PHOTO BY BARON VISUALS

WORKING PAPER #17 / OCTOBER 2014

THE COMPARATIVE ADVANTAGE OF CITIES

+ DONALD R. DAVIS AND JONATHAN I. DINGEL

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 21 manufacturing industries. The results provide support for our theory's predictions.

Thanks to Arnaud Costinot, Gilles Duranton, Vernon Henderson, Corinne Low, Joan Monras, Gianmarco Ottaviano, Keeyoung Rhee, John Romalis, Bernard Salanie, 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 SummerWorkshop, UCLA, and the Urban Economics Association for helpful comments and suggestions. We thank Yuxiao Huang and especially 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 nancial support.

CONTACT

Donald R. Davis
Department of Economics
Columbia University and NBER
drdavis@columbia.edu

Jonathan I. Dingel
Department of Economics
Chicago Booth School of Business
jdingel@chicagobooth.edu

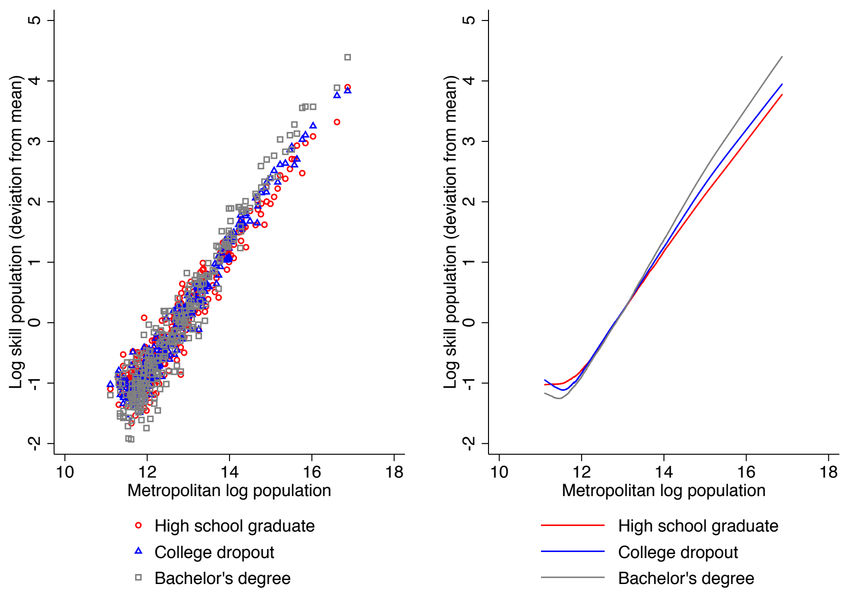
NYU Marron Institute of Urban Management

1 Introduction

The distributions of skills, occupations, and industries vary substantially and systematically across US cities. Figures 1 through 3 illustrate this with three selected examples for each.

- Figure 1 plots the population of three educational attainment categories against total metropolitan area population.[1] The left panel plots the data; the right panel plots a locally weighted regression for each category. While each educational category's population rises with metropolitan population, the relative levels also exhibit a systematic relationship with city size. Comparing elasticities, the population with a bachelor's degree rises with city size faster than the population of college dropouts, which in turn rises faster than the population of high-school graduates.

Figure 1: Populations of three educational groups across US metropolitan areas



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

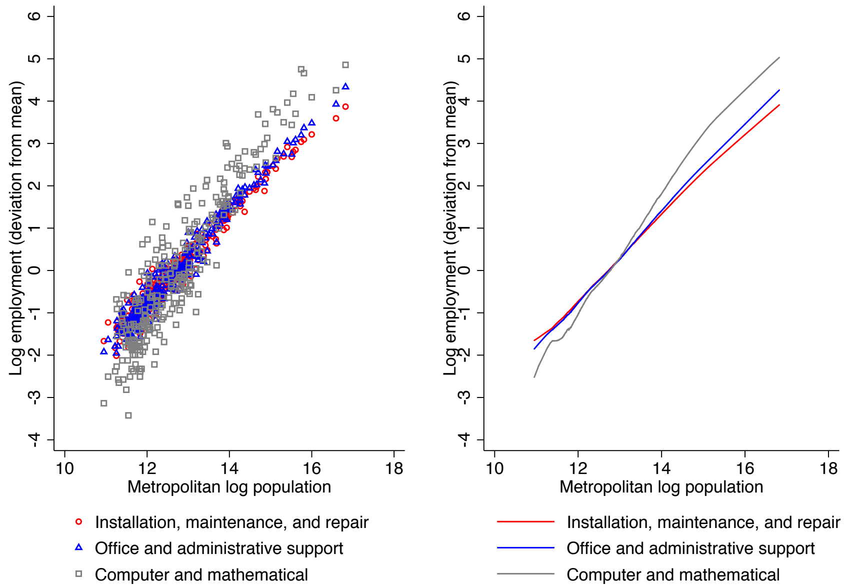
- Figure 2 plots metropolitan area employment in three occupational categories.[2] Computer and mathematical employment rises with city size faster than office and administrative employment, which in turn rises faster than installation, maintenance and repair employment. These sectors also differ in their employee characteristics. Nationally, the average individual in computer and mathematical occupations has about

[1]We use the terms "cities" and "metropolitan areas" interchangeably, as is customary in the literature. These three educational groups comprise about 70 percent of the employed metropolitan population (see Table 1).

[2]The occupations are SOC 49, 43, and 15 in the 2000 Occupational Employment Statistics data.

two more years of schooling than the average individual in office and administrative support and three more years than those in installation, maintenance, and repair.

Figure 2: Employment in three occupations across US metropolitan areas

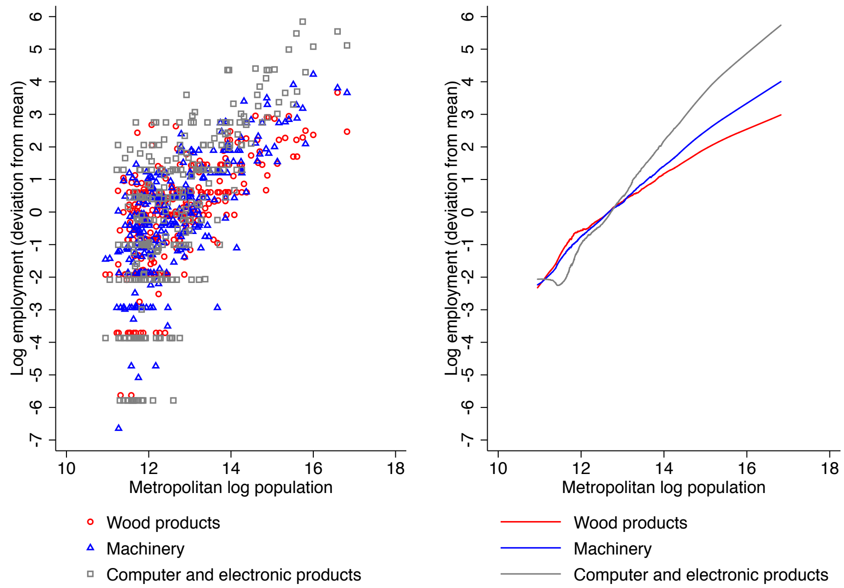


Data source: Occupational Employment Statistics 2000

- Figure 3 plots employment in three manufacturing industries.[3] Employment in computer and electronic products rises with city size faster than machinery, which in turn rises faster than wood products. On average, computer and electronic employees have about one more year of education than machinery employees and two more years of education than wood products employees.

