

The Comparative Advantage of Cities*

Donald R. Davis[†]

Jonathan I. Dingel[‡]

Columbia University and NBER

Chicago Booth and NBER

March 27, 2017

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.

JEL Classification: F11, F14, R12, R13

*Thanks to 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 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.

[†]drdavis@columbia.edu

[‡]jdingel@chicagobooth.edu

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 *ex ante* homogeneous, 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 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 heterogeneity across locations 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 and 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 spatial 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 spatial pattern of skills and sectors. 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

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

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. The relative price of skill varies across cities, as documented by Baum-Snow and Pavan (2013) and Davis and Dingel (2012), 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 Behrens, Duranton, and Robert-Nicoud (2014) and Davis and Dingel (2012).

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, Duranton, and Robert-Nicoud, 2014). Our novel integration is to describe multiple cities with internal geographies when individuals’ valuations of locations within cities differ systematically. 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

²A 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.

³Reconciling 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).

⁴Both Baum-Snow, Freedman, and Pavan (2014) 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.

description of the economic problem and the resulting predicted cross-city skill distributions. 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 introduces the realism needed so that individuals of a given skill level are found in many cities in equilibrium.

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 section 3.5. 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

⁵Sattinger (1993) and Costinot and Vogel (2015) survey the assignment literature.

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.

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).⁶ 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.⁷ In a complementary class of models, cities produce a single composite output, which may be interpreted as perfect diversification 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

⁶An 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, Holmes, and Morgan (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.

⁷Though 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."

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.⁸

In addition to sectoral composition, our theory describes sectoral size. Theories of localization and urbanization economies have contrasting predictions for cities' absolute employment levels.⁹ 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, in urbanization models with a composite output, every sector is (implicitly) larger in larger cities, but there is no information 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, Glaeser, and Kerr, 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.

⁸A 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.

⁹The 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.

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.

3 Model

We develop a general-equilibrium model in which L heterogeneous individuals choose a city, a location within that city, and a sector in which to produce. There are C discrete cities ($c \in \mathbb{C} = \{1, \dots, C\}$), a continuum of skills, and a continuum of sectors. We study the consequences for city total factor productivity and the cross-city distributions of skills and sectors.

3.1 Preferences, production, and places

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 $r(c, \tau)$.

Final-goods producers have a CES production function

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

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

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

Heterogeneous individuals use their labor to produce intermediate goods. There is a mass of L heterogeneous individuals with skills ω that have the cumulative distribution function $F(\omega)$ and density $f(\omega)$ on support $\Omega \equiv [\underline{\omega}, \bar{\omega}]$. The productivity of an individual of skill ω

in sector σ at location τ in city c is

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

$A(c)$ denotes city-level total factor productivity, which results from agglomeration and is taken as given by individuals. $T(\tau)$ reflects the productivity effects of location within the city.¹⁰ We assume that $T(\tau)$ is continuously differentiable and $T'(\tau) < 0$, which is just a normalization that higher- τ locations are less desirable. We assume that the twice-differentiable function $H(\omega, \sigma)$ is strictly log-supermodular in ω and σ and strictly increasing in ω .¹¹ The former governs comparative advantage, so that higher- ω individuals are relatively more productive in higher- σ sectors.¹² 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. The supply of locations with innate desirability of at least τ is $S(\tau)$.¹³ 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.¹⁴ 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

¹⁰As written, $T(\tau)$ indexes the innate desirability of the location for its productive advantages, but a closely related specification makes $T(\tau)$ describe a location's desirability for its consumption value. The production and consumption interpretations yield very similar results but differ slightly in functional form. For expositional clarity, we use the production interpretation given by equations (3) and (4) in describing the model in the main text and present the consumption interpretation in appendix A.1.

¹¹In \mathbb{R}^2 , a function $H(\omega, \sigma)$ is strictly log-supermodular if $\omega > \omega', \sigma > \sigma' \Rightarrow H(\omega, \sigma)H(\omega', \sigma') > H(\omega, \sigma')H(\omega', \sigma)$.

¹²We refer to higher- ω individuals as more skilled and higher- σ sectors as more skill-intensive.

¹³For example, in the canonical monocentric-city model, τ describes physical distance from the central business district and the supply is $S(\tau) = \pi\tau^2$. 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.3 endogenizes it.

¹⁴Models 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, Sarte, and Owens (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.

$r(c, \tau) \geq 0$.

Individuals choose their city c , location τ , and sector σ to maximize utility. An individual's utility depends on their consumption of the numeraire final good, which is their income after paying their locational cost:

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

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. $A(c)$ 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 $L \times f(\omega, c) \equiv L \times \int_{\sigma \in \Sigma} \int_{\tau \in \mathcal{T}} f(\omega, c, \tau, \sigma) d\tau d\sigma$. Total factor productivity is

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

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

$$f(\omega, c, \tau, \sigma) > 0 \iff \{c, \tau, \sigma\} \in \arg \max U(c, \tau, \sigma; \omega). \quad (6)$$

Profit maximization by final-good producers yields demands for intermediates

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

where $I \equiv L \sum_c \int_{\sigma} \int_{\omega} \int_{\tau} q(\omega, c, \tau, \sigma) p(\sigma) f(\omega, c, \tau, \sigma) d\tau d\omega d\sigma$ 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,

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

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.

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

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

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

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

3.3 An autarkic city

We begin by considering a single city, denoted c , with exogenous population $L(c)$ and skill distribution $F(\omega)$. With fixed population, autarky TFP is fixed by equation (5). We describe individuals' choices of sectors and locations to solve for the autarkic equilibrium.

To solve, we exploit the fact that locational and sectoral argument enters individuals' utility functions separably. Individuals' choices of their sectors are independent of their locational decisions:

$$\arg \max_{\sigma} A(c)T(\tau)H(\omega, \sigma)p(\sigma) - r(c, \tau) = \arg \max_{\sigma} H(\omega, \sigma)p(\sigma)$$

Define the assignment function $M(\omega) = \arg \max_{\sigma} H(\omega, \sigma)p(\sigma)$ so that we can write $G(\omega) \equiv H(\omega, M(\omega))p(M(\omega))$. By comparative advantage, $M(\omega)$ is increasing.¹⁵ By absolute advantage, more skilled individuals earn higher nominal incomes and $G(\omega)$ is a strictly increasing function.¹⁶

¹⁵Lemma 1 of Costinot and Vogel (2010) shows that $M(\omega)$ is continuous and strictly increasing in equilibrium.

¹⁶Absolute advantage across all sectors is not necessary. The weaker condition that productivity is increasing in skill at the equilibrium assignments, $\frac{\partial}{\partial \omega} H(\omega, M(\omega)) > 0$, would be sufficient.

Individuals' choices of their locations are related to their sectoral decisions in the sense that willingness to pay for more desirable locations depends on the skill component of income $G(\omega)$. Within the city, individuals choose their optimal location:

$$\max_{\tau} A(c)T(\tau)G(\omega) - r(c, \tau)$$

Competition among landlords ensures that the most desirable locations are those occupied, so the least desirable occupied site $\bar{\tau}(c) \equiv \max_{\tau} \{\tau : f(\omega, c, \tau, \sigma) > 0\}$ in a city of population $L(c)$ is defined by $L(c) = S(\bar{\tau}(c))$. The set of occupied locations is $\bar{\mathcal{T}}(c) \equiv [0, \bar{\tau}(c)]$. More desirable locations have higher rental prices.

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

Individuals of higher skill have greater willingness to pay for more desirable locations. That is, $\frac{\partial^2}{\partial \tau \partial \omega} A(c)T(\tau)G(\omega) < 0$ because locational advantages complement individual productivity. As a result, in equilibrium higher- ω individuals occupy lower- τ locations.

