



Full Length Articles

The comparative advantage of cities[☆]Donald R. Davis^a, Jonathan I. Dingel^{b,*}^a Columbia University and NBER, United States of America^b University of Chicago Booth School of Business and NBER, United States of America

ARTICLE INFO

Article history:

Received 14 April 2019

Received in revised form 27 November 2019

Accepted 13 December 2019

Available online 10 January 2020

Mendeley Data <https://data.mendeley.com/datasets/x2bgryrgbyj/draft?a=9d2c07b1-6ff8-4af0-b60c-1beabba577f8>

JEL classification:

F11

F14

R12

R13

Keywords:

Agglomeration

Assignment models

Cities

Comparative advantage

ABSTRACT

What determines the distributions of skills, occupations, and industries across cities? We develop a theory to jointly address these fundamental questions about the spatial organization of economies. Our model incorporates a system of cities, their internal urban structures, and a high-dimensional theory of factor-driven comparative advantage. It predicts that larger cities will be skill-abundant and specialize in skill-intensive activities according to the monotone likelihood ratio property. We test the model using data on 270 US metropolitan areas, 3 to 9 educational categories, 22 occupations, and 19 industries. The results provide support for our theory's predictions.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The level of economic activity varies considerably across cities. This motivates the study of the agglomeration forces that give rise to concentration. The composition of economic activity also varies considerably across cities. This is evident from *Forbes* ranking “America’s smartest cities” and the use of place names like Wall Street and Silicon Valley as shorthand for industries. Who lives in more productive cities and what they do there should inform our understanding of how

agglomeration economies work. But most theories of cities, to the extent that they consider the spatial distributions of skills and sectors, are dichotomous in nature: individuals are either skilled or unskilled and cities either perform one function or all functions. These low-dimensional accounts are not well suited to interpret data exhibiting much richer variation.

What governs the distributions of skills and sectors across cities? Who lives where and what do they do there? These are challenging questions. In this paper, we develop an analytic approach we term the comparative advantage of cities. Our aim is to provide the simplest theoretical framework capable of providing unified answers to these questions and then to take our model’s predictions to the data. We introduce a model of many cities, many skills, and many sectors, and we show that it fits the cross-section of US cities well.

In our spatial-equilibrium model, the comparative advantage of cities is jointly governed by the comparative advantage of individuals and their locational choices. Cities are symmetric *ex ante*, so cross-city heterogeneity is an emergent outcome. Agglomeration economies make cities with larger, more skilled populations exhibit higher total factor productivity (TFP). Locations within cities exhibit heterogeneity in their innate desirability, as is customary in land-use models (Fujita and Thisse, 2002, Ch 3). These cities are populated by heterogeneous

[☆] Thanks to Rodrigo Adao, Arnaud Costinot, Gilles Duranton, Vernon Henderson, Corinne Low, Joan Monras, Gianmarco Ottaviano, Keeyoung Rhee, John Romalis, Bernard Salanié, Kurt Schmidheiny, Will Strange, Bruno Strulovici, Daniel Sturm, Matt Turner, Jonathan Vogel, and seminar audiences at Barcelona GSE, Berkeley, CURE, NBER ITI, NBER URB, New York Fed, NYU Stern, Penn State, Philadelphia Fed, Princeton IES Summer Workshop, UCLA, and the Urban Economics Association for helpful comments and suggestions. We thank Dylan Clarke, Yuxiao Huang, and especially Luis Moreira da Costa and Antonio Miscio for research assistance. Dingel thanks the Institute for Humane Studies, the Program for Economic Research at Columbia University, and the Kathryn and Grant Swick Faculty Research Fund at the University of Chicago Booth School of Business for financial support.

* Corresponding author.

E-mail addresses: drdavis@columbia.edu (D.R. Davis), jdingel@chicagobooth.edu (J.I. Dingel).

individuals with a continuum of skill types, and these individuals may be employed in a continuum of sectors. Comparative advantage causes more skilled individuals to work in more skill-intensive sectors, as in Ricardo-Roy models (Costinot and Vogel, 2015). While those models of international trade assume exogenous endowments, we study the comparative advantage of cities in an environment in which cities' factor supplies are endogenously determined by the choices of mobile individuals. There is a complementarity between individual income and locational attractiveness, so more skilled individuals are more willing to pay for more attractive locations and occupy these locations in equilibrium, as in the differential-rents model of Sattinger (1979).

In equilibrium, agglomeration, individuals' comparative advantage, and locational heterogeneity within cities combine to deliver a rich set of novel predictions. Agglomeration causes larger cities to have higher TFP, which makes a given location within a larger city more attractive than a location otherwise of the same innate desirability within a smaller city. For example, the best location within a larger city is more attractive than the best location within a smaller city due to the difference in TFP. Since more skilled individuals occupy more attractive locations, larger cities are skill-abundant. The most skilled individuals in the population live only in the largest city and more skilled individuals are relatively more prevalent in larger cities. By individuals' comparative advantage, the most skill-intensive sectors are located exclusively in the largest cities and larger cities specialize in the production of skill-intensive output. Our model therefore predicts an urban hierarchy of skills and sectors. Under slightly stronger assumptions, larger cities will be absolutely larger in all sectors.

As we discuss in Section 2, prior theories describing cities' sectoral composition have overwhelmingly focused on the polar cases in which cities are either completely specialized "industry towns" or all produce the same composite output. The former is starkly at odds with the data, while the latter makes no substantive prediction. Our theory predicts that cities are incompletely specialized across sectors due to incomplete skill sorting: the trade-off between city-level TFP differences and within-city locational desirability differences can make individuals of a given skill level indifferent between locations in different cities. At the same time, our model relates the pattern of specialization to cities' observable characteristics. It makes strong, testable predictions about the distributions of skills and sectors across cities.

We examine the model's predictions about the distribution of skills and sectors across US cities using data from the 2000 Census of Population, County Business Patterns, and Occupational Employment Statistics described in Section 4. We use two empirical approaches to characterize cities' skill and sectoral distributions. The first regresses a city's log sectoral employment on its log total population. More skilled groups and more skill-intensive sectors should exhibit higher population elasticities. The second examines whether the distributions exhibit the monotone likelihood ratio property, as per Costinot (2009), by comparing a pair of skills or sectors across a pair of cities.¹ For example, comparing two cities and two sectors, the skill-intensive sector should have relatively larger employment in the larger city. To characterize sectoral size, we simply compare sectors' employment levels across cities.

Section 5 reports the results, which provide support for our model's predictions about the pattern of skills and sectors across cities. Characterizing skills in terms of three or nine educational groups, we find that larger cities are skill-abundant. Characterizing sectors in terms of 19 industrial categories or 22 occupational categories, we find that larger cities specialize in skill-intensive sectors. Our model's predictions are generally borne out by the data and are statistically significant. If deviations from our model's predictions are idiosyncratic rather than systematic, the empirical success rate should increase with aggregation over groups of cities. This is indeed the case. There is systematic

variation in cities' skill and sectoral distributions that is consistent with the predictions of our theory.

Our account of the distributions of skills and sectors across cities has implications for both research and policy. We find systematic sorting of finely differentiated educational attainment levels across cities. This means that researchers examining spatial variation in binary college/non-college terms report outcomes that partly reflect cross-city compositional differences within these broad educational categories. Similarly, we find that a city's sectoral composition is systematically related to its population size. This means that research designs exploiting cross-city variation in sectoral composition, à la Bartik (1991), that do not control for population size are relying upon size-driven variation, not just idiosyncratic circumstances, for identification. Our model also has potentially interesting implications for welfare. For example, our model of spatial sorting across locations means that biased technical change has asymmetric welfare consequences. In our setting, skill-biased technical change raises attractive locations' rental prices without increasing less attractive locations' prices, while unskilled-biased technical change raises rental prices everywhere. Finally, due to the most skilled individuals residing only in the largest cities, the range of skills and thus the range of welfare is greater in larger cities. These differences likely shape local policymakers' challenges and choices.

In sum, we develop a theory of the distributions of skills and sectors across cities of different sizes and show that it fits the US data well. The model combines elements of urban land-use theory and assignment models of comparative advantage. These are a natural foundation for a theory of a system of cities because locational heterogeneity within and across cities leads the distribution of skills to exhibit the monotone likelihood ratio property. This arises not by assumption but instead as an emergent property of the equilibrium that brings in its wake similar implications for the distribution of sectors. This high-dimensional theory provides a basis for what is, relative to the prior literature, a richer empirical examination of the distributions of skills, occupations, and industries across cities of different sizes.

2. Related literature

Our contributions are related to a diverse body of prior work. Our focus on high-dimensional labor heterogeneity is related to recent developments in labor and urban economics. Our theoretical approach integrates elements from the systems-of-cities literature, land-use theory, and international trade. Our model yields estimating equations and pairwise inequalities describing the comparative advantage of cities that are related to prior reduced-form empirical work in urban economics, despite a contrast in theoretical underpinnings.

Our theory describes a continuum of heterogeneous individuals. A large share of systems-of-cities theories describe a homogeneous population (Abdel-Rahman and Anas, 2004). Most previous examinations of heterogeneous labor have only described two skill levels, typically labeled skilled and unskilled.² Understanding the distribution of skills across cities with more than two homogeneous types is valuable for a number of reasons. First, a prominent literature in labor economics has emphasized the need to move beyond models with just two homogeneous skill groups if we want to understand important empirical developments such as wage polarization, job polarization, and simultaneous changes in between- and within-group inequality (Acemoglu and Autor, 2011). Second, using only two skill groups is problematic for understanding spatial variation in relative prices and quantities.

² A long line of empirical work describes cross-city variation in the share of residents who have a college degree (Glaeser, 2008). Most closely related to our work is Hendricks (2011), who finds a weak relationship between cities' industries and college shares. We focus on theories in which labor is heterogeneous in an asymmetric sense (e.g., more skilled individuals have absolute advantage in tasks or more skilled individuals generate greater human-capital spillovers). There are also models describing matching problems, such as Helsley and Strange (1990) and Duranton and Puga (2001), in which labor is heterogeneous in a horizontal dimension.

¹ The distribution $f_c(\sigma)$ likelihood ratio dominates $f_c(\sigma')$ if, for any $\sigma > \sigma'$, $\frac{f_c(\sigma)}{f_c(\sigma')} \geq \frac{f_c(\sigma')}{f_c(\sigma)}$.

The relative price of skill varies across cities, as documented by Baum-Snow and Pavan (2013) and Davis and Dingel (2019), and this is inconsistent with standard spatial-equilibrium models with two homogeneous skill types and homothetic preferences.³ The relative quantity of skill varies across cities, but whether this cross-city variation predicts city growth depends on how the researcher partitions educational attainment into two skill groups.⁴

To overcome the difficulties of interpreting the world through a two-skill model, we assume a continuum of skills, like the spatial models in Behrens et al. (2014) and Davis and Dingel (2019). Relative to those papers, we introduce multiple tradable sectors so that we can analyze the pattern of sectoral specialization, and we introduce intra-city geographic heterogeneity that causes cities' skill distributions to overlap rather than exhibit strict sorting.

Our model integrates systems-of-cities theory with land-use theory. The canonical model of land rents is one in which locations within a city have heterogeneous innate desirability (Fujita and Thisse, 2002, Ch 3). Prior models of a system of cities have only incorporated urban structure as a city-level congestion mechanism, assuming that in equilibrium individuals are indifferent across all locations within a city (Abdel-Rahman and Anas, 2004; Behrens et al., 2014). Our model describes multiple cities with internal geographies when individuals' valuations of locations within cities differ systematically by income. The essential idea is that individuals choosing between living in Chicago or Des Moines simultaneously consider in what parts of Chicago and what parts of Des Moines they might locate.

Considering both dimensions simultaneously is more realistic in both the description of the economic problem and the resulting predicted cross-city skill distributions. Large cities like Chicago contain very heterogeneous neighborhoods, and within cities higher-income individuals tend to choose more attractive, more expensive locations. The complementarity of skills and city size alone would lead to perfect sorting across cities by skill if there were indifference within cities. Adding the dimension that agents sort by the innate desirability of locations within cities means that individuals of a given skill level are found in many cities in equilibrium. This is more realistic than generating such overlap by simply perturbing perfect sorting with idiosyncratic noise, since that approach would yield no relationship between individuals' incomes and their housing prices.

In our theory, we assume heterogeneity in the innate desirability of locations within each city without imposing a particular geography. We assume two restrictions on the relevant heterogeneity. First, the distribution of innate locational desirability within each potential city is *ex ante* identical, consistent with our aim of characterizing an emergent equilibrium. Second, for our key theoretical results, this distribution must exhibit a regularity condition introduced in subsection 3.4. Thus, we follow the long tradition of treating the distribution of the innate desirability of locations as exogenous, without assuming a functional form. We show that our model makes predictions about the distribution of land prices consistent with available evidence while being sufficiently flexible to accommodate a number of potential spatial configurations. While we think heterogeneous valuations of distinct locations within cities is an important subject in its own right, it also involves considerable complications, as made clear by Duranton and Puga (2015), so in the present paper we limit our inquiry in this dimension.

Our model belongs to a long theoretical tradition describing factor-supply-driven comparative advantage, as in the two-good-two-factor

Heckscher-Ohlin theory formalized by Samuelson (1948). In international contexts, theorists have typically taken locations' factor supplies as exogenously endowed. Since individuals are mobile across cities, our theory endogenizes cities' factor supplies while describing how the composition of output is governed by comparative advantage. Our approach to comparative advantage with a continuum of factors and a continuum of sectors follows a large assignment literature and is most closely related to Costinot (2009).⁵ To obtain definite results about the distribution of outputs across countries, Costinot (2009) assumes that the exogenous distribution of factor endowments across countries satisfies the monotone likelihood ratio property. Any economic mechanism that might generate that pattern is beyond the scope of his theory. By contrast, in the present paper the fact that the distribution of skills across cities satisfies this property is a result rather than an assumption. Similarly, while factor endowments, country size, and country productivity are independent in Costinot (2009), they are closely linked in our model such that the distribution of endowments is tied to an observable characteristic, city size. Thus, from a theoretical perspective, cities within a country constitute a natural setting to examine these theories of comparative advantage. Moreover, the assumption of a common production technology is likely more appropriate within than between economies, and data from a single economy are likely more consistent and comparable than data combined across countries.⁶

The Heckscher-Ohlin model has been the subject of extensive empirical investigation in international economics. A pair of papers describe regional outputs using this framework. Davis and Weinstein (1999) run regressions of regional outputs on regional endowments, employing the framework of Leamer (1984), but they abstract from the issue of labor mobility across regions. Bernstein and Weinstein (2002) consider the two-way links between endowments and outputs, concluding that if we know regions' outputs, we know with considerable precision the inputs used, but not vice versa. We move beyond these papers in two important respects. First, cities' factor supplies are a feature to be explained. Second, our model leads us to explore a dimension of the data not even contemplated in the prior work, namely that the endogenously determined size of the city is itself systematically related to skill and sectoral structure.

Our theory predicts systematic variation in sectoral composition in the form of an urban hierarchy of sectors. Prior systems-of-cities theories have overwhelmingly described sectoral composition in polarized terms (Abdel-Rahman and Anas, 2004).⁷ One class of models develops explicitly multi-sector economies in which each city specializes in a single traded good due to external economies of scale that are sector-specific.⁸ In a complementary class of models, cities produce a single composite output, which may be interpreted as perfect diversification

⁵ Sattinger (1993) and Costinot and Vogel (2015) survey the assignment literature. While it might seem that cities abundant in skilled labor must employ those factors in skill-intensive sectors, factor market clearing alone does not imply such a result. There is considerable distance between factor supplies and production outcomes in a high-dimensional environment, and mild assumptions tend to deliver weak results, as established by Jones and Scheinkman (1977) and Ethier (1984). Indeed, Costinot (2009) focuses precisely on establishing conditions sufficient to tightly link factor supplies to production outcomes.

⁶ In our spatial-equilibrium setting, we make assumptions such that the sectoral assignment function is common across cities, as explained in our discussions of Lemma 5 and Proposition 1.

⁷ An exception is central place theory, and our model relates to that theory's results in interesting ways. Our model's equilibrium exhibits a hierarchy of cities and sectors, as larger cities produce a superset of the goods produced in smaller cities. Models in central place theory, dating from Christaller (1933) through Hsu et al. (2014), have attributed this hierarchy property to the interaction of industry-specific scale economies and geographic market access based on the distance between firms located in distinct city centers. Our model yields the hierarchy property in the absence of both. Our theory links the hierarchy of sectors to a hierarchy of skills shaped by the internal geography of cities, neither of which has been considered in central place theory.

⁸ Though Henderson (1974) characterizes the case of single-industry specialized cities, he suggests that "cities will probably specialize in bundles of goods, where, within each bundle, the goods are closely linked in production."

³ Reconciling this fact with a two-skill model would require a departure from homotheticity such that more skilled people find large cities less attractive than the less skilled. This is at odds with results on both income-specific price indices (Handbury, 2012) and endogenous amenities (Diamond, 2016).

⁴ Both Baum-Snow et al. (2017) and Diamond (2016) use two skill groups, but they report contrasting results about cities' divergence in skills over time due to contrasting assignments of the "some college" population to "skilled" and "unskilled", respectively.

without specifying sectors' relative sizes. Neither complete specialization nor perfect diversification provides a propitious starting point for empirical investigations.

The small empirical literature describing variation in cities' sectoral composition is therefore only loosely informed by theory. For example, Holmes and Stevens (2004) survey the spatial distribution of economic activities in North America. They show that agriculture, mining, and manufacturing are disproportionately in smaller cities, while finance, insurance, real estate, professional, and management activities are disproportionately in larger cities. However, they do not reference a model or theoretical mechanism that predicts this pattern to be an equilibrium outcome. Similarly, Henderson (1983, 1991) reports regressions of employment shares on population sizes. These are motivated by theories of specialization linked to city size, but the regression specifications don't follow directly from those models. Our theory provides an explicit and distinct microfoundation for these regressions for an arbitrary number of sectors. Moreover, it predicts that the estimated population elasticities will be ordered by skill intensity.

A recent exception to the polarized view of sectoral specialization is Helsley and Strange (2014), who examine whether the equilibrium level of coagglomeration is efficient. Our papers have quite distinct objectives. While Helsley and Strange (2014) make minimal assumptions in order to demonstrate that Nash equilibria are generically inefficient when there are interindustry spillovers, we make strong assumptions that yield testable implications about the distribution of sectoral activity across cities. This account of cities' sectoral composition may inform a broader body of empirical work using sectoral composition as a source of identifying variation.⁹

In addition to sectoral composition, our theory describes sectoral size. Theories of localization and urbanization economies have contrasting predictions for cities' absolute employment levels.¹⁰ In the canonical model of pure localization in Henderson (1974), specialized cities of different sizes host different sectors, yielding "textile cities" and "steel cities". Sectoral specialization is the very basis for the city-size distribution, and one wouldn't expect large cities to be larger in all sectors. By contrast, urbanization models with a composite output make no prediction about spatial variation in sectoral composition. Our paper both introduces a multi-sector urbanization model in which larger cities are relatively larger in skill-intensive sectors and identifies conditions under which larger cities are absolutely larger in all sectors.

A recent empirical literature has demonstrated significant agglomeration and coagglomeration of industries relative to the null hypothesis of locations being (uniformly) randomly assigned in proportion to local population (Ellison and Glaeser, 1997; Duranton and Overman, 2005; Ellison et al., 2010). Our model's predictions are consistent with these findings. Since our theory says that sectors are ranked in terms of their relative employment levels, at most one sector could exhibit employment proportionate to total population. All other sectors will exhibit geographic concentration. Similarly, since sectors more similar in skill intensity will exhibit more similar relative employment levels, the cross-city distribution of sectoral employment will be consistent with skill-related coagglomeration. Prior studies have interpreted the agglomeration and coagglomeration of industries as evidence of within-industry and industry-pair-specific interactions or spillovers. In

our framework, significant measured agglomeration and coagglomeration will arise even absent these forces.

In sum, we know of no prior spatial-equilibrium theory that makes the predictions yielded by our model. Guided by our theoretical framework, our empirical investigation documents a cross-city pattern of skills-driven comparative advantage not revealed by prior empirical work.

3. Model

We develop a general-equilibrium model that makes predictions about the distributions of skills and sectors across cities of different sizes. The theory has few moving parts. Individuals vary in skill levels, and skills govern comparative advantage across sectors. Cities are identical ex ante, but agglomeration forces produce asymmetric cities of different sizes that differ in their equilibrium composition of skills and sectors of varying skill intensity.

Individuals freely choose their production sector, a city, and a location within that city. The sector that a person of a given skill type chooses to work in depends on goods prices but is independent of locational choices. The pattern of locational choices reflects the fact that a location can be more attractive if it is in a city with higher TFP or in a more desirable location within a given city. Facing this tradeoff, individuals are indifferent to the relative contributions of these two margins to a scalar index of attractiveness. In equilibrium, all locations of a given attractiveness are occupied by individuals of the same skill level, who are all employed in the same sector. There is thus an isomorphism between the distribution of locational attractiveness across cities and the distributions of skills and sectors across cities. Since equally attractive locations can be found in multiple cities, cities' skill and sectoral distributions exhibit overlap rather than strict sorting.