[3]The industries are NAICS 321, 333, and 334 in the 2000 County Business Patterns data. Employment levels cluster at particular values due to censored observations. See appendix D describing the data.

Figure 3: Employment in three manufacturing industries across US metropolitan areas



Data source: County Business Patterns 2000

Together, these three figures suggest that larger cities are skill-abundant and specialize in skill-intensive activities. Explaining these patterns involves fundamental questions about the spatial organization of economic activity. What determines the distribution of skills across cities? What determines the distribution of occupations and industries across cities? How are these two phenomena interrelated? In this paper, we develop a theory describing the comparative advantage of cities that predicts such a pattern of skills and sectors in a manner amenable to empirical investigation.

As we describe 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 perfectly diversified hosts of all economic activities (Helsley and Strange, 2012). Yet Figures 2 and 3 make clear both that reality falls between these poles and that sectoral employment shares are systematically related to cities' sizes. In this paper, we integrate modern trade theory with urban economics by introducing a spatial-equilibrium model in which the comparative advantage of cities is jointly governed by the comparative advantage of individuals and their locational choices. Our theory both describes the intermediate case in which 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 that we take to the data.

Section 3 introduces our model of a system of cities with heterogeneous internal geographies. Cities are *ex ante* homogeneous, so cross-city heterogeneity is an emergent outcome of the choices made by freely mobile individuals. Agglomeration economies make cities with larger, more skilled populations exhibit higher total factor productivity (TFP). Locations within cities exhibit heterogeneity in their 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 Sattinger (1975), Costinot (2009), and Costinot and Vogel (2010). 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 internal locations within cities combine to deliver a rich set of novel predictions. Agglomeration causes larger cities to have higher TFP, which makes a location within a larger city more attractive than a location 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 more prevalent in larger cities, consistent with the pattern shown in Figure 1. 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. More skill-intensive sectors exhibit higher population elasticities of sectoral employment, as suggested in Figures 2 and 3. Our model therefore predicts an urban hierarchy of skills and sectors. Under slightly stronger assumptions, larger cities will be absolutely larger in all sectors.

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 involves regression estimates of the population elasticities of educational, occupational, and industrial populations akin to those shown in Figures 1 through 3. The second involves pairwise comparisons governed by the monotone likelihood ratio property, as per Costinot (2009).[4] 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

[4]The distributions $f_c(\sigma)$ and $f_{c'}(\sigma)$ exhibit the monotone likelihood ratio property if, for any $\sigma > \sigma'$, $\frac{f_c(\sigma)}{f_{c'}(\sigma)} \geq \frac{f_c(\sigma')}{f_{c'}(\sigma')}$.

groups, we find that larger cities are skill-abundant. Among US-born individuals, cities' skill distributions typically exhibit the monotone likelihood ratio property.[5] Characterizing sectors in terms of 21 manufacturing industries or 22 occupational categories, we find that larger cities specialize in skill-intensive sectors. While sectors do not exhibit the monotone likelihood ratio property as reliably as skills, there is systematic variation in cities' sectoral distributions that is consistent with the novel predictions of our theory.

In short, when mobile individuals optimally choose locations and sectors, larger cities will have more skilled populations and thereby comparative advantage in skilled activities. These features are consistent with US data.

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.[6] To describe greater heterogeneity, we assume a continuum of skills, like Behrens, Duranton, and Robert-Nicoud (2014) and Davis and Dingel (2012).[7] Understanding the distribution of skills across cities with more than two types is valuable for at least three reasons. First, a large literature in labor economics has described important empirical developments such as wage polarization, job polarization, and simultaneous changes in between- and within-group inequality that cannot be explained by a model with two homogeneous skill groups (Acemoglu and Autor, 2011). Second, these developments have counterparts in cross-city variation in inequality and skill premia (Baum-Snow and

[5]Relative to our theory, foreign-born individuals with less than a high-school education tend to disproportionately locate in large US cities. Data from 1980, when foreign-born individuals were a substantially smaller share of the US population, suggest this reflects particular advantages that large cities offer foreign-born individuals rather than a general tendency for the unskilled to locate in large cities. See section 5.1.2.

[6]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 characteristic.

[7]Eeckhout, Pinheiro, and Schmidheiny (2014) describe a model with three skill types.

Pavan, 2013; Davis and Dingel, 2012). Third, we document systematic patterns in the cross-city distribution of skills at high levels of disaggregation, which suggests that individuals within broad skill categories are imperfect substitutes.[8]

Our model is a novel integration of systems-of-cities theory with land-use theory. The Alonso-Muth-Mills model of a single city describes a homogeneous population of residents commuting to a central business district (Brueckner, 1987). In that model, higher rents for locations with shorter commutes equalize utility across locations in equilibrium. When individuals are heterogeneous and value the rent-distance tradeoff differently, the single city's equilibrium rent schedule is the upper envelope of individuals' bid-rent functions (von Thünen, 1826; Fujita and Thisse, 2002, Ch 3). Models of a system of cities have incorporated the Alonso-Muth-Mills urban structure in which all individuals are indifferent across all locations within a city as a city-level congestion mechanism (Abdel-Rahman and Anas, 2004; Behrens, Duranton, and Robert-Nicoud, 2014). Our novel contribution is to describe multiple cities with internal geographies when individuals are not spatially indifferent across all locations.[9] 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. Though these tradeoffs appear obvious, we are not aware of a prior formal analysis. Considering both dimensions simultaneously is more realistic in both the description of the economic problem and the resulting predicted cross-city skill distributions. Since we have a continuum of heterogeneous individuals, we obtain equilibrium rent schedules that are integrals rather than upper envelopes of a discrete number of bid-rent functions.[10]

Our model belongs to a long theoretical tradition describing factor-supply-driven comparative advantage, dating from the 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 the recent work of Costinot (2009) and Costinot and Vogel (2010).[11] While these recent papers assume that countries' factor endowments exhibit the monotone likelihood ratio property, we obtain the result that cities' skill distributions exhibit this property as an equilibrium outcome. Thus,

[8]A long line of empirical work describes cross-city variation in skill distributions in terms of 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.

[9]In order to tractably characterize multiple cities with internal geographies and heterogeneous agents, we neglect the business-vs-residential land-use problem studied by Lucas and Rossi-Hansberg (2002).

[10]Our continuum-by-continuum approach to a differential rents model is in the spirit of Sattinger (1979).

[11]Sattinger (1993) and Costinot and Vogel (2014) survey the assignment literature.

from a theoretical perspective, cities within a country constitute a natural setting to examine these theories of comparative advantage. Moreover, the assumption of a common production technology is likely more appropriate within than between economies, and data from a single economy are likely more consistent and comparable than data combined across countries.

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. For these reasons, traditional Heckscher-Ohlin models did not appear a promising way to explain regional differences in sectoral composition.