Lemma 2 (Autarky locational assignments). *In autarkic equilibrium, there exists a continuous and strictly decreasing locational assignment function $N : \bar{\mathcal{T}}(c) \rightarrow \Omega$ such that $f(\omega, c, \tau, \sigma) > 0 \iff N(\tau) = \omega$, $N(0) = \bar{\omega}$ and $N(\bar{\tau}(c)) = \underline{\omega}$.*

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

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

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 $r(c, \bar{\tau}(c)) = 0$.

Lemma 3 (Autarky locational prices). *In autarkic equilibrium, $r(c, \tau)$ is continuously differentiable on $\tau \geq 0$ and given by $r(c, \tau) = -A(c) \int_{\tau}^{\bar{\tau}(c)} T'(t)G(N(t))dt$ for $\tau \leq \bar{\tau}(c)$.*

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

3.4 A system of cities

The previous section described a single city with an exogenous population. We now describe a system of cities in which these populations are endogenously determined in spatial equilibrium. A crucial element is that cities' TFPs will differ in equilibrium. We have already specified a process in equation (5) by which a city that is larger and more skilled (in likelihood ratio terms) will have higher TFP. We now need to complete the circle by showing that if one city has higher TFP than another, then its population will be larger and more skilled. Without loss of generality, order and label the cities from 1 to C so that $A(C) \geq A(C-1) \geq \dots \geq A(2) \geq A(1)$. In autarky, τ was a sufficient statistic for the attractiveness of a location. In a system of cities, we must clearly distinguish between a location's *attractiveness* and its *innate desirability*. Individuals care about a location's attractiveness, which we denote by γ . This depends both on city-level TFP and a location's innate desirability within the city.

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

Cities with higher TFP have larger populations. Consider two cities, c and c' , that differ in productivity, with $A(c) > A(c')$. The city with greater TFP will have greater population, $L(c) > L(c')$. If it did not, the least desirable occupied location in city c would be more desirable than the least desirable occupied location in city c' , $\bar{\tau}(c) \leq \bar{\tau}(c')$, since the supply of locations, $S(\tau)$, is common across cities. Since TFP is also higher in c , this would make the least attractive occupied location in city c more attractive than the least attractive occupied location in city c' , $A(c)T(\bar{\tau}(c)) > A(c')T(\bar{\tau}(c'))$. In equilibrium, the least desirable occupied location in each city has a price of zero, $r(c, \bar{\tau}(c)) = r(c', \bar{\tau}(c')) = 0$, by lemma 1. In that case, every individual would agree that living in c at $\bar{\tau}(c)$ is strictly better than living in c' at $\bar{\tau}(c')$ (because $A(c)T(\bar{\tau}(c))G(\omega) > A(c')T(\bar{\tau}(c'))G(\omega)$), which contradicts 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' . 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 $r_\Gamma(\gamma)$.

To characterize locational assignments and prices in the system of cities, we first characterize assignments and prices in terms of γ . The solution is analogous to that derived in the autarkic case. We then translate these assignments and prices into functions of c and τ .

More skilled individuals occupy more attractive locations. 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 4 (Locational assignments). *In equilibrium, there exists a continuous and strictly increasing locational assignment function $K : \Gamma \rightarrow \Omega$ such that (i) $f(\omega, c, \tau, M(\omega)) > 0 \iff A(c)T(\tau) = \gamma$ 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 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 4 allow us to say that $S_\Gamma(\gamma) = L \int_\gamma^{\bar{\gamma}} f(K(x))K'(x)dx$, 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 5 (Locational prices). *In equilibrium, the rent schedule $r_\Gamma(\gamma)$ is increasing and continuously differentiable on $[\underline{\gamma}, \bar{\gamma}]$ and given by $r_\Gamma(\gamma) = \int_\gamma^{\bar{\gamma}} G(K(x))dx$.*

Therefore, the determination of locational assignments and prices within the system of cities is analogous to determining these locational assignments and prices for an autarkic city with a supply of locations that is the sum of locations across the system of cities.

3.5 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 the rental price of a location depends only on its attractiveness, which is the product of city TFP and innate desirability, the rental price of a location with innate desirability τ in city c is $r(c, \tau) = r_\Gamma(A(c)T(\tau))$.¹⁷ In any city c , the supply of locations with attractiveness γ , and thus the set of locations with rental price $r_\Gamma(\gamma)$, is

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

where $V(z) \equiv -\frac{\partial}{\partial z} S(T^{-1}(z))$ is the supply of locations with innate desirability $T^{-1}(z)$.

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

Lemma 6 (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 section 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 7 (A city's sectoral employment distribution). *The population of individuals em-*

¹⁷This should also make clear why we get incomplete sorting across cities in our model. The value of a location to any individual ω (and thus its rent) depends only on the product of city TFP and innate desirability, so will be the same for a given ω when comparing cities c and c' and locations τ and τ' just when $A(c)T(\tau) = A(c')T(\tau')$. Hence an individual ω will be indifferent between these locations in different cities.

ployed 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. Equation (12) demonstrates a hierarchy of locational attractiveness, since the most attractive locations are found exclusively in the highest-TFP city. Amongst levels of attractiveness that are supplied in multiple cities, equation (12) shows that cities' TFPs shape the supply schedule $s(\gamma, c)$ through both a scaling effect ($\frac{1}{A(c)}$) and a dilation of $V(\frac{\gamma}{A(c)})$. Comparisons of relative supplies ($s(\gamma, c)s(\gamma', c') \stackrel{?}{>} s(\gamma', c)s(\gamma, c')$) depend only on the dilation.

Our main result, Proposition 1, is a sufficient condition for the ordering of city TFPs to govern the ordering of locational supplies in any equilibrium.

Proposition 1 (Locational attractiveness distribution). *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.*

Proposition 1 links our assumption about each city's distribution of locations, $V(z)$, to endogenous equilibrium locational characteristics, $s(\gamma, c)$. Its proof is in appendix B.¹⁸ 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. Equation (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

¹⁸Lemma 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 $A(c)$.

$V(z)$ schedules satisfy this decreasing-elasticity condition.¹⁹²⁰

The distributions of skills and sectoral employment across cities follow from Proposition 1. The skill distribution follows immediately through the locational assignment function, as shown by Lemma 6. The employment distribution follows in turn through the sectoral assignment function, as shown by Lemma 7. 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.

Corollary 1 (Skill and employment distributions). *If $V(z)$ has a decreasing elasticity, then $f(\omega, c)$ and $f(\sigma, c)$ are log-supermodular.*

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 are endogenously determined and there is within-city heterogeneity in productivity.²¹ 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.²² 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 Corollary 2, which characterizes cities' sectoral outputs.

Corollary 2 (Output and revenue distributions). *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.*

¹⁹For 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. The canonical monocentric city model also satisfies the condition of Proposition 1. See Appendix A.2 for details.

²⁰If $V(z)$ does not have a decreasing elasticity, the model still predicts a hierarchy of locational attractiveness, as shown in equation 12, and imperfect sorting.

²¹In particular, our Corollary 2 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. Corollary 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.

²²In the production 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 consumption interpretation of $T(\tau)$ described in Appendix A.1, occupied locations' productivities $q(c, \tau, \sigma; \omega) = A(c)H(\omega, \sigma)$ differ across cities in a Hicks-neutral fashion.

These results characterize the pattern of comparative advantage across cities.²³

When does the more productive city have a larger population of every skill type? By lemma 6, 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. Corollary 3 applies this result to the least-attractive locations, thereby identifying a sufficient condition for larger cities to have larger populations of all skill types and therefore employ more people in every sector.

Proposition 2. *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 $s(\gamma, c) \geq s(\gamma, c')$.*

Corollary 3. *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 and their corollaries 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

We examine the predictions of our model using two approaches to characterize the outcomes described by Corollary 1.²⁴ 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.

²³A 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).

