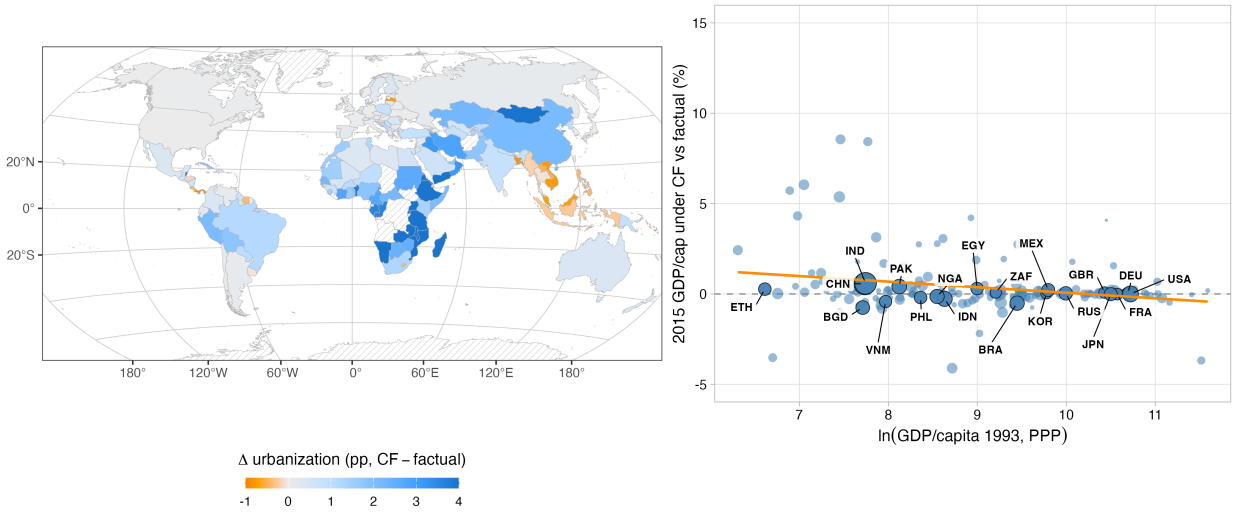

Figure A3: Macroeconomic effects of lowering urban costs to the East Asian level under 1993-2015 growth shocks. Left: map of urbanization changes, right: GDP/capita changes vs log GDP/cap in 2015 PPP.

A.2 Additional tables

	$\ln(\text{Volume}/\text{Person})$	$\ln(\text{Height})$	$\ln(\text{Dev. Area}/\text{Person})$	$\ln(\text{Height}) - \ln(\text{Dev. Area}/\text{Person})$
Raw	0.332	0.222	0.110	0.111
	(0.027)	(0.019)	(0.024)	(0.034)
	[0.408]	[0.442]	[0.086]	[0.053]
Pop-adjusted	0.319	0.215	0.105	0.110
	(0.025)	(0.019)	(0.022)	(0.032)
	[0.442]	[0.428]	[0.094]	[0.056]
Territories			14,996 (pop adjustment)	
Countries	188	188	188	188

Table A1: Regressions of average city characteristics on country-level log GDP/capita. 'Raw' is the simple average. Pop-adjusted residualizes on a polynomial population plus an indicate for largest city in a country before taking the average. Robust standard errors in parentheses. R-squared in brackets. Column (4) effectively tests the null that height and area respond equally to income, which is rejected at conventional levels of significance.

	Cross-country		Within country (country FE)		
	(1)	(2)	(3)	(4)	(5)
Panel A. Dependent variable: $\ln \hat{\tau}_i$					
ln country GDP/cap	-0.208	-0.218			
	(0.063)	(0.061)			
ln territory pop		-0.064	-0.055	-0.170	-0.163
		(0.017)	(0.015)	(0.036)	(0.033)
ln territory VIIRS/cap			-0.071	0.032	0.062
			(0.039)	(0.043)	(0.045)
ln territory pop density				0.591	0.607
				(0.139)	(0.134)
Primate territory					-0.093
					(0.057)
Coastal (≤ 50 km from coast)					-0.174
					(0.048)
Polycentric city ($K^\star \geq 2$)					-0.104
					(0.029)
Observations	14,867	14,866	14,853	14,853	14,850
R^2 (within for FE cols)	0.042	0.047	0.011	0.094	0.106
Panel B. Dependent variable: $\ln \hat{\kappa}_i$					
ln country GDP/cap	-0.127	-0.136			
	(0.043)	(0.042)			
ln territory pop		-0.056	-0.053	-0.113	-0.112
		(0.010)	(0.007)	(0.019)	(0.018)
ln territory VIIRS/cap			-0.041	0.013	0.027
			(0.023)	(0.026)	(0.028)
ln territory pop density				0.313	0.320
				(0.074)	(0.073)
Primate territory					0.001
					(0.030)
Coastal (≤ 50 km from coast)					-0.124
					(0.041)
Polycentric city ($K^\star \geq 2$)					-0.041
					(0.020)
Observations	14,867	14,866	14,853	14,853	14,850
R^2 (within for FE cols)	0.045	0.057	0.023	0.102	0.116
Country FE	–	–	✓	✓	✓

Table A2: Cross-sectional correlations of the (log) commuting cost elasticity τ_i (Panel A) and the urban cost elasticity κ_i (Panel B). Standard errors are clustered at the country level.

Component	Level	Share of $\mathrm{Var}(\kappa)$
Panel A. Between countries		
V_B^{int}	0.0065	2.2%
V_B^{ext}	0.0999	34.2%
$2\,\mathrm{Cov}_B$	0.0120	4.1%
Subtotal (between)	0.1185	40.6%
Panel B. Within countries		
V_W^{int}	0.0074	2.5%
V_W^{ext}	0.1710	58.5%
$2\,\mathrm{Cov}_W$	-0.0048	-1.6%
Subtotal (within)	0.1736	59.4%
Total $\mathrm{Var}(\kappa)$	0.2921	100.0%
Aggregate margin shares (Panel A + Panel B by margin)		
Intensive ($V_B^{\text{int}} + V_W^{\text{int}}$)	—	4.8%
Extensive ($V_B^{\text{ext}} + V_W^{\text{ext}}$)	—	92.8%
Cross-margin covariance	—	2.5%

Table A3: Population-weighted variance decomposition of κ by margin and country. I decompose κ into the 'intensive' margin $1/[(1+\rho)(1+\gamma)]$ and the extensive margin, $\rho/(1+\rho) \times \tau/2$ and the variance of κ into these components and the covariance, within and between countries. Weights are 2015 city level population, with $N = 14869$ cities in 190 countries.

	N	γ	ρ	τ	κ
Panel A. By World Bank income group					
Low income	723	0.28 (0.18–0.51)	1.18 (0.72–1.79)	3.73 (2.43–6.53)	1.36 (0.97–2.01)
Lower middle income	3996	0.55 (0.37–0.86)	1.34 (0.91–1.96)	2.12 (1.21–3.50)	0.88 (0.58–1.30)
Upper middle income	4155	0.51 (0.38–0.63)	1.16 (0.73–1.97)	2.00 (1.37–2.88)	0.84 (0.65–1.12)
High income	1819	0.51 (0.40–0.59)	0.63 (0.51–0.85)	1.54 (1.09–2.10)	0.70 (0.60–0.83)
Panel B. By World Bank region					
East Asia & Pacific (excl. China)	1140	0.81 (0.60–1.10)	1.30 (0.87–1.80)	1.14 (0.76–1.60)	0.56 (0.44–0.70)
China	1931	0.51 (0.42–0.60)	1.11 (0.70–2.77)	2.15 (1.64–2.86)	0.87 (0.72–1.11)
Europe & Central Asia	1496	0.52 (0.44–0.59)	0.60 (0.50–0.78)	1.70 (1.28–2.23)	0.73 (0.64–0.85)
Latin America & Caribbean	1060	0.47 (0.29–0.64)	1.44 (0.98–2.01)	1.95 (1.35–2.98)	0.85 (0.65–1.21)
Middle East & North Africa	851	0.40 (0.22–0.58)	0.90 (0.71–1.20)	2.47 (1.68–3.81)	0.98 (0.75–1.30)
North America	393	0.46 (0.34–0.55)	0.78 (0.61–1.35)	1.19 (0.75–1.71)	0.64 (0.52–0.78)
South Asia	2393	0.65 (0.46–0.92)	1.35 (0.87–2.01)	1.89 (1.07–2.91)	0.79 (0.53–1.12)
Sub-Saharan Africa	1429	0.21 (0.15–0.31)	1.43 (0.93–2.11)	4.70 (2.95–7.26)	1.70 (1.19–2.36)

Table A4: Distribution of structrual elasticities by income group (Panel A) and by World Bank region (Panel B). N reports the number of cities in each cell. Columns report medians with 25th-75th percentils in parentheses.