We develop two propositions. Proposition 1 has three key elements. The first identifies a regularity condition under which the distribution of locational attractiveness across cities is log supermodular. The second shows that this regularity condition implies that skills and sectoral employment are also log supermodular. The third says the same for sectoral outputs and revenues. In simple terms, Proposition 1 says that larger cities will be skill abundant and specialize in skill-intensive sectors. Proposition 2 addresses variation in the absolute (not only relative) size of skills and sectors across cities. It provides a (stronger) regularity condition sufficient for all skills and all sectors to be absolutely larger in larger cities. These two propositions provide the foundation for our empirical work in Section 5.

3.1. Preferences, production, and places

A measure L of heterogeneous individuals choose a city, a location within that city, and a sector in which to produce. There are discrete city sites, a continuum of skills, and a continuum of sectors. As in Davis and Dingel (2019), our model features symmetric fundamentals, but skill-biased agglomeration generates cities of heterogeneous sizes.

Individuals consume a freely traded final good. This final good is the numeraire and produced by combining a continuum of freely traded, labor-produced intermediate goods indexed by $\sigma \in \Sigma \equiv [\underline{\sigma}, \bar{\sigma}]$. These have prices $p(\sigma)$ that are independent of city c because trade costs are zero. Locations are characterized by their city c and their (inverse) innate desirability $\tau \in \mathcal{T} \equiv [0, \infty)$, so they have rental prices $r(c, \tau)$. In the main text, desirability stems from productivity benefits, while Appendix A.2 covers the case in which a location's desirability stems from its amenity value.¹¹

⁹ A number of papers in urban economics posit theoretical models with no sectoral specialization, in which all locations produce a homogeneous good, yet empirically estimate the model parameters exploiting cross-city variation in industrial structure (Diamond, 2016; Notowidigdo, 2013; Yagan, 2014). Our multi-sector model links sectoral composition to city size and skill composition. We hope this framework may help interpret such variation and accompanying identifying assumptions.

¹⁰ The literature traditionally distinguishes two types of external economies of scale (Henderson, 1987, p.929). Localization economies are within-industry, reflecting the scale of activity in that industry in that location. Urbanization economies are general, reflecting the scale of all economic activity in a location. Beyond scale, Lucas (1988) has stressed the composition of a location's human capital. The agglomeration process generating city-level productivities in our theory incorporates both scale and composition effects.

¹¹ These productivity and amenity interpretations of desirability differ slightly in functional form but yield the same predictions for the quantities that we examine empirically. For expositional clarity, we employ the productivity interpretation in the main text. As discussed further below, the canonical model of a monocentric city with commuting costs is a special case of our very general framework. In the amenity interpretation, higher-income individuals are more willing to pay for better amenities. Future work should seek to empirically identify which features within cities govern locational desirability.

Final-goods producers have a CES production function

$$Q = \left\{ \int_{\sigma \in \Sigma} B(\sigma) [Q(\sigma)]^{\frac{\varepsilon-1}{\varepsilon}} d\sigma \right\}^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1}$$

where $Q(\sigma) \geq 0$ is the quantity of intermediate good σ , $\varepsilon > 0$ is the elasticity of substitution between intermediates, and $B(\sigma) > 0$ is an exogenous technological parameter. The profits of final-goods producers are given by

$$\Pi = Q - \int_{\sigma \in \Sigma} p(\sigma) Q(\sigma) d\sigma. \tag{2}$$

Heterogeneous individuals use their labor to produce intermediate goods. There is a measure of $L > 0$ heterogeneous individuals with skills ω that have the cumulative distribution function $F(\omega)$ and density $f(\omega) > 0$ on support $\Omega \equiv [\underline{\omega}, \bar{\omega}]$. The productivity of an individual of skill ω in sector σ who chooses location τ in city c is

$$q(c, \tau, \sigma; \omega) = A(c)T(\tau)H(\omega, \sigma). \tag{3}$$

$A(c) \geq 0$ denotes city-level total factor productivity, which results from agglomeration and is taken as given by individuals. $T(\tau) \geq 0$ reflects the productivity effects of location within the city. We assume that $T(\tau)$ is continuously differentiable and decreasing, so that higher- τ locations are less desirable. We assume that the twice-differentiable function $H(\omega, \sigma) \geq 0$ is strictly log-supermodular in ω and σ and strictly increasing in ω .¹² The former governs comparative advantage, so that higher- ω individuals are relatively more productive in higher- σ sectors.¹³ The latter says that absolute advantage is indexed by ω , so that higher- ω individuals are more productive than lower- ω individuals in all sectors. Each individual inelastically supplies one unit of labor, so her income is her productivity times the price of the output produced, $q(c, \tau, \sigma; \omega)p(\sigma)$.

Locations within each city are heterogeneous, with the innate desirability of a location indexed by $\tau \geq 0$. The most desirable location is denoted $\tau = 0$, and higher values of τ denote less desirable places.¹⁴ The supply of locations with innate desirability of at least τ is $S(\tau)$. This is a strictly increasing function, since the supply of available locations increases as one lowers one's minimum standard of innate desirability. $S(0) = 0$, since there are no locations better than the ideal. We assume $S(\tau)$ is twice continuously differentiable. Locations are owned by absentee landlords who spend their rental income on the final good.¹⁵ The city has sufficient land capacity that everyone can reside in the city and the least desirable locations are unoccupied. We normalize the reservation value of unoccupied locations to zero, so $r(c, \tau) \geq 0$.

Individuals choose their city c , location τ , and sector σ to maximize utility. An individual's utility depends on their consumption

¹² In \mathbb{R}^2 , a function $H(\omega, \sigma)$ is strictly log-supermodular if $\omega > \omega', \sigma > \sigma' \Rightarrow H(\omega, \sigma)H(\omega', \sigma') > H(\omega, \sigma')H(\omega', \sigma)$.

¹³ We refer to higher- ω individuals as more skilled and higher- σ sectors as more skill-intensive.

¹⁴ For example, in the canonical monocentric-city model in which everyone works in the central business district and commuting to the center costs time, τ describes the residential location's physical distance from the center, $T(\tau)$ is labor time net of commuting, and the supply of locations is $S(\tau) = \pi r^2$. In our model, τ captures a wide variety of exogenous amenities that vary across locations within a city, but that provide a common distribution of the innate desirability of locations across cities. While the main text takes the supply of heterogeneous locations $S(\tau)$ as exogenously given, Appendix A.4 endogenizes it.

¹⁵ Models in the systems-of-cities literature typically treat the distribution of rents as a secondary issue, dispensed with in a variety of ways. Sometimes they allow for absentee landlords; sometimes they distribute the land rents on a per capita basis to local residents; sometimes it is assumed that all agents have a proportional share in an economy-wide mutual fund in land. See, for example, Henderson (1987), Helpman (1998), Rossi-Hansberg et al. (2010), or Diamond (2016). For simplicity we assume that rents accrue to landlords who live outside the cities of interest and spend it on the numeraire final good.

of the numeraire final good, which is their income after paying their locational cost:

$$U(c, \tau, \sigma; \omega) = A(c)T(\tau)H(\omega, \sigma)p(\sigma) - r(c, \tau). \tag{4}$$

Denote the endogenous quantity of individuals of skill ω residing in city c at location τ and working in sector σ by $L \times f(\omega, c, \tau, \sigma)$.

City-level TFP, $A(c)$, reflects agglomeration gains derived from both population size and composition. Note that city sites are *ex ante* identical: city-level TFPs differ in equilibrium due to individuals' locational choices. $A(c)$ is higher when a city contains a larger and more skilled population. Denote the endogenous quantity of individuals of skill ω residing in city c by $L \times f(\omega, c) \equiv L \times \int_{\sigma \in \Sigma} \int_{\tau \in T} f(\omega, c, \tau, \sigma) d\tau d\sigma$. Total factor productivity is

$$A(c) = J \left(L \int_{\omega \in \Omega} j(\omega) f(\omega, c) d\omega \right), \tag{5}$$

where $J(\cdot)$ is a positive, strictly increasing function and $j(\omega)$ is a positive, non-decreasing function. Numerous agglomeration processes may generate such productivity benefits, and we do not attempt to distinguish between them here.

3.2. Equilibrium

In a competitive equilibrium, individuals maximize utility, final-good producers and landowners maximize profits, and markets clear. Individuals maximize their utility by their choices of city, location, and sector such that

$$f(\omega, c, \tau, \sigma) > 0 \Leftrightarrow \{c, \tau, \sigma\} \in \operatorname{argmax} U(c, \tau, \sigma; \omega). \tag{6}$$

Profit maximization by final-good producers yields demands for intermediates

$$Q(\sigma) = I \left(\frac{p(\sigma)}{B(\sigma)} \right)^{-\varepsilon}, \tag{7}$$

where $I \equiv L \sum_{c \in \mathbb{C}} \int_{\omega \in \Omega} \int_{\tau \in T} q(\omega, c, \tau, \sigma) p(\sigma) f(\omega, c, \tau, \sigma) d\tau d\omega d\sigma$ denotes total income and these producers' profits are zero. Profit maximization by absentee landlords engaged in Bertrand competition causes unoccupied locations to have rental prices of zero,

$$r(c, \tau) \times \left(S'(\tau) - L \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, \tau, \sigma) d\omega d\sigma \right) = 0 \quad \forall c \quad \forall \tau. \tag{8}$$

Market clearing requires the endogenous quantity of individuals of skill ω residing in city c at location τ and working in sector σ , $L \times f(\omega, c, \tau, \sigma)$, to be such that the supply of a location type is greater than or equal to its demand, the demand and supply of intermediate goods are equal, and every individual is located somewhere. Denoting the set of city sites by \mathbb{C} :

$$S'(\tau) \geq L \int_{\omega \in \Omega} \int_{\sigma \in \Sigma} f(\omega, c, \tau, \sigma) d\sigma d\omega \quad \forall c \quad \forall \tau \tag{9}$$

$$Q(\sigma) = \sum_{c \in \mathbb{C}} Q(\sigma, c) = L \sum_{c \in \mathbb{C}} \int_{\omega \in \Omega} \int_{\tau \in T} q(c, \tau, \sigma; \omega) f(\omega, c, \tau, \sigma) d\omega d\tau \quad \forall \sigma \tag{10}$$

$$f(\omega) = \sum_{c \in \mathbb{C}} f(\omega, c) = \sum_{c \in \mathbb{C}} \int_{\sigma \in \Sigma} \int_{\tau \in T} f(\omega, c, \tau, \sigma) d\tau d\sigma \quad \forall \omega \tag{11}$$

A competitive equilibrium is a set of functions $Q : \Sigma \rightarrow \mathbb{R}^+$, $f : \Sigma \times \mathbb{C} \times T \times \Omega \rightarrow \mathbb{R}^+$, $r : \mathbb{C} \times T \rightarrow \mathbb{R}^+$, and $p : \Sigma \rightarrow \mathbb{R}^+$ such that conditions (6) through (11) hold.

3.3. A system of cities

Cities' populations are endogenously determined in spatial equilibrium. Two equilibrium arrangements are possible: either all cities are identical in their TFPs and population sizes, or cities are heterogeneous. The latter is the empirically relevant case. Our predictions about the distributions of skills and sectors across cities of different sizes apply to any equilibrium in which cities are heterogeneous.¹⁶

First, we solve for occupational assignments by exploiting the fact that locational and sectoral arguments are separable in individuals' utility functions. Given goods prices $p(\sigma)$, individuals' sectoral choices are independent of their locational choices:

$$\operatorname{argmax}_{\sigma} A(c)T(\tau)H(\omega, \sigma)p(\sigma) - r(c, \tau) = \operatorname{argmax}_{\sigma} H(\omega, \sigma)p(\sigma)$$

Define the assignment function $M(\omega) = \operatorname{argmax}_{\sigma} H(\omega, \sigma)p(\sigma)$ so that we can define the income associated with each skill's optimal occupational choice as $G(\omega) \equiv H(\omega, M(\omega))p(M(\omega))$. By comparative advantage, $M(\omega)$ is increasing.¹⁷ By absolute advantage, more skilled individuals earn higher nominal incomes and $G(\omega)$ is a strictly increasing function.¹⁸

Second, we solve for locational assignments by introducing a notion of locational attractiveness. Individuals value the product $A(c)T(\tau)$, which we dub attractiveness.

Definition 1. The attractiveness of a location in city c of (inverse) innate desirability τ is $\gamma = A(c)T(\tau)$.

Note that individuals are indifferent to the relative contributions of endogenous city-level TFP $A(c)$ and a location's innate desirability within the city $T(\tau)$ to attractiveness.¹⁹ In equilibrium, two locations of equal attractiveness must have the same price, so we can describe the rental price of a location of attractiveness γ as $r_{\Gamma}(\gamma)$. Thus, individuals' locational choices can be characterized in terms of attractiveness:

$$\max_{\gamma} \gamma G(\omega) - r_{\Gamma}(\gamma)$$

In equilibrium, more skilled individuals occupy more attractive locations. More attractive locations have higher rental prices. Since $G(\omega)$ is strictly increasing, locational attractiveness complements individual skill. Since more skilled individuals are more willing to pay for more attractive locations, equilibrium locational assignments feature positive matching between skill ω and attractiveness γ .

We have considered a location's attractiveness without regard to which city that location belongs. If the entire population lived in a single city, then desirability τ would be a sufficient statistic for attractiveness γ . In that case, equilibrium locational assignments and prices can be characterized as in standard land-use models (Fujita and Thisse, 2002, Ch 3), as we show in Appendix A.1. For a system of cities, we first

¹⁶ In fact, for a special case in which TFP $A(c)$ depends on population size alone with constant elasticity α and locational supply $V(z)$, defined below, exhibits a constant elasticity of $\frac{1-\alpha}{\alpha}$, it can be shown that any collection of city sizes summing to L is indeed an equilibrium of the model.

¹⁷ Lemma 1 of Costinot and Vogel (2010) shows that $M(\omega)$ is continuous and strictly increasing in equilibrium, given equation (7). It is worth noting here, as well, the role played by our assumption of zero trade costs. If trade between cities were costly, prices and sectoral assignments would be city-specific, hence $p(\sigma, c)$ and $M(\omega, c)$. We follow a vast factor-price-equalization literature in assuming no trade costs. To the extent that trade between cities is less costly than trade between countries, this is a weaker assumption than in the international trade literature, e.g. in Costinot and Vogel (2010). Intercity trade costs are one reason the data might reject our model's predictions making those predictions non-obvious.

¹⁸ Absolute advantage across all sectors is not necessary to insure income rises with skill. The weaker condition that productivity is increasing in skill at the equilibrium assignments, $\frac{d}{d\omega} H(\omega, M(\omega)) > 0$, would be sufficient.

¹⁹ This is why our model generates incomplete sorting across cities. Individuals of a given skill type will be indifferent between equally attractive locations that are in different cities.

characterize locational assignments and prices in terms of attractiveness γ using similar tools. We then translate these assignments and prices into functions of c and τ to characterize the system of cities.

Within each city, more desirable (low τ) locations are more attractive. Competition among landlords ensures that the most desirable locations are those occupied, so the least desirable occupied site $\bar{\tau}(c) \equiv \sup_{\tau} \{ \tau : f(\omega, c, \tau, \sigma) > 0 \}$ in a city of population $L(c)$ is defined by $L(c) = S(\bar{\tau}(c))$. Denote each city's set of occupied locations by $\bar{T}(c) \equiv [0, \bar{\tau}(c)]$. Less desirable locations have lower rental prices, and the least desirable occupied site has a rental price of zero.

Lemma 1 (Populated locations). In equilibrium, $S(\tau) = L \int_0^{\tau} \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, x, \sigma) d\omega d\sigma dx \quad \forall \tau \in \bar{T}(c)$, $r(c, \tau)$ is strictly decreasing in $\tau \quad \forall \tau < \bar{\tau}(c)$, and $r(c, \bar{\tau}(c)) = 0$.

These results suffice to allow us to make some cross-city comparisons prior to solving for locational assignments and prices.

In equilibrium, cities with higher TFP have larger populations. Consider two cities, c and c' , that differ in productivity, with $A(c) > A(c')$. If c were less populous than c' , then its least desirable occupied location would be more desirable, $\bar{\tau}(c) \leq \bar{\tau}(c')$, since the supply of locations, $S(\tau)$, is increasing (and common across cities). Since $A(c) > A(c')$, this would make the least attractive occupied location in city c more attractive than the least attractive occupied location in city c' , $A(c)T(\bar{\tau}(c)) > A(c')T(\bar{\tau}(c'))$. By lemma 1, each of these locations has a rental price of zero. Every individual would strictly prefer the more attractive location at the same price, so no one would choose to live in c' at $\bar{\tau}(c')$, contradicting the definition of $\bar{\tau}(c')$ as an occupied location. So the city with higher TFP must have a larger population.

A smaller city's locations are a subset of those in a larger city in terms of attractiveness. For every location in the less populous city, there is a location in the more populous city that is equally attractive. The location in city c' of innate desirability τ' is equivalent to a location τ in city c , given by $A(c)T(\tau) = A(c')T(\tau')$. The equally attractive location in the larger city has higher TFP but lower innate desirability. The more populous city also has locations that are strictly more attractive than the best location in the less populous city; there are locations of attractiveness $\gamma \in (A(c')T(0), A(c)T(0))$ found in c and not in c' .

Across and within cities, more skilled individuals occupy more attractive locations. Without loss of generality, order and label positive-population cities from 1 to C so that $A(C) \geq A(C-1) \geq \dots \geq A(2) \geq A(1)$. Denote the set of attractiveness levels occupied in equilibrium by $\Gamma \equiv [\underline{\gamma}, \bar{\gamma}]$, where $\underline{\gamma} \equiv A(C)T(\bar{\tau}(C))$ and $\bar{\gamma} \equiv A(C)T(0)$. Individuals of higher skill have greater willingness to pay for more attractive locations, so in equilibrium higher- ω individuals occupy higher- γ locations.

Lemma 2 (Locational assignments). In equilibrium, there exists a continuous and strictly increasing locational assignment function $K : \Gamma \rightarrow \Omega$ such that (i) $f(\omega, c, \tau, M(\omega)) > 0 \Leftrightarrow A(c)T(\tau) = \gamma$ and $K(\gamma) = \omega$, and (ii) $K(\underline{\gamma}) = \underline{\omega}$ and $K(\bar{\gamma}) = \bar{\omega}$.

To obtain an explicit expression for $K : \Gamma \rightarrow \Omega$, we can denote the supply of locations across all cities combined with attractiveness γ or greater as $S_{\Gamma}(\gamma)$. The supply function is

$$S_{\Gamma}(\gamma) = \sum_{c: A(c)T(0) \geq \gamma} S\left(T^{-1}\left(\frac{\gamma}{A(c)}\right)\right).$$

By definition $S_{\Gamma}(\bar{\gamma}) = 0$ and by the fact that the best locations are populated $S_{\Gamma}(\underline{\gamma}) = L$. Lemmas 1 and 2 imply that $S_{\Gamma}(\gamma) = L \int_{\gamma}^{\bar{\gamma}} f(K(x))$

$$K'(x) dx, \text{ so } K(\gamma) = F^{-1}\left(\frac{L - S_{\Gamma}(\gamma)}{L}\right).$$

These locational assignments yield an expression for equilibrium locational prices, which are increasing in attractiveness and given by the upper envelope of individuals' bid-rent schedules.

Lemma 3 (Locational prices). *In equilibrium, the rent schedule $r_r(\gamma)$ is increasing and continuously differentiable on $[\underline{\gamma}, \bar{\gamma}]$ and given by $r_r(\gamma) = \int_{\underline{\gamma}}^{\gamma} G(K(x)) dx$.*

This determination of locational assignments and prices within the system of cities in terms of attractiveness is analogous to determining these locational assignments and prices for a single autarkic city with a supply of locations that is the sum of locations across the system of cities. Next, we translate these assignments and prices stated in terms of attractiveness to locations within each city.

3.4. The distributions of skills and sectors across cities

We can now characterize the distributions of rents, skills, and sectoral employment in a system of cities. We first show how the distribution of locations across cities governs the distributions of skills and sectoral employment across cities through the locational and sectoral assignment functions. We then identify a sufficient condition under which these distributions are log-supermodular. Finally, we identify conditions under which larger cities will have larger populations of all skill types and employ more people in all sectors.

Since attractiveness is the product of city-level TFP and innate desirability, a city's supply of locations of a given attractiveness depends on both its endogenous TFP and the exogenous supply of locations of particular desirability. The supply of locations with attractiveness γ in city c is

$$s(\gamma, c) \equiv \frac{\partial}{\partial \gamma} \left[S(\bar{\tau}(c)) - S\left(T^{-1}\left(\frac{\gamma}{A(c)}\right)\right) \right] \text{ if } \gamma \leq A(c)T(0) \\ = \begin{cases} \frac{1}{A(c)} V\left(\frac{\gamma}{A(c)}\right) & \text{if } \gamma \leq A(c)T(0) \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

where $V(z) \equiv -\frac{\partial}{\partial z} S(T^{-1}(z))$ is the supply of locations with innate desirability $T^{-1}(z)$. $V(z)$ is a composite fundamental that depends only on the exogenous functions $T(\tau)$ and $S(\tau)$. For example, if each city were a disc, $S(\tau) = \pi\tau^2$, and desirability reflected linear costs of commuting to the center, $T(\tau) = d_1 - d_2\tau$, this supply of locations within cities would be $V(z) = \frac{2\pi}{d_2^2}(d_1 - z)$. The rental price of a location depends only on its attractiveness, so the rental price of a location with innate desirability τ in city c is $r(c, \tau) = r_r(A(c)T(\tau))$.

The distribution of skills follows from $s(\gamma, c)$ and locational assignments $K : \Gamma \rightarrow \Omega$.