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 polarized sectoral composition: *specialized cities* that have only one tradable sector and *perfectly diversified cities* that have all the tradable sectors (Abdel-Rahman and Anas, 2004; Helsley and Strange, 2012).[12] A recent exception is Helsley and Strange (2012), who examine whether the equilibrium level of coagglomeration is efficient. While Helsley and Strange (2012) 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.

In addition to sectoral composition, our theory describes sectoral size. Theories of localization and urbanization economies have contrasting predictions for cities' absolute employment levels.[13] In the canonical model of pure localization in Henderson (1974), specialized cities of different sizes host different industries, yielding "textile cities" and "steel cities". Industrial specialization is the very basis for the city-size distribution, and one wouldn't

[12] An exception is central place theory, and our model relates to that theory's results in interesting ways. Our model's equilibrium exhibits a hierarchy of cities and sectors, as larger cities produce a superset of the goods produced in smaller cities. Models in central place theory, dating from Christaller (1933) through Hsu, 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 have been considered in central place theory.

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

expect large cities to be larger in all industries. By contrast, in urbanization models with a composite output, every industry is (implicitly) larger in larger cities. In addition to introducing a multi-sector urbanization model in which larger cities are relatively larger in skill-intensive sectors, we identify 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. 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. We obtain these results in the absence of industry-specific scale economies and industry-pair-specific interactions or spillovers.

Our empirical work follows directly from our model's predictions about the cross-city distribution of sectoral activity relating cities' and sectors' characteristics. There is a small empirical literature describing variation in cities' sectoral composition, but this work has not been tightly tied to theory. This is likely because theories describing specialized or perfectly diversified cities provide limited guidance to empirical investigations of data that fall between the extremes. Holmes and Stevens (2004) survey the spatial distribution of economic activities in North America. In examining the empirical pattern of specialization, 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 the equilibrium outcome. Seminal work by Vernon Henderson explores theoretically and empirically the relationship between city size and industrial composition (Henderson, 1991). Henderson (1974) theoretically describes the polar cases of specialized and perfectly diversified cities (Helsley and Strange, 2012), while our model predicts incomplete industrial specialization. Henderson has argued that localization economies link cities' and industries' sizes, while our theory relies on urbanization economies and individuals' comparative advantage.

Despite these contrasts, our theory yields estimating equations for the population elasticities of sectoral employment that are closely related to the reduced-form regressions of employment shares on population that Henderson (1983) estimated for a few select industries. Our theory provides an explicit microfoundation for these regressions for an arbitrary number of sectors. Moreover, it predicts that we can order these elasticities by skill intensity. It allows all these elasticities to be positive, consistent with larger cities being larger in all sectors. It also describes how to compare the sectoral composition of groups of cities ordered by size, nesting the comparison of large and medium-size cities made by Henderson (1997). While our urbanization-based theory abstracts from the localization economies emphasized by Henderson, we believe future work should seek to integrate these distinct approaches.

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 ($c \in \mathbb{C} = \{1, \ldots, C\}$), 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 location 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

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

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

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

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

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

$A(c)$ 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, which in a canonical case is the cost of commuting to the central business district.[14] 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.[15] The former governs comparative advantage, so that higher-ω individuals are relatively more productive in higher-σ sectors.[16] 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 $\tau = 0$, so higher values of τ denote greater distance from the ideal location. The supply of locations with innate desirability of at least τ is $S(\tau)$.[17] 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. The city has sufficient land capacity that everyone can reside in the city and the least desirable locations are unoccupied. We normalize the reservation value of unoccupied locations to zero, so $r(c, \tau) \geq 0$.

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

[14] As 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.

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

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

[17] In the special case of the classical von Thünen model, τ describes physical distance from the central business district and the supply is $S(\tau) = \pi\tau^2$.

after paying their locational cost:

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

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

City-level TFP, $A(c)$, reflects agglomeration gains derived from both population size and composition. $A(c)$ 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

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

where $J(\cdot)$ is a positive, strictly increasing function and $j(\omega)$ is a positive, non-decreasing function.

3.2 Equilibrium

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

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

Profit maximization by final-good producers yields demands for intermediates

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

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,

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

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

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

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

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

A competitive equilibrium is a set of functions $Q : \Sigma \to \mathbb{R}^+$, $f : \Sigma \times \mathbb{C} \times \mathcal{T} \times \Omega \to \mathbb{R}^+$, $r : \mathbb{C} \times \mathcal{T} \to \mathbb{R}^+$, and $p : \Sigma \to \mathbb{R}^+$ 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.[18] By absolute advantage, more skilled individuals earn higher nominal incomes and $G(\omega)$ is a strictly increasing function.[19]

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, individual 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,

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

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

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 τ $\forall \tau < \bar{\tau}(c)$, 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 $N : \bar{\mathcal{T}}(c) \to \Omega$ 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 $M : \Omega \to \Sigma$ and $N : \bar{\mathcal{T}}(c) \to \Omega$. 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. Take cities' TFPs, which will be endogenously determined in equilibrium, as given for now and label the cities so that $A(C) \geq A(C-1) \geq \cdots \geq A(2) \geq A(1)$.[20] In

[20]Individuals take these TFPs as given. For now, we can assume these differences in total factor productivity are exogenously given. We describe their endogenous determination in section 3.6.

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*. A location's attractiveness, which we denote by γ, depends both on city-level TFP and its 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. That is, an individual who is indifferent between c and c' lives closer to the most desirable location in c' than the most desirable location in c, $\tau' = T^{-1}\left(\frac{A(c)}{A(c')}T(\tau)\right) < \tau$. 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* $K : \Gamma \to \Omega$ *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 $K : \Gamma \to \Omega$, 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.

Lemma 5 (Locational prices)**.** *In equilibrium,* $r_\Gamma(\gamma)$ *is increasing and continuously differentiable on* $[\underline{\gamma}, \bar{\gamma}]$ *and given by* $r_\Gamma(\gamma) = \int_{\underline{\gamma}}^{\gamma} G(K(x))dx$*.*

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 necessary and 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 $r(c, \tau) = r_\Gamma(A(c)T(\tau))$. In any city c, the supply of locations with attractiveness γ, and thus the set of locations with rental price $r_\Gamma(\gamma)$, is

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

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

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

Lemma 6 (A city's skill distribution)**.** *The population of individuals of skill ω in city c is*

$$L \times f(\omega, c) = \begin{cases} K^{-1'}(\omega) s(K^{-1}(\omega), c) & \text{if } A(c)T(0) \geq K^{-1}(\omega) \\ 0 & \text{otherwise} \end{cases}.$$

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

Lemma 7 (A city's sectoral employment distribution)**.** *The population of individuals employed in sector σ in city c is*

$$L \times f(\sigma, c) = \begin{cases} M^{-1'}(\sigma) K^{-1'}(M^{-1}(\sigma)) s(K^{-1}(M^{-1}(\sigma)), c) & \text{if } A(c)T(0) \geq K^{-1}(M^{-1}(\sigma)) \\ 0 & \text{otherwise} \end{cases}.$$

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. The sectoral assignments $M : \Omega \to \Sigma$, which are common across locations, translate this hierarchy of skills across cities into a hierarchy of sectors across cities.