			Dep. var.: $\log \hat{\kappa}_i$		
Group	N_c	N_{city}	$\log y_{j,93}$	$\log y_{j,15}$	$\Delta \log y_j$
Panel A: Income group					
High income	65	1819	$+0.033$	$+0.012$	-0.068
			(0.061)	(0.055)	(0.076)
Upper middle (ex. China)	49	2224	-0.092	-0.071	$+0.066$
			(0.042)	(0.038)	(0.063)
China	1	1931	-0.163	-0.149	$+0.110$
			(0.016)	(0.017)	(0.020)
Lower middle (ex. India)	48	2199	-0.141	-0.124	$+0.057$
			(0.056)	(0.057)	(0.055)
India	1	1797	$+0.093$	$+0.068$	-0.002
			(0.022)	(0.018)	(0.039)
Low income	26	723	-0.277	-0.271	$+0.037$
			(0.073)	(0.077)	(0.140)
Panel B: World Bank region					
East Asia & Pacific (ex. China)	26	1140	-0.096	$+0.008$	$+0.164$
			(0.077)	(0.076)	(0.092)
Europe & Central Asia	51	1496	-0.006	-0.025	-0.057
			(0.042)	(0.037)	(0.026)
Latin America & Caribbean	32	1060	-0.106	-0.069	$+0.132$
			(0.039)	(0.051)	(0.061)
Middle East & North Africa	22	851	-0.101	-0.127	-0.060
			(0.098)	(0.096)	(0.128)
North America	2	393	$+0.032$	-0.011	-0.070
			(0.044)	(0.018)	(0.048)
South Asia (ex. India)	6	596	-0.009	-0.091	-0.125
			(0.073)	(0.045)	(0.074)
Sub-Saharan Africa	49	1429	-0.221	-0.254	-0.009
			(0.060)	(0.061)	(0.111)
China	1	1931	-0.163	-0.149	$+0.110$
			(0.016)	(0.017)	(0.020)
India	1	1797	$+0.093$	$+0.068$	-0.002
			(0.022)	(0.018)	(0.039)

Table A5: Correlations of κ_i and city-level GDP per capita in 1993, 2015, and on GDP per capita growth. Panel A: pooled by income group; Panel B: by World Bank region. All regressions contain coutnry fixed effects. Standard errors clustered at the country level in parentheses.

	Dependent variable: $\ln \tau$				
	(1)	(2)	(3)	(4)	(5)
Share of roads paved	-1.047	-1.011	-0.566	-0.556	0.173
	(0.152)	(0.156)	(0.112)	(0.107)	(0.086)
ln population density			0.181	0.175	0.162
			(0.033)	(0.029)	(0.040)
ln road density (km/km^2)		0.140	0.184	0.184	-0.086
		(0.149)	(0.119)	(0.128)	(0.108)
ln intersection density		-0.055	-0.009	-0.021	0.146
		(0.069)	(0.073)	(0.072)	(0.045)
Circuity		-0.021	0.208	0.151	0.117
		(0.197)	(0.160)	(0.143)	(0.086)
Alpha index (network connectivity)		1.904	2.202	1.866	1.700
		(0.573)	(0.468)	(0.446)	(0.418)
Orientation entropy		0.035	0.036	0.030	-0.056
		(0.117)	(0.091)	(0.091)	(0.057)
Share of roads with surface tag	0.003	0.000	0.008	0.007	0.008
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
(squared)	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	14,704	14,676	14,676	12,672	14,661
R^2	0.066	0.091	0.176	0.171	0.332
Sample	Full	Full	Full	$\geq$10% tagged	Full
Country fixed effects					✓

Table A6: $\ln \tau_i$ versus road network characteristics. Standard errors clustered at the country level.

	(1)	(2)	(3)	(4)
$1\{\text{fac} > 0\}$	-0.222	-0.075	-0.054	-0.087
	(0.043)	(0.043)	(0.027)	(0.018)
$\log(\text{facilities})$	0.052	0.094	0.028	0.042
	(0.012)	(0.011)	(0.008)	(0.009)
$1\{\text{major} > 0\}$	-0.171	-0.089	0.048	-0.071
	(0.055)	(0.036)	(0.040)	(0.035)
$\log(\text{major road km})$	0.016	0.060	-0.049	0.024
	(0.025)	(0.013)	(0.017)	(0.011)
$\log(\text{NTL / cap})$		-0.642	-0.623	-0.553
		(0.017)	(0.014)	(0.021)
$\log(\text{pop})$			0.224	0.064
			(0.027)	(0.023)
$\log(\text{pop density})$				0.664
				(0.021)
N	10659	10659	10659	10659
Within R^2	0.026	0.578	0.619	0.759
Country FE	✓	✓	✓	✓

Table A7: City-level regressions of $\log A_i$ on infrastructure measures in the SUDS (Hsiao, 2025) data. Standard errors clustered at the country level in parentheses. All regressions include country fixed effects, which control for differences in the scale of amenities across countries (which is unidentified from model inversion). Columns (2)-(4) control for endogenous city level characteristics to assuage concerns of reverse causality (high amenity cities have more population, or income, or density, and therefore more infrastructure).

	(1)	(2)	(3)	(4)
	Dep. var.: country $\Delta Y\%$ at 2 m SLR			
Exposure (%)	-0.159	-0.171	-0.163	-0.098
	(0.082)	(0.018)	(0.011)	(0.020)
Exposure $\times$ (Urbanization $-$ median)		-0.005	-0.004	-0.002
		(0.001)	(0.001)	(0.001)
Exposure $\times$ (κ_{tech} $-$ median)			-0.206	-0.145
			(0.054)	(0.041)
Constant	0.114	-0.015	0.014	-0.069
	(0.208)	(0.102)	(0.088)	(0.090)
N	129	129	129	127
R^2	0.501	0.849	0.896	0.438
Drops $\Delta Y < -10\%$				✓

Table A8: Country GDP loss (model-derived) at 2m SLR versus SLR exposure (% of urban land inundated at 2m), with urbanization and κ interactions. Robust standard errors in parentheses

B CBD detection: algorithm details

This appendix describes the CBD detection algorithm of Section 2.

B.1 Data Processing

The WSF-3D building volume raster is delivered at 90 m resolution. I divide each pixel's volume by its developable area and aggregate to coarse blocks of approximately 500 m. For an agglomeration with mean latitude $\bar{\phi}$, the angular spacing is:

$$\Delta\text{lat} = \frac{500}{111{,}000}, \qquad \Delta\text{lon} = \frac{500}{111{,}000 \cdot \cos(\bar{\phi} \cdot \pi/180)}. \tag{12}$$

Each 90 m cell is assigned to its nearest block centroid. The developable fraction f_i for block i is the share of constituent 10 m pixels below a 15% slope grade (Copernicus DEM GLO-30) and not classified as water, ice, or wetland (ESA WorldCover 2021). Blocks with $f_i = 0$ contribute zero likelihood. Block centroids are rescaled to the unit square; all distances are in normalized coordinates.

B.2 Model

Building height at coarse block i follows:

$$H_i \sim \text{Poisson}(\lambda_i)\,, \qquad \lambda_i = f_i \sum_{k=1}^{K} A_k\, x_{ik}^{-\theta_k}, \tag{13}$$

where $A_k > 0$ is the scale and $\theta_k > 0$ is the distance-decay exponent for city k and $x_{ik} = \|\mu_i - \mu_k\|_2$ is geodesic distance of cell i at location μ_i to downtown k, where μ_k contains the location of downtown k in the unit square. The log-likelihood is:

$$\ell(\Omega) = \sum_{i=1}^{N} \left[H_i \ln \lambda_i - \lambda_i\right]. \tag{14}$$

The additivity of (13) follows from Poisson superposition: if $H_{ik} \sim \text{Poisson}(f_i A_k x_{ik}^{-\theta_k})$ independently, then $H_i = \sum_k H_{ik}$ follows $\text{Poisson}(\sum_k f_i A_k x_{ik}^{-\theta_k})$. The log-likelihood (14) is nonconvex in $\{\mu_k\}$. The latent per-city flows H_{ik} are unobserved; the EM recovers their conditional expectation in the E-step.

B.3 Initialization

I initialize μ_k randomly, and run the model for different restarts, pruning to get the best estimation run. I use I independent random restarts, scaled with agglomeration size (see Section B.6), and select the restart with the highest penalized log-likelihood. Each restart seeds center locations by K-means++ on the coarse-block point cloud, weighted by $w_i = \max(f_i H_i, 1)$. Initial θ_k are drawn from $[0.3, 1.5]$ with small per-center perturbations; initial A_k are set by a single-cluster optimization over each Voronoi cell.

B.4 The EM Algorithm

B.4.1 E-Step: Attribution Weights

Given $\Omega^{(t)}$ (the parameters at step t), the expected contribution of block i to city k is:

$$\hat{H}_{ik} = a_{ik} H_i, \qquad a_{ik} = \frac{A_k^{(t)}\, x_{ik}^{-\theta_k^{(t)}}}{\sum_{j=1}^{K} A_j^{(t)}\, x_{ij}^{-\theta_j^{(t)}}}. \tag{15}$$

The weights satisfy $\sum_k a_{ik} = 1$ and do not depend on f_i.

B.4.2 M-Step: Intensity Parameters A_k

The first-order condition for A_k is closed form:

$$\hat{A}_k^{(t+1)} = \frac{\sum_{i=1}^{N} \hat{H}_{ik}}{\sum_{i=1}^{N} f_i\, x_{ik}^{-\theta_k^{(t)}}}. \tag{16}$$

B.4.3 M-Step: Distance-Decay Parameters θ_k

There is no closed-form update for θ_k. Profiling out A_k via (16), the profile first-order condition

$$\text{FOC}(\theta_k) \;=\; \sum_{i=1}^{N} \left[\hat{H}_{ik} - f_i\, \hat{A}_k(\theta_k)\, x_{ik}^{-\theta_k}\right] \ln(x_{ik}) = 0 \tag{17}$$

is solved by Brent's method.

B.4.4 M-Step: CBD Locations μ_k

The gradient of the expected log-likelihood with respect to μ_k is:

$$g_k = \theta_k \sum_{i=1}^{N} \left[\hat{H}_{ik} - f_i\, A_k\, x_{ik}^{-\theta_k}\right] \frac{\mu_i - \mu_k}{x_{ik}^2}, \tag{18}$$

with expected Fisher information:

$$\mathcal{I}_k = \theta_k^2 \sum_{i=1}^{N} f_i \, A_k \, x_{ik}^{-\theta_k} \cdot \frac{(\mu_i - \mu_k)(\mu_i - \mu_k)^\top}{x_{ik}^4}. \tag{19}$$

I update μ_k via $\Delta\mu_k = \mathcal{I}_k^{-1} g_k$, clamping steps exceeding 0.10 normalized units. This is basically gradient descent.