Lemma 4 (A city's skill distribution). *The population of individuals of skill ω in city c is*

$$L \times f(\omega, c) = \begin{cases} K^{-1}(\omega) s(K^{-1}(\omega), c) & \text{if } A(c)T(0) \geq K^{-1}(\omega) \\ 0 & \text{otherwise} \end{cases}$$

Thus, the relative population of individuals of skill ω in two cities depends on the relative supply of locations of attractiveness $K^{-1}(\omega)$. Since higher- ω individuals occupy more attractive locations and the most attractive locations are found exclusively in the larger city, there is an interval of high- ω individuals who reside exclusively in the larger city. Individuals of abilities below this interval are found in both cities, and their relative quantity $\frac{f(\omega, c)}{f(\omega, c')}$ is proportionate to the relative supply of locations of attractiveness $K^{-1}(\omega)$.

The distribution of sectoral employment follows from $s(\gamma, c)$, locational assignments $K : \Gamma \rightarrow \Omega$, and sectoral assignments $M : \Omega \rightarrow \Sigma$. As

established in subsection 3.3, the sectoral assignment function $M(\omega)$ is common across c and τ because locations' productivity advantages are Hicks-neutral. Thus, the employment distribution $f(\sigma, c)$ closely follows the skill distribution $f(\omega, c)$.

Lemma 5 (A city's sectoral employment distribution). *The population of individuals employed in sector σ in city c is*

$$L \times f(\sigma, c) = \begin{cases} M^{-1}(\sigma) K^{-1}(M^{-1}(\sigma)) s(K^{-1}(M^{-1}(\sigma)), c) & \text{if } A(c)T(0) \geq K^{-1}(M^{-1}(\sigma)) \\ 0 & \text{otherwise} \end{cases}$$

As a result, two cities' relative employment levels in sector σ depend on their relative supplies of locations with attractiveness $K^{-1}(M^{-1}(\sigma))$.

We now identify the condition under which the distributions of rents, skills, and sectoral employment across cities are log-supermodular functions. When the distribution of locational attractiveness is log-supermodular, so are the distributions of skills and sectoral employment. The first result follows from more skilled individuals occupying more attractive locations in equilibrium. The second result follows from the fact that sectoral assignments are common across locations, so that sectoral composition is governed by skill composition.

Since the distribution of locations in terms of innate desirability τ is common across cities, cross-city differences in the distributions of locational attractiveness γ reflect differences in cities' TFPs. Eq. (12) demonstrates a hierarchy of locational attractiveness, since the most attractive locations are found exclusively in the highest-TFP city. Among levels of attractiveness that are supplied in multiple cities, eq. (12) shows that cities' TFPs shape the supply schedule $s(\gamma, c)$ through both a scaling effect $\left(\frac{1}{A(c)}\right)$ and a dilation of $V\left(\frac{\gamma}{A(c)}\right)$. Comparisons of relative supplies $(s(\gamma, c) s(\gamma', c')) \stackrel{?}{\geq} s(\gamma', c) s(\gamma, c')$ depend only on the dilation.

Our main result, Proposition 1, states a sufficient condition for the ordering of city TFPs to govern the ordering of locational supplies in any equilibrium. In turn, these govern the distributions of skills and sectoral employment across cities in any equilibrium.

Proposition 1 (Cross-city distributions of attractiveness, skills, and sectors).

- (a) *The supply of locations of attractiveness γ in city c , $s(\gamma, c)$, is log-supermodular if the supply of locations with innate desirability $T^{-1}(z)$ within each city, given by $V(z)$, has a decreasing elasticity.*
- (b) *If $V(z)$ has a decreasing elasticity, then $f(\omega, c)$ and $f(\sigma, c)$ are log-supermodular.*
- (c) *If $V(z)$ has a decreasing elasticity, then sectoral output $Q(\sigma, c)$ and revenue $R(\sigma, c) \equiv p(\sigma)Q(\sigma, c)$ are log-supermodular.*

Proposition 1a links our assumption about each city's exogenous distribution of locations, $V(z)$, to endogenous equilibrium locational characteristics, $s(\gamma, c)$. The proof is in Appendix B.²⁰ Heuristically, note that a higher-TFP city is relatively abundant in more attractive locations when the elasticity $\frac{\partial \ln s(\gamma, c)}{\partial \ln \gamma}$ is larger in the higher-TFP city. Eq. (12)

implies that the γ -elasticity of $s(\gamma, c)$ is the elasticity of $V(z)$ at $z = \frac{\gamma}{A(c)}$.

When this elasticity is higher at lower values of z , an ordering of cities' TFPs (and thus cities' sizes) is an ordering of these elasticities, and thus an ordering of relative supplies at equilibrium. A number of conceivable $V(z)$ schedules satisfy this decreasing-elasticity condition.²¹ For example, in the canonical monocentric-city model mentioned

²⁰ Lemma 8 in Appendix B also shows that the sufficient condition is necessary for $s(\gamma, c)$ to be log-supermodular for all possible values of $A(c)$.

²¹ For example, if each city's supply of locations with innate desirability $T^{-1}(z)$ is the exponential, Weibull, gamma, or log-normal distribution, $V(z)$ exhibits this decreasing-elasticity property. See Appendix A.3 for details.

earlier, $V(z) = \frac{2\pi}{d_2^2}(d_1 - z)$ has an elasticity of $-\frac{z}{d_1 - z}$, which is decreasing in z .

An illustrative boundary case is when $V(z)$ is constant.²² Per eq. (12), there is a hierarchy of locational attractiveness in which the most attractive locations are found only in the largest cities, independent of $V(z)$. For attractiveness levels that are attained in both cities, there is no difference in relative supplies when $V(z)$ is constant, because $\frac{s(\gamma, c)}{s(\gamma', c)}$ is a constant when both $s(\gamma, c)$ and $s(\gamma', c)$ are non-zero. Thus, $s(\gamma, c)$ is weakly log-supermodular everywhere and strictly log-supermodular for the highest values of γ . For the more general result, the decreasing-elasticity condition in Proposition 1 ensures that intensive-margin variation in relative supplies aligns with the extensive margin of locational attractiveness.

The distributions of skills and sectoral employment across cities in Proposition 1b follow the distribution of locational attractiveness. The skill distribution follows immediately through the locational assignment function (lemma 4) and the employment distribution follows in turn through the sectoral assignment function (lemma 5). Note that these lemmas show that the most skilled individuals working in the most skill-intensive sectors are found only in the most populous cities. The skills and sectors found in a smaller city are a strict subset of those found in a larger city. This hierarchy of skills across cities at the top of the skill distribution does not depend on the shape of $V(z)$. Since $K: \Gamma \rightarrow \Omega$ and $M: \Omega \rightarrow \Sigma$ are strictly increasing functions, $f(\omega, c)$ and $f(\sigma, c)$ are log-supermodular if and only if $s(\gamma, c)$ is log-supermodular. Thus, Proposition 1 links the log-supermodularity of these distributions over the full range of skills and sectors to the properties of $V(z)$. By eq. (5), a city that is larger and more skilled has higher TFP. In Section 3.3, we established that cities with higher TFP have larger populations. Proposition 1b completes the circle by showing that if one city has higher TFP than another, then its population is more skilled. TFP differences driven by population size and skill composition are self-sustaining.

Given our assumptions about technologies and spatial equilibrium, the fact that larger cities are skill-abundant and more skilled individuals work in more skill-intensive sectors implies that larger cities produce relatively more in skill-intensive sectors. This pattern of specialization is closely related to the high-dimensional model of endowment-driven comparative advantage introduced by Costinot (2009), but in our setting cities' populations and skill composition are endogenously determined and there is within-city heterogeneity in productivity.²³ The distribution of output follows the distribution of skills because in spatial equilibrium individuals employed in the same sector occupy locations of the same productivity.²⁴ Thus, a city's sectoral output is proportionate to its sectoral employment. Since that output is freely traded, a city's sectoral revenue is also

proportionate to its sectoral employment. We therefore obtain part c of Proposition 1, which completes our characterization of the pattern of comparative advantage across cities.²⁵

When does the more productive city have a larger population of every skill type? By lemma 4, whenever it has a larger supply of every attractiveness level, $s(\gamma, c) \geq s(\gamma, c') \forall \gamma$. This is trivially true for $\gamma > A(c')T(0)$. What about attractiveness levels found in both cities? Proposition 2 identifies a sufficient condition under which a larger city has a larger supply of locations of a given attractiveness. Its proof appears in Appendix B. Applying this result to the least-attractive locations yields a sufficient condition for larger cities to have larger populations of all skill types and therefore employ more people in every sector.

Proposition 2 (City size and absolute size of local skills and sectors).

For any $A(c) > A(c')$, if $V(z)$ has a decreasing elasticity that is less than -1 at $z = \frac{\gamma}{A(c)}$, then $s(\gamma, c) \geq s(\gamma, c')$. If $V(z)$ has a decreasing elasticity that is less

than -1 at $z = \frac{K^{-1}(\omega)}{A(c)} = \frac{\gamma}{A(c)}$, then $A(c) > A(c')$ implies $f(\omega, c) \geq f(\omega, c')$ and $f(M(\omega), c) \geq f(M(\omega), c') \forall \omega \in \Omega$.

Our two propositions characterize the distribution of skills and sectors across cities. If $V(z)$ has a decreasing elasticity, then larger cities are more skill-abundant and specialize in skill-intensive activities. If the elasticity is sufficiently negative, then larger cities are larger in terms of all skill and sectors. We now turn to the data to see how well these predictions describe US metropolitan areas.

4. Data description and empirical approach

Our model describes distributions within and across cities. While we have employed an abstract idea of within-city heterogeneity in locational desirability, Proposition 1 makes concrete predictions about the distributions of skills and sectors across cities.²⁶ We examine the predictions of part b of Proposition 1 using two approaches. The first involves regression estimates of the population elasticities of educational, occupational, and industrial populations. The second involves pairwise comparisons governed by the monotone likelihood ratio property.

These tests require data on cities' skill distributions, sectors' skill intensities, and cities' sectoral employment. We use public-use microdata from the US Census of Population to identify the first two. The latter is described by data from County Business Patterns and Occupational Employment Statistics. The Census of Population describes individuals' educational attainments, geographic locations, places of birth, occupations, and industries. County Business Patterns describes cities' industrial employment. Occupational Employment Statistics describes cities' occupational employment. We combine these various data at the level of (consolidated) metropolitan statistical areas (MSAs); see Appendix C for details.

²² For example, a linear city $S(\tau) = 2\tau$ with linear commuting costs $T(\tau) = d_1 - d_2\tau$ makes $V(z)$ a constant.

²³ In particular, our Proposition 1c result that $Q(\sigma, c)$ is log-supermodular is similar to Theorem 2 in Costinot (2009), but the economic environment and relevant assumptions differ. Assumption 2 in Costinot (2009)'s factor-endowment model is that countries' exogenous endowments are such that countries can be ranked according to the monotone likelihood ratio property. Our Proposition 1 identifies a sufficient condition for cities' equilibrium skill distributions to exhibit this property. Definition 4 of Costinot (2009) requires that factor productivity vary across countries (cities) in a Hicks-neutral fashion. Since productivity $A(c)T(\tau)$ varies both across and within cities, our production function $q(c, \tau, \sigma; \omega)$ does not satisfy this requirement for arbitrary locational assignments.

²⁴ In the productivity interpretation of $T(\tau)$, equilibrium productivity $q(c, \tau, \sigma; \omega) = K^{-1}(\omega)H(\omega, \sigma)$ does not vary across ω -occupied locations. In the amenity interpretation of $T(\tau)$ described in Appendix A.2, occupied locations' productivities $q(c, \tau, \sigma; \omega) = A(c)H(\omega, \sigma)$ differ across cities in a Hicks-neutral fashion.

²⁵ A traditional definition of comparative advantage refers to locations' autarkic prices. In our setting, autarky means prohibiting both trade of intermediate goods and migration between cities. Since individuals are spatially mobile, cities do not have "factor endowments", and we must specify the autarkic skill distributions. If we consider an autarkic equilibrium with the skill distributions from the system-of-cities equilibrium, then larger cities have lower relative autarkic prices for higher- σ goods because they are skill-abundant, as shown by Costinot and Vogel (2010, p. 782).

²⁶ Proposition 1 also makes predictions about other economic outcomes, such as sectoral outputs (part c) and rental prices (implied by part a). We examine occupational employment levels, which are readily available, since occupational output is not typically observed. The available evidence on urban costs, which shows that the maximum, mean, and range of unimproved land prices are greater in larger cities (Combes et al., 2012), is all consistent with our model's predictions. We are not aware of a representative sample of (unimproved) land prices to examine stochastic or likelihood-ratio dominance in rental prices across cities. Investigating the hedonic determinants of locational desirability, which our theoretical model has abstractly treated as potentially production benefits or consumption amenities, is a first-order question about the internal structure of cities that is quite beyond the scope of this paper.

Subsections 4.1 through 4.3 describe the observable measures of skills, sectors, and skill intensities that we employ in the two empirical approaches defined in subsection 4.4.

4.1. Skills

Following a large literature, we use observed educational attainment as a proxy for individuals' skills. Educational attainment is a coarse measure, but it is the best measure available in data describing many people across detailed geographic locations.²⁷ We do not assume that individuals with the same educational attainment are equally skilled. We map the continuum of skills in our theory to the discrete set of educational levels observed in the data by assuming that the distribution of skills is increasing with educational attainment, such that the distribution of educational attainment across cities is log-supermodular if $f(\omega, c)$ is log-supermodular.²⁸ To describe cities' skill distributions, we aggregate individual-level microdata to the level of metropolitan statistical areas. A large literature in urban economics describes variation in terms of two skill groups, typically college and non-college workers. Following Acemoglu and Autor (2011), we use a minimum of three skill groups. The 2000 Census of Population microdata identify 16 levels of educational attainment, from "no schooling completed" to "doctoral degree". We define three skill groups of approximately equal size among the working population: high-school degree or less; some college or associate's degree; and bachelor's degree or more. In a more ambitious approach, we also consider nine skill groups, ranging from individuals who never reached high school (4% of the population) to those with doctoral degrees (1%). Table 1 shows the population shares and percentage US-born for each of these skill groups in 2000.²⁹ Foreign-born individuals are disproportionately in the tails of the educational distribution.

4.2. Sectors

In our model, workers produce freely traded sectoral outputs indexed by σ that are used to produce the final good. In the international trade literature, it is common to interpret sectors in models of comparative advantage as industries. Recent work in both international and labor economics has emphasized a perspective focused on workers completing tasks, which empirical work has frequently operationalized as occupations (Grossman and Rossi-Hansberg, 2008; Acemoglu and Autor, 2011). We will implement empirical tests using each. We define sectors to be the 19 private-sector industries in the two-digit stratum of the North American Industry Classification System (NAICS) or the 22 occupational categories in the two-digit stratum of the Standard Occupational Classification (SOC).³⁰

We measure industrial employment in a metropolitan area using data from the 2000 County Business Patterns, which counts paid employees in almost all non-farm, non-government establishments. We measure occupational employment in a metropolitan area using estimates from the 2000 BLS Occupational Employment Statistics, which cover full-time and part-time employees in all non-farm establishments. See Appendix C for details.

²⁷ In Appendix E.4, we infer skills from nominal wages and obtain similar results to those using educational attainment.

²⁸ Costinot and Vogel (2010, 774) show that log-supermodularity of factor supplies in an observed characteristic and unobserved skill ω is sufficient for mapping a theory with a continuum of skills to data with discrete observed characteristics.

²⁹ This table describes labor-force participants 25 and older. See Appendix E for similar results using other inclusion criteria.

³⁰ Per Costinot and Vogel (2010, 773-774), for mapping a continuum of sectors to coarse categories in the data, it is sufficient that more skill-intensive tasks are relatively more prevalent in sectors employing more educated workers. Our industry results are not driven by our choice of NAICS aggregation: we find similar empirical patterns when using three-digit industry definitions rather than two-digit definitions.

4.3. Skill intensities

Our theory makes the strong assumption that $H(\omega, \sigma)$ is strictly log-supermodular so that sectors are ordered with respect to their skill intensities. In our empirical work, we infer sectors' skill intensities from the data using the observable characteristics of the workers employed in them. We use microdata from the 2000 Census of Population, which contains information about workers' educational attainments, industries, and occupations. We use the average years of schooling of workers employed in a sector as a measure of its skill intensity.³¹ In doing so, we control for spatial differences by regressing years of schooling on both sectoral and city fixed effects, but we have found that omitting the city fixed effects has little effect on the estimated skill intensities. Table 2 reports the five least skill-intensive and five most skill-intensive sectors among both the 19 industrial categories and the 22 occupational categories. There is considerably greater variation in average years of schooling across occupational categories than across industries.³²

4.4. Empirical tests

Proposition 1b says that the distribution of skills across cities, $f(\omega, c)$, and the distribution of sectoral employment across cities, $f(\sigma, c)$, are log-supermodular functions. Log-supermodularity has many implications; we focus on two that are amenable to empirical testing. If the function $f(\nu, c)$ is log-supermodular, then

1. a linear regression $\ln f(\nu, c) = \alpha_\nu + \beta_\nu \ln L(c) + \varepsilon_{\nu, c}$ in which α_ν are fixed effects and $L(c)$ is city population yields $\beta_\nu \geq \beta_{\nu'} \Leftrightarrow \nu \geq \nu'$;
2. if C and C' are distinct sets and C is greater than C' ($\inf_{c \in C} L(c) > \sup_{c \in C'} L(c')$) and n_c ($n_{c'}$) is the number of elements in C (C'),

$$\frac{1}{n_C} \sum_{c \in C} \ln f(\nu, c) + \frac{1}{n_{C'}} \sum_{c \in C'} \ln f(\nu', c') \geq \frac{1}{n_C} \sum_{c \in C} \ln f(\nu', c) + \frac{1}{n_{C'}} \sum_{c \in C'} \ln f(\nu, c')$$

The first implication, which we will refer to as the "elasticity test," says that the city-population elasticity of the population of a skill type in a city $f(\omega, c)$ is increasing in skill ω .³³ Similarly, the population elasticity of sectoral employment $f(\sigma, c)$ is increasing in skill intensity σ . Our theory thus provides a structure to interpret previous work describing the population elasticities of sectoral employment, such as Henderson (1983) and Holmes and Stevens (2004).³⁴ Standard econometric tests are available to assess whether our estimated population elasticities exhibit the property that $\nu \geq \nu' \Rightarrow \beta_\nu \geq \beta_{\nu'}$.

³¹ Autor and Dorn (2013) rank occupations by their skill level according to their mean log wage. Our assumption of absolute advantage is consistent with such an approach. Using average log wages as our measure of skill intensity yields empirical success rates comparable to and slightly higher on average than those reported in Section 5. We use years of schooling rather than wages as our measure of sectoral skill intensities since nominal wages may also reflect compensating differentials or local amenities.

³² The standard deviations of average years of schooling across occupational and industrial categories are 2.3 and 1.2, respectively.

³³ The linear regression may understood as a first-order Taylor approximation:

$$\ln f(\nu, c) \approx \ln f(\nu, c^*) + \frac{\partial \ln f(\nu, c)}{\partial \ln L(c)} \Big|_{c=c^*} (\ln L(c) - \ln L(c^*)) + \varepsilon = \alpha_\nu + \beta_\nu \ln L(c) + \varepsilon_{\nu, c},$$

where $\beta_\nu = \frac{\partial \ln f(\nu, c)}{\partial \ln L(c)} \Big|_{c=c^*}$ is increasing in ν by log-supermodularity of $f(\nu, c)$.

³⁴ Henderson (1983) regresses employment shares on population levels, but reports "percent Δ share / percent Δ population", which is equal to $\beta_\nu - 1$ in our notation. Similarly, [46] describe how location quotients, a city's share of industry employment divided by its share of total employment, vary with city size. In our notation, a location quotient is $LQ(\sigma, c) = \frac{f(\sigma, c) / \sum_{c'} f(\sigma, c')}{L(c)/L}$, so the $L(c)$ -elasticity of $LQ(\sigma, c)$ is $\beta_\nu - 1$.

Table 1
Skill groups by educational attainment.

Skill (3 groups)	Population share	Share US-born	Skill (9 groups)	Population share	Share US-born
High school or less	0.37	0.78	Less than high school	0.04	0.29
			High school dropout	0.08	0.73
			High school graduate	0.25	0.88
Some college	0.31	0.89	College dropout	0.23	0.89
			Associate's degree	0.08	0.87
			Bachelor's degree	0.20	0.86
Bachelor's or more	0.32	0.85	Master's degree	0.08	0.84
			Professional degree	0.03	0.81
			Doctorate	0.01	0.72

NOTES: Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas. Data source: 2000 Census of Population microdata via IPUMS-USA.

Table 2
Sectoral skill intensities.

SOC	Occupational category	Skill intensity	NAICS	Industry	Skill intensity
45	Farming, fishing & forestry	9.2	11	Forestry, fishing, hunting & agriculture support	10.9
37	Cleaning & maintenance	10.9	72	Accommodation & food services	11.8
35	Food preparation & serving related	11.5	23	Construction	11.9
47	Construction & extraction	11.5	56	Admin, support, waste mgt, remediation	12.2
51	Production	11.5	48	Transportation & warehousing	12.6
29	Healthcare practitioners & technical	15.6	52	Finance & insurance	14.1
21	Community & social services	15.8	51	Information	14.1
25	Education, training & library	16.3	55	Management of companies & enterprises	14.5
19	Life, physical & social science	17.1	54	Professional, scientific & technical services	15.3
23	Legal occupations	17.3	61	Educational services	15.6

NOTES: Skill intensity is average years of schooling for individuals 25 and older after removing metropolitan-area fixed effects. Data source: 2000 Census of Population microdata via IPUMS-USA.