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 comes from larger cities' TFP advantages being sector-neutral, 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 $(s(\gamma, c)s(\gamma', c') \overset{?}{>} s(\gamma', c)s(\gamma, c'))$ depend only on the dilation.

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

Proposition 1 (Locational attractiveness distribution)**.** *The supply of locations of attractiveness γ in city c, $s(\gamma, c)$, is log-supermodular if and only if the supply of locations with innate desirability $T^{-1}(z)$ within each city, $V(z)$, 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. 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.[21]

The distributions of skills and sectoral employment across cities follow straightforwardly from Proposition 1. The skill distribution follows immediately through the locational assignment function. The employment distribution follows in turn through the sectoral assignment function. Since $K : \Gamma \to \Omega$ and $M : \Omega \to \Sigma$ 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.*

Since productivity differences across locations are Hicks-neutral in our economy's equilibrium, the employment distribution across cities governs the output distribution across cities. Larger cities are skill-abundant and more skilled individuals work in more skill-intensive sectors, so larger cities produce relatively more in skill-intensive sectors. These patterns of specialization and trade are 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.[22] Since at equilibrium larger cities' productivity advantages are sector-neutral differences in total factor productivity, $f(\omega, c)$ is log-supermodular, and $H(\omega, \sigma)$ is log-supermodular, our economy's equilibrium satisfies Definition 4, Assumption 2,

[21] For example, for the von Thünen 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. Therefore, this canonical case satisfies the condition of Proposition 1.

[22] 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.

and Assumption 3 of Costinot (2009).[23] 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.*

These results characterize the pattern of comparative advantage across cities.[24]

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 $z = \frac{\gamma}{A(c)}$, $s(\gamma, c) \geq s(\gamma, c')$.*

Corollary 3. *If $V(z)$ has a decreasing elasticity that is less than -1 at $z = \frac{K^{-1}(\omega)}{A(c)} = \frac{\gamma}{A(c)}$, $A(c) > A(c')$ implies $f(\omega, c) \geq f(\omega, c')$ and $f(M(\omega), c) \geq f(M(\omega), c') \; \forall \omega \in \Omega$.*

3.6 Endogenizing cities' total factor productivities

Our exposition of equilibrium in sections 3.4 and 3.5 took cities' total factor productivities as given. When the condition of Proposition 1 is satisfied, a city that has higher total factor productivity $A(c)$ is larger and has a skill distribution $f(\omega, c)$ that likelihood ratio dominates those of cities with lower TFPs. Thus, this spatial pattern can be supported by endogenous productivity processes that make the city-level characteristic $A(c)$ higher when

[23] 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. However, in equilibrium, our economy does exhibit this property. In 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 and is log-supermodular in ω and σ. In the notation of equation (6) in Costinot (2009), $a(\gamma) = 1$ and $h(\omega, \sigma) = K^{-1}(\omega)H(\omega, \sigma)$, satisfying Definition 4 and Assumption 3.

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

the city contains a larger and more skilled population, such as the class of agglomeration functions described by equation (5). Numerous agglomeration processes may generate such productivity benefits, and we do not attempt to distinguish between them here.

4 Empirical approach and data description

We examine the predictions of our model using two approaches to characterize the outcomes described by Corollary 1.[25] 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.

Empirically testing Corollary 1 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 D for details.

4.1 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

- 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 $\beta_\nu \geq \beta_{\nu'} \iff \nu \geq \nu'$;
- if $\mathcal{C}$ and $\mathcal{C}'$ are distinct sets and $\mathcal{C}$ is greater than $\mathcal{C}'$ ($\inf_{c\in\mathcal{C}} L(c) > \sup_{c'\in\mathcal{C}'} L(c')$), then $\sum_{c\in\mathcal{C}} \ln f(\nu, c) + \sum_{c'\in\mathcal{C}'} \ln f(\nu', c') \geq \sum_{c\in\mathcal{C}} \ln f(\nu', c) + \sum_{c'\in\mathcal{C}'} \ln f(\nu, c')\ \forall \nu > \nu'$.

[25] Proposition 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.

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.[26] Similarly, the population elasticity of sectoral employment $f(\sigma, c)$ is increasing in skill intensity σ. The elasticity test examines the patterns suggested by Figures 1 through 3, where steeper slopes correspond to higher elasticities. 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).[27] 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.[28]

These two empirical tests are not independent, since they are both implied by log-supermodularity. Appendix C 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. Figures 1 through 3 suggest that they do. We also implement a test for systematic deviations proposed by Sattinger (1978) and examine whether the pairwise comparisons success rate increases with the number of cities per bin.[29]

4.2 Skills

Following a large literature, we use observed educational attainment as a proxy for individuals' skills.[30] 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

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

[27] Henderson (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 $\beta_\sigma - 1$.

[28] Provided $f(\nu, c) > 0 \ \forall \nu \forall c$, log-supermodularity means $\nu > \nu', c > c' \Rightarrow \ln f(\nu, c) + \ln f(\nu', c') \geq \ln f(\nu', c) + \ln f(\nu, c')$. The pairwise comparisons test follows from taking sums twice of each side of this inequality given $c > c' \ \forall c \in C \ \forall c' \in C'$.

[29] Appendix C shows that, if $\ln f(\nu, c)$ is the sum of a log-supermodular function and idiosyncratic errors, the probability of obtaining the correct inequality increases with aggregation. If the deterministic component is log-submodular (log-modular), this probability decreases with (is invariant to) aggregation.

[30] Costinot 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.

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 Census 2000 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 (3 percent of the population) to those with doctoral degrees (1 percent).[31] Table 1 shows the population shares and percentage US-born for each of these skill groups in 2000. Foreign-born individuals are disproportionately in the tails of the educational distribution.

Table 1: Skill groups by educational attainment

Skill (3 groups)	Population share	Percent US-born	Skill (9 groups)	Population share	Percent US-born
High school or less	.35	.77	Less than high school	.03	.21
			High school dropout	.07	.69
			High school graduate	.24	.87
Some college	.32	.88	College dropout	.24	.89
			Associate's degree	.08	.87
Bachelor's or more	.33	.85	Bachelor's degree	.21	.86
			Master's degree	.08	.83
			Professional degree	.03	.81
			Doctoral degree	.01	.69

Population shares and percentage US-born are percentages of full-time, full-year prime-age workers.
Source: Census 2000 microdata via IPUMS-USA

4.3 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 21 manufacturing industries in the three-digit stratum of the

[31] Individuals with doctorates typically earn less than individuals with professional degrees, so it may be inappropriate to treat PhDs as higher-ω individuals than professionals.

North American Industry Classification System (NAICS) or the 22 occupational categories in the two-digit stratum of the Standard Occupational Classification (SOC). We suspect that the assignment of workers to sectors is better characterized as assignments to occupations than assignments to industries, since virtually all industries employ both skilled and unskilled workers. Our measures of cross-sectoral variation in skill intensities in the following section are consistent with this conjecture.

We measure industrial employment in a metropolitan area using data from the 2000 County Business Patterns. We measure occupational employment in a metropolitan area using estimates from the 2000 BLS Occupational Employment Statistics. See appendix D for details.