B.5 Constraints on Center Locations

During EM estimation, I make a number of adjustments so the solver does not get stuck.

Keeping centers on land inside the UCDB polygon When a centers update falls outside the convex hull of coarse-block centroids, I projected them back to the nearest boundary point to stay inside the UCDB shapefile convex hull. When a center is placed on a location with $f_i = 0$ – the solver likes to do this in bays, where it appears as if construction encircles a phantom CBD in the middle of the water – I snap back the μ_k update to the nearest block with $f_i > 0$.

Separation penalty Centers occasionally run into each other and then collapse (or get stuck in a dog-chase-tail loop). To prevent center collapse I add:

$$\text{Pen}(\mu) = -\lambda_{\text{sep}} \sum_{j<k} \ln\left(1 + \exp\left(-\frac{\ln(d_{jk}/d_{\min})}{s_{\text{sep}}}\right)\right), \tag{20}$$

where $d_{jk} = \|\mu_j - \mu_k\|$ (normalized), $d_{\min} = 0.12$, and $\lambda_{\text{sep}} = 100 \cdot \ln(N_c)$. The penalty gradient enters g_k before the Newton step. This has almost no effect on center updating until centers get within 0.1 normalized units of each other, then they repel like magnets.

B.6 Convergence, Restarts, and Screening

The number of restarts scales with agglomeration size (N_c = coarse blocks):

$N_c < 1{,}000$	$I = 16$
$1{,}000 \le N_c < 5{,}000$	$I = 20$
$5{,}000 \le N_c < 10{,}000$	$I = 28$
$N_c \ge 10{,}000$	$I = 36$

I use two-phase screening: all I restarts run for 15 EM iterations, the top 4 run to full convergence (patience: 30 iterations without improvement), and I take the best result.

B.7 Model Selection

I select K^* by the Bayesian information criterion:

$$\text{BIC}(K) = -2\,\hat{\ell}_K + 4K \cdot \ln(N_c), \tag{21}$$

where $\hat{\ell}_K$ is the best penalized log-likelihood for that K and $4K$ counts the free parameters ($K_{\max} = 8$). I halt the sweep at the first K where $\text{BIC}(K) > \text{BIC}(K-1)$ and set $K^* = K - 1$.

B.8 City Assignment and Post-Processing

Pruning. I drop centers with $\hat{\theta}_k \leq 0$, scale $\hat{A}_k < \hat{A}_{\max}/20$, or coverage below 2% of pixels. At least the dominant center is always kept.

Hard assignment via Potts model I assign 100 m pixels to cities via a Potts model combining per-pixel Poisson log-likelihoods with a spatial smoothness prior, solved by Iterated Conditional Modes (ICM). The smoothness weight is twice the median per-pixel log-likelihood gap between the best and second-best assignments. This is a fast way of finding the labeling $z = z_i{}_{i=1}^{N}$, $z_i \in 1, \ldots, K^*$, that minimizes the energy

$$E(z) = -\sum_i \ell_{z_i}(i) + \eta \sum_{\langle i,j \rangle} 1(z_i \neq z_j)$$

where the first term is the negative Poisson log-likelihood at the EM estimates,

$$\ell_k(i) = H_i \ln\left(f_i \hat{A}_k x_{ik}^{-\hat{\theta}_k}\right) - f_i \hat{A}_k x_{ik}^{-\hat{\theta}_k},$$

and $\langle i, j \rangle$ runs over all 8-connected pixel pairs. Off-the-shelf tools automatically adjust the smoothness weight, η. The basic idea is the labeling should best fit the Poisson model, subject to a penalty if your neighbors, spatially, don't match your assignment rule. It is a form of spatial smoothing for discrete labels.

Watershed for cleanup The Potts model solution can produce disconnected city components. I retain only the largest connected component per city. This can then leave orphaned islands, which I via BFS watershed – a fast off-the-shelf tool – or by $\arg\min_k x_{ik}^{\hat{\theta}_k}$ if no adjacent component exists. This basically assigns any orphaned clusters to the nearest CBD, in θ_k-adjusted terms. In all transparency, this particular idea was developed by Claude Opus 4.6.

B.9 Second-Stage Estimation of θ_k at 100 m

The EM estimates θ_k at 500 m, sufficient to locate CBDs but subject to within-block averaging. I reestimate at 100 m via PPML (`fixest::fepois`) within each city:

$$H_i \sim \text{Poisson}\Big(\exp\Big(-\hat{\theta}_k \cdot \ln x_{ik} + \ln f_i\Big)\Big), \tag{22}$$

with $\ln f_i$ as an offset. The preferred `donut_area_large_theta` specification excludes the CBD core within radius

$$r_{\text{inner}}(c) = 2.0\,\text{km} \times \left(\frac{\text{Area}_c}{3{,}128\,\text{km}^2} \right)^{0.25}, \tag{23}$$

where c indexes the agglomeration and the reference area is Shanghai. Each observation is weighted by $w_i = 1/n_{b(i)}$, where $n_{b(i)}$ counts pixels sharing the same cell of an 8×10 angle-by-log-distance grid, removing the mechanical growth in cell count with distance. Standard errors are clustered by this grid cell. I use the resulting $\hat{\theta}_k$ for all downstream structural quantities (τ, κ, counterfactuals).

C Full model derivation

This appendix states and solves the urban system model. The model has two 'levels.' First is a within-city partial equilibrium: city wages w_i, city populations L_i, factory-gate prices $\{p_j\}$, and firm-side delivered price indices $\{P_j\}$ are held fixed, and the floorspace market, residential location, and urban-radius problems are solved. In the 'upper level,' is the general-equilibrium intercity equilibrium: cities are embedded in a closed national economy with cross-city Fréchet labor supply, roundabout production, and a rural-to-urban migration margin with Roy-like sorting. Notation matches Section 3 throughout.

Primitives belong to a single nation, modelled as a closed economy, containing N urban cities indexed $i = 1, \ldots, N$ and an rural (or agricultural) sector $i = R$. Total population is $\bar{L}$. All country-level subscripts are suppressed.

C.1 City partial equilibrium

C.1.1 Environment

A city i is a disk of endogenous radius X_i, with locations in polar coordinates (x, ϕ). Allocations are symmetric in ϕ, so a location is summarised by its distance to the CBD, $x \in [0, X_i]$. All urban production occurs at the CBD; commuting is radial.

City wages w_i and labor stock L_i are taken as given. The household composite-good price index $P \equiv 1$ is the numéraire, set at the GE level (Section C.2). City primitives are the citywide amenity A_i, floorspace supply TFP $\tilde{Z}_i^H$, and land-supply TFP $\tilde{Z}_i^X$.

City-level structural parameters are the floorspace-supply elasticity $\gamma_i \in (0, \infty)$, the land-supply elasticity $\rho_i \in (0, \infty)$, and the commuting-cost elasticity $\tau_i \in (0, \infty)$. The household housing-expenditure share is $\beta \in (0, 1)$; in calibrations β may differ across countries but is constant within them. The CES elasticity across urban varieties is $\sigma > 1$.

Endogenous at the city partial-equilibrium level are the floorspace price function $q_i(x)$, floorspace supply $H_i(x)$, residential population density $L_i(x)$, urban radius X_i, common indirect utility V_i, and the local floor-rent gradient.

The amenity supply function combines citywide amenities and commuting costs:

$$A_i(x) \;=\; \tilde{A}_i \cdot x^{-\tau_i}, \qquad \tilde{A}_i \;=\; A_i \left(1 - \frac{\tau_i \,(1+\gamma_i)}{2}\right)^{1/(1+\gamma_i)}, \tag{24}$$

For the floorspace and labor-market clearing integrals to converge over $[0, X_i]$:

$$\frac{1}{1+\gamma_i} \;>\; \frac{\tau_i}{2}, \qquad \forall i, \tag{25}$$

which is the equilibrium-existence condition stated in Section 3.

C.1.2 Preferences (households)

A household choosing city i at distance x, given wage w_i, floorspace price $q_i(x)$, and factory-gate prices $\{p_j\}$, solves:

$$\max_{\{c_j\},\, h,\, x \in (0, X_i]} \big(A_i(x)\, h/\beta\big)^{\beta} \big(C/(1-\beta)\big)^{1-\beta}\, \epsilon_i^{\nu}, \qquad C \;=\; \Big(\sum_{j=1}^{N} c_j^{(\sigma-1)/\sigma}\Big)^{\sigma/(\sigma-1)}, \tag{26}$$

subject to:

$$\sum_{j=1}^{N} p_j\, c_j \;+\; q_i(x)\, h \;\leq\; w_i. \tag{27}$$

β is the expenditure share on floorspace; $1-\beta$ is spent on the CES bundle C. The household CES aggregate is frictionless: households face factory-gate prices $\{p_j\}$ at every location. The shock ϵ_i^{ν} is idiosyncratic, drawn iid across cities and the agricultural sector from a Fréchet distribution; it does not vary across x within a city. Commuting costs enter utility through $A_i(x)$ rather than as a wage deduction, keeping the household's expenditure decision spatially invariant within the city.

Alternatively, we can assume households consume C, the numeraire good, which is produced by a final goods firm that constructs C using a CES aggregator of all urban varieties.

C.1.3 Technology (construction firms)

Two technologies operate within the city: floorspace construction (vertical) and land incorporation (horizontal), both operated by competitive developers.

Floorspace technology A developer at location x produces floorspace per unit land:

$$H_i(x) \;=\; \tilde{Z}_i^{H}\, M^{\,\gamma_i/(1+\gamma_i)}\, T_i(x,\phi)^{1/(1+\gamma_i)}, \qquad T_i(x,\phi) \;\equiv\; 1. \tag{28}$$

The parameter γ_i is the elasticity of floorspace supplied per unit land with respect to price; heterogeneity across cities captures geophysical and regulatory barriers to vertical development.