The second implication, which we will refer to as the “pairwise comparisons test”, says that if cities are divided into bins ordered by population sizes, then in any pairwise comparison of two bins and two skills/sectors, the bin containing more populous cities will have relatively more of the more skilled type. Our theory therefore implies thousands of pairwise comparisons for skills and millions for sectors. Appendix D shows that, in the presence of additive random errors to $\ln f(\nu, c)$, the likelihood of a successful pairwise comparison increases with the difference in population size, the difference in skill (intensity), and the number of cities assigned to each bin. To summarize this test, we report the fraction of pairwise inequalities matching the predicted sign, weighted by the product of the two cities' difference in log population and two sectors' difference in skill intensity.³⁵ When comparing narrow educational categories, we also weight by the product of the educational categories' population shares.³⁶ To assess the statistical significance of the fraction of these pairwise comparisons that yield the expected inequality, we compute the probability of obtaining the observed success rate under the null hypothesis that skills and sectors are uniformly distributed across cities.³⁷

These two empirical tests are not independent, since they are both implied by log-supermodularity. Appendix D describes how they are

³⁵ We report unweighted success rates in Appendix E. An unweighted statistic assigns equal weight to correctly predicting that Chicago (population 9 million) is relatively more skilled than Des Moines (456 thousand) and correctly predicting that Des Moines is relatively more skilled than Kalamazoo (453 thousand). Given the numerous idiosyncratic features of the real world omitted from our parsimonious theory, the former comparison is much more informative about the relevance of our theory than the latter. Stated differently, a failure to correctly order Chicago and Des Moines should be much more damning for our theory than a failure to correctly order Des Moines and Kalamazoo. Weighting accomplishes this.

³⁶ An unweighted average, reported in Appendix E, treats comparisons involving high school graduates (25% of the workforce) and those involving PhDs (1%) equally, while these differ in their economic import.

³⁷ Appendix D provides details of a permutation test in which we shuffle our observations 1000 times to compute the distribution of success rates under the null. Power to reject the null decreases as the number of bins decreases.

related. In short, success of one test implies success of the other, to the extent that the first-order approximations of $\ln f(\nu, c)$ fit the data well. We also implement a test for systematic deviations from log-supermodularity proposed by [Sattinger \(1978\)](#).³⁸

5. Empirical results

In this section, we test our predictions relating cities' sizes to their distributions of skill, occupational employment, and industrial employment. First, we examine whether populations are log-supermodular in educational attainment and city size. This prediction is a much stronger characterization of cities' skill distributions than the well-known fact that larger cities typically have a greater share of college graduates. Second, we examine whether the pattern of sectoral employment is strongly ordered by this pattern of skills. Our theory's predictions are more realistic than those which say cities will be completely specialized in an industry or produce in fixed proportions across all industries, as implicit in models with a single output. They are also more specific than theories allowing arbitrary patterns of interindustry spillovers. Finally, we examine whether larger cities are larger in all skill and sectoral categories or whether, as might be suggested by theories of industrial localization, different skills and sectors attain their maximal presence at different points in the city-size distribution.

The data are broadly consistent with our predictions. Larger cities are more skilled than smaller cities when comparing narrowly defined educational categories, although in the lowest educational category, international migrants play an important role that is omitted from our model. More skill-intensive sectors are relatively larger in more populous cities: sectoral population elasticities rise with skill intensity, and pairwise comparisons yield statistically significant results in the direction predicted by our theory. We show that there are not systematic violations of our

³⁸ These cross-sectional results should be thought of as examining a long-run equilibrium. Such a focus is valuable as well if one wants to think about comparative statics for long-run outcomes. Of course, if one wants to think explicitly about transitional dynamics, one would need to augment our model with adjustment frictions appropriate to the shock contemplated.

Table 3
Population elasticities of three skill groups.

Dependent variable: $\ln f(\omega, c)$	(1)	(2)	Population	Share
	All	US-born	share	US-born
$\beta_{\omega 1}$ High school or less \times log population	0.954 (0.0108)	0.895 (0.0153)	0.37	0.78
$\beta_{\omega 2}$ Some college \times log population	0.996 (0.0105)	0.969 (0.0122)	0.31	0.89
$\beta_{\omega 3}$ Bachelor's or more \times log population	1.086 (0.0153)	1.057 (0.0162)	0.32	0.85

NOTES: Standard errors, clustered by MSA, in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas.

predicted pattern of comparative advantage. Consistent with our model, there is a strong tendency for larger cities to be larger in all skills and sectors.

5.1. Larger cities are relatively more skilled

This subsection tests our prediction that larger cities have relatively more skilled populations. We empirically describe skill abundance using the two tests described in subsection 4.4. We first do these exercises using three skill groups defined by educational attainment levels and then repeat them using nine very disaggregated skill groups. In Appendix E.4, we infer skills from nominal wages and obtain similar results.

5.1.1. Three skill groups

The elasticity test applied to the three skill groups across 270 metropolitan areas is reported in Table 3. The results match our theory's prediction that larger cities will have relatively more people from higher skill groups. The population elasticities are monotonically increasing in educational attainment and the elasticities differ significantly from each other.³⁹ In anticipation of issues related to international immigration that arise when we examine nine skill groups, the second column of the table reports the population elasticities of US-born individuals for these three educational categories. The estimated elasticities are slightly lower, since foreign-born individuals are more concentrated in larger cities, but the differences between the elasticities are very similar. Larger cities' populations are more skilled, on average.

The pairwise comparison test examines ordered groups of cities to see if the relative population of the more skilled is greater in larger cities. Following subsection 4.4, implementing this test involves defining bins of cities. Ordering cities by population, we partition the 270 metropolitan areas in our data into 2, 3, 5, 10, 30, 90, and 270 bins of cities. Making pairwise comparisons between three skill groups and as many as 270 metropolitan areas involves computing up to 108,945 inequalities.⁴⁰ Note that prior work typically describes a contrast between large and small cities for skilled and unskilled, whereas our *most aggregated* comparison is between large and small cities for three skill groups.

Table 4 reports the success rates for these pairwise comparisons for both the population as a whole and those individuals born in the United States. When making pairwise comparisons across all 270 metropolitan areas, the success rates are 67 and 69%, respectively, and highly statistically significant. As we decrease the number of bins (increase the number of cities per bin), the success rates increase, consistent with binning reducing the influence of idiosyncratic errors. When using five bins or fewer, these success rates exceed 97%.⁴¹ These results are all statistically

³⁹ Younger cohorts have higher average educational attainment. The results in Tables 3 and 5 are robust to estimating the elasticities for educational groups within 10-year age cohorts. Thus, our results are not due to the young being both more educated and more likely to live in large cities.

⁴⁰ With n city bins and m skill groups, we make $\frac{n(n-1)m(m-1)}{2}$ comparisons. For example, $\frac{270 \times 2693 \times 2}{2} = 108,945$.

⁴¹ Our comparisons of two or five bins of cities are analogous to the empirical exercises presented in Eeckhout et al. (2014) and Bacolod et al. (2009).

Table 4
Pairwise comparisons of three skill groups.

Bins	Total comparisons	Success rate	Success rate
		All	US-born
270	108,945	0.67 (0.00)	0.69 (0.00)
90	12,015	0.74 (0.00)	0.77 (0.00)
30	1305	0.82 (0.00)	0.86 (0.00)
10	135	0.89 (0.00)	0.91 (0.00)
5	30	0.97 (0.00)	0.99 (0.00)
3	9	1.00 (0.00)	1.00 (0.00)
2	3	1.00 (0.17)	1.00 (0.17)

NOTES: P -values for uniform null hypothesis in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas. Outcomes weighted by log-population differences.

Table 5
Population elasticities of nine skill groups.

Dependent variable: $\ln f(\omega, c)$	(1)	(2)	Population	Share
	All	US-born	share	US-born
$\beta_{\omega 1}$ Less than high school \times log population	1.089 (0.0314)	0.858 (0.0239)	0.04	0.29
$\beta_{\omega 2}$ High school dropout \times log population	1.005 (0.0152)	0.933 (0.0181)	0.08	0.73
$\beta_{\omega 3}$ High school graduate \times log population	0.925 (0.0132)	0.890 (0.0163)	0.25	0.88
$\beta_{\omega 4}$ College dropout \times log population	0.997 (0.0111)	0.971 (0.0128)	0.23	0.89
$\beta_{\omega 5}$ Associate's degree \times log population	0.997 (0.0146)	0.965 (0.0157)	0.08	0.87
$\beta_{\omega 6}$ Bachelor's degree \times log population	1.087 (0.0149)	1.059 (0.0164)	0.20	0.86
$\beta_{\omega 7}$ Master's degree \times log population	1.095 (0.0179)	1.063 (0.0181)	0.08	0.84
$\beta_{\omega 8}$ Professional degree \times log population	1.113 (0.0168)	1.082 (0.0178)	0.03	0.81
$\beta_{\omega 9}$ PhD \times log population	1.069 (0.0321)	1.021 (0.0303)	0.01	0.72

NOTES: Standard errors, clustered by MSA, in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas.

significant at the 1% level, except for the 2-bin case that involves only three comparisons, severely reducing the test's power.⁴²

Both our empirical tests show a clear central tendency in the data across three skill groups. More skilled individuals are relatively more prevalent in more populous cities. Moreover, these patterns are difficult to reconcile with a "skilled-unskilled" dichotomy. Individuals with "some college" are distinct from both those with high school or less education and those with a bachelor's or higher education.

⁴² Under a uniform null hypothesis, the probability that three tossed coins all turn up heads is 1/8, so the minimum possible p -value is 0.125. Similarly, our 1000 simulations of the three comparisons under the null yield three correct comparisons in 17.8% of simulations, bounding the p -value.

5.1.2. Nine skill groups

We next examine our tests for the case with nine skill groups. Per Lemma 4, our predictions about the top of the skill distribution are stronger in the sense that the skills found in a larger city are a superset of those found in a smaller city, independent of the shape of the locational supply function $V(z)$. The results of the elasticity test are presented in Table 5. The more educated skill groups generally have higher population elasticities, as would be expected from the three-skill-group results. This pattern is very clear for high school graduates through professional degree holders, an interval that accounts for 87% of the US population. The extreme tails of the distribution, however, do not conform to the prediction of our model.

The population elasticity of PhDs is somewhat below that of other college-educated categories. This is problematic for all theories in which skill-biased agglomeration should cause the most skilled to concentrate in larger cities. On the other hand, only 1% of US workers possess a doctoral degree, so their spatial distribution has limited impact on sectoral employment patterns and aggregate outcomes. The lower-than-expected population elasticity reflects the cross-city distribution of institutions of higher education. The 27% of PhDs whose occupation is “postsecondary teacher” exhibit a population elasticity of 1.01, while the remainder of PhDs exhibit a population elasticity of 1.14. Thus, outside of higher education, which may be poorly described by a competitive model with agglomeration economies, the distribution of PhDs across cities is well predicted by our theory.⁴³

Those with less than a high-school education, who account for 12% of the working population, exhibit higher-than-expected population elasticities. Within the context of our model, this could be attributed to the locational supply function $V(z)$ failing to exhibit a decreasing elasticity, causing $s(\gamma, c)$ to not be log-supermodular everywhere. As noted previously, while our predictions about the top of the skill distribution are independent of the shape of $V(z)$, our predictions about lower skills depend on $V(z)$ exhibiting a decreasing elasticity. Other assumptions about this function can deliver non-monotonic population elasticities. For example, it is possible to make assumptions on $V(z)$ such that population elasticities decrease with skill among the least-skilled segment of the population and increase with skill among the most-skilled segment of the population.⁴⁴

Looking outside our model's mechanics, foreign-born individuals are more concentrated in larger cities, regardless of their educational attainment. The high population elasticity of the least-educated group is attributable to 71% of those with less than a high school education being immigrants. The second column of Table 5 shows that if we restrict attention to US-born individuals, this population elasticity drops from 1.089 to 0.858, below that of all other skill groups.⁴⁵ An alternative to restricting attention to US-born individuals is to estimate the population elasticities using data from 1980, when immigrants were a much smaller share of the US population. In 1980 data, the least skilled group has the lowest population elasticity, and the difference between the 1980 and 2000 population elasticities is almost entirely attributable to the rising share of the foreign-born in this least-skilled population. We discuss mechanisms – mechanisms omitted from our model – that may

⁴³ About two-thirds of faculty in higher education are employed by public institutions, which are not profit-maximizing and for practical and political reasons are likely to be proportionate to local population. Furthermore, universities are conventionally land-intensive enterprises.

⁴⁴ In particular, suppose that $V(z)$ had an increasing elasticity at low values of z and a decreasing elasticity at high values of z , $\frac{\partial \ln V(z)}{\partial \ln z} > 0$ if $z < z^*$ and $\frac{\partial \ln V(z)}{\partial \ln z} < 0$ if $z > z^*$. It can then be shown that $s(\gamma, c)$ would be log-submodular for $\gamma < A(1)z^*$ and log-supermodular for $\gamma > A(C)z^*$. Hence, $f(\omega, c)$ would be log-submodular for $\omega < K(A(1)z^*)$ and log-supermodular for $\omega > K(A(C)z^*)$.

⁴⁵ Interestingly, among US-born individuals, the nine estimated elasticities naturally break into the three more aggregate educational attainment categories that we used: $\beta_{\omega_1}, \beta_{\omega_2}, \beta_{\omega_3} \in (0.85, 0.94)$; $\beta_{\omega_4}, \beta_{\omega_5} \in (0.96, 0.98)$; $\beta_{\omega_6}, \beta_{\omega_7}, \beta_{\omega_8}, \beta_{\omega_9} \in (1.02, 1.09)$.

Table 6
Pairwise comparisons of nine skill groups.

Bins	Total comparisons	Success rate	
		All	US-born
270	1,307,340	0.61 (0.00)	0.64 (0.00)
90	144,180	0.66 (0.00)	0.71 (0.00)
30	15,660	0.72 (0.00)	0.77 (0.00)
10	1620	0.74 (0.00)	0.81 (0.00)
5	360	0.76 (0.00)	0.85 (0.00)
3	108	0.75 (0.02)	0.88 (0.00)
2	36	0.75 (0.08)	0.90 (0.01)

NOTES: P-values for uniform null hypothesis in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas. Outcomes weighted by product of log-population differences and educational population shares.

cause immigrants of all skill levels to concentrate in larger cities in Appendix E.1.⁴⁶

We now turn to the pairwise comparisons for the case with nine skill groups in 2000. These comparisons, presented in Table 6, check the predicted inequalities for as many as 36,315 city pairs for each pairing of the nine skill groups, separately for the population as a whole and for those individuals born in the United States. These comparisons ask a lot of the data, so we should not expect perfection. They predict, for example, that the number of associate's degree divided by the number of college dropouts will be higher in Des Moines than Kalamazoo, because the former's population is 3000 residents larger. When making pairwise comparisons across all 270 metropolitan areas, the success rates are 61 and 64%, respectively, and highly statistically significant. Aggregating raises the success rates, consistent with binning reducing the influence of idiosyncratic errors. When using five bins or fewer, these success rates exceed 75%. As suggested by the elasticity test, the success rate is higher when restricting attention to US-born individuals. These results are all statistically significant at the 1% level, except for the 2-bin case with all individuals, which is significant at the 10% level.

Both our empirical tests demonstrate strong support for our theory's predictions. In general, more educated individuals are relatively more prevalent in larger cities. The nine-skill-group predictions are far more detailed than prior descriptions of cities' skill distributions. The high population elasticity of those not reaching high school

⁴⁶ The fact that the extreme tails of the skill distribution do not conform to our predictions invites a comparison to Eeckhout et al. (2014). Eeckhout et al. (2014) introduce a model in which larger cities' skill distributions have the same median skill but exhibit thicker tails. In particular, their model predicts that skill types' population elasticities will be U-shaped, with the median skill type exhibiting the lowest population elasticity. The population elasticity of a skill type should increase with its distance from the median skill. By symmetry, percentiles equally distant from the median should exhibit the same population elasticity. For example, the 1st and 99th percentile skill types should exhibit the (same) highest population elasticity. In their empirical work, Eeckhout et al. (2014) use cost-of-living-adjusted wages to measure skill and identify thicker tails in larger cities at the 10th and 90th percentiles. However, their predictions do not hold throughout the skill distribution. Any partition of skills into three groups with equal-sized low and high groups should yield U-shaped population elasticities. Table 2 rejects this prediction for a benchmark case of three groups of approximately equal size: the population elasticities of the high-school-or-less and bachelors-or-more groups are not close to equal. Instead, population elasticities are monotonically increasing in educational attainment, consistent with our prediction. Looking at the nine-skill case in Table 3, we see that the median skill in the United States is a college dropout, which has a considerably higher population elasticity than high school graduates (0.997 vs. 0.925). This is at odds with the Eeckhout et al. (2014) prediction. Moreover, Table 4 shows that the elasticities are monotonically increasing for high school graduates through professional degrees, which jointly account for 87% of the labor force. While the population elasticity of professional degrees is greater than that for PhDs (1.113 vs. 1.069), neither their theory nor ours predicts this modest decline.

does not match our model; however, these individuals are a small fraction of the population and are overwhelmingly foreign-born. International immigrants are particularly attracted to large cities; we show that these outliers are absent in 1980, when US-born individuals were a much larger share of the least-skilled group. The outliers in 2000 dampen the success rate of the pairwise comparisons, but our model's predictions are an apt description of the broad pattern and highly statistically significant. Thus, our theory provides stronger and more detailed predictions about skill patterns across cities than prior work and, on the whole, the US data strongly support those predictions.

5.2. Larger cities specialize in skill-intensive sectors

This subsection examines the pattern of sectoral employment across cities. In our theory, larger cities are relatively more skilled and the sectoral assignment function is common across cities, so larger cities employ relatively more labor in skill-intensive sectors. Section 5.1 established that larger cities are relatively more skilled, but it need not be the case that larger cities specialize in skill-intensive sectors, since our theory relied on a number of assumptions to obtain this result. We now examine whether larger cities are relatively specialized in skill-intensive sectors. Since employment levels in both industries and occupations are readily available in the data, we test the employment implications of Proposition 1b.

5.2.1. The distribution of occupations across cities

We first implement the elasticities test and the pairwise comparisons test interpreting sectors as occupations.⁴⁷ We begin with a visualization of the elasticity results. Define the skill intensity of an occupation as the average years of schooling of individuals employed in that occupation. Fig. 1 plots the 22 occupational categories' estimated population elasticities of employment against these occupational skill intensities.⁴⁸ Our theory says the population elasticity of occupational employment should rise with skill intensity and indeed we see a clear positive relationship.⁴⁹ Outliers in the figure include close-to-unitary elasticities for the relatively skilled occupations in education, healthcare, and social services, which may reflect non-traded status. On the other side, computer and mathematical occupations have an elasticity that is quite high relative to their average years of schooling.

We can also look at this more formally. With the population elasticities of occupations in hand, the hypothesis that $\beta_{\sigma} \geq \beta_{\sigma'} \Leftrightarrow \sigma \geq \sigma'$ involves 231 ($=22 \times 21/2$) comparisons of the estimated coefficients.⁵⁰ This hypothesis is rejected at the 5% significance level in only 46 of the 231 such comparisons. The occupational elasticity results are broadly consistent with our prediction that larger cities specialize in skill-intensive activities.

The results for pairwise comparisons for occupations appear in Table 7.⁵¹ Using 276 cities and 22 occupational categories to make more than 8 million pairwise comparisons yields an average success rate of 57%.⁵² While far from perfect, our model's predictive power over these millions of pairs of cities and occupations is highly

statistically significant. As we increase the number of cities per bin, the success rate increases up to 75%, as idiosyncratic errors are averaged out to reveal the central tendency of the data. These results are statistically significant across all levels of aggregation.

Thus, both the estimated population elasticities and pairwise comparisons reveal a broad and strong tendency for more populous metropolitan areas to employ relatively more individuals in skill-intensive occupations.

5.2.2. The distribution of industries across cities

We again implement the elasticities test and the pairwise comparisons test, now interpreting sectors as industries. Define the skill intensity of an industry as the average years of education of those employed in that industry. A visualization of the elasticity test appears in Fig. 2.⁵³ Again, as predicted by our theory, there is a clear positive relationship so that the population elasticity of industrial employment is rising with the skill intensity of the industry.⁵⁴ Testing the hypothesis that $\beta_{\sigma} \geq \beta_{\sigma'} \Leftrightarrow \sigma \geq \sigma'$ for the 19 industries involves 171 ($=19 \times 18/2$) comparisons of these estimated elasticities.⁵⁵ This hypothesis is rejected in only 33 comparisons, so the elasticity implication holds true for industries about 80% of the time.

The results for pairwise comparisons for industries appear in Table 8. Making more than 6 million pairwise comparisons yields a statistically significant success rate of 61%. Raising the number of cities per bin raises the success rate monotonically and always at a high level of statistical significance. The success rate rises to 77% when contrasting two bins of large and small cities.

Thus, both the estimated population elasticities and pairwise comparisons reveal a precise and systematic pattern in which more populous metropolitan areas employ relatively more individuals in skill-intensive industries.⁵⁶ Prior work, including Henderson (1983) and Holmes and Stevens (2004), estimated industries' population elasticities, although without having a theoretical foundation for interpreting them. Our theory provides a basis for such estimation and predicts the ordering of the elasticities based on skill intensities. Other prior work, namely Henderson (1997), contrasted industrial patterns among large- and medium-size cities. Our theory implies such comparisons between two or many more groups defined by cities' sizes, and our results show that there is precise and systematic empirical content even as we look at these finer comparisons between (groups of) cities.