4.4 Skill intensities

Our theory makes the strong assumption that $H(\omega, \sigma)$ is strictly log-supermodular so that sectors are ordered with respect to their skill intensities. In our empirical work, we infer sectors' skill intensities from the data using the observable characteristics of the workers employed in them. We use microdata from the 2000 Census of Population, which contains information about workers' educational attainments, industries, and occupations. We use the average years of schooling of workers employed in a sector as a measure of its skill intensity.[32] 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 21 manufacturing industries and the 22 occupational categories. There is considerably greater variation in average years of schooling across occupational categories than across industries.[33] This may suggest that the "assignment to occupations" interpretation of our model will be a more apt description of the data than the "assignments to industries" interpretation.

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

[33] The standard deviations of average years of schooling across occupational categories, industries, and manufacturing industries are 2.2, 1.0, and 0.9, respectively.

Table 2: Sectoral skill intensities

SOC	Occupational category	Skill intensity	NAICS	Manufacturing industry	Skill intensity
45	Farming, Fishing, and Forestry	9.3	315	Apparel	10.7
37	Building & Grounds Cleaning	10.9	314	Textile Product Mills	11.4
35	Food Preparation and Serving	11.4	316	Leather and Allied Product	11.7
47	Construction and Extraction	11.5	313	Textile Mills	11.7
51	Production	11.6	337	Furniture and Related Products	11.7
29	Healthcare Practitioners and Technical	15.6	312	Beverage and Tobacco Products	13.1
21	Community and Social Services	15.8	336	Transportation Equipment	13.2
25	Education, Training, and Library	16.5	324	Petroleum and Coal Products	13.5
19	Life, Physical, and Social Science	17.1	334	Computer & Electronic Products	14.1
23	Legal	17.3	325	Chemical	14.1

Source: Census 2000 microdata via IPUMS-USA

4.5 Pairwise weights

The most disaggregate implications of Corollary 1 are inequalities describing the number of individuals residing (employed) in two cities and two skill groups (sectors). Empirically testing these pairwise predictions involves evaluating as many as eight million of these inequalities and summarizing the results. An important question is whether each of these comparisons should be considered equally informative.

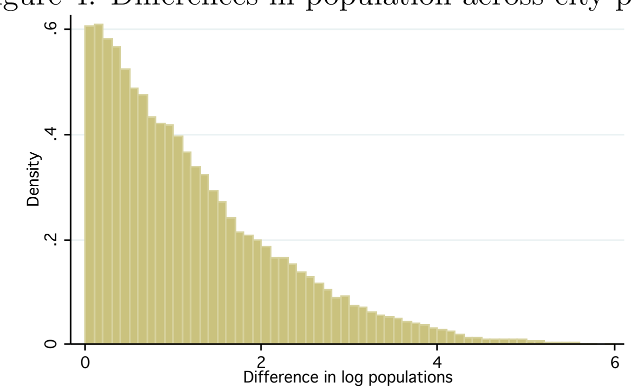
An unweighted summary statistic assigns equal weight to correctly predicting that Chicago (population 9 million) is relatively more skilled than Des Moines (population 456 thousand) and correctly predicting that Des Moines is relatively more skilled than Kalamazoo (population 453 thousand). Given the numerous idiosyncratic features of the real world omitted from our parsimonious theory, the former comparison seems much more informative about the relevance of our theory than the latter. Similarly, an unweighted summary statistic treats comparisons involving high school graduates (24 percent of the workforce) and comparisons involving PhDs (1 percent of the workforce) equally, while these differ in their economic import.

Following Trefler (1995), we report weighted averages of success rates in addition to unweighted statistics. In describing skill distributions, we weight each pairwise comparison by the two cities' difference in log population.[34] When we consider nine skill groups, we also report a case where we weight by the product of the two skill groups' population shares. Figure 4 shows the distribution of differences in log population across city pairs. Since the majority of city pairs have quite small differences in log population, the unweighted and

[34]Appendix C shows that, if $\ln f(\nu, c)$ is the sum of a log-supermodular function and idiosyncratic errors, the probability of obtaining the correct inequality is increasing in the difference in log population.

weighted statistics may yield substantially different results. In describing sectoral distributions, we weight pairwise comparisons by the product of the two cities' difference in log population and the two sectors' difference in skill intensity.[35] Figure 5 shows the distribution of differences in skill intensity across occupational pairs. While not as right-skewed as the distribution of differences in log population, this distribution may cause the unweighted and weighted statistics to differ. Figure 6 shows the distribution of differences in skill intensity across industries. The median difference between occupations is 2.3 years while the median difference between manufacturing industries is only 0.9 years. This relative compression in skill differences in the industrial data suggests that it may prove harder to make strong statements about differences across cities in industries than in occupations. Figures 4 through 6 underscore the importance of looking at weighted comparisons.

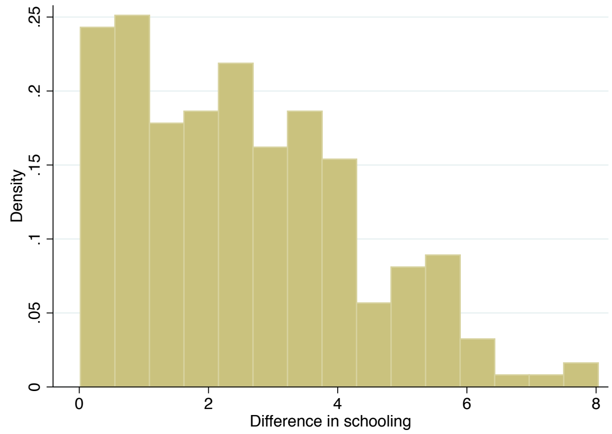
Figure 4: Differences in population across city pairs



Data source: Census 2000 PHC-T-3

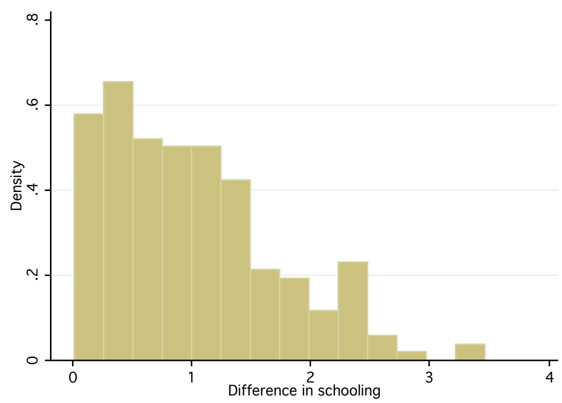
[35]Appendix C shows that, if $\ln f(\nu, c)$ is the sum of a log-supermodular function and idiosyncratic errors, the probability of obtaining the correct inequality is increasing in the product of the difference in log population and the difference in population elasticities. The latter are increasing in the difference in skill intensities in our theory.