Land-incorporation technology To incorporate marginal land at distance x, a developer pays the fixed cost:

$$F_i(x) \;=\; \tilde{Z}_i^{X}\, x^{2/\rho_i}. \tag{29}$$

Fixed costs rise in x at rate $2/\rho_i$.

Developer's problem Taking floorspace prices as given, a developer at x solves:

$$\max_{M} \; q_i(x)\, H_i(x) \;-\; M \;-\; F_i(x), \tag{30}$$

subject to (28). Land rents are $r_i(x) = \max_M [q_i(x)H_i(x) - M]$. Free entry implies developers operate at x if and only if $r_i(x) \geq F_i(x)$; the urban radius X_i satisfies $r_i(X_i) = F_i(X_i)$.

C.1.4 Equilibrium definition

A city partial equilibrium of city i, given (w_i, L_i), $\{p_j\}_{j=1}^N$, and primitives $(A_i, Z_i^H, Z_i^X, \beta, \gamma_i, \tau_i, \rho_i, \sigma)$, is a tuple $\big(q_i(x),\, H_i(x),\, L_i(x),\, X_i,\, V_i\big)$ such that:

1. households solve (26) subject to (27);
2. developers solve (30) subject to (28);
3. spatial equilibrium holds, $V_i(x) = V_i$ for all $x \in (0, X_i]$;
4. the floorspace market clears at every x, $H_i(x) = h_i(x)\, L_i(x)$;
5. the city population is housed, $2\pi \int_0^{X_i} x\, L_i(x)\, dx = L_i$;
6. developers freely enter, pinning X_i via $r_i(X_i) = F_i(X_i)$.

C.1.5 Equilibrium solution

Floorspace demand. The Cobb-Douglas expenditure share on housing gives:

$$h_i(x) \;=\; \frac{\beta\, w_i}{q_i(x)}. \tag{31}$$

Floorspace price gradient. Indirect utility at (i, x) is $V_i(x) = A_i(x)^{\beta}\, w_i\, (P)^{-(1-\beta)}\, q_i(x)^{-\beta}$. Spatial equilibrium ($V_i(x) = V_i$) with $P \equiv 1$ gives:

$$q_i(x) \;=\; A_i(x) \left(\frac{w_i}{V_i}\right)^{1/\beta}, \qquad x \in (0, X_i]. \tag{32}$$

Floorspace supply. Inverting the FOC on M in (30) and substituting into (28) gives:

$$H_i(x) \;=\; Z_i^H\, q_i(x)^{\gamma_i}. \tag{33}$$

Local rent share. Cobb-Douglas pricing gives $q_i(x)\, H_i(x) = (1+\gamma_i)\, r_i(x)$, while floorspace-market clearing gives $q_i(x)\, H_i(x) = \beta\, w_i\, L_i(x)$. Together:

$$r_i(x) \;=\; \frac{\beta\, w_i\, L_i(x)}{1 + \gamma_i}. \tag{34}$$

Population density. Combining (32), (33), and floorspace market clearing gives $L_i(x) \propto A_i(x)^{1+\gamma_i}$. Substituting $A_i(x) = \tilde{A}_i\, x^{-\tau_i}$ and using the population-clearing integral to pin the constant of proportionality:

$$L_i(x) \;=\; \left(1 - \frac{\tau_i\,(1+\gamma_i)}{2}\right)\left(\frac{x}{X_i}\right)^{-\tau_i(1+\gamma_i)} \frac{L_i}{\pi X_i^2}. \tag{35}$$

The integral converges under restriction (25).

Urban radius. Substituting (34), (35) at $x = X_i$, and (29) into the periphery free-entry condition gives:

$$\pi X_i^2 \;=\; \pi\, Z_i^X \left(\beta\, w_i\, L_i\right)^{\rho_i/(1+\rho_i)}. \tag{36}$$

Aggregate floorspace volume. Integrating $H_i(x)$ over the disk using (32)–(35):

$$\frac{H_i}{\pi X_i^2} \;=\; (Z_i^H)^{1/(1+\gamma_i)} \left(\frac{\beta\, w_i\, L_i}{\pi X_i^2}\right)^{\gamma_i/(1+\gamma_i)} \frac{(1-(\tau_i/2)(1+\gamma_i))^{\gamma_i/(1+\gamma_i)}}{1-(\tau_i/2)\gamma_i}. \tag{37}$$

Average-height and area regressions. Equations (37) and (36) identify γ_i and ρ_i respectively. Taking logs and absorbing city-specific constants into residuals e_i, f_i:

$$\bar{H}_i \;\equiv\; \log\frac{H_i}{\pi X_i^2} \;=\; \frac{\gamma_i}{1+\gamma_i}\Big(\log w_i + \log\frac{L_i}{\pi X_i^2}\Big) \;+\; e_i, \tag{38}$$

$$\log \pi X_i^2 \;=\; \frac{\rho_i}{1+\rho_i}\,\log(w_i L_i) \;+\; f_i. \tag{39}$$

The residuals (e_i, f_i) encode the structural shifters $Z_i^H, Z_i^X, \beta, \tau_i, \rho_i, \gamma_i$ via the constants in (37)–(36).

Indirect utility. Spatial equilibrium implies $V_i = V_i(X_i)$. Substituting (32) and (34) at $x = X_i$, then using (35) and (36) to eliminate $L_i(X_i)$ and X_i:

$$V_i \;=\; A_i^{\beta}\,(Z_i^H)^{\beta/(1+\gamma_i)}\,(Z_i^X)^{\beta/(1+\gamma_i)-\beta\tau_i/2}\,\frac{w_i}{(P)^{1-\beta}}\,\pi^{\beta/(1+\gamma_i)}\,\left(\beta w_i L_i\right)^{-\beta\kappa_i}, \tag{40}$$

where $P \equiv 1$ (defined above) and the urban cost elasticity is:

$$\kappa_i \;=\; \frac{1}{1+\rho_i}\,\frac{1}{1+\gamma_i} \;+\; \frac{\rho_i}{1+\rho_i}\,\frac{\tau_i}{2}. \tag{41}$$

The product $\beta\kappa_i$ is the welfare-relevant urban-cost elasticity in the sense of Combes et al. (2019).

Effective floor price. An equivalent statement of (40), used in equation (5) of Section 3, introduces the effective city-level floor price $\check{q}_i = q_i(x)/A_i(x)$, which by (32) is spatially invariant. Up to city-specific constants absorbed into A_i:

$$\log \check{q}_i \;=\; \kappa_i\,\big(\log L_i + \log \mathcal{E}_i\big) \;+\; \tilde{c}_0, \tag{42}$$

where $\mathcal{E}_i$ is per-household floorspace expenditure and $\tilde{c}_0$ is a function of primitives. Under Cobb-Douglas preferences, $\mathcal{E}_i = \beta w_i$, so (42) reduces to $\log \check{q}_i = \kappa_i\,\log(w_i L_i) + \text{const}$, the form in equation (5).

C.2 Intercity general equilibrium

C.2.1 Environment

The nation contains N cities and a rural sector R, linked by iceberg trade costs $\delta_{ij} \geq 1$ with $\delta_{ii} = 1$. There is a mass $\bar{L}$ of ex-ante identical households; L_U are urban, $L_R = \bar{L} - L_U$ rural.

Exogenous at the GE level: the trade-cost matrix $\{\delta_{ij}\}$, city-level raw TFP shifters $\bar{Z}_i$, amenities A_i, within-city primitives (Z_i^H, Z_i^X), the rural reservation utility U_R, and parameters $(\beta, \sigma, \eta, \zeta, \psi, \varepsilon, \bar{L})$.

Endogenous at the GE level: city wages $\{w_i\}$, city populations $\{L_i\}$, factory-gate prices $\{p_i\}$, firm-tier price indices $\{P_i\}$, urban population L_U, and (through L_i) the within-city objects of Section C.1. $P \equiv 1$ is the country-level numéraire (defined above).

GE parameters: $\sigma > 1$ is the CES elasticity across origins in both the household and firm intermediate bundle; $\eta \in (0, 1)$ is the firm cost share on the intermediate bundle ($1 - \eta$ is the labor share); $\zeta > 0$ is the agglomeration elasticity; $\psi > 0$ is the within-urban-system Fréchet shape; $\varepsilon > 0$ is the rural-urban Fréchet shape which governs skill dispersion and Roy sorting.

C.2.2 Preferences (households)

A household first chooses urban or rural; conditional on urban, it chooses a city; conditional on a city, it solves (26). Idiosyncratic preference shocks are drawn iid Fréchet across discrete choices but not across within-city locations.

Within-urban indirect utility. Plugging the within-city solution (40) into the city-choice problem gives, *in expectation*,

$$V_i \;=\; A_i\, w_i\, S\, \left(w_i L_i\right)^{-\beta\kappa_i}, \qquad S \;=\; \left(L_U/\bar{L}\right)^{-1/\varepsilon}, \tag{43}$$

where S measures the average skill of urban workers. This is Lagakos and Waugh (2013)/Bryan and Morten (2019) machinery without the skill correlation copula.

Within-urban Fréchet sorting. City shares of urban labor follow:

$$\frac{L_i}{L_U} \;=\; \frac{V_i^{\,\psi}}{\sum_{k=1}^{N} V_k^{\,\psi}}, \qquad \mathcal{V}_U \;\equiv\; \left(\sum_{k=1}^{N} V_k^{\,\psi}\right)^{1/\psi}. \tag{44}$$

Rural-urban margin. With Bryan-Morten Fréchet draws of shape ε, the urban share is:

$$\frac{L_U}{\bar{L}} \;=\; \frac{\mathcal{V}_U^{\varepsilon}}{\mathcal{V}_U^{\varepsilon} + U_R^{\varepsilon}}, \tag{45}$$

where $\mathcal{V}_U$ is the within-urban aggregator in (44) and U_R encodes rural amenity A_R and rural productivity w_R (see Section C.3).