5.3. Testing for systematic failures of comparative advantage

Our results for the cross-city distributions of skills, industries, and occupations demonstrate systematic patterns in line with our theory's predictions. While demonstrating predictive power, the pairwise comparisons also fall well short of 100% success. This is not surprising, given that our model's parsimony stems from making strong assumptions that omit various features that influence the real world.

An important question is whether our theory's unsuccessful pairwise predictions are merely idiosyncratic deviations from the pattern of comparative advantage or are systematic violations of our predicted pattern. The fact that our pairwise comparisons success rates increase with the number of cities per bin is consistent with aggregation over idiosyncratic errors, as shown in Appendix D. We also implement a formal test for systematic violations of the predicted patterns.

Sattinger (1978) develops an approach to test for such systematic violations in the form of systematic intransitivity in the pattern of comparative advantage. It is possible for the data to exhibit, for $c > c' > c''$

⁴⁷ While there is a long line of literature looking at cities' industrial composition, occupational composition has been little studied. A notable exception is Duranton and Puga (2005), who document that larger cities have more managers per production worker. Our investigation extends their inquiry to many more occupational categories.

⁴⁸ These elasticities are estimated without including zero-employment observations. The results obtained when including those observations are similar.

⁴⁹ Studies of international trade have characterized an analogous relationship by regressing country-sectoral exports on country fixed effects, sectoral fixed effects, and the interaction of measures of sectoral factor intensity and country factor abundance (e.g., Romalis, 2004; Nunn, 2007). Estimating $\ln f(\sigma, c) = \alpha_{\sigma} + \alpha_c + \beta(\text{skill}(\sigma) \times \ln L(c)) + \varepsilon_{\sigma, c}$ in our setting yields a β coefficient of 0.030 with a standard error of 0.003.

⁵⁰ The elasticity estimates appear in Appendix Table E.2.

⁵¹ With 276 metropolitan areas and bins that are factors of 270, the number of cities per bin may differ by one.

⁵² These pairwise comparisons omit zero-employment observations. The results obtained when including those observations are similar.

⁵³ As for occupations, these elasticities are estimated without including zero-employment observations. The results obtained when including those observations are similar.

⁵⁴ A regression on the interaction of skill intensity and log population, as described in footnote 49, yields a coefficient of 0.080 with a standard error of 0.008.

⁵⁵ The elasticity estimates appear in Appendix Table E.3.

⁵⁶ We have found broadly similar results when examining industries at a finer level of disaggregation and when restricting attention to manufacturing industries.

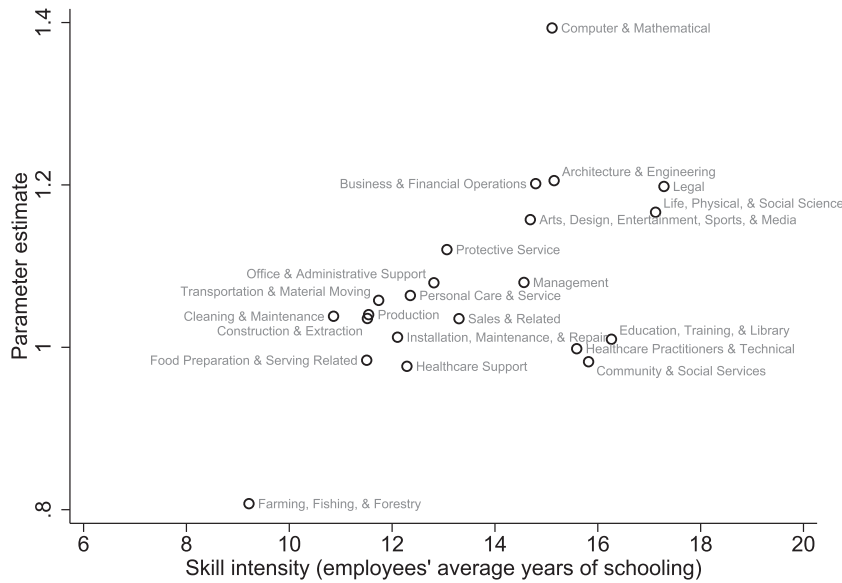


Fig. 1. Occupations' population elasticities and skill intensities.

and $\sigma > \sigma' > \sigma''$, $\frac{f(\sigma, c)}{f(\sigma', c)} \geq \frac{f(\sigma, c')}{f(\sigma', c')}$ and $\frac{f(\sigma' c')}{f(\sigma'' c')} \geq \frac{f(\sigma' c'')}{f(\sigma'' c'')}$ without exhibiting $\frac{f(\sigma c)}{f(\sigma'' c)} \geq \frac{f(\sigma c'')}{f(\sigma'' c'')}$. With hundreds of metropolitan areas and dozens of sectors, it is easy to find three cities and three sectors in the data exhibiting such intransitivity. But do intransitivities arise systematically? [Sattinger \(1978\)](#) shows that if $\ln f(\sigma, c)$ is a polynomial function of $\hat{\beta}_\sigma$ and $\ln L(c)$, then there can be systematic intransitivity only if $\ln f(\sigma, c)$ is a function of higher-order interactions of $\hat{\beta}_\sigma$ and $\ln L(c)$. We therefore added quadratic terms and their interactions to our elasticity regressions. These did little to improve the regression's adjusted R^2 , and F-tests yielded p -values that did not come close to rejecting the null that these additional terms were uninformative. There is no evidence of systematic intransitivity in comparative advantage. While our theory's predictive successes are systematic, the empirical departures from our theory appear to be idiosyncratic.

5.4. Larger cities are larger in all skills and sectors

As described in [Section 2](#), different agglomeration theories have different implications for the relationship between city size and sectoral employment levels. Localization theories make the trade-off between industry-specific agglomeration economies and general congestion costs the foundation of the city-size distribution. Formally, steel cities

Table 7
Pairwise comparisons of occupations.

Bins	Total comparisons	Success rate
276	8,073,382	0.57 (0.00)
90	925,155	0.62 (0.00)
30	100,485	0.63 (0.00)
10	10,395	0.65 (0.00)
5	2310	0.71 (0.00)
3	693	0.74 (0.00)
2	231	0.75 (0.03)

NOTES: P-values for uniform null hypothesis in parentheses. Outcomes weighted by product of log-population differences and skill-intensity differences.

would come in one size, textile cities in another, and so on. Localization theories militate against the idea that larger cities will be larger in all sectors. Our theory, in its baseline form, does not require that larger cities are larger in all sectors. However, by focusing on urbanization economies, our theory allows that large cities may be the largest site of economic activity for all sectors ([Proposition 2](#)). Our empirical exercise in this sub-section asks what weight should be placed on the predictions flowing from the localization and urbanization archetypes.

We have already shown that larger cities have relatively larger numbers of skilled workers. We now investigate whether larger cities also tend to have larger populations of all skill types. They certainly do, as all the population elasticities reported in [Tables 3 and 5](#) are strongly positive. Among the nine educational categories, the prediction that $c > c' \Rightarrow f(\omega, c) \geq f(\omega, c')$ is true in 88% of 326,835 cases. The largest metropolitan area, New York, has the largest population in all educational categories, except the least skilled “less than high school” category, which is most populous in the second-largest metropolitan area, Los Angeles. Among US-born individuals, New York has the largest population in all nine educational categories.

Turning to sectors, larger cities tend to have larger sectoral employment in all activities. This tendency is clear from the population elasticities plotted in [Figs. 1 and 2](#), as they are all strongly positive. Among the 19 2-digit NAICS industries, the prediction that $c > c' \Rightarrow f(\sigma, c) \geq f(\sigma, c')$ is true in 86% of 721,050 cases. Sixteen industries attain their maximal size in the largest metropolitan area (New York). The three exceptions are manufacturing (second-largest Los Angeles), mining (tenth-largest Houston), and forestry, fishing, hunting, and agriculture support (thirteenth-largest Seattle). The analogous results for occupational categories show a similar tendency for larger cities to have higher employment levels in all occupations. The $c > c' \Rightarrow f(\sigma, c) \geq f(\sigma, c')$ prediction holds true in 89% of occupational comparisons, and 19 of the 22 occupations attain their maximal size in the largest metropolitan area, New York. The exceptions are production occupations (Los Angeles), architecture and engineering occupations (fifth-largest San Francisco), and farming, fishing, and forestry occupations (51st-largest Fresno).

These findings are more consistent with urbanization economies than localization mechanisms at the city level. While particular examples such as San Francisco's concentration of architecture and engineering occupations may be consistent with localization economies, the very large majority of sectors exhibit larger employment levels in more populous cities. Our theory, which parsimoniously assumes only

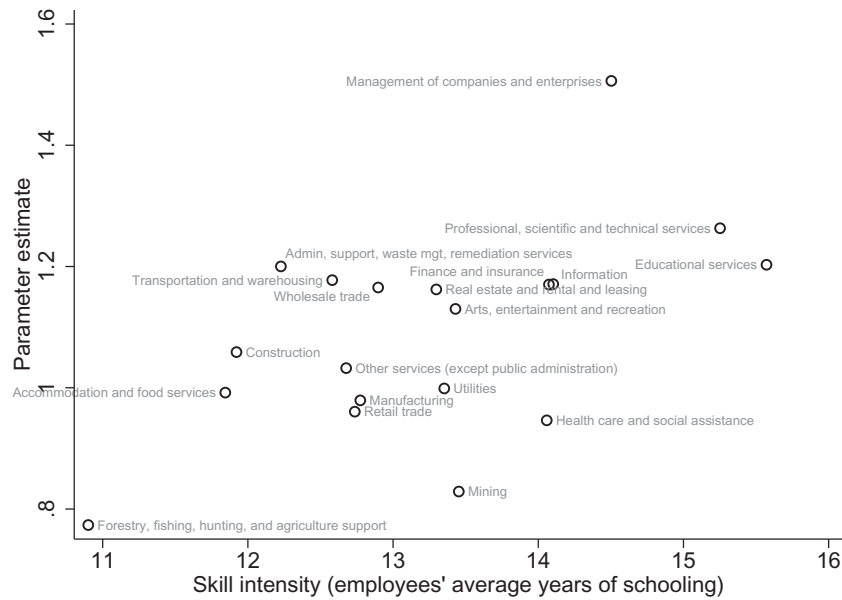


Fig. 2. Industries' population elasticities and skill intensities.

urbanization economies, matches the data on cities' sectoral composition and sectoral sizes quite well relative to existing models.

6. Discussion and conclusions

In this paper, we introduce a model that simultaneously characterizes the distribution of skills and sectors across cities. We describe a high-dimensional economic environment that is a system of cities in which cities' internal geographies exhibit substantive heterogeneity and individuals' comparative advantage governs the distribution of sectoral employment. Our model achieves two aims. First, we obtain "smooth" predictions, in the sense that cities' skill and sectoral distributions will be highly overlapping. These are more realistic than prior theories describing cities that are perfectly sorted along skills or polarized in terms of sectoral composition. Second, we obtain "strong" predictions, in the sense that cities' skill and sectoral distributions will exhibit systematic variation according to the monotone likelihood ratio property. These are much finer than the predictions of many prior theories of the spatial organization of economy activity and guide our empirical investigation.

Examining data on US metropolitan areas' populations, occupations, and industries in the year 2000 reveals systematic variation in the cross-city distribution of skills and sectors that is consistent with our theory. Larger cities are skill-abundant. Our results using three roughly

equal-sized categories of educational attainment are quite strong. Even disaggregated to nine educational categories, the cross-city distribution of skills is broadly consistent with our theory. Larger cities specialize relatively in skill-intensive activities. More skill-intensive occupations and industries tend to have higher population elasticities of employment, and in pairwise comparisons, the more skill-intensive sector tends to employ relatively more individuals in the larger city. Consistent with our approach based on urbanization economies, larger cities tend to have larger absolute employment in all sectors.

We believe that our framework is amenable to both theoretical and empirical applications and extensions. The "strong" character of our predictions and their demonstrated relevance for describing US cities in 2000 suggest that their examination in other settings, such as economies at different stages of development or in different historical periods, would be interesting. The "smoothness" resulting from the simultaneous consideration of cross- and within-city heterogeneity in a continuum-by-continuum environment would make our model amenable to theoretical analyses of the consequences of commuting costs, globalization, and skill-biased technical change. For example, skill-biased technological change has consequences for both the wage distribution and the land-rent distribution in our model. Technological change, in the form of a change from $B(\sigma)$ to $B'(\sigma)$, is skill-biased if $B'(\sigma)B(\sigma) \geq B'(\sigma)B(\sigma) \quad \forall \sigma' \geq \sigma$. Such a shift increases both wage inequality and rent inequality, in the sense that $\frac{G'(\omega')}{G'(\omega)} \geq \frac{G(\omega')}{G(\omega)} \quad \forall \omega' \geq \omega$ and $\frac{r'_\Gamma(\gamma')}{r'_\Gamma(\gamma)} \geq \frac{r_\Gamma(\gamma')}{r_\Gamma(\gamma)} \quad \forall \gamma' \geq \gamma$.⁵⁷ In

models in which individuals are indifferent across all locations within a city, the rent gradient is invariant to the wage distribution. In our assignment model of differentiated locations, the composition of income growth governs the consequences for less-skilled individuals' housing costs.⁵⁸

In short, our theoretical approach provides a new perspective on the comparative advantage of cities that aligns well with the patterns of specialization documented in our empirical work.

Table 8
Pairwise comparisons of industries.

Bins	Total comparisons	Success rate
276	6,469,758	0.61 (0.00)
90	684,855	0.64 (0.00)
30	74,385	0.66 (0.00)
10	7695	0.71 (0.00)
5	1710	0.74 (0.00)
3	513	0.74 (0.01)
2	171	0.77 (0.05)

Notes: P-values for uniform null hypothesis in parentheses. Outcomes weighted by product of log-population differences and skill-intensity differences.

⁵⁷ Analogous to our result linking rent inequality to wage inequality, Clemens et al. (2018) propose an assignment model in which greater income inequality among consumers causes greater income inequality among their producers of vertically differentiated services.

⁵⁸ Couture et al. (2019) quantify such welfare implications of income growth by introducing idiosyncratic preferences shocks to a model of urban spatial sorting. In line with our logic, they find that income growth at the top of the distribution raises low-income households' rents much less than broad-based income growth.

Appendix A. Theory

A.1. Autarkic equilibrium

Suppose the entire population lives in a single city, denoted c . It has an exogenous population $L(c)$ and skill distribution $F(\omega)$. With fixed population, autarky TFP is fixed by eq. (5).

Lemma 6 (Autarky locational assignments). *In autarkic equilibrium, there exists a continuous and strictly decreasing locational assignment function $N : \bar{T}(c) \rightarrow \Omega$ such that $f(\omega, c, \tau, M(\omega)) > 0 \Leftrightarrow N(\tau) = \omega$, $N(0) = \bar{\omega}$ and $N(\bar{T}(c)) = \underline{\omega}$.*

This assignment function is obtained by equating supply and demand of locations:

$$S(\tau) = L \int_0^{\bar{T}} \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, \tau, \sigma) d\omega d\sigma dx$$

$$\Rightarrow N(\tau) = F^{-1} \left(\frac{L(c) - S(\tau)}{L(c)} \right)$$

Given individuals' equilibrium locations within the city, the schedule of locational rental prices supporting these assignments comes from combining individuals' utility-maximizing decisions and the boundary condition $r(c, \bar{T}(c)) = 0$.⁵⁹

Lemma 7 (Autarky locational prices). *In autarkic equilibrium, $r(c, \tau)$ is continuously differentiable on $\tau \geq 0$ and given by*

$$r(c, \tau) = -A(c) \int_{\bar{T}(c)}^{\tau} T'(t)G(N(t))dt \text{ for } \tau \leq \bar{T}(c).$$

The properties of interest in a competitive equilibrium are characterized by the assignment functions $M : \Omega \rightarrow \Sigma$ and $N : \bar{T}(c) \rightarrow \Omega$. In the autarkic equilibrium of a single city with an exogenous population, more skilled individuals work in more skill-intensive sectors and occupy more desirable locations.

A.2. Amenity interpretation of desirability

The productivity and amenity interpretations of desirability yield very similar results but differ slightly in notation. In the amenity interpretation, an individual's productivity and utility are

$$q(c, \tau, \sigma; \omega) = A(c)H(\omega, \sigma) \tag{A.1}$$

$$U(c, \tau, \sigma; \omega) = T(\tau)[A(c)H(\omega, \sigma)p(\sigma) - r(c, \tau)] \tag{A.2}$$

where amenity $T(\tau)$ determines the value of the individual's disposable income after paying his or her locational price.⁶⁰ In this interpretation, preferences are non-homothetic in a manner akin to that of Gabszewicz et al. (1981). Higher-income individuals are more willing to pay for higher-amenity locations because a more desirable location complements their higher consumption of tradables.

In this case, instead of $\gamma = A(c)T(\tau) = A(c')T(\tau') \Leftrightarrow r(c, \tau) = r(c', \tau') = r_{\Gamma}(\gamma)$, the appropriate equivalence between two locations is their "amenity-amplified price", which is $T(\tau)r(c, \tau)$. So the equivalence statement is now $\gamma = A(c)T(\tau) = A(c')T(\tau') \Leftrightarrow T(\tau)r(c, \tau) = T(\tau')r(c', \tau') = r_{\Gamma}(\gamma)$. The results in lemma 1 are unaltered, though the proof is modified to use the relevant $U(c, \tau, \sigma; \omega)$. The expressions for $K : \Gamma \rightarrow \Omega$, $\bar{\gamma}$, and $\underline{\gamma}$ are unaltered. This

leaves the conclusions of lemmas 2 and 3 intact. The locational price schedule is given by $r(c, \tau) = \frac{r_{\Gamma}(A(c)T(\tau))}{T(\tau)} = A(c) \frac{r_{\Gamma}(\gamma)}{\gamma}$.

These locational prices do not appear in the endogenous definition of $A(c)$ nor the proofs of Lemma 4 and subsequent results. When evaluated at equilibrium, occupied locations' productivities $q(c, \tau, \sigma; \omega) = A(c)H(\omega, \sigma)$ differ across cities in a Hicks-neutral fashion. Thus, Proposition 1c holds true. As a result, the predictions about cities' population, sectors, and productivities described in subsection 3.4 are unaltered by interpreting $T(\tau)$ as describing consumer amenity benefits rather than productivity benefits.

The wage distribution does differ between the productivity and amenity interpretations of desirability. In the productivity interpretation, productivity $q(c, \tau, \sigma; \omega)$ does not vary across ω -occupied locations in equilibrium, as reported in footnote 24, and $p(\sigma)$ is common across locations, so individuals of the same skill earn the same income everywhere. In the amenity interpretation, an individual's income $p(\sigma)q(c, \tau, \sigma; \omega)$ is equal to $A(c)G(\omega)$, so individuals of the same skill earn higher nominal incomes in larger cities in equilibrium.

A.3. Cities' internal geographies

A number of conceivable $V(z)$ schedules satisfy the decreasing-elasticity condition of Proposition 1. We provide some examples here:

Monocentric-city model: For the monocentric city's disc geography, $S(\tau) = \pi\tau^2$, with linear transportation costs, $T(\tau) = d_1 - d_2\tau$, the supply of locations within cities $V(z) = \frac{2\pi}{d_2}(d_1 - z)$ has an elasticity of $-\frac{z}{d_1 - z}$, which is decreasing in z .

Exponential family: The exponential family of distributions has PDFs that can be written in the (canonical) form $V(z|\eta) = v_1(z) \exp(\eta \cdot v_2(z) - v_3(\eta))$, where η and $v_2(z)$ may be vectors. Thus, if $V(z)$ is a member of the exponential family, we are interested in its elasticity

$$\frac{\partial \ln V(z)}{\partial \ln z} = \frac{\partial \ln v_1(z)}{\partial \ln z} + \eta \cdot \nabla_{\ln z} v_2(z)$$

⁵⁹ Note that this boundary condition implies that in equilibrium every individual's final-good consumption is strictly positive, provided that $H(\underline{\omega}, \underline{\omega}) > 0$.

⁶⁰ Recall that the final good is the numeraire.

- Exponential: $v_1(z) = 1$ and $v_2(z) = z$. Therefore $\frac{\partial \ln V(z)}{\partial \ln z} = \eta z$ and the elasticity is decreasing because $\eta < 0$ for the exponential distribution.
- Weibull: $v_1(z) = z^{k-1}$ and $v_2(z) = z^k$. Therefore $\frac{\partial \ln V(z)}{\partial \ln z} = k-1 + \eta k z^k$ and the elasticity is decreasing because $\eta = \frac{-1}{\lambda^k} < 0$ in the standard expression of Weibull parameters.
- Gamma: $v_1(z) = 1$ and $v_2(z) = [\ln z, z]$. Therefore $\frac{\partial \ln V(z)}{\partial \ln z} = \eta_1 + \eta_2 z$ and the elasticity is decreasing because $\eta_2 = -\beta < 0$ in the standard expression of Gamma parameters.
- Log-normal: $v_1(z) = \frac{1}{\sqrt{2\pi z}}$ and $v_2(z) = [\ln z, (\ln z)^2]$. Therefore $\frac{\partial \ln V(z)}{\partial \ln z} = -1 + \eta_1 + 2\eta_2 \ln z$ and the elasticity is decreasing because $\eta_2 = -\frac{1}{2\sigma^2} < 0$ in the standard expression of log-normal parameters.