Figure 5: Differences in skill intensities across occupational pairs



231 pairwise comparisons between 22 2-digit SOC occupational categories
Data source: 2000 Census of Population microdata via IPUMS-USA

Figure 6: Differences in skill intensities across industrial pairs



210 pairwise comparisons between 21 3-digit NAICS manufacturing industries
Data source: 2000 Census of Population microdata via IPUMS-USA

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 governed by this spatial pattern of skills. Our theory's predictions are more realistic than completely specialized or perfectly diversified cities and more specific than theories allowing arbitrary patterns of interindustry spillovers. Finally, we examine whether larger cities are larger in all industries or whether different industries attain their maximal employment at different points in the city-size distribution.

The data are broadly consistent with our novel predictions. Skill distributions regularly exhibit the monotone likelihood ratio property, although international migration plays an important role in the largest US cities that is omitted from our model. More skill-intensive sectors are relatively larger in more populous cities, on average. However, cities' sectoral distributions do not exhibit the monotone likelihood ratio property as often as cities' skill distributions do. One interpretation of this result could be that skill-driven comparative advantage plays an important role in determining the spatial pattern of production, but localization and coagglomeration economies may also play some role.[36] We show that there are not systematic violations of our predicted pattern of comparative advantage, and consistent with our model there is a strong tendency for larger cities to be larger in all industries.

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 4.1. We first 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.[37] 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.

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 4.1, implementing this test involves defining bins of cities. Ordering cities by population, we partition the

[36] Since the employment data do not distinguish employees by birthplace, another possibility is that the disproportionate presence of low-skill foreign-born individuals in larger cities influences sectoral composition in a manner not described by our theory. See footnote 52.

[37] Younger cohorts have higher average educational attainment. The results in Tables 3 and 4 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.

Table 3: Population elasticities of educational groups

Dep var: $\ln f(\omega, c)$	All	US-born
β_{ω_1} HS or less	0.96	0.90
$\times$ log population	(0.011)	(0.016)
β_{ω_2} Some college	1.00	0.97
$\times$ log population	(0.010)	(0.012)
β_{ω_3} BA or more	1.10	1.07
$\times$ log population	(0.015)	(0.017)

Standard errors, clustered by MSA, in parentheses.

Sample is all full-time, full-year employees residing in 270 metropolitan areas.

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.[38] Note that prior work typically describes a contrast between large and small cities for skilled and unskilled, whereas our most aggregated comparison is between large and small cities for three skill groups.

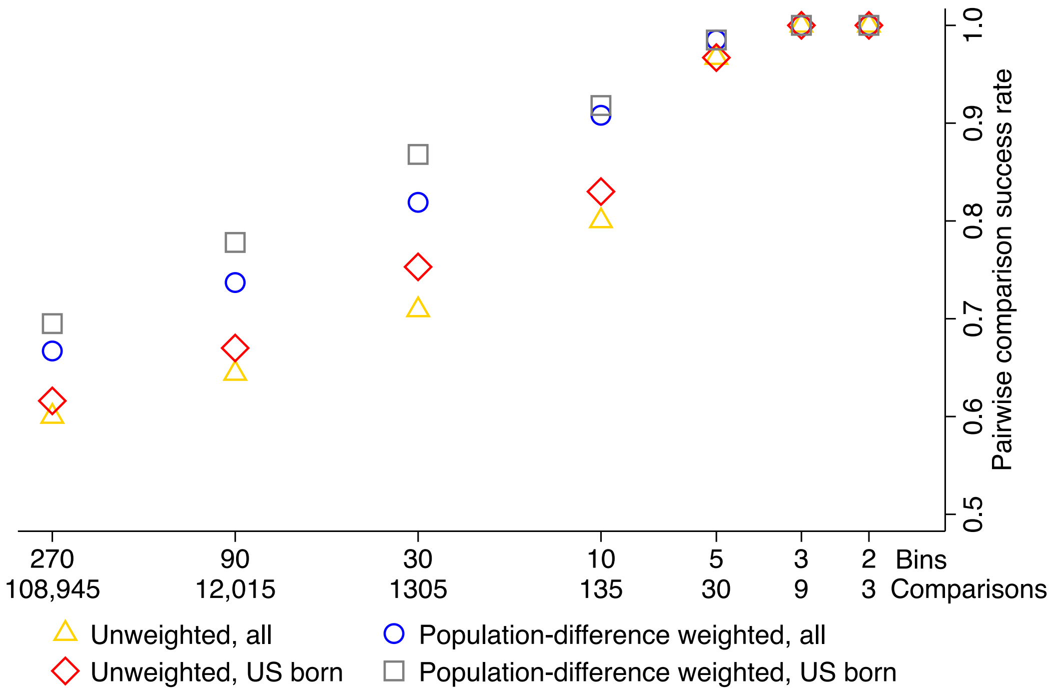
Figure 7 and Appendix Table 6 summarize the results of these tests using various sets of cities, weights, and birthplaces. In the unweighted comparisons, the success rate ranges from 60 percent when comparing individual cities to 97 percent when comparing five bins of cities to 100 percent for the standard case of two groups of cities. Weighting the comparisons by the population difference generally yields a higher success rate.[39] When we weight by population differences, the success rate is 67 percent when comparing individual cities, 98 percent for five bins of cities, and 100 percent for the simple comparison of large versus small cities.[40]

[38] With n city bins and m skill groups, we make $\frac{n(n-1)}{2}\frac{m(m-1)}{2}$ comparisons. For example, $\frac{270\times 269}{2}\frac{3\times 2}{2} = 108,945$.

[39] Despite the fact that the success rate of the Des-Moines-Kalamazoo comparisons is actually higher than the Chicago-Des-Moines comparisons.

[40] Our 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).

Figure 7: Pairwise comparisons of three skill groups



5.1.2 Nine skill groups

We next examine our tests for the case with nine skill groups. Starting with the elasticity test, Table 4 shows, contrary to our model's prediction, that those not completing high school are highly prevalent in larger cities. The second column reveals that this result is due to the presence of foreign-born individuals with low educational attainment in larger cities. If we restrict attention to US-born individuals, we can only reject the hypothesis that $\beta_\omega \geq \beta_{\omega'} \iff \omega \geq \omega'$ in one of thirty-six comparisons, the case where $\beta_{\omega_2} = 0.94 > 0.90 = \beta_{\omega_3}$.[41] In short, the elasticity test provides strong support for our theory when we examine the US-born population.[42]

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

[41] The estimated elasticities for the tails of the skill distribution have larger standard errors. This likely reflects greater sampling noise for scarce educational categories; for example, the median (C)MSA had 34 observations of full-time, full-year employees with a PhD in the 5 percent public-use 2000 Census microdata.

[42] Interestingly, among US-born individuals, the nine estimated elasticities naturally break into the three more aggregate educational attainment categories that we used above: $\beta_{\omega_1}, \beta_{\omega_2}, \beta_{\omega_3} \in (0.90, 0.94)$; $\beta_{\omega_4}, \beta_{\omega_5} \in (0.96, 0.98)$; $\beta_{\omega_6}, \beta_{\omega_7}, \beta_{\omega_8}, \beta_{\omega_9} \in (1.06, 1.09)$.