²⁴Proposition 1 and Corollary 2 make predictions about other economic outcomes, such as rental prices and sectoral outputs, that are difficult to empirically characterize due to data constraints. For example, data on occupations describe employment levels, not occupational output or revenue. Similarly, examining stochastic or likelihood-ratio dominance in rental prices across cities would require representative samples of (unimproved) land prices, which are not available. The available evidence on urban costs, which shows that the maximum, mean, and range of unimproved land prices are greater in larger cities (Combes, Duranton, and Gobillon, 2012), are consistent with our model's predictions.

These tests requires 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.

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.²⁵ Educational attainment is a coarse measure, but it is the best measure available in data describing a large number of people across detailed geographic locations. 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 percent of the population) to those with doctoral degrees (1 percent). Table 1 shows the population shares and percentage US-born for each of these skill groups in 2000.²⁶ Foreign-born individuals are disproportionately in the tails of the educational distribution.

²⁵Costinot and Vogel (2010) 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 characteristics.

²⁶This table describes labor-force participants 25 and older. See appendix E for similar results using other inclusion criteria.

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 | .37 | .78 | Less than high school | .04 | .29 |
| | | | High school dropout | .08 | .73 |
| | | | High school graduate | .25 | .88 |
| Some college | .31 | .89 | College dropout | .23 | .89 |
| | | | Associate’s degree | .08 | .87 |
| Bachelor’s or more | .32 | .85 | Bachelor’s degree | .20 | .86 |
| | | | Master’s degree | .08 | .84 |
| | | | Professional degree | .03 | .81 |
| | | | Doctorate | .01 | .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

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).

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.

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

Table 2: Sectoral skill intensities

| SOC | Occupational category | Skill intensity | NAICS | Industry | Skill intensity |
|-----|--------------------------------------|-----------------|-------|--|-----------------|
| 45 | Farming, Fishing & Forestry | 8.7 | 11 | Forestry, fishing, hunting & agriculture support | 10.5 |
| 37 | Cleaning & Maintenance | 10.8 | 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 | 14.6 |
| 19 | Life, Physical & Social Science | 17.2 | 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

intensity.²⁷ 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.²⁸

4.4 Empirical tests

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

$$\frac{1}{n_{\mathcal{C}}} \sum_{c \in \mathcal{C}} \ln f(\nu, c) + \frac{1}{n_{\mathcal{C}'}} \sum_{c' \in \mathcal{C}'} \ln f(\nu', c') \geq \frac{1}{n_{\mathcal{C}}} \sum_{c \in \mathcal{C}} \ln f(\nu', c) + \frac{1}{n_{\mathcal{C}'}} \sum_{c' \in \mathcal{C}'} \ln f(\nu, c') \quad \forall \nu > \nu'.$$

²⁷Autor 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.

²⁸The standard deviations of average years of schooling across occupational and industrial categories are 2.3 and 1.2, respectively.

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 .²⁹ 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).³⁰ Standard econometric tests are available to assess whether our estimated population elasticities exhibit the property that $\nu \geq \nu' \Rightarrow \beta_\nu \geq \beta_{\nu'}$.

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.³¹ When comparing narrow educational categories, we also weight by the product of the educational categories’ population shares.³² 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.³³

²⁹The linear regression may be understood as a first-order Taylor approximation: $\ln f(\nu, c) \approx \ln f(\nu, c^*) + \frac{\partial \ln f(\nu, c)}{\partial \ln L(c)} \Big|_{c=c^*} (\ln L(c) - \ln L(c^*)) + \epsilon = \alpha_\nu + \beta_\nu \ln L(c) + \epsilon_{\nu, c}$, where $\beta_\nu = \frac{\partial \ln f(\nu, c)}{\partial \ln L(c)} \Big|_{c=c^*}$ is increasing in ν by log-supermodularity of $f(\nu, c)$.

³⁰Henderson (1983) regresses employment shares on population levels, but reports “percent Δ share / percent Δ population”, which is equal to $\beta_\sigma - 1$ in our notation. Similarly, Holmes and Stevens (2004) 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 $\beta_\sigma - 1$.

³¹We 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.

³²An unweighted average, reported in appendix E, treats comparisons involving high school graduates (25 percent of the workforce) and those involving PhDs (1 percent) equally, while these differ in their economic import.

³³Appendix D provides details of a permutation test in which we shuffle our observations 1,000 times to compute the distribution of success rates under the null. Power to reject the null decreases as the number

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

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 spatial pattern of sectoral employment is strongly ordered by this spatial 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 novel 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 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 section 4.4. We first

of bins decreases.

Table 3: Population elasticities of three skill groups

| Dependent variable: $\ln f(\omega, c)$ | (1) All | (2) US-born | Population share | Share US-born |
|--|-------------------|-------------------|---------------------|------------------|
| β_{ω_1} High school or less \times log population | 0.954 (0.0108) | 0.895 (0.0153) | .37 | .78 |
| β_{ω_2} Some college \times log population | 0.996 (0.0105) | 0.969 (0.0122) | .31 | .89 |
| β_{ω_3} Bachelor's or more \times log population | 1.086 (0.0153) | 1.057 (0.0162) | .32 | .85 |

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

do these exercises using three skill groups defined by educational attainment levels and then repeat them using nine very disaggregated skill groups.

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.³⁴ 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 section 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.³⁵ Note that prior work typically describes a

³⁴Younger 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.

³⁵With n city bins and m skill groups, we make $\frac{n(n-1)}{2} \frac{m(m-1)}{2}$ comparisons. For example, $\frac{270 \times 269}{2} \frac{3 \times 2}{2} =$

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 percent, 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 percent.³⁶ These results are all statistically significant at the 1-percent level, except for the 2-bin case that involves only three comparisons, severely reducing the test’s power.³⁷

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.

108,945.

³⁶Our comparisons of two or five bins of cities are analogous to the empirical exercises presented in Eeckhout, Pinheiro, and Schmidheiny (2014) and Bacolod, Blum, and Strange (2009).

³⁷Under 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 1,000 simulations of the three comparisons under the null yield three correct comparisons in 17.8% of simulations, bounding the p-value.

Table 4: Pairwise comparisons of three skill groups

| Bins | Total comparisons | Success rate All | Success rate US-born |
|------|-------------------|------------------|----------------------|
| 270 | 108,945 | .67 (.00) | .69 (.00) |
| 90 | 12,015 | .74 (.00) | .77 (.00) |
| 30 | 1,305 | .82 (.00) | .86 (.00) |
| 10 | 135 | .89 (.00) | .91 (.00) |
| 5 | 30 | .97 (.00) | .99 (.00) |
| 3 | 9 | 1.00 (.01) | 1.00 (.01) |
| 2 | 3 | 1.00 (.18) | 1.00 (.18) |

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.

5.1.2 Nine skill groups

We next examine our tests for the case with nine skill groups. 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 percent 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 one percent of US workers possess a doctoral degree, so their spatial distribution may have limited impact on sectoral employment patterns and aggregate outcomes. The lower-than-expected population elasticity reflects the spatial distribution of institutions of higher education. The 27 percent 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

Table 5: Population elasticities of nine skill groups

| Dependent variable: $\ln f(\omega, c)$ | (1) All | (2) US-born | Population share | Share US-born |
|--|-------------------|-------------------|---------------------|------------------|
| β_{ω_1} Less than high school \times log population | 1.089 (0.0314) | 0.858 (0.0239) | .04 | .29 |
| β_{ω_2} High school dropout \times log population | 1.005 (0.0152) | 0.933 (0.0181) | .08 | .73 |
| β_{ω_3} High school graduate \times log population | 0.925 (0.0132) | 0.890 (0.0163) | .25 | .88 |
| β_{ω_4} College dropout \times log population | 0.997 (0.0111) | 0.971 (0.0128) | .23 | .89 |
| β_{ω_5} Associate's degree \times log population | 0.997 (0.0146) | 0.965 (0.0157) | .08 | .87 |
| β_{ω_6} Bachelor's degree \times log population | 1.087 (0.0149) | 1.059 (0.0164) | .20 | .86 |
| β_{ω_7} Master's degree \times log population | 1.095 (0.0179) | 1.063 (0.0181) | .08 | .84 |
| β_{ω_8} Professional degree \times log population | 1.113 (0.0168) | 1.082 (0.0178) | .03 | .81 |
| β_{ω_9} PhD \times log population | 1.069 (0.0321) | 1.021 (0.0303) | .01 | .72 |

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

economies, the spatial distribution of PhDs is well predicted by our theory.³⁸

An examination of those with less than a high-school education, who account for 12 percent of the working population, suggests that their higher-than-expected elasticities likely reflect international migrants' particular preferences for large US cities for reasons outside of our model. First, immigrants account for 71 percent of the less-than-high-school group, the largest outlier. Second, 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.³⁹ Finally, an alternative to restricting attention to US-born individuals is to estimate the population elasticities using data from 1980, a time when immigrants were a much smaller share of the US population. In 1980 data, the least skilled group has the lowest

³⁸About 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.