A.4. Endogenous supply of heterogeneous locations

In the main text, the supply of locations with innate desirability of at least τ is $S(\tau)$. This section relaxes the assumption of an inelastic supply schedule and shows that our main results still hold.

Let the cost (in units of the numeraire good) of building s units at a location of (inverse) innate desirability τ be $C(s; \tau)$. Profit-maximizing, perfectly competitive landlords build additional units until rent equals marginal cost.

$$\pi(s; \tau, c) = r_\Gamma(A(c)T(\tau))s - C(s; \tau)$$

$$\pi'(s; \tau, c) = 0 \Rightarrow r_\Gamma(A(c)T(\tau)) = C'(s; \tau)$$

Let $C(s; \tau) = \frac{s^\beta}{\beta h(\tau)}$, where $\beta > 1$ and $h(\tau)$ is a supply shifter. When $h(\tau)$ is greater, the cost of building a given number of units at τ is lower. Given this functional form, the supply of units in city c of attractiveness γ is

$$\ln s(\gamma, c) = \frac{1}{\beta-1} \ln r_\Gamma(\gamma) + \frac{1}{\beta-1} \ln \left(h \left(T^{-1} \left(\frac{\gamma}{A(c)} \right) \right) \right).$$

Thus, $s(\gamma, c)$ is log-supermodular if and only if $h(T^{-1}(\gamma/A(c)))$ is log-supermodular in (γ, A) . Note that $\frac{\gamma}{A}$ is submodular and log-modular in (γ, A) on \mathbb{R}_+^2 .

Thus, this function is log-supermodular if and only if $h(T^{-1}(z))$ has a decreasing elasticity by lemma 8.

Thus, with endogenous supplies of locations, the decreasing-elasticity sufficient condition of Proposition 1 applies to the schedule of exogenous supply shifters rather than the inelastic supply schedule. A sufficient condition to obtain the result in Proposition 2 is that $h(\tau)$ is increasing.

Appendix B. Proofs

Proof of Lemma 1:

Proof. Suppose that $\exists \tau' < \bar{\tau}(c) : S(\tau') > L \int_0^{\tau'} \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, x) d\omega d\sigma dx$. Then $\exists \tau \leq \tau' : S(\tau) > L \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, \tau) d\omega d\sigma$. Then $r(c, \tau) = 0 \leq r(c, \bar{\tau}(c))$, so $U(c, \tau, \sigma; \omega) > U(c, \bar{\tau}(c), \sigma; \omega) \forall \omega \forall \sigma$ since $T(\tau)$ is strictly decreasing. This contradicts the definition of $\bar{\tau}(c)$, since $\bar{\tau}(c)$ is a location that maximizes utility for some individual. Therefore $S(\tau) = L \int_0^\tau \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, x) d\omega d\sigma dx \forall \tau \leq \bar{\tau}(c)$.

Suppose that $\exists \tau', \tau'' : \tau' < \tau'' \leq \bar{\tau}(c)$ and $r(c, \tau') \leq r(c, \tau'')$. Then $U(c, \tau', \sigma; \omega) > U(c, \tau'', \sigma; \omega) \forall \omega \forall \sigma$ since $T(\tau)$ is strictly decreasing. This contradicts the result that τ'' maximizes utility for some individual. Therefore $r(c, \tau)$ is strictly decreasing in $\tau \forall \tau \leq \bar{\tau}(c)$.

Suppose $r(c, \bar{\tau}(c)) > 0$. Then by its definition as a populated location, $\exists \omega : A(c)T(\bar{\tau}(c))G(\omega) - r(c, \bar{\tau}(c)) \geq A(c)T(\bar{\tau}(c) + \varepsilon)G(\omega) \forall \varepsilon > 0$. This inequality is false for all ω for sufficiently small ε , by the continuity of $T(\tau)$. Therefore $r(c, \bar{\tau}(c)) = 0$. \square

Proof of Lemma 2:

This proof is analogous to the proof of lemma 6 below.

Proof of Lemma 3:

Proof. By utility maximization

$$\gamma G(K(\gamma)) - r_\Gamma(\gamma) \geq (\gamma + d\gamma)G(K(\gamma)) - r_\Gamma(\gamma + d\gamma)$$

$$(\gamma + d\gamma)G(K(\gamma + d\gamma)) - r_\Gamma(\gamma + d\gamma) \geq \gamma G(K(\gamma + d\gamma)) - r_\Gamma(\gamma)$$

Together, these inequalities imply

$$\frac{(\gamma + d\gamma)G(K(\gamma + d\gamma)) - \gamma G(K(\gamma + d\gamma))}{d\gamma} \geq \frac{r_\Gamma(\gamma + d\gamma) - r_\Gamma(\gamma)}{d\gamma} \geq \frac{(\gamma + d\gamma)G(K(\gamma)) - \gamma G(K(\gamma))}{d\gamma}$$

Taking the limit as $d\gamma \rightarrow 0$, we obtain $\frac{\partial r_\Gamma(\gamma)}{\partial \gamma} = G(K(\gamma))$. Integrating from $\underline{\gamma}$ to γ and using the boundary condition $r_\Gamma(\underline{\gamma}) = 0$ yields $r_\Gamma(\gamma) = \int_{\underline{\gamma}}^\gamma G(K(x)) dx$. \square

Proof of Lemma 4:

Proof. In city c , the population of individuals with skills between ω and $\omega + d\omega$ is

$$L \int_{\omega}^{\omega+d\omega} f(x, c) dx = S \left(T^{-1} \left(\frac{K^{-1}(\omega)}{A(c)} \right) \right) - S \left(T^{-1} \left(\frac{K^{-1}(\omega + d\omega)}{A(c)} \right) \right).$$

Taking the derivative with respect to $d\omega$ and then taking the limit as $d\omega \rightarrow 0$ yields the population of ω in c . Using the definition of $s(\gamma, c)$ yields the desired expression. \square

Proof of Lemma 5:

Proof. In city c , the population of individuals employed in sectors between σ and $\sigma + d\sigma$ is

$$L \int_{\sigma}^{\sigma+d\sigma} f(x, c) dx = S \left(T^{-1} \left(\frac{K^{-1}(M^{-1}(\sigma))}{A(c)} \right) \right) - S \left(T^{-1} \left(\frac{K^{-1}(M^{-1}(\sigma + d\sigma))}{A(c)} \right) \right).$$

Taking the derivative with respect to $d\sigma$ and then taking the limit as $d\sigma \rightarrow 0$ yields the population employed in σ in c . Using the definition of $s(\gamma, c)$ yields the desired expression. \square

Proof of Lemma 6:

Proof. Nearly all of our argument follows the proof of Lemma 1 in Costinot and Vogel (2010). Define $f(\omega, c, \tau) \equiv \int_{\sigma \in \Sigma} f(\omega, c, \tau, \sigma) d\sigma$. Define $\Omega(\tau) \equiv \{\omega \in \Omega \mid f(\omega, c, \tau) > 0\}$ and $\mathcal{T}(\omega) \equiv \{\tau \in [0, \bar{\tau}(c)] \mid f(\omega, c, \tau) > 0\}$.

1. $\mathcal{T}(\omega) \neq \emptyset$ by eq. (11) and $f(\omega) > 0$. $\Omega(\tau) \neq \emptyset \forall \tau \leq \bar{\tau}(c)$ by lemma 1.

2. $\Omega(\tau)$ is a non-empty interval for $\tau \in [0, \bar{\tau}(c)]$. Suppose not, such that $\omega < \omega' < \omega''$ with $\omega, \omega'' \in \Omega(\tau)$ and $\omega' \notin \Omega(\tau)$. $\exists \tau' : \omega' \in \Omega(\tau')$. Suppose $\tau' > \tau$. By utility maximization

$$\begin{aligned} A(c)T(\tau')G(\omega') - r(c, \tau') &\geq A(c)T(\tau)G(\omega') - r(c, \tau) \\ A(c)T(\tau)G(\omega) - r(c, \tau) &\geq A(c)T(\tau')G(\omega) - r(c, \tau') \end{aligned}$$

These jointly imply $(T(\tau') - T(\tau))(G(\omega') - G(\omega)) \geq 0$, contrary to $\tau' > \tau$ and $\omega' > \omega$. The $\tau' < \tau$ case is analogous, using ω' and ω'' . Therefore $\Omega(\tau)$ is a non-empty interval. The same pair of inequalities proves that for $\tau < \tau' \leq \bar{\tau}(c)$, if $\omega \in \Omega(\tau)$ and $\omega' \in \Omega(\tau')$, then $\omega \geq \omega'$.

3. $\Omega(\tau)$ is a singleton for all but a countable subset of $[0, \bar{\tau}(c)]$. Since $\Omega(\tau) \subset \Omega$ is a non-empty interval for any $\tau \in [0, \bar{\tau}(c)]$, $\Omega(\tau)$ is measurable for any $\tau \in [0, \bar{\tau}(c)]$. Let \mathcal{T}_0 denote the subset of locations τ such that $\mu[\Omega(\tau)] > 0$, where μ is the Lebesgue measure over \mathbb{R} . \mathcal{T}_0 is a countable set. For any $\tau \in \mathcal{T}_0$, define $\underline{\omega}(\tau) \equiv \inf \Omega(\tau)$ and $\bar{\omega}(\tau) \equiv \sup \Omega(\tau)$. Because $\mu[\Omega(\tau)] > 0$, we know $\bar{\omega}(\tau) > \underline{\omega}(\tau)$. Thus, for any $\tau \in \mathcal{T}_0$, there exists a $j \in \mathbb{N}$ such that $\bar{\omega}(\tau) - \underline{\omega}(\tau) \geq (\bar{\omega} - \underline{\omega})/j$. From the last result in step 2, we know that for any $\tau \neq \tau'$, $\mu[\Omega(\tau) \cap \Omega(\tau')] = 0$. Thus, for any $j \in \mathbb{N}$, there can be at most j elements $\{\tau_1, \dots, \tau_j\} \equiv \mathcal{T}_0^j \subset \mathcal{T}_0$ for which $\bar{\omega}(\tau_i) - \underline{\omega}(\tau_i) \geq (\bar{\omega} - \underline{\omega})/j$ for $i = 1, \dots, j$. By construction, $\mathcal{T}_0 = \cup_{j \in \mathbb{N}} \mathcal{T}_0^j$, where \mathcal{T}_0^j is a countable set. Since the union of countable sets is countable, \mathcal{T}_0 is a countable set. The fact that $\Omega(\tau)$ is a singleton for all but a countable subset of $[0, \bar{\tau}(c)]$ follows from the fact that \mathcal{T}_0 is a countable set and the fact that only the nonempty intervals of Ω with measure zero are singletons.

4. $\mathcal{T}(\omega)$ is a singleton for all but a countable subset of Ω . This follows from the same arguments as in steps 2 and 3.

5. $\Omega(\tau)$ is a singleton for $\tau \in [0, \bar{\tau}(c)]$. Suppose not, such that there exists $\tau \in [0, \bar{\tau}(c)]$ for which $\Omega(\tau)$ is not singleton. By step two, $\Omega(\tau)$ is an interval, so $\mu[\Omega(\tau)] > 0$, where μ is the Lebesgue measure over \mathbb{R} . By step four, we know that $\mathcal{T}(\omega) = \{\tau\}$ for μ -almost all $\omega \in \Omega(\tau)$. Hence condition (11) implies

$$f(\omega, c, \tau) = f(\omega) \delta^{Dirac} [1 - 1_{\Omega(\tau)}] \quad \text{for } \mu\text{-almost all } \omega \in \Omega(\tau), \tag{A.3}$$

where δ^{Dirac} is a Dirac delta function. Combining eqs. (9) and (3) with $\mu[\Omega(\tau)] > 0$ yields $S'(\tau) = +\infty$, which contradicts our assumptions about $S(\tau)$. Step 5 means there is a function $N : \mathcal{T} \rightarrow \Omega$ such that $f(\omega, c, \tau) > 0 \Leftrightarrow N(\tau) = \omega$. Step 2 says N is weakly decreasing. Since $\Omega(\tau) \neq \emptyset \forall \tau \leq \bar{\tau}(c)$, N is continuous and satisfies $N(0) = \bar{\omega}$ and $N(\bar{\tau}(c)) = \underline{\omega}$. Step 4 means that N is strictly decreasing on $(0, \bar{\tau}(c))$. \square

Proof of the explicit expression of $N(\tau)$ that follows Lemma 6:

$$\begin{aligned} S(\tau) &= L \int_0^{\tau} \int_{\sigma \in \Sigma} \int_{\omega \in \Omega} f(\omega, c, x, \sigma) d\omega d\sigma dx \\ &= L \int_0^{\tau} \int_{\omega \in \Omega} f(\omega) \delta^{Dirac} [x - N^{-1}(\omega)] d\omega dx \\ &= L \int_0^{\tau} \int_{\tau'}^{\tau} f(N(\tau')) \delta^{Dirac} [x - \tau'] N'(\tau') d\tau' dx \\ &= -L \int_0^{\tau} f(N(x)) N'(x) dx = L(1 - F(N(\tau))) \\ \Rightarrow N(\tau) &= F^{-1} \left(\frac{L - S(\tau)}{L} \right) \end{aligned}$$

Proof of Lemma 7:

Proof. By utility maximization

$$\begin{aligned} A(c)T(\tau)G(N(\tau)) - r(c, \tau) &\geq A(c)T(\tau + d\tau)G(N(\tau)) - r(c, \tau + d\tau) \\ A(c)T(\tau + d\tau)G(N(\tau + d\tau)) - r(c, \tau + d\tau) &\geq A(c)T(\tau)G(N(\tau + d\tau)) - r(c, \tau) \end{aligned}$$

Together, these inequalities imply

$$\frac{A(c)T(\tau + d\tau)G(N(\tau)) - A(c)T(\tau)G(N(\tau))}{d\tau} \leq \frac{r(c, \tau + d\tau) - r(c, \tau)}{d\tau} \leq \frac{A(c)T(\tau + d\tau)G(N(\tau + d\tau)) - A(c)T(\tau)G(N(\tau + d\tau))}{d\tau}$$

Taking the limit as $d\tau \rightarrow 0$, we obtain $\frac{\partial r(c, \tau)}{\partial \tau} = A(c)T'(\tau)G(N(\tau))$. Integrating from τ to $\bar{\tau}(c)$ and using the boundary condition $r(c, \bar{\tau}(c)) = 0$ yields

$$r(c, \tau) = -A(c) \int_{\tau}^{\bar{\tau}(c)} T'(t)G(N(t))dt. \quad \square$$

In the course of proving Proposition 1, we use the following lemma.

Lemma 8. Let $f(z) : \mathbb{R} \rightarrow \mathbb{R}_{++}$ and $g(x,y) : \mathbb{R}_+^2 \rightarrow \mathbb{R}_{++}$ be C^2 functions. If $g(x,y)$ is submodular and log-modular, then $f(g(x,y))$ is log-supermodular in (x,y) if and only if $f(z)$ has a decreasing elasticity.

Proof. $f(g(x,y))$ is log-supermodular in (x,y) if and only if

$$\frac{\partial^2 \ln f(g(x,y))}{\partial x \partial y} = \left[\frac{\partial \ln f(z)}{\partial z} g_{xy} + \frac{\partial^2 \ln f(z)}{\partial z^2} g_x g_y \right]_{z=g(x,y)} \geq 0$$

If $g(x,y)$ is submodular ($g_{xy} < 0$) and log-modular ($g = \frac{g_x g_y}{g_{xy}}$), this condition can be written as

$$\left[\frac{\partial \ln f(z)}{\partial z} + \frac{\partial^2 \ln f(z) g_x g_y}{\partial z^2 g_{xy}} \right]_{z=g(x,y)} = \frac{\partial}{\partial z} \left[\frac{\partial \ln f(z)}{\partial \ln z} \right] \leq 0.$$

□

Proof of Proposition 1:

Proof. Recall that the supply of locations with attractiveness γ in city c is

$$s(\gamma, c) = \begin{cases} \frac{1}{A(c)} V\left(\frac{\gamma}{A(c)}\right) & \text{if } \gamma \leq A(c)T(0) \\ 0 & \text{otherwise} \end{cases}$$

It is obvious that $\gamma > \gamma', c > c' \Rightarrow s(\gamma, c)s(\gamma', c') \geq s(\gamma, c')s(\gamma', c)$ is true when $\gamma > A(c')T(0)$. For $\gamma \leq A(c')T(0)$, the inequality holds true if and only if $V\left(\frac{\gamma}{A(c)}\right)$ is log-supermodular in (γ, c) . Note that $\frac{\gamma}{A}$ is submodular and log-modular in (γ, A) on \mathbb{R}_+^2 . Therefore, by lemma 8, $V\left(\frac{\gamma}{A}\right)$ is log-supermodular in (γ, A) on \mathbb{R}_+^2 if and only if $V(z)$ has a decreasing elasticity. Since $A(c)$ is increasing in c , $V\left(\frac{\gamma}{A(c)}\right)$ is log-supermodular in (γ, c) on $\mathbb{R}_+ \times \mathbb{C}$ if $V\left(\frac{\gamma}{A}\right)$ is log-supermodular in (γ, A) . Thus, $s(\gamma, c)$ is log-supermodular if $V(z)$ has a decreasing elasticity. □

Proof of Proposition 2:

Proof. $s(\gamma, c) \geq s(\gamma, c')$ is trivially true for $\gamma > A(c')T(0)$. For $\gamma \leq A(c')T(0)$,

$$s(\gamma, c) \geq s(\gamma, c') \Leftrightarrow \ln V\left(\frac{\gamma}{A(c)}\right) - \ln V\left(\frac{\gamma}{A(c')}\right) \geq \ln A(c) - \ln A(c')$$

This condition can be rewritten as

$$\int_{\ln A(c')}^{\ln A(c)} \left. \frac{-\partial \ln V(z)}{\partial \ln z} \right|_{z=\frac{\gamma}{x}} d \ln x \geq \int_{\ln A(c')}^{\ln A(c)} d \ln x$$

$$\int_{\ln A(c')}^{\ln A(c)} \left\{ \frac{-\partial \ln V(z)}{\partial \ln z} \right|_{z=\frac{\gamma}{x}} - 1 \right\} d \ln x \geq 0$$

Thus, a sufficient condition for the larger city to have more locations of attractiveness γ when $V(z)$ has a decreasing elasticity is $\frac{\partial \ln V(z)}{\partial \ln z} \leq -1$ at

$$z = \frac{\gamma}{A(c)}. \quad \square$$

Appendix C. Data

Data sources: Our metropolitan population data are from the US Census website (2000). Our data on individuals' demographics, educational attainments, geographic locations, and sectors of employment come from the 5% sample of the 2000 US Census and the 1% metro sample of the 1980 US Census made available by IPUMS-USA (Ruggles et al., 2010). Our data on industrial employment come from the 2000 County Business Patterns, available from the US Census Bureau website. Our data on occupational employment come from the 2000 Occupational Employment Statistics, available from the Bureau of Labor Statistics website.

Geography: We use (consolidated) metropolitan statistical areas as defined by the OMB as our geographic unit of analysis.

The smallest geographic unit in the IPUMS-USA microdata is the public-use microdata area (PUMA), which has a minimum of 100,000 residents. We map the PUMAs to metropolitan statistical areas (MSAs) using the MABLE Geocorr2K geographic correspondence engine from the Missouri Census Data Center. In some sparsely populated areas, a PUMA is larger than a metropolitan area. We drop six MSAs in which fewer than half of the residents of the only relevant PUMA live within the metropolitan area. As a result, there are 270 MSAs when we use these microdata.

The 1980 Census of Population IPUMS-USA microdata do not identify PUMAs, so we use the “metarea” variable describing 270 consolidated MSAs for the regressions in Table E.1.

The County Business Patterns data describe 318 metropolitan statistical areas. These correspond to a mix of OMB-defined primary and consolidated metropolitan statistical areas outside New England and New England county metropolitan areas (NECMAs). We aggregate these into OMB-defined (consolidated) metropolitan statistical areas to obtain 276 MSAs.

The Occupational Employment Statistics data describe 331 (primary) metropolitan statistical areas. We aggregate these into OMB-defined (consolidated) metropolitan statistical areas to obtain observations for 276 MSAs.

Skill distribution: Our sample of individuals in the IPUMS data includes those 25 and older in the labor force. We exclude individuals living in group quarters. Using the “educd” variable from IPUMS, we construct nine levels of educational attainment: less than high school (educd values 2–24), high school dropout (30–61), high school graduate (62), college dropout (65, 71), associate's degree (81), bachelor's degree (101), master's degree (114), professional degree (115), and doctorate (116). There is at least one observation in every educational category in every metropolitan area. In Appendix E, we report robustness checks using a narrower sample of IPUMS observations: full-time full-year (FTFY) workers, defined as individuals 25 and older who reported working at least 35 hours per week and 40 weeks in the previous year.

In Appendix E, we report robustness checks using aggregate tabulations from the Census of Population, Summary File 3, available from the US Census website. These tabulations are less noisy than the IPUMS observations, because they come from the 1-in-6 Census long form rather than the 1-in-20 public-use microdata. Because they are aggregate tabulations, we cannot condition on individual characteristics like labor-force participation or birthplace.