Table 4: Population elasticities of educational groups, 2000

Population share	Percent US-born	Dep var: $\ln f(\omega, c)$	Population elasticities All	Population elasticities US-born
.03	.21	β_{ω_1} Less than high school	1.17	0.91
		$\times$ log population	(0.039)	(0.028)
.07	.69	β_{ω_2} High school dropout	1.03	0.94
		$\times$ log population	(0.017)	(0.020)
.24	.87	β_{ω_3} High school graduate	0.93	0.90
		$\times$ log population	(0.013)	(0.016)
.24	.89	β_{ω_4} College dropout	1.00	0.98
		$\times$ log population	(0.011)	(0.013)
.08	.87	β_{ω_5} Associate's degree	1.00	0.96
		$\times$ log population	(0.014)	(0.016)
.21	.86	β_{ω_6} Bachelor's degree	1.10	1.07
		$\times$ log population	(0.015)	(0.017)
.08	.83	β_{ω_7} Master's degree	1.12	1.09
		$\times$ log population	(0.018)	(0.019)
.03	.81	β_{ω_8} Professional degree	1.12	1.09
		$\times$ log population	(0.018)	(0.019)
.01	.69	β_{ω_9} PhD	1.11	1.06
		$\times$ log population	(0.035)	(0.033)

Standard errors, clustered by MSA, in parentheses.
Sample is all full-time, full-year employees residing in 270 metropolitan areas.

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).[43]

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. Larger cities' benefits for immigrants merely serve as a "tie breaker" that causes the foreign-born to choose larger cities in equilibrium. This theory 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. More importantly, in 2000, foreign-born individuals constituted nearly 80 percent of the lowest skill group, while in

[43]Another potential mechanism is that immigrants may find larger cities' combination of higher nominal wages and higher housing prices more attractive than natives (Diamond, 2012), possibly because they remit their nominal incomes abroad or demand less housing than US-born individuals.

Table 5: Population elasticities of educational groups, 1980

Population share	Percent US-born	Dep var: $\ln f(\omega, c)$	Population elasticities All	Population elasticities US-born
.06	.67	β_{ω_1} Less than grade 9	0.99	0.89
		$\times$ log population	(0.028)	(0.030)
.11	.91	β_{ω_2} Grades 9-11	1.00	0.98
		$\times$ log population	(0.019)	(0.021)
.33	.94	β_{ω_3} Grade 12	0.97	0.95
		$\times$ log population	(0.013)	(0.015)
.08	.94	β_{ω_4} 1 year college	1.04	1.03
		$\times$ log population	(0.018)	(0.018)
.13	.92	β_{ω_5} 2-3 years college	1.09	1.07
		$\times$ log population	(0.018)	(0.018)
.13	.92	β_{ω_6} 4 years college	1.10	1.08
		$\times$ log population	(0.018)	(0.018)
.13	.90	β_{ω_7} 5+ years college	1.13	1.11
		$\times$ log population	(0.022)	(0.022)

Standard errors, clustered by MSA, in parentheses

Sample is full-time, full-year employees residing in 253 metropolitan areas.

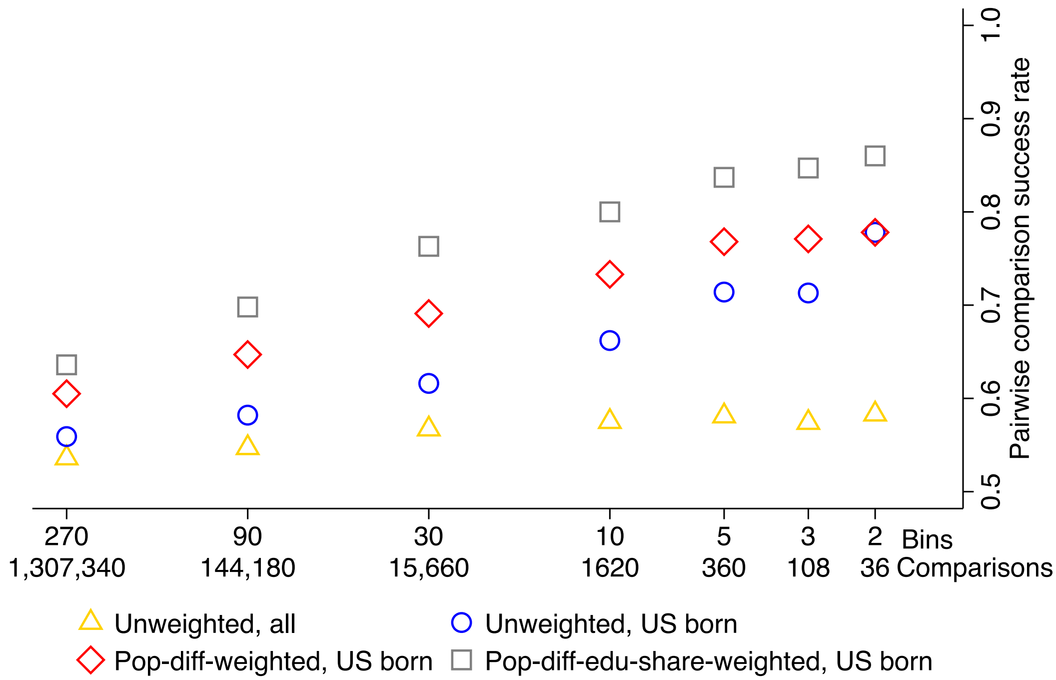
1980 they were only one third of the lowest skill group. If our hypothesis that foreign-born individuals are particularly attracted to larger cities is correct, then the population elasticity of less-skilled types should be lower when foreign-born shares are lower. Table 5 demonstrates that this is the case in 1980.[44] 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. The contrast between 1980 and 2000 for the population as a whole reflects the increasing foreign-born share in the least skilled groups.[45] Reconciling these results with the 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.

We now turn to the pairwise comparisons for the case with nine skill groups in 2000. These test inequalities for 36,315 city pairs for each pairing of the nine skill groups and are summarized in Figure 8 and Appendix Table 7. In both the unweighted and weighted comparisons, our theory does best in predicting comparisons of skill groups that have a high school degree or higher attainment. Fewer than 50 percent of the comparisons yield the correct inequality when the "less than high school" skill group is involved in the comparison.

[44]The educational categories in Table 5 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.

[45]The population elasticities of foreign-born individuals in the two least skilled categories are 1.4 and 1.3 in 1980 and 1.4 and 1.4 in 2000.

Figure 8: Pairwise comparisons of nine skill groups



As would be expected from the previous results, these comparisons are considerably more successful when restricted to the US-born population. When these are also weighted by population differences and education shares, the overall success rate in comparing individual cities rises to 64 percent.

Figure 8 and Appendix Table 8 show how the pairwise comparison success rate varies when we bin cities by size. When we restrict attention to the US-born, the unweighted success rate for individual cities, five bins, and two bins of cities are 56 percent, 71 percent, and 81 percent, respectively. If, in addition, we weight successes by population differences, the success rates for individual cities, five bins, and two bins of cities are 61 percent, 77 percent, and 81 percent, respectively. Weighting by population differences and education shares yields success rates of 64, 87, and 88 percent. In short, for the case of nine skill groups, the raw comparisons for individual cities including the foreign born show very modest success. As in the elasticities test, restricting attention to the US-born population yields significant improvement. Likewise, there is considerably greater success as we group cities and as we weight them by the overall prevalence of the education group in the labor force. Overall, we consider this solid support for our theory.