C.2.3 Technology (firms)

Production. A representative firm in city i produces output using city labor and an intermediate bundle via Cobb-Douglas with external scale economies:

$$y_i \;=\; Z_i\,(S\,L_i)^{1-\eta}\,M_i^{\eta}, \quad Z_i = \bar{Z}_i\,L_i^{\zeta} \tag{46}$$

$\bar{Z}_i$ is raw city productivity; L_i^{ζ} is the agglomeration multiplier; $(S\,L_i)^{1-\eta}$ packages efficiency units of labor at labor share $1-\eta$.

Intermediate CES bundle. The intermediate input M_i is a CES aggregate over city outputs, with delivered prices subject to iceberg costs:

$$M_i \;=\; \left(\sum_{k=1}^{N} m_{ki}^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}, \qquad P_i \;=\; \left(\sum_{k=1}^{N}(\delta_{ki}\,p_k)^{1-\sigma}\right)^{1/(1-\sigma)}. \tag{47}$$

Pricing. Cost minimisation gives the factor shares:

$$w_i\,L_i \;=\; (1-\eta)\,p_i\,y_i, \qquad P_i\,M_i \;=\; \eta\,p_i\,y_i. \tag{48}$$

Free entry plus constant returns set factory-gate price equal to marginal cost:

$$p_i \;=\; \tilde{\eta}\cdot\frac{w_i^{1-\eta}\,P_i^{\eta}}{\bar{Z}_i\,L_i^{\zeta}\,S^{1-\eta}}, \tag{49}$$

where $\tilde{\eta} \equiv (1-\eta)^{-(1-\eta)}\eta^{-\eta}$ is the Cobb-Douglas constant. Minimising $w_iL_i + P_iM_i$ subject to the production constraint gives conditional factor demands with the split (48), and total cost $C(y_i) = \tilde{\eta}\,w_i^{1-\eta}P_i^{\eta}y_i/(\bar{Z}_iL_i^{\zeta}S^{1-\eta})$; CRS plus free entry set $p_i = C(y_i)/y_i$.

Intermediate spending shares. This is the basis of the gravity equation used to recover δ_{ki},

$$s_{ki}^{F} \;=\; \frac{(\delta_{ki}\,p_k)^{1-\sigma}}{P_i^{\,1-\sigma}}. \tag{50}$$

C.2.4 Equilibrium definition

A general equilibrium, given primitives $\{\bar{Z}_i, A_i, \delta_{ij}, Z_i^H, Z_i^X, U_R\}$ and parameters $(\beta, \sigma, \eta, \zeta, \psi, \varepsilon, \bar{L})$, is a tuple $\left(\{w_i, L_i, p_i, P_i\}_{i=1}^{N},\, L_U,\, \mathcal{V}_U,\, V_i\right)$ such that:

1. the city partial equilibrium of Section C.1.4 holds in every city, given (w_i, L_i) and $\{p_j\}$;
2. firms price by (49);
3. the firm-tier delivered price index satisfies (47);

4. the household price index is the numéraire,

$$P \;=\; \left(\sum_{j=1}^{N} p_j^{1-\sigma}\right)^{1/(1-\sigma)} \;\equiv\; 1; \tag{51}$$

5. goods markets clear via (53);

6. city indirect utility is given by (43) and within-urban Fréchet shares satisfy (44);

7. the rural-urban margin satisfies (45);

8. labor adds up: $L_U + L_R = \bar{L}$ and $\sum_{i=1}^{N} L_i = L_U$.

C.2.5 Equilibrium solution

Goods market clearing. City j's output is consumed by households (facing p_j) and firms (facing $\delta_{ji}p_j$ via P_i). Under the housing closure, land rents and developer profits return to households at $P \equiv 1$, so aggregate household demand equals the wage bill $W \equiv \sum_i w_i L_i$, with city j's share $p_j^{1-\sigma}$. Equating j's output to the sum of household and firm demand:

$$p_j\, y_j \;=\; p_j^{1-\sigma}\, W \;+\; \eta\, p_j^{1-\sigma} \sum_{i=1}^{N} \delta_{ji}^{1-\sigma}\, P_i^{\sigma-1}\, p_i\, y_i. \tag{52}$$

Substituting the cost-share identity $p_i y_i = w_i L_i/(1-\eta)$ and multiplying through by $(1-\eta)$ gives the wage equation:

$$w_j L_j \;=\; p_j^{1-\sigma}\Big[(1-\eta)\, W \;+\; \eta\, \mathrm{MA}_j^F\Big], \qquad \mathrm{MA}_j^F \;\equiv\; \sum_{i=1}^{N} \delta_{ji}^{1-\sigma}\, P_i^{\sigma-1}\, w_i L_i. \tag{53}$$

The first term is the national frictionless demand pool; MA_j^F is the firm-side market access, the trade-cost-weighted wage bill reachable from j.

Wage recursion. Substituting (49) into (53) and collecting powers of w_j:

$$w_j^E\, L_j \;\propto\; \left(\bar{Z}_j\, L_j^{\zeta}\, S^{1-\eta}\right)^{\sigma-1} P_j^{-\eta(\sigma-1)} \Big[(1-\eta)W + \eta\, \mathrm{MA}_j^F\Big], \qquad E \;\equiv\; \sigma(1-\eta)+\eta. \tag{54}$$

The exponent $E = \sigma(1-\eta)+\eta = 3$ at the headline calibration.

At $\eta \to 0$, MA_j^F drops out, $E \to \sigma$, and (54) reduces to the single-tier Armington wage recursion. At $\eta \to 1$, $E \to 1$ and the model approaches a pure roundabout structure.

Existence and uniqueness. The parameter restriction (25) is necessary for the integrals defining $L_i(x)$ in (35) to converge. At the GE level, Section 3 establishes existence and uniqueness via the sufficient condition of Allen et al. (forthcoming): existence and uniqueness obtain when a matrix bounding the elasticities of wages with respect to one another has spectral radius ≤ 1. This requires agglomeration (ζ) to be weak relative to congestion forces in $(\kappa_i, \psi, \varepsilon)$; the condition is checked numerically at the headline calibration (coming soon to a theater near you!)

Walras' law. Summing (53) over j and applying (51) gives $W = (1-\eta)W + \eta W$, so one of the N wage equations is redundant. The system in $(\{p_j\}_{j=1}^{N-1}, \{w_j, L_j\}_{j=1}^{N}, L_U)$ is exactly identified given the numéraire (51), the sorting (44)–(45), and the within-city closed forms.

Aggregate GDP. Because $P \equiv 1$, nominal and real GDP coincide:

$$\text{GDP}^{\$} \;=\; \text{GDP}^{\text{real}} \;=\; W \;=\; \sum_{i=1}^{N} w_i L_i, \tag{55}$$

which is the welfare-meaningful aggregate used in all counterfactuals in Section 5.

(A, V) normalisations for counterfactuals. For the counterfactual pipeline, the recovered city primitives are summarised by the (A, V) pair via $A_i = V_i / w_i$ at the data equilibrium, with V_i from (43) after imposing $\text{geomean}(A_i) = 1$. Under this convention, a pure κ_i-shock at the baseline rescales all A_i by a common factor absorbed by the geomean; pure κ-counterfactuals are therefore a no-op, as required by the Greaney (2026) normalisation. This property holds under both the single-tier and two-tier specifications because the inversion maps $\{\bar{Z}_i, A_i\}$ to data without using κ_i outside of amenity recovery.

Inversion from data to primitives. Given observed gross output Y_i^{obs}, employment L_i^{obs}, total labor $\bar{L}$, urban share s^{obs}, trade-cost matrix δ, and parameters $(\eta, \sigma, \zeta, \psi, \varepsilon, \beta, \kappa_i)$, I recover $(\bar{Z}_i, A_i, U_R)$ by:

1. Wages from the cost-share identity: $w_i = (1-\eta)\, Y_i^{\text{obs}} / L_i^{\text{obs}}$, with country wage bill $W = \sum_i w_i L_i^{\text{obs}}$.
2. Selection: $L_U = s^{\text{obs}} \bar{L}$ and $S = (L_U / \bar{L})^{-1/\varepsilon}$.
3. Joint fixed point in $\{p_j, P_i\}$: iterate (47), (53), and (51) via damped Picard until $\max_j |\Delta \log p_j| < 10^{-10}$, with the $\eta = 0$ closed form $p_j^{1-\sigma} = w_j L_j / W$ as the warm start.
4. Raw TFP from (49): $\bar{Z}_i = p_i^{-1} (w_i/(1-\eta))^{1-\eta} (P_i/\eta)^{\eta} / (L_i^{\zeta}\, S^{1-\eta})$.
5. Indirect utility from within-urban Fréchet shares: $V_i \propto (L_i^{\text{obs}} / L_U)^{1/\psi}$ up to a country-level scale.
6. Amenities from (43): $A_i = V_i / [w_i\, S\, (w_i L_i)^{-\beta\kappa_i}]$, normalised to $\text{geomean}(A_i) = 1$.
7. U_R from a scalar root-find on (45) given s^{obs} and $\mathcal{V}_U$.

The inversion is exactly identified given the numéraire and the Walras-redundant wage equation. I verify numerically: the wage bill identity $W = (1-\eta)W + \eta W$ closes to machine precision, $P = 1$ to within 10^{-10}, and $\bar{Z}_i > 0$ throughout.

C.3 Rural sector closure

I derive w_R from the farm-ownership structure and establish $U_R = A_R w_R$ as used in (45).

Rural production is Cobb-Douglas in labor and land, $Y_a = Z_a\, L_a^{1-\mu}\, T_a^{\mu}$. Each rural household owns the unit plot it farms; land rents therefore accrue to the household rather than to an outside landlord. Rural land is supplied at a fixed price q_R (constant marginal cost), and each household purchases one unit, so total farm land equals the rural labor force, $T_a = L_a$. The land-labor ratio is identically one, giving average product per worker:

$$w_R \;=\; \frac{Y_a}{L_a} \;=\; Z_a \left(\frac{T_a}{L_a}\right)^{\mu} \;=\; Z_a, \tag{56}$$

a constant independent of L_a. The rebate confirms this: the competitive wage equals the marginal product $(1-\mu)Z_a$, and the productive-land rent μZ_a flows back to the household, so total income is $(1-\mu)Z_a+\mu Z_a = Z_a = w_R$.