Sectoral skill intensity: Using the same sample of individuals 25 and older in the labor force, we measure a sector's skill intensity by calculating the average years of schooling of its employees after controlling for spatial differences in average schooling. We calculate years of schooling using the educational attainment “educd” variable from IPUMS at its finest level of disaggregation. For instance, this means that we distinguish between those whose highest educational attainment is sixth grade or eighth grade. We use the “indnaics” and “occSOC” variables to assign individuals to their 2-digit NAICS and 2-digit SOC sectors of employment. Aggregating observations to the MSA-sector level, weighted by the IPUMS-provided person weights, we regress the average years of schooling on MSA and sectoral dummies. The sectoral dummy coefficients are our measure of skill intensities.

Industrial employment: There are 19 2-digit NAICS industries covered by both the Census of Population microdata and the County Business Patterns data (the latter omits public administration, NAICS 92). The County Business Patterns data are an almost exhaustive account of US employer establishments. When necessary to protect the confidentiality of individual establishments, employment in an industry in a location is censored and reported as falling within an interval rather than its exact number. In our empirical work, we use the midpoints of these intervals as the level of employment. There are four (C)MSAs that have zero establishments in mining; the remaining 5240 (=19 × 276 – 4) industry-metropolitan pairs have at least one establishment. The County Business Patterns data omit self-employed individuals and employees of private households, railroads, agriculture production, the postal service, and public administrations. See the CBP methodology webpage for details.

Occupational employment: There are 22 2-digit SOC occupations. Across 331 (P)MSAs, there should be 7282 metropolitan-occupation observations. The 2000 BLS Occupational Employment Statistics contain employment estimates for 7129 metropolitan-occupation observations, none of which are zero. The 153 omitted observations “may be withheld from publication for a number of reasons, including failure to meet BLS quality standards or the need to protect the confidentiality of [BLS] survey respondents.”

Appendix D. Empirical tests

D.1. Population elasticities and pairwise comparisons

This section describes the relationship between our two empirical tests in more detail.

If $f(v, c)$ is log-supermodular and $f(v, c) > 0 \quad \forall v \forall c$,

$$v > v', c > c' \Rightarrow \ln f(v, c) + \ln f(v', c') \geq \ln f(v', c) + \ln f(v, c').$$

If C and C' are distinct sets and C is greater than C' ($\inf_{c \in C} L(c) > \sup_{c' \in C'} L(c')$) and n_C is the number of elements in C while $n_{C'}$ is the number of elements in C' , then log-supermodularity of $f(v, c)$ implies

$$\frac{1}{n_C} \sum_{c \in C} \ln f(v, c) + \frac{1}{n_{C'}} \sum_{c' \in C'} \ln f(v', c') \geq \frac{1}{n_C} \sum_{c \in C} \ln f(v', c) + \frac{1}{n_{C'}} \sum_{c' \in C'} \ln f(v, c') \quad \text{for } v > v'$$

Suppose that the world is noisy. Consider the following form for $f(v, c)$, which is a first-order approximation for any form,

$$\ln f(v, c) = \alpha_v + \beta_v \ln L_c + \varepsilon_{v,c}$$

where $\varepsilon_{v,c}$ is an error term with $\mathbb{E}(\varepsilon_{v,c}) = 0$. The probability of obtaining the expected inequality when $v > v', c > c'$ is

$$\begin{aligned} \mathcal{P} &= \Pr \left(\frac{1}{n_C} \sum_{c \in C} \ln f(v, c) + \frac{1}{n_{C'}} \sum_{c' \in C'} \ln f(v', c') \geq \frac{1}{n_C} \sum_{c \in C} \ln f(v', c) + \frac{1}{n_{C'}} \sum_{c' \in C'} \ln f(v, c') \right) \\ &= \Pr \left(\frac{1}{n_C} \sum_{c \in C} [\varepsilon_{v,c} - \varepsilon_{v',c}] + \frac{1}{n_{C'}} \sum_{c' \in C'} [\varepsilon_{v',c'} - \varepsilon_{v,c'}] \leq (\beta_v - \beta_{v'}) [\ln \bar{L}_C - \ln \bar{L}_{C'}] \right) \end{aligned}$$

where $\ln \bar{L}_{c'}$ denotes an average log population, $\ln \bar{L}_{c'} \equiv \frac{1}{n_{c'}} \sum_{c \in c'} \ln L_c$. This probability is higher when there is a larger difference in population size between the two bins and when the difference in population elasticities, $\beta_{\nu} - \beta_{\nu'}$, is larger. Since log-supermodularity implies that β_{ν} is increasing in ν , this probability is higher when the difference $\nu - \nu'$ is larger.

To illustrate the properties of this probability, consider the special case in which the error term is normally distributed, $\varepsilon_{\nu,c} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Then $\frac{1}{n_c} \sum_{c \in c} [\varepsilon_{\nu',c} - \varepsilon_{\nu,c}]$ is distributed $\mathcal{N}\left(0, \frac{2\sigma^2}{n_c}\right)$. Thus, the left side of the inequality inside the probability is distributed $\mathcal{N}\left(0, 2\left[\frac{1}{n_c} + \frac{1}{n_{c'}}\right]\sigma^2\right)$.

Therefore, we can write that the left side divided by $\sqrt{2\left[\frac{1}{n_c} + \frac{1}{n_{c'}}\right]\sigma^2}$ is distributed standard normal, $\mathcal{N}(0, 1)$. Thus, the probability of obtaining the expected inequality is

$$\mathcal{P} = \Omega\left(\frac{(\beta_{\nu} - \beta_{\nu'}) [\ln \bar{L}_c - \ln \bar{L}_{c'}]}{\sqrt{2\left[\frac{1}{n_c} + \frac{1}{n_{c'}}\right]\sigma^2}}\right)$$

where $\Omega(\cdot)$ denotes the cumulative distribution function of $\mathcal{N}(0, 1)$. If n_c equals $n_{c'}$, the expression for \mathcal{P} simplifies to

$$\mathcal{P} = \Omega\left(\frac{\sqrt{n_c}}{2\sqrt{\sigma^2}} \cdot (\beta_{\nu} - \beta_{\nu'}) \cdot [\ln \bar{L}_c - \ln \bar{L}_{c'}]\right).$$

The probability of obtaining the inequality depends on the difference in population size ($\ln \bar{L}_c - \ln \bar{L}_{c'}$), the difference in population elasticities ($\beta_{\nu} - \beta_{\nu'}$), the noisiness (σ^2) of the relationship, and the number of cities aggregated (n_c). When the deterministic function is log-supermodular ($c > c' \Rightarrow L_c \geq L_{c'}; \nu > \nu' \Rightarrow \beta_{\nu} \geq \beta_{\nu'}$), $\mathcal{P} \rightarrow 1$ as $\sigma^2 \rightarrow 0$ (and $\mathcal{P} \rightarrow \frac{1}{2}$ as $\sigma \rightarrow \infty$). When the function is log-modular, $\mathcal{P} \rightarrow \frac{1}{2}$ as $\sigma^2 \rightarrow 0$, and when the function is log-submodular, $\mathcal{P} \rightarrow 0$ as $\sigma^2 \rightarrow 0$. As n_c increases (as we aggregate cities into fewer bins), it becomes more likely that we obtain the expected inequality. However, using fewer bins also decreases the number of times that we evaluate this inequality, so we will tend to have lower power to reject the null hypothesis that $f(\nu, c)$ is log-modular.

Our finding that β_{ν} is increasing in ν when estimated in the population elasticity test implies that this pairwise comparison test will tend to have the correct inequality, and its success rate will increase with differences in city size and aggregation. The success of the elasticity test implies success of the pairwise comparison test (with aggregation) to the extent that the log-linear approximation of $f(\nu, c)$ is a good approximation. At the same time, $\sigma^2 \gg 0$, so we should not expect the pairwise comparison test to have a 100% success rate. We use an exact test to compute the probability of obtaining our observed success rates under the null hypothesis that the deterministic function is log-modular ($\beta_{\nu} = \beta \forall \nu$).

D.2. P-values for pairwise comparisons test

To assess the statistical significance of the fraction of these pairwise comparisons that yield the expected inequality, we compute the probability of obtaining a success rate at least as high as the observed success rate under the null hypothesis that skills and sectors are uniformly distributed across metropolitan areas. To do so, we employ a permutation test, shuffling the microdata observations to construct cities of the true population size with randomly assigned skill and sectoral distributions. For example, we shuffle the 4.4 million individuals in our IPUMS microdata sample to randomly assign them to metropolitan areas. We then perform the relevant pairwise comparisons test and record its success rate. We repeat this process 1000 times, yielding a cumulative distribution for the pairwise comparisons test's success rate under the null hypothesis. This distribution is centered around 0.5, and the p -value assigned to a success rate is the fraction of success rates in the cumulative distribution simulated under the null hypothesis that exceed the observed rate. When there are more observations (more bins, more educational categories, etc), the cumulative distribution exhibits less dispersion, yielding a more powerful test.

Appendix E. Empirical results

E.1. US-born and foreign-born skill distributions

Table 5 shows that more educated skill groups generally have higher population elasticities: the population elasticity of professionals exceeds that of bachelor's degree holders, which exceeds that of associate's degree holders, and so forth. In this appendix subsection, we investigate outliers that deviate from this broad pattern.

In particular, the population elasticity of those never reaching high school is 1.089. This skill group constitutes about 4% of the population and is overwhelmingly foreign-born. The second column of Table 5 reveals that its high population elasticity is attributable to the presence of foreign-born individuals with low educational attainment in larger cities.

How should we interpret the difference between the spatial distribution of skills among the population as a whole and among US-born individuals? One possibility is that immigrants strongly prefer larger cities for reasons omitted from our model, causing less-skilled foreign-born individuals to disproportionately locate in larger cities. This would be consistent with an established literature that describes agglomeration benefits particular to unskilled foreign-born individuals, such as linguistic enclaves (Edin et al., 2003; Bauer et al., 2005).⁶¹

Ecckhout et al. (2014) articulate another possibility, in which an economic mechanism they term "extreme-skill complementarity" causes less skilled individuals, foreign-born or US-born, to disproportionately reside in larger cities. Since this theory is silent with regard to birthplace, it predicts that in the absence of foreign-born low-skilled individuals, US-born low-skilled individuals would disproportionately locate in larger cities.

⁶¹ Another potential mechanism is that immigrants may find larger cities' combination of higher nominal wages and higher housing prices more attractive than natives (Diamond, 2016), possibly because they remit their nominal incomes abroad or demand less housing than US-born individuals. Albert and Monras (2018) study this mechanism in detail.

We attempt to distinguish between these hypotheses by looking at the skill distributions of US cities two decades earlier. In 2000, foreign-born individuals were 11% of the US population, while in 1980 they constituted about 6% (Gibson and Jung, 2006). More importantly, in 2000, foreign-born individuals constituted 71% of the least-skilled group, while in 1980 they were only 27%. If our hypothesis that less-skilled foreign-born individuals are particularly attracted to larger cities is correct, then the population elasticity of less-skilled types should be lower when foreign-born shares are lower. Table E.1 demonstrates that this is the case in 1980.⁶² It does not provide any evidence that the least skilled were overrepresented in larger cities in 1980, among either the population as a whole or US-born individuals. Reconciling these results with the birthplace-neutral model of Eeckhout et al. (2014) would require that the production function changed from top-skill complementarity in 1980 to extreme-skill complementarity in 2000.

The contrast in the least-skilled population elasticities between 1980 and 2000 for the population as a whole overwhelmingly reflects the increasing foreign-born share in the least-skilled groups. For foreign-born individuals with less than high school education, the population elasticities were 1.46 in 1980 and 1.43 in 2000. For US-born individuals, these population elasticities were 0.89 in 1980 and 0.86 in 2000. That is, these birthplace-specific elasticities hardly budged over twenty years. What changed was the foreign-born share: the vast majority of the difference in population elasticities for the less-than-high-school skill group in 1980 and 2000 is due to its increasingly foreign-born composition. If the total less-than-high-school population were its year-2000 size and exhibited the foreign-born and US-born population elasticities estimated for the year 2000, but the share of this group that was US-born were its 1980 share (73%) rather than its 2000 share (29%), the less than high school group's population elasticity would be 0.944. This is close to the population elasticity of 0.974 estimated using the 1980 data. Thus our model's poor prediction of the less-than-high-school population elasticity in 2000 is attributable to large cities being particularly attractive to foreign-born individuals.

Table E.1
Population elasticities of seven skill groups, 1980.

Dependent variable: $\ln f(\omega, c)$	(1)	(2)	Population share	Share
	All	US-born		US-born
β_{o1} Less than high school \times log population	0.974 (0.0244)	0.890 (0.0269)	0.09	0.73
β_{o2} High school dropout \times log population	1.017 (0.0165)	0.995 (0.0187)	0.13	0.92
β_{o3} Grade 12 \times log population	0.993 (0.0101)	0.974 (0.0119)	0.34	0.93
β_{o4} 1 year college \times log population	1.055 (0.0162)	1.041 (0.0166)	0.10	0.94
β_{o5} 2–3 years college \times log population	1.093 (0.0154)	1.074 (0.0155)	0.12	0.92
β_{o6} 4 years college \times log population	1.106 (0.0171)	1.088 (0.0177)	0.11	0.92
β_{o7} 5+ years college \times log population	1.134 (0.0216)	1.114 (0.0215)	0.11	0.90

NOTES: Standard errors, clustered by MSA, in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas in 1980.

E.2. Sectoral population elasticities

Tables E.2 and E.3 report the population elasticities estimates depicted in Figs. 1 and 2, along with the accompanying standard errors. Table E.3 also reports elasticities estimated using only city-industry employment levels that are not censored by being reported as falling within an interval in the County Business Patterns data.

Table E.2
Occupational employment population elasticities.

β_{o1} Farming, Fishing, and Forestry Occupations	0.807 (0.0470)	β_{o12} Sales and Related Occupations	1.035 (0.00917)
β_{o2} Building and Grounds Cleaning and Maintenance Occupations	1.038 (0.0103)	β_{o13} Management Occupations	1.080 (0.0143)
β_{o3} Food Preparation and Serving Related Occupations	0.984 (0.0104)	β_{o14} Arts, Design, Entertainment, Sports, and Media Occupations	1.157 (0.0189)
β_{o4} Construction and Extraction Occupations	1.035 (0.0138)	β_{o15} Business and Financial Operations Occupations	1.202 (0.0177)
β_{o5} Production Occupations	1.040 (0.0250)	β_{o16} Computer and Mathematical Occupations	1.393 (0.0335)
β_{o6} Transportation and Material Moving Occupations	1.058 (0.0135)	β_{o17} Architecture and Engineering Occupations	1.205 (0.0255)
β_{o7} Installation, Maintenance, and Repair Occupations	1.012 (0.0110)	β_{o18} Healthcare Practitioners and Technical Occupations	0.998 (0.0139)
β_{o8} Healthcare Support Occupations	0.977 (0.0130)	β_{o19} Community and Social Services Occupations	0.982 (0.0199)
β_{o9} Personal Care and Service Occupations	1.064 (0.0170)	β_{o20} Education, Training, and Library Occupations	1.010 (0.0168)
β_{o10} Office and Administrative Support Occupations	1.079 (0.00999)	β_{o21} Life, Physical, and Social Science Occupations	1.166 (0.0295)
β_{o11} Protective Service Occupations	1.120 (0.0140)	β_{o22} Legal Occupations	1.198 (0.0221)

NOTES: Standard errors, clustered by MSA, in parentheses.

⁶² The educational categories in Table 9 differ from prior tables because Census microdata collected prior to 1990 identify coarser levels of educational attainment in terms of years of schooling rather than highest degree attained.

Table E.3
Industrial employment population elasticities.

	(1)	(2)		(1)	(2)
β_{01} Forestry, fishing, hunting, and agriculture support	0.774 (0.0546)	0.660 (0.142)	β_{011} Utilities	0.999 (0.0338)	0.917 (0.0545)
β_{02} Accommodation and food services	0.992 (0.0125)	0.991 (0.0129)	β_{012} Arts, entertainment and recreation	1.130 (0.0220)	1.125 (0.0251)
β_{03} Construction	1.059 (0.0158)	1.057 (0.0160)	β_{013} Mining	0.829 (0.0621)	0.633 (0.122)
β_{04} Admin, support, waste mgt, remediation services	1.200 (0.0206)	1.207 (0.0195)	β_{014} Health care and social assistance	0.946 (0.0127)	0.946 (0.0127)
β_{05} Transportation and warehousing	1.178 (0.0253)	1.175 (0.0264)	β_{015} Finance and insurance	1.170 (0.0206)	1.167 (0.0215)
β_{06} Other services (except public administration)	1.032 (0.0109)	1.032 (0.0110)	β_{016} Information	1.171 (0.0216)	1.184 (0.0248)
β_{07} Retail trade	0.961 (0.00739)	0.962 (0.00713)	β_{017} Management of companies and enterprises	1.506 (0.0435)	1.406 (0.0469)
β_{08} Manufacturing	0.979 (0.0293)	0.973 (0.0296)	β_{018} Professional, scientific and technical services	1.263 (0.0174)	1.267 (0.0177)
β_{09} Wholesale trade	1.165 (0.0193)	1.159 (0.0190)	β_{019} Educational services	1.203 (0.0346)	1.205 (0.0391)
β_{010} Real estate and rental and leasing	1.162 (0.0142)	1.160 (0.0145)			
Observations	5240	4295	Observations	5240	4295
Only uncensored observations		Yes	Only uncensored observations		Yes

NOTES: Standard errors, clustered by MSA, in parentheses.

E.3. Skill distribution: alternative inclusion criteria

In the main text, we describe the skill distribution using labor-force participants 25 and older. Tables E.4 through E.6 demonstrate that we obtain similar results when using alternative inclusion criteria. First, we use a broader sample of individuals, counting everyone 25 and older regardless of labor-force participation. For this, we use aggregate tabulations of population over 25 by educational attainment from the Census of Population, Summary File 3. Second, we use a narrower sample of individuals, restricting attention to the population of full-time full-year (FTFY) workers 25 and older in the IPUMS microdata.

Table E.4
Skill groups by educational attainment.

	Summary File 3 population share	IPUMS FTFY population share	Share US-born	Skill (9 groups)	Summary File 3 population share	IPUMS FTFY population share	Share US-born
High school or less	0.46	0.35	0.79	Less than high school	0.07	0.03	0.27
				High school dropout	0.12	0.07	0.73
				High school graduate	0.27	0.24	0.88
Some college	0.28	0.32	0.89	College dropout	0.21	0.24	0.90
				Associate's degree	0.06	0.08	0.88
				Bachelor's degree	0.17	0.21	0.87
Bachelor's or more	0.27	0.34	0.85	Master's degree	0.06	0.08	0.84
				Professional degree	0.02	0.03	0.82
				Doctorate	0.01	0.01	0.72

NOTES: "Summary File 3" denotes 2000 Census of Population Summary File 3 tabulations of populations 25 and older by educational attainment for 270 metropolitan areas. "IPUMS FTFY" denotes full-time, full-year employees 25 and older residing in 270 metropolitan areas in IPUMS Census of Population microdata. The "share US-born" columns are computed using the IPUMS FTFY microdata.

Table E.5
Population elasticities of three skill groups.

Dependent variable: $\ln f(\omega, c)$	(1)	(2)	(3)
	SF3	IPUMS FTFY	IPUMS FTFY
	All	All	US-born
β_{01} High school or less \times log population	0.976 (0.00984)	0.955 (0.0110)	0.899 (0.0160)
β_{02} Some college \times log population	1.010 (0.00770)	0.999 (0.0105)	0.973 (0.0123)
β_{03} Bachelor's or more \times log population	1.093 (0.0125)	1.097 (0.0152)	1.068 (0.0163)

NOTES: Standard errors, clustered by MSA, in parentheses. Column 1 sample is Summary File 3 tabulations of populations by educational attainment. Column 2 sample is full-time, full-year employees 25 and older residing in 270 metropolitan areas in IPUMS Census of Population microdata. Column 3 is the column 2 sample restricted to individuals born in the United States.

Table E.6
Population elasticities of nine skill groups.

Dependent variable: $\ln f(\omega, c)$	(1)	(2)	(3)
	SF3	IPUMS FTFY	IPUMS FTFY
	All	All	US-born
$\beta_{\omega 1}$ Less than high school \times log population	1.026 (0.0221)	1.119 (0.0336)	0.877 (0.0270)
$\beta_{\omega 2}$ High school dropout \times log population	1.012 (0.0146)	1.014 (0.0156)	0.941 (0.0193)
$\beta_{\omega 3}$ High school graduate \times log population	0.953 (0.0111)	0.927 (0.0135)	0.893 (0.0168)
$\beta_{\omega 4}$ College dropout \times log population	1.010 (0.00838)	1.002 (0.0113)	0.977 (0.0131)
$\beta_{\omega 5}$ Associate's degree \times log population	1.013 (0.0121)	0.999 (0.0142)	0.968 (0.0155)
$\beta_{\omega 6}$ Bachelor's degree \times log population	1.096 (0.0119)	1.098 (0.0148)	1.070 (0.0165)
$\beta_{\omega 7}$ Master's degree \times log population	1.094 (0.0155)	1.117 (0.0178)	1.085 (0.0184)
$\beta_{\omega 8}$ Professional degree \times log population	1.115 (0.0139)	1.105 (0.0170)	1.076 (0.0180)
$\beta_{\omega 9}$ PhD \times log population	1.079 (0.0282)	1.080 (0.0334)	1.031 (0.0315)

NOTES: Standard errors, clustered by MSA, in parentheses. Column 1 sample is Summary File 3 tabulations of populations by educational attainment. Column 2 sample is full-time, full-year employees 25 and older residing in 270 metropolitan areas in IPUMS Census of Population microdata. Column 3 is the column 2 sample restricted to individuals born in the United States.