³⁹Interestingly, among US-born individuals, the nine estimated elasticities naturally break into the three more aggregate educational attainment categories that we used: $\beta_{\omega_1}, \beta_{\omega_2}, \beta_{\omega_3} \in (0.85, 0.94)$; $\beta_{\omega_4}, \beta_{\omega_5} \in (0.96, 0.98)$; $\beta_{\omega_6}, \beta_{\omega_7}, \beta_{\omega_8}, \beta_{\omega_9} \in (1.02, 1.09)$.

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 these issues further in appendix E.1.⁴⁰

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 3,000 residents larger. When making pairwise comparisons across all 270 metropolitan areas, the success rates are 61 and 64 percent, 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 percent. 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-percent level, except for the 2-bin case with all individuals, which is significant at the ten-percent 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 does not match our model; however, these individuals are a small fraction of the population and are overwhelmingly

⁴⁰The fact that the extreme tails of the skill distribution do not conform to our predictions invites a comparison to Eeckhout, Pinheiro, and Schmidheiny (2014). Eeckhout, Pinheiro, and Schmidheiny (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, Pinheiro, and Schmidheiny (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 3 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 5, 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, Pinheiro, and Schmidheiny (2014) prediction. Moreover, Table 5 shows that the elasticities are monotonically increasing for high school graduates through professional degrees, which jointly account for 87 percent 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.

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.

Table 6: Pairwise comparisons of nine skill groups

| Bins | Total comparisons | Success rate | Success rate |
|------|-------------------|--------------|--------------|
| | | All | US-born |
| 270 | 1,307,340 | .61 (.00) | .64 (.00) |
| 90 | 144,180 | .66 (.00) | .70 (.00) |
| 30 | 15,660 | .72 (.00) | .76 (.00) |
| 10 | 1,620 | .74 (.00) | .80 (.00) |
| 5 | 360 | .76 (.00) | .85 (.00) |
| 3 | 108 | .75 (.02) | .88 (.00) |
| 2 | 36 | .75 (.08) | .90 (.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.

5.2 Larger cities specialize in skill-intensive sectors

This subsection examines the spatial pattern of sectoral employment. 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.⁴¹ We now examine whether larger cities are relatively specialized in skill-

⁴¹While 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

intensive sectors. Since employment levels in both industries and occupations are readily available in the data, we test the employment implications of Corollary 1.

5.2.1 The spatial distribution of occupations

We first implement the elasticities test and the pairwise comparisons test interpreting sectors as occupations.⁴² 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. Figure 1 plots the 22 occupational categories' estimated population elasticities of employment against these occupational skill intensities.⁴³ Our theory says the population elasticity of occupational employment should rise with skill intensity and indeed we see a clear positive relationship. 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 $\beta_\sigma \geq \beta_{\sigma'} \iff \sigma \geq \sigma'$ involves 231 ($= 22 \times 21/2$) comparisons of the estimated coefficients.⁴⁴ This hypothesis is rejected at the five-percent 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 7.⁴⁵ Using 276 cities and 22 occupational categories to make more than 8 million pairwise comparisons yields an average success rate of 57 percent.⁴⁶ 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 76 percent,

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. 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 7 and Corollary 1.

⁴²While 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.

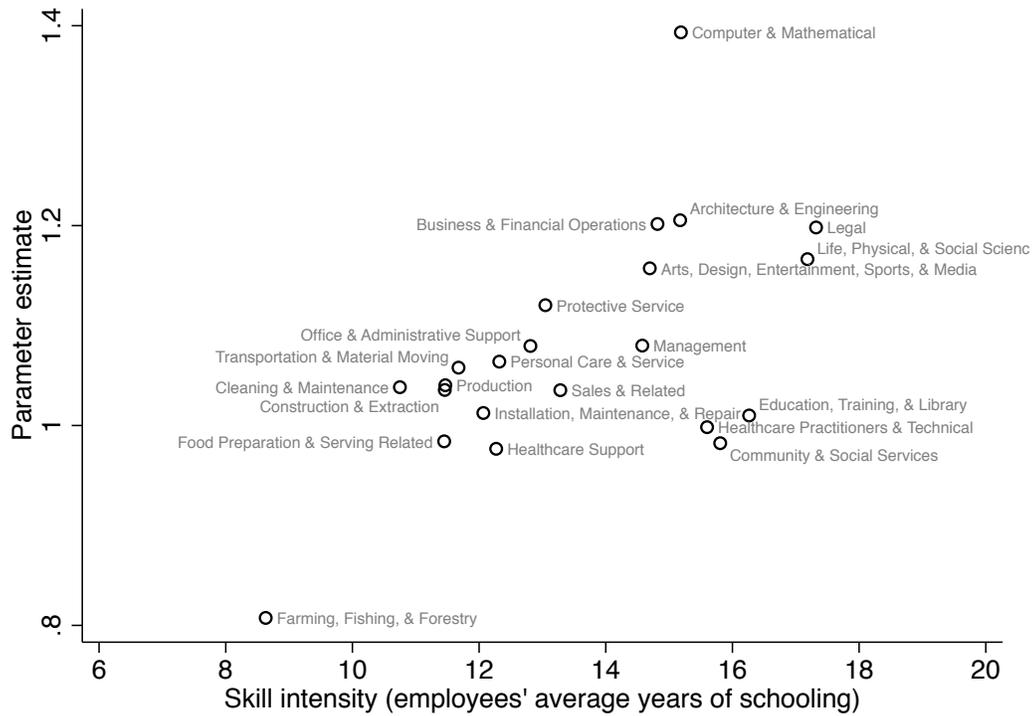
⁴³These elasticities are estimated without including zero-employment observations. The results obtained when including those observations are similar.

⁴⁴The elasticity estimates appear in Appendix Table E.2.

⁴⁵With 276 metropolitan areas and bins that are factors of 270, the number of cities per bin may differ by one.

⁴⁶These pairwise comparisons omit zero-employment observations. The results obtained when including those observations are similar.

Figure 1: Occupations' population elasticities and skill intensities



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.

Table 7: Pairwise comparisons of occupations

| | Total | |
|------|-------------|--------------|
| Bins | comparisons | Success rate |
| 276 | 8,073,382 | .57 (.00) |
| 90 | 925,155 | .62 (.00) |
| 30 | 100,485 | .64 (.00) |
| 10 | 10,395 | .65 (.00) |
| 5 | 2,310 | .71 (.00) |
| 3 | 693 | .74 (.00) |
| 2 | 231 | .76 (.01) |

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

5.2.2 The spatial distribution of industries

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 Figure 2.⁴⁷ 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. Testing the hypothesis that $\beta_\sigma \geq \beta_{\sigma'} \iff \sigma \geq \sigma'$ for the 19 industries involves 171 ($= 19 \times 18/2$) comparisons of these estimated elasticities.⁴⁸ This hypothesis is rejected in only 33 comparisons, so the elasticity implication holds true for industries about 80 percent of the time.