Standard Cobb-Douglas optimization with income w_R and housing-land price q_R gives indirect utility $\mathcal{V}_R \propto A_R\, w_R\, q_R^{-\beta}$. Since q_R is a fixed primitive, absorbing it into A_R:

$$\mathcal{V}_R \;=\; A_R\, w_R, \tag{57}$$

the form used in Section 3. A_R encodes non-wage amenity forces that keep workers in agriculture. Calibrating $U_R = \mathcal{V}_R$ to the observed urban share absorbs both A_R and w_R; the agricultural sector enters the inner-loop solver only through U_R.

D Cost elasticity estimation details

This appendix details the estimation of the floorspace supply elasticity γ_i, the land supply elasticity ρ_i and the commuting cost elasticity, τ_i.

D.1 Building-height gradient θ_i

The building height gradient, θ_i, the coefficient on floorspace density against distance to the CBD, estimated using a Poisson regression. I estimate it city-by-city using `fixest::fepois` in R on teh 100m raster. I make a number of adjustments to improve large-scale estimation.

'Donut' estimation I cut out a 'donut hole' within each city that scales with urban area. I do so to better estimate the *residential* gradient, and not construction in the CBD. This is important as in rich-world cities, the $\approx 1 - 2km$ donut hole has a 'flat top' in which there are many tall buildings of similar height and no decay. This affects estimation, creating a steep drop-off between the commercial and residential gradient in the rich world. The cut-out radius is, $r_k(\text{city}) = 2.0\,\text{km} \cdot (A_{\text{city}}/3128)^{1/4}$, where $3128\,\text{km}^2$ is Shanghai's area, so the donut adapts to city size while pinning the reference at the city.

Shrinkage to enforce a sign restriction As $\theta_i > 0$ is required by the structural model (it underlies $\tau_i = \theta_i/\gamma_i > 0$ and hence κ_i), I shrink $\hat{\theta}_i$ shrunk onto $(0, \infty)$. I do so with an empirical Bayes routine following Morris (1983). I fit the hierarchical model,

$$\begin{aligned} \hat{\theta}_i \mid \theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \mathcal{N}_{(0,\infty)}(\mu_c, \sigma_c^2) \end{aligned}$$

where $\mathcal{N}_{(0,\infty)}$ denotes a truncated normal distribution. I estimate the parameters of the 'prior' following the fixed point iteration outlined in Morris (1983). Given these parameters, the posterior mean is given in closed form, as a normal likelihood is conjugate with a truncated normal prior. The closed form is the standard normal-normal shrinkage estimate with an inverse-Mills correction for the truncation.

Figure A4 plots the cross-city distribution of $\hat{\theta}_i$ together with the smoothed posterior density.

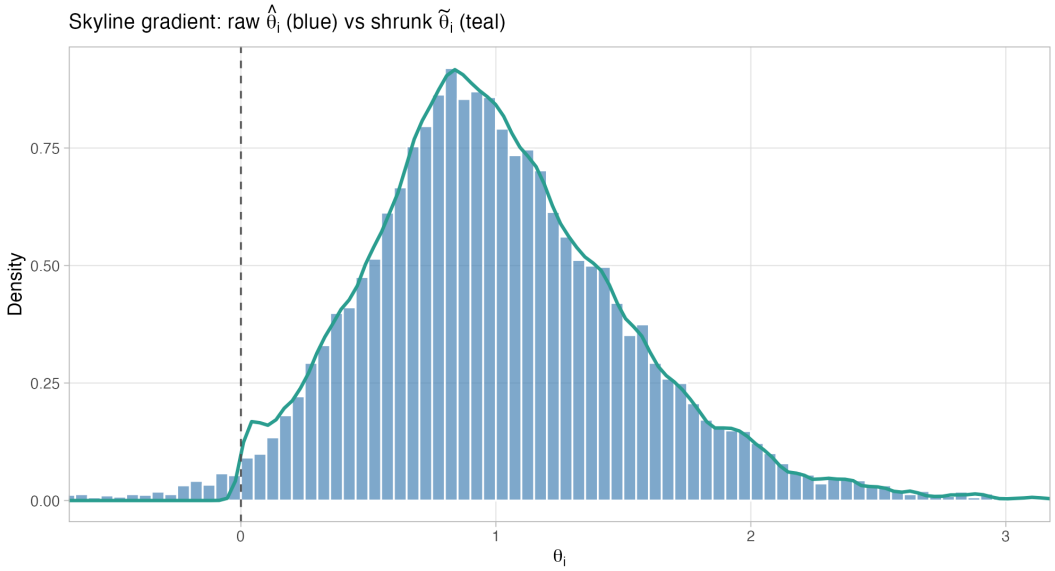

Figure A4: θ_i estimates: raw Poisson estimates (blue histogram) and empirical Bayes posterior means (teal density).

	p10	Median	p90	Pop-weighted mean	% theory-consistent
Floorspace supply elasticities $\hat{\gamma}_i$ (per territory)					
Pooled DML-PLIV-RF: $\hat{\gamma} = 0.44$ (0.05), first-stage $F = 95.3$.					
Raw $\hat{\Gamma}(G_i)/(1-\hat{\Gamma}(G_i))$	0.24	0.52	0.91	0.64	96%
Bayesian posterior (uniform prior)	0.22	0.51	0.90	0.62	100%
Land supply elasticities $\hat{\rho}_j$ (per city)					
Pooled DML-PLIV-RF: $\hat{\rho} = 0.91$ (0.11), first-stage $F = 75.7$.					
Raw $\hat{B}(G_j)/(1-\hat{B}(G_j))$	0.50	1.10	2.89	1.52	99%
Bayesian posterior (uniform prior)	0.50	1.10	2.77	1.39	100%
Commuting cost elasticities $\hat{\tau}_i = \hat{\theta}_i/\hat{\gamma}_i$ (per territory; derived, no DML)					
Raw $\hat{\theta}_i^{\text{raw}}/\hat{\gamma}_i^{\text{raw}}$	0.70	1.91	4.53	2.15	95%
Plug-in from shrunk inputs ($\tilde{\theta}^{\text{shrunk}}/\tilde{\gamma}^{\text{shrunk}}$)	0.67	1.93	4.63	2.05	100%
Urban cost elasticities $\hat{\kappa}_i$ (per territory; assembled, no DML, no shrinkage)					
$\hat{\kappa}_i = \frac{1}{(1+\hat{\rho}_i)(1+\hat{\gamma}_i)} + \frac{\hat{\rho}_i}{1+\hat{\rho}_i} \cdot \frac{\hat{\tau}_i}{2}$	0.46	0.81	1.65	0.85	100%

Table A9: Distribution of elasticity estimates from the DML-PLIV and instrumental forest estimates.

D.2 Estimation of γ_i and ρ_i by DML-PLIV and instrumental forests

The floorspace-supply elasticity γ_i and the land-supply elasticity ρ_i are estimated with a common pipeline: a pooled DML-PLIV stage that delivers a single point estimate to understand instrument strength, and then an instrumental-forest stage that recovers per-territory ratios $\Gamma(G_i) = \gamma_i/(1+\gamma_i)$ and $\varrho(G_j) = \rho_j/(1+\rho_j)$.

Instrumental forest for heterogeneity For both the floorspace supply, and land supply, and for each covariate value g, the forest estimates $\Gamma(g)$ as the solution to the local IV moment condition

$$\sum_i \alpha_i(g)\, \tilde{z}_i\, (\tilde{y}_i - \Gamma(g)\tilde{x}_i) = 0,$$

where $\alpha_i(g)$ are forest weights learned from the covariates G_i, where $\tilde{y}_i, \tilde{x}_i, \tilde{z}_i$ are net of country fixed effects. I additionally control for covariates G_i. Land supply elasticities are estimated similarly. To do so I use `grf::instrumental_forest` (Athey et al., 2019). Both forests use 2,000 trees, minimum node size 20, sample fraction 0.5, and honesty groups of size 2. For γ, Y_i, D_i, Z_i are pre-residualised against the primate indicator before entering the forest so that primate status is not a heterogeneity dimension; the forest's feature matrix is the raw (un-demeaned) geographic levels and pairwise interactions, which lets the forest split on absolute terrain and soil levels. Forest standard errors $\hat{\sigma}_i$ are recovered from `predict(..., estimate.variance=TRUE)`.

Figure A5 reports the top 20 features by split frequency in each forest. Figure A6 shows the sorted per-city $\widehat{\Gamma(G_i)}$ for γ (left panel) and per-city $\widehat{\varrho(G_i)}$ for ρ (right panel)

Shrinkage The model predicts that the instrumental forest estimates must lie within (0,1). I enforce this by postulating a uniform prior on the estimates and shrinking the instrumental forest estimates by recovering the Bayesian posterior means for each estimate. On A6 within each panel the blue line is the Uniform$(0,1)$ posterior mean $\tilde{\beta}_i$, the orange line is the raw $\hat{\beta}_i$, and the shaded band is the 95% pointwise CI from the truncated-normal posterior variance.