E.4. Skill distribution: inferring skills from wages

In the main text, we use educational attainment as a proxy for skill, assuming that the distribution of skills increases with educational attainment. In this section, we infer skills from nominal wages. The theoretically appropriate wage measure differs between the productivity and amenity interpretations of a location's desirability.

In the productivity interpretation of the desirability of a location, individuals of the same skill earn the same income, $K^{-1}(\omega)G(\omega)$, everywhere they locate in equilibrium. Since both locational assignments K and the skill component of income G are strictly increasing functions, there is a one-to-one mapping between wages and skill. Thus, if the skill distribution $f(\omega, c)$ is log-supermodular, the number of people earning a given wage in a city is also log-supermodular in the wage and city population size. If we divide individuals into twenty wage ventiles, there should be relatively more people in the higher wage ventiles in larger cities.

In the amenity interpretation of desirability, individuals of the same skill earn different incomes in different cities. As shown in Appendix A.2, an individual of skill ω in city c has a nominal income of $A(c)G(\omega)$, which obviously varies with city c . Thus, nominal wages alone cannot be used to infer an individual's skill. Within a city, the relative income of two skill types is independent of the city, $\frac{A(c)G(\omega)}{A(c)G(\omega')} = \frac{G(\omega)}{G(\omega')}$. In our model, the least skilled type ω is present in all cities. Thus, if we normalize all observed incomes by the least skilled type's income in each city, we can compare the distributions of these normalized incomes across cities. To do so empirically, we normalize wages in each city relative to the fifth percentile wage in that city and compute national ventiles of this normalized wage distribution. There should be relatively more people in the higher normalized-wage ventiles in larger cities.

Table E.7 reports the estimated population elasticities of these wage ventiles for the productivity and amenity interpretations for both the entire population and US-born individuals. These population elasticities are almost always monotonically increasing across wage ventiles. The sole statistically significant exception is the first ventile. The 5% of the population that earns the lowest nominal wages are overrepresented in larger cities relative to our model's predictions. But the overwhelming majority of the wage distribution exhibits log-supermodularity, in line with our theory.

Table E.7
Population elasticities of wage ventiles.

	(1)	(2)	(3)	(4)
	Productivity interpretation		Amenity interpretation	
	All	US-born	All	US-born
$\beta_{\omega 1}$ wage ventile 1 \times log population	0.898 (0.0171)	0.816 (0.0164)	1.013 (0.00843)	0.930 (0.0172)
$\beta_{\omega 2}$ wage ventile 2 \times log population	0.869 (0.0167)	0.780 (0.0180)	0.958 (0.0108)	0.875 (0.0204)
$\beta_{\omega 3}$ wage ventile 3 \times log population	0.868 (0.0133)	0.795 (0.0166)	0.951 (0.0121)	0.882 (0.0217)
$\beta_{\omega 4}$ wage ventile 4 \times log population	0.878 (0.0127)	0.814 (0.0182)	0.955 (0.0136)	0.899 (0.0214)
$\beta_{\omega 5}$ wage ventile 5 \times log population	0.916 (0.0111)	0.862 (0.0162)	0.962 (0.0154)	0.911 (0.0229)
$\beta_{\omega 6}$ wage ventile 6 \times log population	0.929 (0.0123)	0.881 (0.0176)	0.968 (0.0153)	0.925 (0.0216)
$\beta_{\omega 7}$ wage ventile 7 \times log population	0.945 (0.0120)	0.902 (0.0168)	0.982 (0.0163)	0.942 (0.0218)

Table E.7 (continued)

	(1)	(2)	(3)	(4)
	Productivity interpretation		Amenity interpretation	
	All	US-born	All	US-born
$\beta_{\omega 8}$ wage ventile 8 \times log population	0.975 (0.0111)	0.934 (0.0154)	0.997 (0.0145)	0.959 (0.0196)
$\beta_{\omega 9}$ wage ventile 9 \times log population	0.993 (0.0119)	0.958 (0.0160)	1.006 (0.0135)	0.971 (0.0181)
$\beta_{\omega 10}$ wage ventile 10 \times log population	1.009 (0.0110)	0.972 (0.0146)	1.014 (0.0121)	0.982 (0.0158)
$\beta_{\omega 11}$ wage ventile 11 \times log population	1.023 (0.0124)	0.992 (0.0157)	1.014 (0.0117)	0.984 (0.0149)
$\beta_{\omega 12}$ wage ventile 12 \times log population	1.035 (0.0110)	1.004 (0.0138)	1.020 (0.0125)	0.991 (0.0148)
$\beta_{\omega 13}$ wage ventile 13 \times log population	1.046 (0.0115)	1.020 (0.0140)	1.021 (0.0136)	0.994 (0.0164)
$\beta_{\omega 14}$ wage ventile 14 \times log population	1.062 (0.0116)	1.036 (0.0138)	1.035 (0.0123)	1.008 (0.0136)
$\beta_{\omega 15}$ wage ventile 15 \times log population	1.073 (0.0129)	1.046 (0.0152)	1.020 (0.0124)	0.994 (0.0133)
$\beta_{\omega 16}$ wage ventile 16 \times log population	1.103 (0.0134)	1.078 (0.0154)	1.040 (0.0114)	1.014 (0.0121)
$\beta_{\omega 17}$ wage ventile 17 \times log population	1.123 (0.0142)	1.097 (0.0155)	1.057 (0.0134)	1.031 (0.0136)
$\beta_{\omega 18}$ wage ventile 18 \times log population	1.178 (0.0155)	1.157 (0.0167)	1.076 (0.0151)	1.052 (0.0149)
$\beta_{\omega 19}$ wage ventile 19 \times log population	1.238 (0.0183)	1.215 (0.0192)	1.115 (0.0173)	1.091 (0.0165)
$\beta_{\omega 20}$ wage ventile 20 \times log population	1.268 (0.0175)	1.250 (0.0181)	1.163 (0.0165)	1.143 (0.0157)

NOTES: Standard errors, clustered by MSA, in parentheses. Sample is individuals 25 and older earning hourly wages greater than \$2/h residing in 270 metropolitan areas. Wages in columns 3 and 4 are normalized by city as described in the text.

E.5. Pairwise comparisons: alternative weighting schemes

In the main text, we report the success rate for the pairwise comparisons test as a weighted average of the share of pairwise comparisons that yield the predicted inequality. As shown in Appendix D, in the presence of idiosyncratic errors, the predicted inequality should hold with greater probability when there are larger differences in population size and skill intensity. There is more at stake when evaluating our model's prediction comparing the skill distributions of Chicago (population 9 million) and Des Moines (population 456 thousand) than its prediction comparing Des Moines and Kalamazoo (population 453 thousand). Nonetheless, we report success rates with alternative weighting schemes, including no weights at all. Since there is significant variation in differences in populations and skill intensities, as depicted in Fig. E.1, the unweighted success rates may differ substantially from those weighted by the relevant criteria.

Tables E.8 through E.11 present the pairwise comparisons for skill groups, occupations, and industries with a variety of weights for the same set of bins presented in the main text. The unweighted success rates show patterns similar to those of the weighted success rates in terms of birthplace and number of bins. The unweighted success rates are generally lower than the weighted success rates, demonstrating that our model's predictions are borne out in the data more frequently when the relevant differences in population size and skill intensity are larger. Consistent with the results of subsection 5.3, our theory's predictive successes are systematic and the empirical departures from our theory appear to be idiosyncratic. These results are highly statistically significant, confirming the model's predictive power.

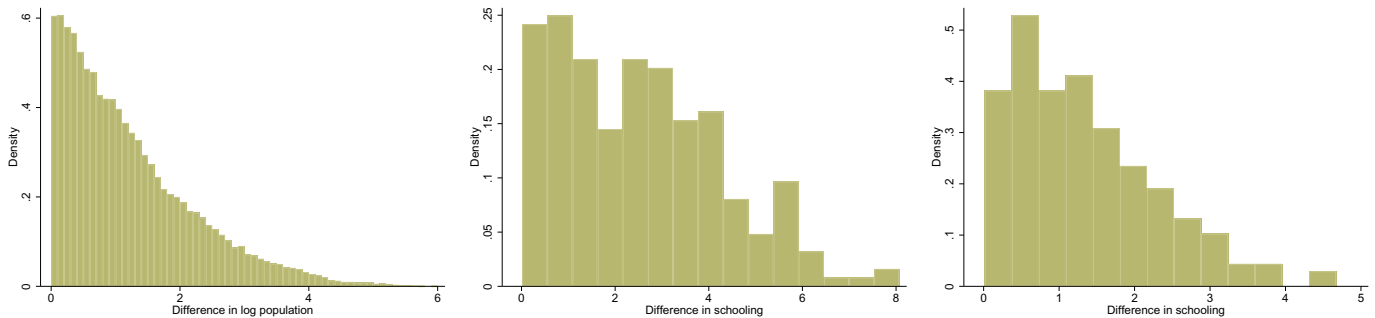


Fig. E.1. Differences in populations and skill intensities.

Table E8
Pairwise comparisons of three skill groups.

Pairwise comparisons of three skill groups						
Bins	Total comparisons	Unweighted		Pop-diff-weighted		
		All	US-born	All	US-born	
2	3	1.00 (0.17)	1.00 (0.17)	1.00 (0.17)	1.00 (0.17)	
3	9	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	
5	30	0.93 (0.01)	0.97 (0.00)	0.97 (0.00)	0.99 (0.00)	
10	135	0.77 (0.00)	0.81 (0.00)	0.89 (0.00)	0.91 (0.00)	
30	1305	0.70 (0.00)	0.74 (0.00)	0.82 (0.00)	0.86 (0.00)	
90	12,015	0.64 (0.00)	0.67 (0.00)	0.74 (0.00)	0.77 (0.00)	
270	108,945	0.60 (0.00)	0.61 (0.00)	0.67 (0.00)	0.69 (0.00)	

NOTES: P-values for uniform null hypothesis in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas.

Table E.9
Pairwise comparisons of nine skill groups.

Pairwise comparisons of nine skill groups							
Bins	Total comparisons	Unweighted		Pop-diff-weighted		Pop-diff x edu-share	
		All	US-born	All	US-born	All	US-born
2	36	0.61 (0.38)	0.86 (0.02)	0.61 (0.38)	0.86 (0.02)	0.75 (0.08)	0.90 (0.01)
3	108	0.58 (0.33)	0.80 (0.00)	0.61 (0.26)	0.83 (0.00)	0.75 (0.02)	0.88 (0.00)
5	360	0.58 (0.21)	0.76 (0.00)	0.62 (0.12)	0.80 (0.00)	0.76 (0.00)	0.85 (0.00)
10	1620	0.58 (0.07)	0.67 (0.00)	0.63 (0.02)	0.75 (0.00)	0.74 (0.00)	0.81 (0.00)
30	15,660	0.57 (0.01)	0.62 (0.00)	0.62 (0.00)	0.70 (0.00)	0.72 (0.00)	0.77 (0.00)
90	144,180	0.55 (0.00)	0.59 (0.00)	0.59 (0.00)	0.66 (0.00)	0.66 (0.00)	0.71 (0.00)
270	1,307,340	0.54 (0.00)	0.56 (0.00)	0.57 (0.00)	0.61 (0.00)	0.61 (0.00)	0.64 (0.00)

NOTES: P-values for uniform null hypothesis in parentheses. Sample is individuals 25 and older in the labor force residing in 270 metropolitan areas.

Table E.10
Pairwise comparisons of occupations.

Bins	Total comparisons	Unweighted success rate	Pop-diff weighted weighted success rate	Pop-diff x skill-diff weighted success rate
2	231	0.70 (0.02)	0.70 (0.02)	0.75 (0.03)
3	693	0.66 (0.00)	0.68 (0.00)	0.74 (0.00)
5	2310	0.64 (0.00)	0.67 (0.00)	0.71 (0.00)
10	10,395	0.59 (0.00)	0.62 (0.00)	0.65 (0.00)
30	100,485	0.56 (0.00)	0.60 (0.00)	0.63 (0.00)
90	925,155	0.55 (0.00)	0.59 (0.00)	0.62 (0.00)
276	8,073,382	0.54 (0.00)	0.55 (0.00)	0.57 (0.00)

NOTES: P-values for uniform null hypothesis in parentheses.

Table E.11

Pairwise comparisons of industries.

Bins	Total comparisons	Unweighted success rate	Pop-diff weighted weighted success rate	Pop-diff x skill-diff weighted success rate
2	171	0.65 (0.09)	0.65 (0.09)	0.77 (0.05)
3	513	0.64 (0.02)	0.64 (0.03)	0.74 (0.01)
5	1710	0.62 (0.00)	0.64 (0.00)	0.74 (0.00)
10	7695	0.59 (0.00)	0.62 (0.00)	0.71 (0.00)
30	74,385	0.57 (0.00)	0.60 (0.00)	0.66 (0.00)
90	684,855	0.56 (0.00)	0.59 (0.00)	0.64 (0.00)
276	6,469,758	0.54 (0.00)	0.57 (0.00)	0.61 (0.00)

Notes: P-values for uniform null hypothesis in parentheses.

References

- Abdel-Rahman, H.M., Anas, A., 2004. Theories of systems of cities. In: Henderson, J.V., Thisse, J.F. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2293–2339 Chap. 52.
- Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: implications for employment and earnings. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1043–1171.
- Albert, C., Monras, J., 2018. Immigration and spatial equilibrium: the role of expenditures in the country of origin. CEPR Discussion Papers 12842, C.E.P.R. Discussion Papers.
- Autor, D.H., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US labor market. *Am. Econ. Rev.* 103 (5), 1553–1597.
- Bacolod, M., Blum, B.S., Strange, W.C., 2009. Skills in the city. *J. Urban Econ.* 65 (2), 136–153.
- Bartik, T.J., 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute for Employment Research
- Bauer, T., Epstein, G., Gang, I., 2005. Enclaves, language, and the location choice of migrants. *J. Popul. Econ.* 18 (4), 649–662.
- Baum-Snow, N., Pavan, R., 2013. Inequality and city size. *Rev. Econ. Stat.* 95 (5), 1535–1548.
- Baum-Snow, N., Freedman, M., Pavan, R., 2017. Why has urban inequality increased? mimeo
- Behrens, K., Duranton, G., Robert-Nicoud, F., 2014. Productive cities: sorting, selection, and agglomeration. *J. Polit. Econ.* 122 (3), 507–553.
- Bernstein, J.R., Weinstein, D.E., 2002. Do endowments predict the location of production?: evidence from national and international data. *J. Int. Econ.* 56 (1), 55–76.
- Christaller, W., 1933. *Die zentralen Orte in Suddeutschland*. Gustav Fischer, Jena.
- Clemens, J., Gottlieb, J., Hemous, D., Olsen, M., 2018. The Spillover Effects of Top Income Inequality.
- Combes, P.-P., Duranton, G., Gobillon, L., 2012. The costs of agglomeration: land prices in French cities. CEPR Discussion Paper 9240.
- Costinot, A., 2009. An elementary theory of comparative advantage. *Econometrica* 77 (4), 1165–1192.
- Costinot, A., Vogel, J., 2015. Beyond Ricardo: assignment models in international trade. *Ann. Rev. Econ.* 7 (1).
- Costinot, A., Vogel, J., 2010. Matching and inequality in the world economy. *J. Polit. Econ.* 118 (4), 747–786.
- Couture, V., Gaubert, C., Handbury, J., Hurst, E., 2019. Income Growth and the Distributional Effects of Urban Spatial Sorting.
- Davis, D.R., Dingel, J.I., 2019. A spatial knowledge economy. *Am. Econ. Rev.* 109 (1), 153–170.
- Davis, D.R., Weinstein, D.E., 1999. Economic geography and regional production structure: an empirical investigation. *Eur. Econ. Rev.* 43 (2), 379–407.
- Diamond, R., 2016. The determinants and welfare implications of US Workers' diverging location choices by skill: 1980–2000. *Am. Econ. Rev.* 106 (3), 479–524.
- Duranton, G., Puga, D., 2005. From sectoral to functional urban specialisation. *J. Urban Econ.* 57 (2), 343–370.
- Duranton, G., Puga, D., 2015. Urban land use. In: Duranton, G., Henderson, J.V., Strange, W.S. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 5A. North-Holland, pp. 467–560 Chap. 8.
- Duranton, G., Overman, H.G., 2005. Testing for localization using micro-geographic data. *Rev. Econ. Stud.* 72 (4), 1077–1106.
- Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. *Am. Econ. Rev.* 91 (5), 1454–1477.
- Edin, P.-A., Fredriksson, P., Aslund, O., 2003. Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *Q. J. Econ.* 118 (1), 329–357.
- Eeckhout, J., Pinheiro, R., Schmidheiny, K., 2014. Spatial sorting. *J. Polit. Econ.* 122 (3), 554–620.
- Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. *J. Polit. Econ.* 105 (5), 889–927.
- Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *Am. Econ. Rev.* 100 (3), 1195–1213.
- Ethier, W., 1984. Higher dimensional issues in trade theory. In: Jones, R.W., Kenen, P.B. (Eds.), *Handbook of International Economics*, 1 edn Vol. 1. Elsevier, pp. 131–184 Chap. 03.
- Fujita, M., Thisse, J.-F., 2002. *Economics of Agglomeration*. Cambridge University Press.
- Gabszewicz, J.J., Shaked, A., Sutton, J., Thisse, J., 1981. International trade in differentiated products. *Int. Econ. Rev.* 22 (3), 527–534.
- Gibson, C., Jung, K., 2006. Historical census statistics on the foreign-born population of the United States: 1850–2000. Population Division Working Paper 81, US Census Bureau <http://www.census.gov/population/www/documentation/twps0081/twps0081.html>.
- Glaeser, E.L., 2008. *Cities, Agglomeration, and Spatial Equilibrium*, The Lindahl Lectures. Oxford University Press.
- Grossman, G.M., Rossi-Hansberg, E., 2008. Trading tasks: a simple theory of offshoring. *Am. Econ. Rev.* 98 (5), 1978–1997.
- Handbury, J., 2012. Are poor cities cheap for everyone? Non-homotheticity and the cost of living across U.S. cities. mimeo.
- Helpman, E., 1998. The size of regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), *Topics in Public Economics: Theoretical and Applied Analysis*. Cambridge University Press Cambridge, pp. 33–54.
- Helsley, R.W., Strange, W.C., 2014. Coagglomeration, clusters, and the scale and composition of cities. *J. Polit. Econ.* 122 (5), 1064–1093.
- Helsley, R.W., Strange, W.C., 1990. Matching and agglomeration economies in a system of cities. *Reg. Sci. Urban Econ.* 20 (2), 189–212.
- Henderson, J.V., 1974. The sizes and types of cities. *Am. Econ. Rev.* 64 (4), 640–656.
- Henderson, J.V., 1983. Industrial bases and City sizes. *Am. Econ. Rev.* 73 (2), 164–168.
- Henderson, J.V., 1987. General equilibrium modeling of systems of cities. In: Mills, E.S. (Ed.), *Handbook of Regional and Urban Economics*. Vol. 2. Elsevier, pp. 927–956 Chap. 23.
- Henderson, J.V., 1991. *Urban Development: Theory, Fact, and Illusion*. Oxford University Press.
- Henderson, J.V., 1997. Medium size cities. *Reg. Sci. Urban Econ.* 27 (6), 583–612.
- Hendricks, L., 2011. The skill composition of U.S. cities. *Int. Econ. Rev.* 52 (1), 1–32.
- Holmes, T.J., Stevens, J.J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J.V., Thisse, J.F. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2797–2843 Chap. 63.
- Hsu, W.-T., Holmes, T.J., Morgan, F., 2014. Optimal city hierarchy: a dynamic programming approach to central place theory. *J. Econ. Theory* 154 (0), 245–273.
- Jones, R.W., Scheinkman, J.A., 1977. The relevance of the two-sector production model in trade theory. *J. Polit. Econ.* 85 (5), 909–935.
- Leamer, E.E., 1984. *Sources of International Comparative Advantage: Theory and Evidence*. MIT Press.
- Lucas, R.E., 1988. On the mechanics of economic development. *J. Monet. Econ.* 22 (1), 3–42.
- Notowidigdo, M.J., 2013. The incidence of local labor demand shocks. mimeo.
- Nunn, N., 2007. Relationship-specificity, incomplete contracts, and the pattern of trade. *Q. J. Econ.* 122 (2), 569–600.
- Romalís, J., 2004. Factor proportions and the structure of commodity trade. *Am. Econ. Rev.* 94 (1), 67–97.
- Rossi-Hansberg, E., Sarte, P.-D., Owens, R., 2010. Housing externalities. *J. Polit. Econ.* 118 (3), 485–535.
- Ruggles, S., Alexander, J.T., Genadek, K., Goeken, R., Schroeder, M.B., Sobek, M., 2010. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minnesota Population Center, Minneapolis, MN.
- Samuelson, P.A., 1948. International trade and the equalisation of factor prices. *Econ. J.* 58 (230), 163–184.
- Sattinger, M., 1978. Comparative advantage in individuals. *Rev. Econ. Stat.* 60 (2), 259–267.
- Sattinger, M., 1979. Differential rents and the distribution of earnings. *Oxf. Econ. Pap.* 31 (1), 60–71.
- Sattinger, M., 1993. Assignment models of the distribution of earnings. *J. Econ. Lit.* 31 (2), 831–880.
- Yagan, D., 2014. Moving to opportunity? Migratory insurance over the great recession. mimeo.