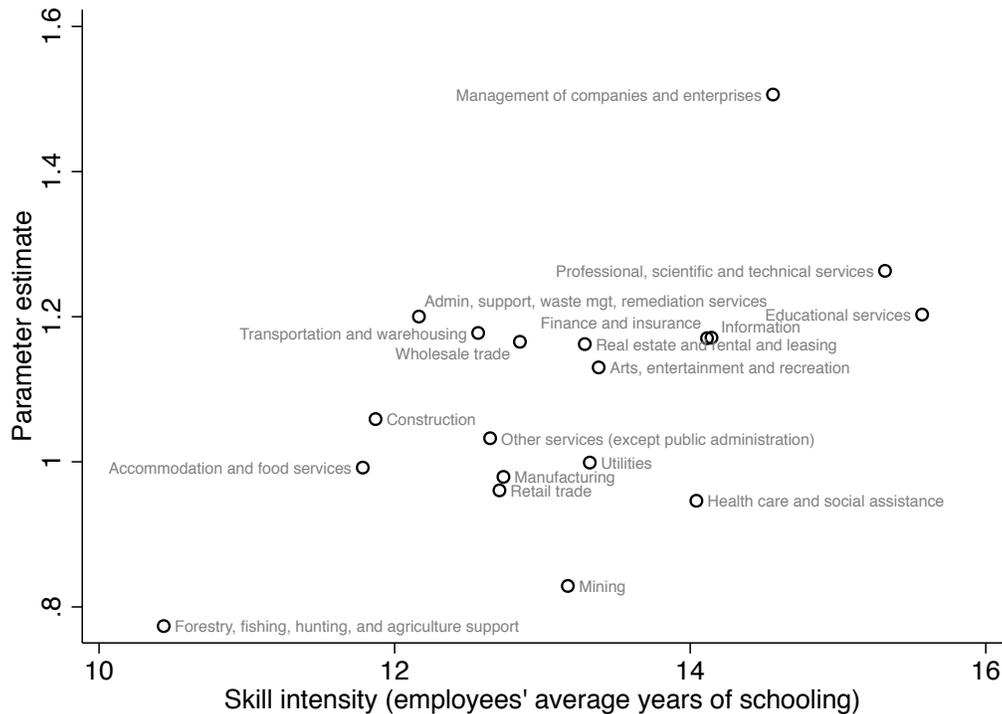
The results for pairwise comparisons for industries appear in Table 8. Making more than 6 million pairwise comparisons yields a statistically significant success rate of 61 percent. Raising the number of cities per bin raises the success rate monotonically and always at a high level of statistical significance. The success rate rises to 79 percent when contrasting two bins of large and small cities.

⁴⁷As for occupations, these elasticities are estimated without including zero-employment observations. The results obtained when including those observations are similar.

⁴⁸The elasticity estimates appear in Appendix Table E.3.

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.⁴⁹ 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.

Figure 2: Industries' population elasticities and skill intensities



⁴⁹We have found broadly similar results when examining industries at a finer level of disaggregation and when restricting attention to manufacturing industries.

Table 8: Pairwise comparisons of industries

| | Total | |
|------|-------------|--------------|
| Bins | comparisons | Success rate |
| 276 | 6,469,758 | .61 (.00) |
| 90 | 684,855 | .65 (.00) |
| 30 | 74,385 | .67 (.00) |
| 10 | 7,695 | .72 (.00) |
| 5 | 1,710 | .75 (.00) |
| 3 | 513 | .76 (.01) |
| 2 | 171 | .78 (.03) |

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

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 percent 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 $c > c' > c''$ 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 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 (Corollary 3). 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. Amongst the nine educational categories, the prediction that $c > c' \Rightarrow f(\omega, c) \geq f(\omega, c')$ is true in 88 percent 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 Figures 1 and 2, as they are all strongly positive. Amongst the 19 2-digit NAICS industries, the prediction that $c > c' \Rightarrow f(\sigma, c) \geq f(\sigma, c')$ is true in 86 percent 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 percent 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 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 “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.⁵⁰ 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.

⁵⁰For 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 $B'(\sigma')B(\sigma) \geq B'(\sigma)B(\sigma') \forall \sigma' \geq \sigma$. Such a shift increases both wage inequality and rent inequality, in the sense that $\frac{G'(\omega')}{G'(\omega)} \geq \frac{G(\omega')}{G(\omega)} \forall \omega' \geq \omega$ and $\frac{r'_\Gamma(\gamma')}{r'_\Gamma(\gamma)} \geq \frac{r_\Gamma(\gamma')}{r_\Gamma(\gamma)} \forall \gamma' \geq \gamma$. In models in which individuals are indifferent across all locations within a city, the rent gradient is invariant to the wage distribution.

References

- ABDEL-RAHMAN, H. M., AND A. ANAS (2004): “Theories of systems of cities,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4, chap. 52, pp. 2293–2339. Elsevier.
- ACEMOGLU, D., AND D. AUTOR (2011): “Skills, Tasks and Technologies: Implications for Employment and Earnings,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4, pp. 1043–1171. Elsevier.
- AUTOR, D. H., AND D. DORN (2013): “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review*, 103(5), 1553–97.
- BACOLOD, M., B. S. BLUM, AND W. C. STRANGE (2009): “Skills in the city,” *Journal of Urban Economics*, 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., G. EPSTEIN, AND I. GANG (2005): “Enclaves, language, and the location choice of migrants,” *Journal of Population Economics*, 18(4), 649–662.
- BAUM-SNOW, N., M. FREEDMAN, AND R. PAVAN (2014): “Why Has Urban Inequality Increased?,” mimeo.
- BAUM-SNOW, N., AND R. PAVAN (2013): “Inequality and City Size,” *The Review of Economics and Statistics*, 95(5), 1535–1548.
- BEHRENS, K., G. DURANTON, AND F. ROBERT-NICOUD (2014): “Productive Cities: Sorting, Selection, and Agglomeration,” *Journal of Political Economy*, 122(3), 507 – 553.
- BERNSTEIN, J. R., AND D. E. WEINSTEIN (2002): “Do endowments predict the location of production?: Evidence from national and international data,” *Journal of International Economics*, 56(1), 55–76.
- CHRISTALLER, W. (1933): *Die zentralen Orte in Suddeutschland*. Gustav Fischer, Jena.
- COMBES, P.-P., G. DURANTON, AND L. GOBILLON (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., AND J. VOGEL (2010): “Matching and Inequality in the World Economy,” *Journal of Political Economy*, 118(4), 747–786.
- (2015): “Beyond Ricardo: Assignment Models in International Trade,” *Annual Review of Economics*, 7(1).
- DAVIS, D. R., AND J. I. DINGEL (2012): “A Spatial Knowledge Economy,” NBER Working Papers 18188, National Bureau of Economic Research, Inc.
- DAVIS, D. R., AND D. E. WEINSTEIN (1999): “Economic geography and regional production structure: An empirical investigation,” *European Economic Review*, 43(2), 379–407.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106(3), 479–524.
- DURANTON, G., AND H. G. OVERMAN (2005): “Testing for Localization Using Micro-Geographic Data,” *Review of Economic Studies*, 72(4), 1077–1106.
- DURANTON, G., AND D. PUGA (2001): “Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products,” *American Economic Review*, 91(5), 1454–1477.
- (2005): “From sectoral to functional urban specialisation,” *Journal of Urban Economics*, 57(2), 343–370.
- (2015): “Urban Land Use,” in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. S. Strange, vol. 5A, chap. 8, pp. 467–560. North-Holland.
- EDIN, P.-A., P. FREDRIKSSON, AND O. ASLUND (2003): “Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment,” *The Quarterly Journal of Economics*, 118(1), 329–357.
- ECKHOUT, J., R. PINHEIRO, AND K. SCHMIDHEINY (2014): “Spatial Sorting,” *Journal of Political Economy*, 122(3), 554 – 620.
- ELLISON, G., AND E. L. GLAESER (1997): “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 105(5), 889–927.
- ELLISON, G., E. L. GLAESER, AND W. R. KERR (2010): “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns,” *American Economic Review*, 100(3), 1195–1213.