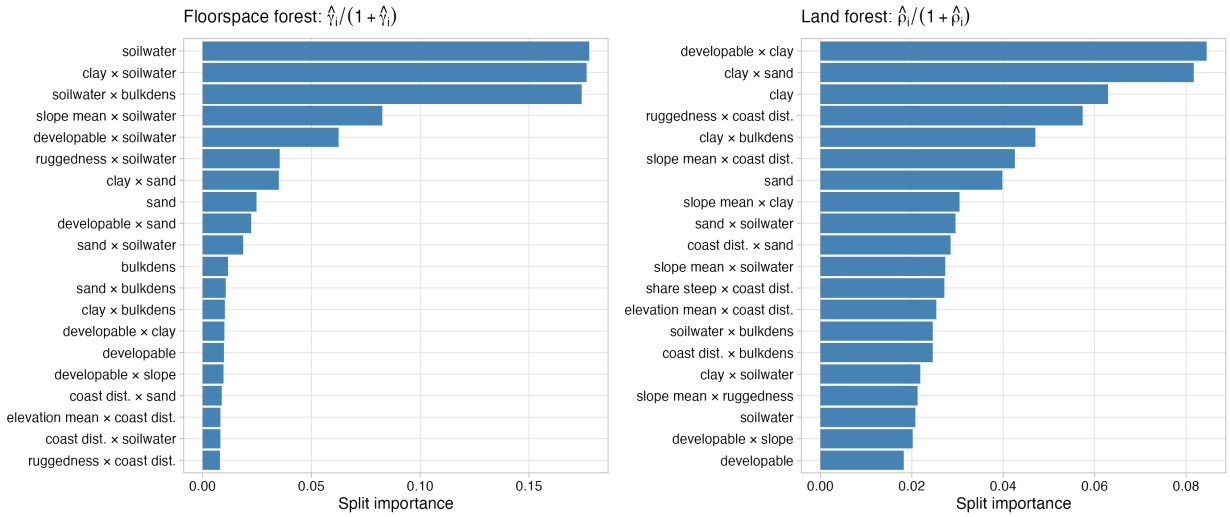

Figure A5: Variable importance in the instrumental forest for floorspace supply elasticities (left) and land supply elasticities (right). Split importance is the share of forest splits made on each covariate.

With a Normal likelihood $\hat{\beta}_i \mid \beta_i \sim \mathcal{N}(\beta_i, \hat{\sigma}_i^2)$, the posterior is truncated-Normal on $(0, 1)$ and the closed-form posterior mean

$$\tilde{\beta}_i \;=\; \hat{\beta}_i + \hat{\sigma}_i \cdot \frac{\phi(a_i) - \phi(b_i)}{\Phi(b_i) - \Phi(a_i)}, \qquad a_i = \frac{0 - \hat{\beta}_i}{\hat{\sigma}_i},\; b_i = \frac{1 - \hat{\beta}_i}{\hat{\sigma}_i},$$

is taken per-city with no cross-unit pooling.

Figure A7 shows, for $\Gamma(G_i) = \gamma_i/(1+\gamma_i)$ (left panel) and $\varrho(G_i) = \rho_i/(1+\rho_i)$ (right panel), the raw $\hat{\beta}_i$ histogram (blue) together with the Uniform$(0,1)$ prior (orange step function) and the truncated-normal posterior density (teal curve) overlaid in each panel.

Plug-in assembly of τ_i and κ_i The commuting-cost elasticity τ_i is not estimated directly. The model implies $\theta_i = \tau_i\,\gamma_i$, so τ_i is assembled by the plug-in

$$\tilde{\tau}_i \;=\; \frac{\tilde{\theta}_i^{\text{shrunk}}}{\tilde{\gamma}_i^{\text{shrunk}}}.$$

Both inputs are strictly positive by construction.

I then assemble the urban-cost elasticity from the shrunken inputs as

$$\hat{\kappa}_i \;=\; \frac{1}{(1+\tilde{\rho}_j)(1+\tilde{\gamma}_i)} \;+\; \frac{\tilde{\rho}_j}{1+\tilde{\rho}_j} \cdot \frac{\tilde{\tau}_i}{2},$$

the first term being the intensive (crowding-within-structures) margin and the second being the extensive (peripheral-expansion) margin. Both terms are strictly positive given $\tilde{\rho}_j, \tilde{\gamma}_i, \tilde{\tau}_i > 0$, so $\hat{\kappa}_i > 0$ for every city.

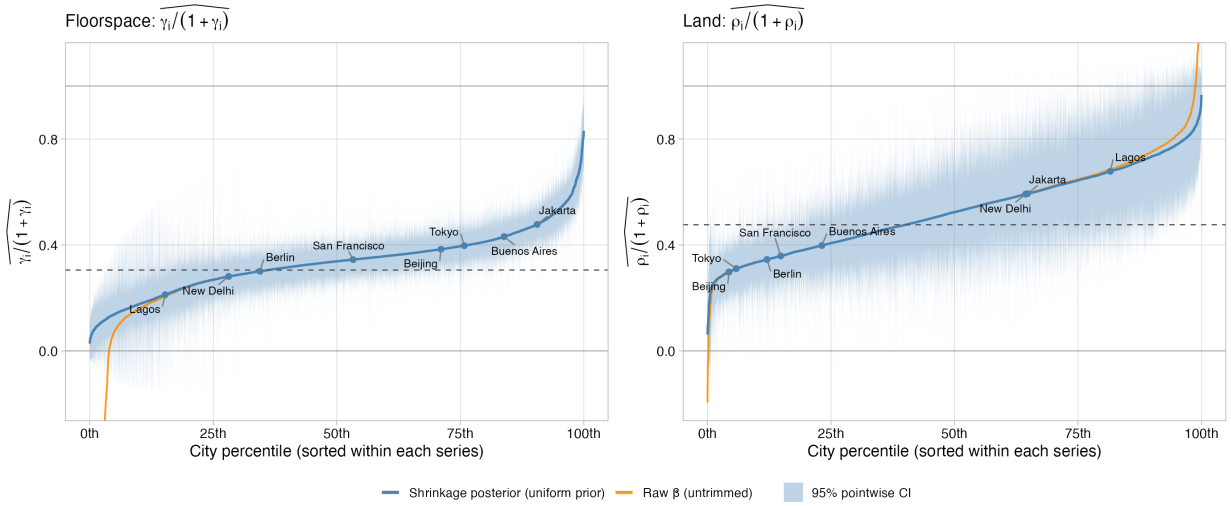

Figure A6: Sorter per-city supply elasticty estimates. Left: $\gamma/(1+\gamma)$ estimates. Right: $\rho/(1+\rho)$ estimates. Orange is raw estimate, blue is posterior mean from Bayesian shrinkage. Some major cities labeled.

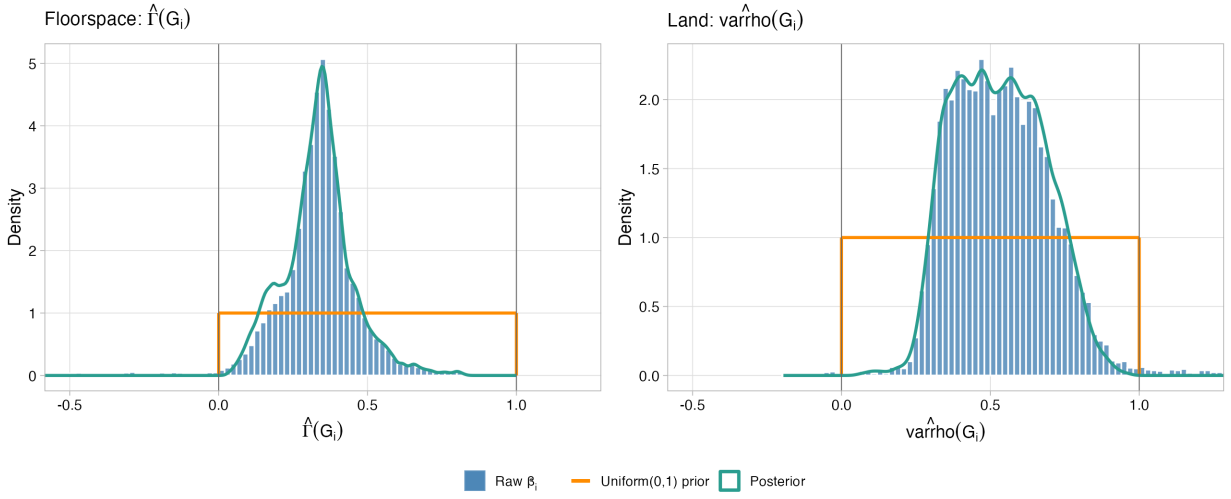

Figure A7: Prior (orange), likelihood (data, blue histogram), and posterior (teal density estimate) for floorspace and land supply elasticity ratios. Left: $\gamma/(1+\gamma)$. Right: $\rho/(1+\rho)$

D.3 External validation of τ_i, γ_i, ρ_i

A natural external check on the commuting cost elasticity is whether τ_i covaries with directly observable measures of urban mobility. Akbar et al. (2023b) construct driving-speed indices for over 1,000 cities globally using Google Maps queries, splitting observed travel times into a *freeflow* (midnight) and a *congested* (midday) component. Table A10 reports correlations of $\ln \tau_i$ with these indices, both unconditionally and conditional on country fixed effects and city characteristics.

	Dependent variable: $\ln \tau$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Downtown 5km speed measurements						
ln(speed, midnight)	-0.081		-1.565	-0.372		-0.442
	(0.373)		(0.724)	(0.210)		(0.301)
ln(speed, midday)		0.332	1.499		-0.236	0.068
		(0.275)	(0.547)		(0.158)	(0.214)
R^2	0.036	0.045	0.085	0.446	0.445	0.446
Panel B. City-wide speed indices						
Uncongested speed (z)	0.101		-2.221	-0.404		-0.729
	(0.416)		(1.294)	(0.291)		(0.674)
Speed (z)		0.376	2.414		-0.278	0.296
		(0.356)	(1.157)		(0.301)	(0.751)
R^2	0.036	0.043	0.071	0.445	0.444	0.445
Observations	1,072	1,072	1,072	1,027	1,027	1,027
Controls	✓	✓	✓	✓	✓	✓
Country fixed effects				✓	✓	✓

Table A10: $\ln \tau_i$ versus Google-Maps queried driving speeds from Akbar et al. (2023b).