- ETHIER, W. (1984): “Higher dimensional issues in trade theory,” in *Handbook of International Economics*, ed. by R. W. Jones, and P. B. Kenen, vol. 1, chap. 03, pp. 131–184. Elsevier, 1 edn.
- FUJITA, M., AND J.-F. THISSE (2002): *Economics of Agglomeration*. Cambridge University Press.
- GABSZEWICZ, J. J., A. SHAKED, J. SUTTON, AND J. THISSE (1981): “International Trade in Differentiated Products,” *International Economic Review*, 22(3), pp. 527–534.
- GIBSON, C., AND K. JUNG (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., AND E. ROSSI-HANSBERG (2008): “Trading Tasks: A Simple Theory of Offshoring,” *American Economic Review*, 98(5), 1978–97.
- 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 *Topics in Public Economics: Theoretical and Applied Analysis*, ed. by D. Pines, E. Sadka, and I. Zilcha, pp. 33–54. Cambridge University Press Cambridge.
- HELSLEY, R. W., AND W. C. STRANGE (1990): “Matching and agglomeration economies in a system of cities,” *Regional Science and Urban Economics*, 20(2), 189–212.
- (2014): “Coagglomeration, Clusters, and the Scale and Composition of Cities,” *Journal of Political Economy*, 122(5), 1064 – 1093.
- HENDERSON, J. V. (1974): “The Sizes and Types of Cities,” *American Economic Review*, 64(4), 640–56.
- (1983): “Industrial Bases and City Sizes,” *American Economic Review*, 73(2), 164–68.
- (1987): “General equilibrium modeling of systems of cities,” in *Handbook of Regional and Urban Economics*, ed. by E. S. Mills, vol. 2, chap. 23, pp. 927–956. Elsevier.

- (1991): *Urban Development: Theory, Fact, and Illusion*. Oxford University Press.
- (1997): “Medium size cities,” *Regional Science and Urban Economics*, 27(6), 583–612.
- HENDRICKS, L. (2011): “The Skill Composition Of U.S. Cities,” *International Economic Review*, 52(1), 1–32.
- HOLMES, T. J., AND J. J. STEVENS (2004): “Spatial distribution of economic activities in North America,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4, chap. 63, pp. 2797–2843. Elsevier.
- HSU, W.-T., T. J. HOLMES, AND F. MORGAN (2014): “Optimal city hierarchy: A dynamic programming approach to central place theory,” *Journal of Economic Theory*, 154(0), 245 – 273.
- JONES, R. W., AND J. A. SCHEINKMAN (1977): “The Relevance of the Two-Sector Production Model in Trade Theory,” *Journal of Political Economy*, 85(5), 909–35.
- LEAMER, E. E. (1984): *Sources of International Comparative Advantage: Theory and Evidence*. MIT Press.
- LUCAS, R. E. (1988): “On the mechanics of economic development,” *Journal of Monetary Economics*, 22(1), 3–42.
- NOTOWIDIGDO, M. J. (2013): “The Incidence of Local Labor Demand Shocks,” mimeo.
- ROSSI-HANSBERG, E., P.-D. SARTE, AND R. OWENS (2010): “Housing Externalities,” *Journal of Political Economy*, 118(3), 485–535.
- RUGGLES, S., J. T. ALEXANDER, K. GENADEK, R. GOEKEN, M. B. SCHROEDER, AND M. SOBEK (2010): “Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database],” Minneapolis, MN: Minnesota Population Center.
- SAMUELSON, P. A. (1948): “International trade and the equalisation of factor prices,” *The Economic Journal*, 58(230), 163–184.
- SATTINGER, M. (1978): “Comparative Advantage in Individuals,” *The Review of Economics and Statistics*, 60(2), 259–67.
- (1979): “Differential Rents and the Distribution of Earnings,” *Oxford Economic Papers*, 31(1), 60–71.

——— (1993): “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature*, 31(2), 831–80.

YAGAN, D. (2014): “Moving to Opportunity? Migratory Insurance over the Great Recession,” mimeo.

A Appendix: Theory

A.1 Consumption interpretation

The production and consumption interpretations yield very similar results but differ slightly in notation. In the consumption interpretation, an individual’s productivity and utility are

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

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

where $T(\tau)$ determines the value of the individual’s disposable income after paying his or her locational price.⁵¹ In this interpretation, preferences are non-homothetic in a manner akin to that of Gabszewicz, Shaked, Sutton, and Thisse (1981). Higher-income individuals are more willing to pay for higher-quality locations because a more desirable location complements their higher consumption of tradables.

In this case, instead of $\gamma = A(c)T(\tau) = A(c')T(\tau') \iff r(c, \tau) = r(c', \tau') = r_\Gamma(\gamma)$, the appropriate equivalence between two locations is their “amenity-amplified price”, which is $T(\tau)r(c, \tau)$. So the equivalence statement is now $\gamma = A(c)T(\tau) = A(c')T(\tau') \iff T(\tau)r(c, \tau) = T(\tau')r(c', \tau') = r_\Gamma(\gamma)$. 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 4 and 5 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 6 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, Corollary 2 holds true. As a result, the predictions about cities’ population, sectors, and productivities described in sections 3.5 are unaltered by interpreting $T(\tau)$ as describing consumption benefits rather than production benefits.

⁵¹Recall that the final good is the numeraire.

A.2 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 $V(z) = \frac{2\pi}{d_2^2}(d_1 - z)$ 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

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

- 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: $v_1(z) = \frac{1}{\sqrt{2\pi}} \frac{1}{z}$ 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.3 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 $\mathcal{C}(s; \tau)$. Profit-maximizing, perfectly competitive landlords build

additional units until rent equals marginal cost.

$$\begin{aligned}\pi(s; \tau, c) &= r_{\Gamma}(A(c)T(\tau))s - \mathbf{C}(s; \tau) \\ \pi'(s; \tau, c) = 0 &\Rightarrow r_{\Gamma}(A(c)T(\tau)) = \mathbf{C}'(s; \tau)\end{aligned}$$

Let $\mathbf{C}(s; \tau) = \frac{s^{\beta}}{\beta h(\tau)}$, 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. Corollaries 1 and 2 follow immediately. A sufficient condition to obtain the result in Proposition 2 is that $h(\tau)$ is increasing.

B Appendix: 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, $\exists \omega : A(c)T(\bar{\tau}(c))G(\omega) - r(c, \bar{\tau}(c)) \geq A(c)T(\bar{\tau}(c) + \epsilon)G(\omega) \forall \epsilon > 0$. 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:

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 $\Omega(\tau) \equiv \{\omega \in \Omega | f(\omega, c, \tau) > 0\}$ and $\mathcal{T}(\omega) \equiv \{\tau \in [0, \bar{\tau}(c)] | f(\omega, c, \tau) > 0\}$.

1. $\mathcal{T}(\omega) \neq \emptyset$ by equation (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

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

where δ^{Dirac} is a Dirac delta function. Combining equations (9) and (A.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 $f(\omega, c, \tau) > 0 \iff N(\tau) = \omega$. 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 2:

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

Proof of Lemma 3:

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

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

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

Proof of Lemma 4:

This proof is analogous to the proof of lemma 2.

Proof of Lemma 5:

Proof. By utility maximization

$$\begin{aligned} \gamma G(K(\gamma)) - r_{\Gamma}(\gamma) &\geq (\gamma + d\gamma)G(K(\gamma)) - r_{\Gamma}(\gamma + d\gamma) \\ (\gamma + d\gamma)G(K(\gamma + d\gamma)) - r_{\Gamma}(\gamma + d\gamma) &\geq \gamma G(K(\gamma + d\gamma)) - r_{\Gamma}(\gamma) \end{aligned}$$

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 6:

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 7:

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

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

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

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

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

□

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 $\gamma \leq A(c')T(0)$,

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

This condition can be rewritten as

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

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

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

C Appendix: 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 percent sample of the 2000 US Census and the 1 percent metro sample of the 1980 US Census made available by IPUMS-USA (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek, 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 255 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 hour 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 ($= 19 \times 276 - 4$) 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 ob-

servations, 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.”