To validate γ_i and ρ_i, I compare them to estimates in Baum-Snow and Han (2024) and Saiz (2010) in Figure A8.

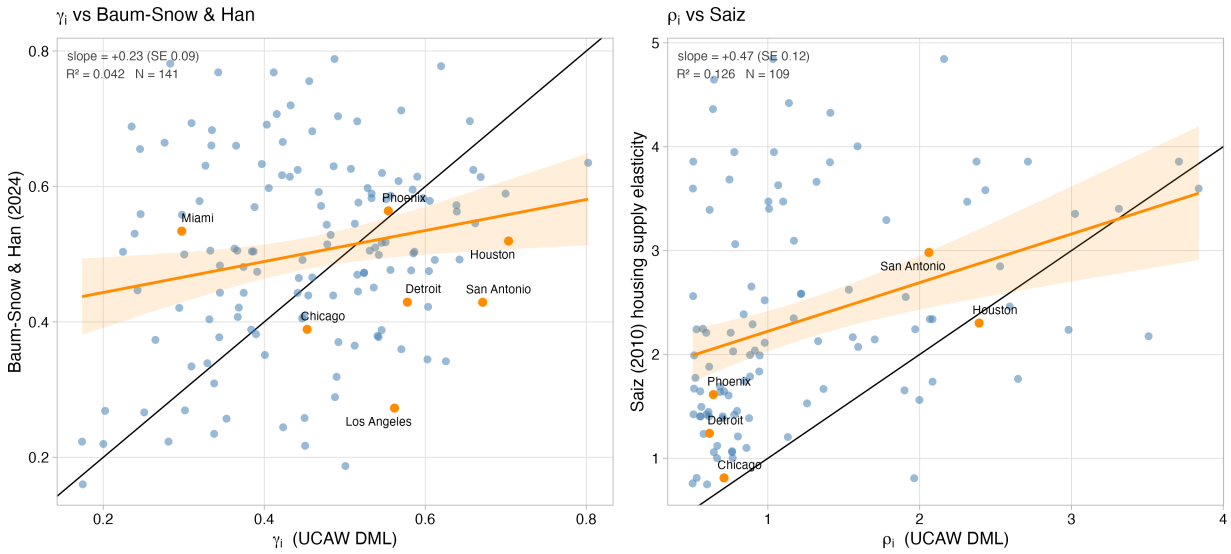

Figure A8: Left: correlation between MSA housing supply elasticities from Baum-Snow and Han (2024) and $\hat{\gamma}_i$. Right: estimated land supply elasticities from Saiz (2010) for US MSAs versus $\hat{\rho}_i$. Black line is the 45-degree line. Unweighted OLS fit with 95% CI in orange on both plots. Some major cities labeled.

E Model quantification: calibration details

This appendix documents the details of the model quantification as discussed in Section 5.

Floorspace expenditure share β from ICP 2017

I calibrate the parameter β country-by-country using the World Bank ICP 2017 round (source 78). In particular, I use,

- Series `9060000`: housing and utilities expenditure, classification `ZS` (share of GDP = 100%).
- Series `9020000`: total individual consumption, classification `ZS`.

For each country, I compute $\beta_n =$ the ratio of these series. These data do not cover all countries.

Imputation The ICP 2017 round covers 162 countries. For the remaining countries in the master panel, β is imputed from a quadratic fit in log GDP per capita (PPP):

$$\hat{\beta}_n = \hat{a}_0 + \hat{a}_1 \ln(\text{GDP/cap}_n) + \hat{a}_2[\ln(\text{GDP/cap}_n)]^2, \tag{58}$$

estimated by OLS on the ICP data.

Figure A9: OSRM road and ferry routes for Indonesia. Solid blue: road legs. Dashed orange: ferry-classified legs (route passes > 500 m outside the Indonesia land polygon). Orange polygons: UCDB city footprints. Black outline: Jakarta agglomeration boundary. Of 428 unique city-to-city legs, 97 (23%) are ferry-classified. Source: OSRM routing on OpenStreetMap

Recovering bilateral trade costs

This appendix discusses how I estimate bilateral trade costs between cities. At a high level, I model bilateral trade costs as a function of distance, measure distances between all city pairs within each country in the data with the Open Source Routing Machine (OSRM). I then estimate a mapping of distance to trade costs in U.S. shipping data, and apply that mapping to estimate bilateral frictions at the city-pair level in my UCDB-derived data.

Data sources Three data sources combine to estimate trade costs.

1. The primary one is the 2017 Commodity Flow Survey (CFS) public use microdata sample. This contains establishment level records of shipments between firms within the United States in 2017. Each record contains an origin and destination metropolitan area as well as a value of the shipment (in USD), how the shipment was routed, and the reported travel distance.

2. To map the trade elasticity onto travel distances between cities in each country, I query pairwise road distances and drive durations between every intra-country city pair on the OSRM server, using CBD to CBD routing with my estimated CBDs.

3. The final item is the Census 2022 CFS Areas file. I use the TIGER/Line county polygons and the mapping of county FIPS to CFS17 Areas to construct polygons for the CFS17 metro areas.

CFS Sample construction I do the following:

1. Keep road-based shipments only ie truck based or multimodal that includes truck.
2. I drop shipments between non-CONUS states and territories.
3. I collapse to origin-destination metro area pairs, yielding 4,817 pairs.

Figure A9 illustrates the OSRM city-pair driving input in Indonesia as an instructive archipelago case. The figure plots 428 unique city-to-city legs that decompose the 412 OSRM routes from Jakarta to every other Indonesian UCDB territory centroid. Ferry legs are displayed as dashed lines. I prefer drive time instead of distance because the travel duration is much longer in the presence of ferries.

CFS to UCDB crosswalk I assign each UCDB city to the US county whose polygon contains its population-weighted centroid, and then map those counties to CFS Areas.

OSRM aggregation from cities to metros I aggregate intercity distances between metro m and n pairs by population-weighting drive distance or duration,

$$\bar{d}_{mn} = \frac{\sum_{i \in m, j \in n} \text{pop}_i \text{pop}_j d_{ij}}{\sum_{i \in m, j \in n} \text{pop}_i \text{pop}_j}$$

using 2015 city populations.

Specifications and estimates I model the bilateral trade cost between cities i and j in the same country as,

$$\delta_{ij} = (1 + d_{ij})^{\phi}, \quad \delta_{ii} = 1$$

where d_{ij} is a measure of travel duration (driving on roads; OSRM multiplies this for sea/ferry voyages). In the model, trade-share expressions in the urban varieties block of the model only enter through $\delta^{1-\sigma}$, so the parameter I seek to estimate is $\beta = -\phi(\sigma - 1)$.

I embed this model into the estimation equation,

$$\mathbb{E}\left[\text{shipmt\_value}_{mn}\right] \;=\; \exp\!\big(\beta \, \log(1 + \bar{d}_{mn}) + \gamma \, \mathbf{1}\{m = n\} + \alpha_m + \alpha_n\big)\,, \tag{59}$$

which I estimate via Poisson pseudo-maximum likelihood (Silva and Tenreyro, 2006; Correia et al., 2020).

Estimates are available in Table A11. I use the third column for calibration.

	(1) CFS dist	(2) OSRM dist	(3) OSRM duration
log(1 + CFS dist km)	-0.973		
	(0.045)		
log(1 + OSRM dist km)		-0.837	
		(0.035)	
log(1 + OSRM duration h)			-1.270
			(0.064)
Self-flow (orig = dest)	-0.164	0.099	0.190
	(0.103)	(0.097)	(0.117)
Observations	4,817	4,225	4,225
Pseudo R^2	0.958	0.952	0.952
Origin MA FE	✓	✓	✓
Destination MA FE	✓	✓	✓

Table A11: 2017 Commodity Flow Survey shipment values versus travel costs between metropolitan area pairs. All specifications use a PPML estimator with origin and destination fixed effects and include an indicator for whether the origin equals the destination. Column (1): CFS-reported road distance. (2) OSRM drive distance, pop-weighted by cities within each CFS metropolitan area. (3) OSRM drive duration, preferred specification. Robust standard errors in parentheses.

Calibration Moments: WDI Aggregates

The model inversion requires matching a set of macroeconomic aggregates in order to facilitate international comparisons. Table A12 describes these moments, and the data I match in the World Development Indicators. All data is for 2015.

Model object	Formula from data	WDI indicator(s)	Notes
City-level output Y_i	$Y_i = Y_n^{\text{non-ag}} \cdot \frac{\text{NTL}_i^{\varepsilon_{\text{NTL}}}}{\sum_j \text{NTL}_j^{\varepsilon_{\text{NTL}}}}$	`NY.GDP.MKTP.CD` (GDP, current USD); `NV.AGR.TOTL.ZS` (ag VA share)	Non-ag GDP = GDP $\times(1-$ ag share); NTL from VIIRS 2015 (`ntl_sum_2015`)
City-level wage w_i	$w_i = (1-\eta)\, Y_i / L_i$	(derived from Y_i, L_i)	Firm FOC cost-share identity; not directly observed
City population L_i	$L_i = \tilde{p}_i \cdot \frac{s_n \bar{L}_n}{\sum_j \tilde{p}_j}$	`SP.POP.TOTL` ($\bar{L}_n$); `SP.URB.TOTL.IN.ZS` (s_n)	$\tilde{p}_i$ = WorldPop 2015 city headcount; $s_n \bar{L}_n$ is observed urban population
Urban share s_n	$s_n = L_{U,n} / \bar{L}_n$	`SP.URB.TOTL.IN.ZS` (as fraction)	Used directly to calibrate U_R
Rural wage w_R	–	`NV.AGR.TOTL.ZS` (2015)	Agricultural productivity

Table A12: Calibration moments and data sources for macro aggregates in 2015.