D Appendix: Empirical Tests

D.1 Population elasticities and pairwise comparisons

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

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

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

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

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

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

$$\ln f(\nu, c) = \alpha_{\nu} + \beta_{\nu} \ln L_c + \epsilon_{\nu, c}$$

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

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

where $\ln \bar{L}_{\mathcal{C}'}$ denotes an average log population, $\ln \bar{L}_{\mathcal{C}'} \equiv \frac{1}{n_{\mathcal{C}'}} \sum_{c' \in \mathcal{C}'} \ln L_{c'}$. 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, $\epsilon_{\nu,c} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Then $\frac{1}{n_c} \sum_{c \in \mathcal{C}} [\epsilon_{\nu',c} - \epsilon_{\nu,c}]$ 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

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

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 $\mathcal{P} \rightarrow 1/2$ as $\sigma \rightarrow \infty$). When the function is log-modular, $\mathcal{P} \rightarrow 1/2$ 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-percent 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.

E Appendix: 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 percent 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, Fredriksson, and Aslund, 2003; Bauer, Epstein, and Gang, 2005).⁵²

⁵²Another potential mechanism is that immigrants may find larger cities’ combination of higher nominal

Eeckhout, Pinheiro, and Schmidheiny (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.

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 percent of the US population, while in 1980 they constituted about 6 percent (Gibson and Jung, 2006). More importantly, in 2000, foreign-born individuals constituted 71 percent of the least-skilled group, while in 1980 they were only 28 percent. 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 1980.⁵³ 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, Pinheiro, and Schmidheiny (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 (72%) 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 0.975 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.

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.

⁵³The educational categories in Table E.1 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.1: Population elasticities of seven skill groups, 1980

| Dependent variable: $\ln f(\omega, c)$ | (1) All | (2) US-born | Population share | Share US-born |
|--|--------------------|-------------------|---------------------|------------------|
| $\beta_{\omega 1}$ Less than high school \times log population | 0.975 (0.0236) | 0.892 (0.0255) | .09 | .72 |
| $\beta_{\omega 2}$ High school dropout \times log population | 1.006 (0.0157) | 0.983 (0.0179) | .12 | .91 |
| $\beta_{\omega 3}$ Grade 12 \times log population | 0.989 (0.00936) | 0.971 (0.0111) | .33 | .93 |
| $\beta_{\omega 4}$ 1 year college \times log population | 1.047 (0.0144) | 1.033 (0.0151) | .10 | .94 |
| $\beta_{\omega 5}$ 2-3 years college \times log population | 1.095 (0.0153) | 1.076 (0.0155) | .12 | .91 |
| $\beta_{\omega 6}$ 4 years college \times log population | 1.091 (0.0153) | 1.073 (0.0157) | .12 | .92 |
| $\beta_{\omega 7}$ 5+ years college \times log population | 1.113 (0.0202) | 1.093 (0.0196) | .12 | .90 |

NOTES: Standard errors, clustered by MSA, in parentheses. Sample is individuals 25 and older in the labor force residing in 255 metropolitan areas in 1980.

E.2 Sectoral population elasticities

Tables E.2 and E.3 report the population elasticities estimates depicted in Figures 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

| | | | |
|--|--------------------|--|--------------------|
| $\beta_{\sigma 1}$ Farming, Fishing, and Forestry Occupations | 0.807 (0.0470) | $\beta_{\sigma 12}$ Sales and Related Occupations | 1.035 (0.00917) |
| $\beta_{\sigma 2}$ Building and Grounds Cleaning and Maintenance Occupations | 1.038 (0.0103) | $\beta_{\sigma 13}$ Management Occupations | 1.080 (0.0143) |
| $\beta_{\sigma 3}$ Food Preparation and Serving Related Occupations | 0.984 (0.0104) | $\beta_{\sigma 14}$ Arts, Design, Entertainment, Sports, and Media Occupations | 1.157 (0.0189) |
| $\beta_{\sigma 4}$ Construction and Extraction Occupations | 1.035 (0.0138) | $\beta_{\sigma 15}$ Business and Financial Operations Occupations | 1.202 (0.0177) |
| $\beta_{\sigma 5}$ Production Occupations | 1.040 (0.0250) | $\beta_{\sigma 16}$ Architecture and Engineering Occupations | 1.205 (0.0255) |
| $\beta_{\sigma 6}$ Transportation and Material Moving Occupations | 1.058 (0.0135) | $\beta_{\sigma 17}$ Computer and Mathematical Occupations | 1.393 (0.0335) |
| $\beta_{\sigma 7}$ Installation, Maintenance, and Repair Occupations | 1.012 (0.0110) | $\beta_{\sigma 18}$ Healthcare Practitioners and Technical Occupations | 0.998 (0.0139) |
| $\beta_{\sigma 8}$ Healthcare Support Occupations | 0.977 (0.0130) | $\beta_{\sigma 19}$ Community and Social Services Occupations | 0.982 (0.0199) |
| $\beta_{\sigma 9}$ Personal Care and Service Occupations | 1.064 (0.0170) | $\beta_{\sigma 20}$ Education, Training, and Library Occupations | 1.010 (0.0168) |
| $\beta_{\sigma 10}$ Office and Administrative Support Occupations | 1.079 (0.00999) | $\beta_{\sigma 21}$ Life, Physical, and Social Science Occupations | 1.166 (0.0295) |
| $\beta_{\sigma 11}$ Protective Service Occupations | 1.120 (0.0140) | $\beta_{\sigma 22}$ Legal Occupations | 1.198 (0.0221) |

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

Table E.3: Industrial employment population elasticities

| | (1) | (2) | | (1) | (2) |
|--|--------------------|--------------------|---|-------------------|-------------------|
| $\beta_{\sigma 1}$ Forestry, fishing, hunting, and agriculture support | 0.774 (0.0546) | 0.660 (0.142) | $\beta_{\sigma 11}$ Real estate and rental and leasing | 1.162 (0.0142) | 1.160 (0.0145) |
| $\beta_{\sigma 2}$ Accommodation and food services | 0.992 (0.0125) | 0.991 (0.0129) | $\beta_{\sigma 12}$ Utilities | 0.999 (0.0338) | 0.917 (0.0545) |
| $\beta_{\sigma 3}$ Construction | 1.059 (0.0158) | 1.057 (0.0160) | $\beta_{\sigma 13}$ Arts, entertainment and recreation | 1.130 (0.0220) | 1.125 (0.0251) |
| $\beta_{\sigma 4}$ Admin, support, waste mgt, remediation services | 1.200 (0.0206) | 1.207 (0.0195) | $\beta_{\sigma 14}$ Health care and social assistance | 0.946 (0.0127) | 0.946 (0.0127) |
| $\beta_{\sigma 5}$ Transportation and warehousing | 1.178 (0.0253) | 1.175 (0.0264) | $\beta_{\sigma 15}$ Finance and insurance | 1.170 (0.0206) | 1.167 (0.0215) |
| $\beta_{\sigma 6}$ Other services (except public administration) | 1.032 (0.0109) | 1.032 (0.0110) | $\beta_{\sigma 16}$ Information | 1.171 (0.0216) | 1.184 (0.0248) |
| $\beta_{\sigma 7}$ Retail trade | 0.961 (0.00739) | 0.962 (0.00713) | $\beta_{\sigma 17}$ Management of companies and enterprises | 1.506 (0.0435) | 1.406 (0.0469) |
| $\beta_{\sigma 8}$ Manufacturing | 0.979 (0.0293) | 0.973 (0.0296) | $\beta_{\sigma 18}$ Professional, scientific and technical services | 1.263 (0.0174) | 1.267 (0.0177) |
| $\beta_{\sigma 9}$ Wholesale trade | 1.165 (0.0193) | 1.159 (0.0190) | $\beta_{\sigma 19}$ Educational services | 1.203 (0.0346) | 1.205 (0.0391) |
| $\beta_{\sigma 10}$ Mining | 0.829 (0.0621) | 0.633 (0.122) | | | |
| Observations | 5,240 | 4,295 | Observations | 5,240 | 4,295 |
| 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

| Skill (3 groups) | 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 | .46 | .35 | .79 | Less than high school | .07 | .03 | .27 |
| | | | | High school dropout | .12 | .07 | .73 |
| | | | | High school graduate | .27 | .24 | .88 |
| Some college | .28 | .32 | .89 | College dropout | .21 | .24 | .90 |
| | | | | Associate's degree | .06 | .08 | .88 |
| Bachelor's or more | .27 | .34 | .85 | Bachelor's degree | .17 | .21 | .87 |
| | | | | Master's degree | .06 | .08 | .84 |
| | | | | Professional degree | .02 | .03 | .82 |
| | | | | Doctorate | .01 | .01 | .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 All | IPUMS FTFY All | IPUMS FTFY US-born |
| $\beta_{\omega 1}$ High school or less \times log population | 0.976 (0.00984) | 0.955 (0.0110) | 0.899 (0.0160) |
| $\beta_{\omega 2}$ Some college \times log population | 1.010 (0.00770) | 0.999 (0.0105) | 0.973 (0.0123) |
| $\beta_{\omega 3}$ 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) SF3 All | (2) IPUMS FTFY All | (3) IPUMS FTFY 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) |

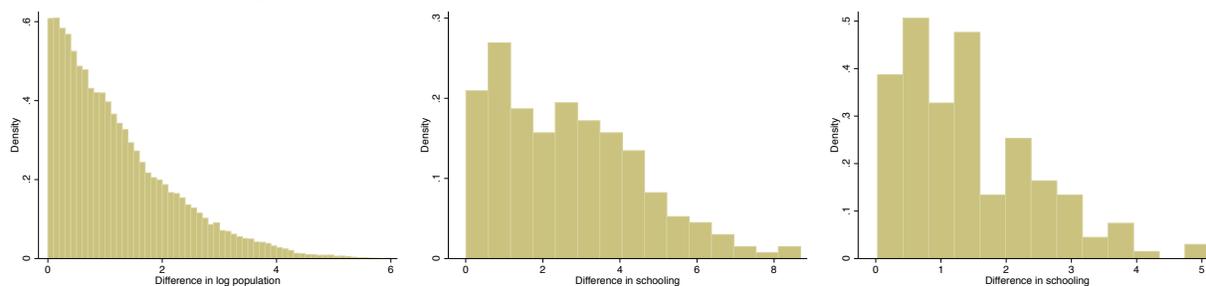
NOTES: Standard errors, clustered by MSA, in parentheses. Column 1 sample is Summary File 3 tabulations of populations by educational attainment. Column 2 sample is full-time, full-year employees 25 and older residing in 270 metropolitan areas in IPUMS Census of Population microdata. Column 3 is the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

E.4 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 Figure E.1, the unweighted success rates may differ substantially from those weighted by the relevant criteria.

Tables E.7 through E.10 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 section 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.

Figure E.1: Differences in populations and skill intensities



37,950 pairwise comparisons
between 276 MSA populations

Data source: Census 2000
PHC-T-3

231 pairwise comparisons between
22 2-digit SOC occupations

Data source: 2000 Census of
Population microdata via
IPUMS-USA

171 pairwise comparisons between
19 2-digit NAICS industries

Data source: 2000 Census of
Population microdata via
IPUMS-USA

Table E.7: Pairwise comparisons of three skill groups

| Bins | Total comparisons | Unweighted | | Pop-diff-weighted | |
|------|-------------------|---------------|---------------|-------------------|---------------|
| | | All | US-born | All | US-born |
| 2 | 3 | 1.00 (.16) | 1.00 (.18) | 1.00 (.16) | 1.00 (.18) |
| 3 | 9 | 1.00 (.01) | 1.00 (.01) | 1.00 (.01) | 1.00 (.01) |
| 5 | 30 | .93 (.00) | .97 (.00) | .97 (.00) | .99 (.00) |
| 10 | 135 | .77 (.00) | .81 (.00) | .89 (.00) | .91 (.00) |
| 30 | 1,305 | .70 (.00) | .74 (.00) | .82 (.00) | .86 (.00) |
| 90 | 12,015 | .64 (.00) | .67 (.00) | .74 (.00) | .77 (.00) |
| 270 | 108,945 | .60 (.00) | .61 (.00) | .67 (.00) | .69 (.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.8: Pairwise comparisons of nine skill groups

| Bins | Total comparisons | Unweighted | | Pop-diff-weighted | | Pop-diff×edu-share | |
|------|-------------------|--------------|--------------|-------------------|--------------|--------------------|--------------|
| | | All | US-born | All | US-born | All | US-born |
| 2 | 36 | .61 (.42) | .86 (.02) | .61 (.42) | .86 (.02) | .75 (.08) | .90 (.01) |
| 3 | 108 | .58 (.37) | .80 (.00) | .61 (.26) | .83 (.00) | .75 (.02) | .88 (.00) |
| 5 | 360 | .58 (.24) | .76 (.00) | .62 (.14) | .80 (.00) | .76 (.00) | .85 (.00) |
| 10 | 1,620 | .58 (.08) | .67 (.00) | .63 (.02) | .75 (.00) | .74 (.00) | .80 (.00) |
| 30 | 15,660 | .57 (.00) | .62 (.00) | .62 (.00) | .70 (.00) | .72 (.00) | .76 (.00) |
| 90 | 144,180 | .55 (.00) | .59 (.00) | .59 (.00) | .66 (.00) | .66 (.00) | .70 (.00) |
| 270 | 1,307,340 | .54 (.00) | .56 (.00) | .57 (.00) | .61 (.00) | .61 (.00) | .64 (.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 occupations

| Bins | Total comparisons | Unweighted success rate | Pop-diff weighted success rate | Pop-diff x skill-diff weighted success rate |
|------|-------------------|-------------------------|--------------------------------|---|
| 2 | 231 | .71 (.01) | .71 (.01) | .76 (.01) |
| 3 | 693 | .66 (.00) | .69 (.00) | .74 (.00) |
| 5 | 2,310 | .64 (.00) | .67 (.00) | .71 (.00) |
| 10 | 10,395 | .59 (.00) | .63 (.00) | .65 (.00) |
| 30 | 100,485 | .56 (.00) | .61 (.00) | .64 (.00) |
| 90 | 925,155 | .55 (.00) | .59 (.00) | .62 (.00) |
| 276 | 8,073,382 | .54 (.00) | .55 (.00) | .57 (.00) |

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

Table E.10: Pairwise comparisons of industries

| Bins | Total comparisons | Unweighted success rate | Pop-diff weighted success rate | Pop-diff x skill-diff weighted success rate |
|------|-------------------|-------------------------|--------------------------------|---|
| 2 | 171 | .67 (.06) | .67 (.06) | .78 (.03) |
| 3 | 513 | .65 (.01) | .65 (.02) | .76 (.01) |
| 5 | 1,710 | .63 (.00) | .65 (.00) | .75 (.00) |
| 10 | 7,695 | .60 (.00) | .63 (.00) | .72 (.00) |
| 30 | 74,385 | .57 (.00) | .61 (.00) | .67 (.00) |
| 90 | 684,855 | .56 (.00) | .60 (.00) | .65 (.00) |
| 276 | 6,469,758 | .54 (.00) | .57 (.00) | .61 (.00) |

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