5.2 Larger cities specialize in skill-intensive sectors

This section examines the spatial pattern of sectoral employment. In our theory, larger cities are relatively more skilled, cities' equilibrium productivity differences are Hicks-neutral, and sectors can be ordered by their skill intensity, so larger cities employ relatively more labor in skill-intensive sectors. We established that larger cities are relatively more skilled in section 5.1. We now examine whether larger cities are relatively specialized in skill-intensive sectors. Since employment levels in both industries and occupations are readily available in the data, we test the employment implications of Corollary 1.[46]

5.2.1 The spatial distribution of occupations

We first implement the elasticities test and the pairwise comparisons test interpreting sectors as occupations. We begin with a visualization of the elasticity results. Figure 9 plots the 22 occupational categories' estimated population elasticities of employment against their skill intensities, measured as the average years of schooling of individuals employed in that occupation.[47] There is 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 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.[48] This hypothesis is rejected at the five-percent significance level in 46 comparisons, so the success rate is 80 percent.

The results for pairwise comparisons for occupations appear in Figure 10 and Appendix Table 10. When we do this for 276 cities and 22 occupations, we have a total of more than 8 million pairwise comparisons, of which 54 percent are correct.[49] This is low compared to our results for skills. When we stay with individual cities but weight by population and skill differences, this rises above 59 percent. We can maintain the weighting and consider it for cities grouped by size into, for example, 30, 5, or 2 bins. The corresponding proportion of successes rises respectively to 66, 76, and 78 percent. While the results for occupations

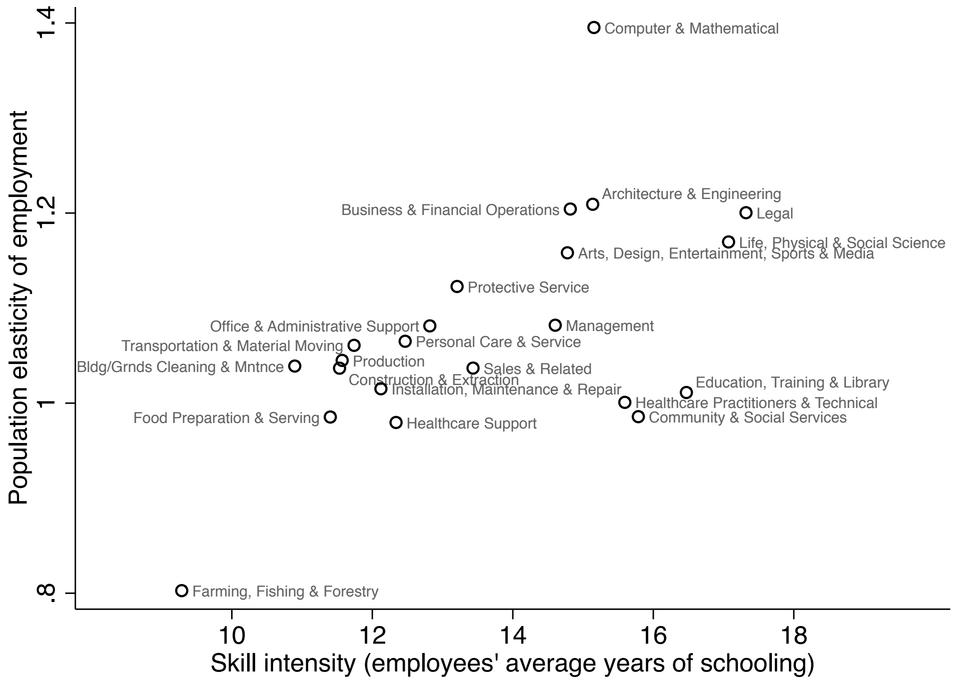
[46]Section 5.1 showed that US-born individuals better match our model's predictions about the distribution of skills. Unfortunately, the County Business Patterns and Occupational Employment Statistics data describe employment counts, not individual employees' characteristics, so we cannot address the birthplace issues in this section.

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

[48]The elasticity estimates appear in Appendix Table 9.

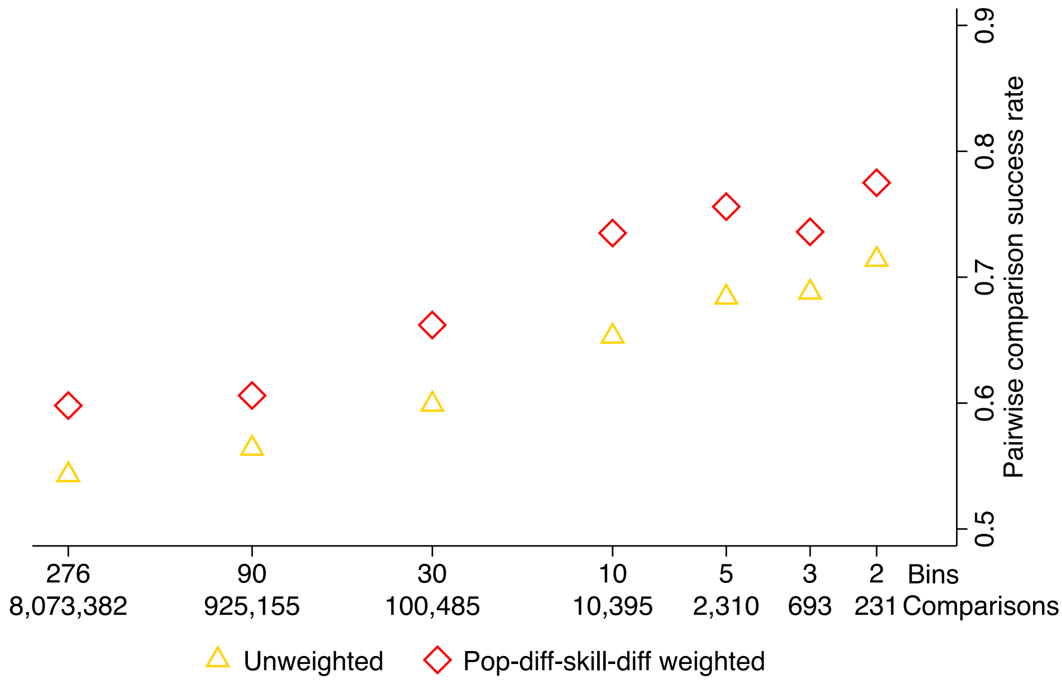
[49]These pairwise comparisons omit zero-employment observations. The results obtained when including those observations are similar.

Figure 9: Occupations' population elasticities and skill intensities



are not as strong as the results for skills, there are nonetheless quite informative patterns – even when we group cities into five size-based bins, we get three quarters of the pairwise comparisons correct across the 22 occupational categories.

Figure 10: Pairwise comparisons of 22 occupational categories



5.2.2 The spatial distribution of industries

We now implement the elasticities test and the pairwise comparisons test interpreting sectors as manufacturing industries.[50] A visualization of the elasticity test appears in Figure 11.[51] Again, as predicted by our theory, there is a clear positive relationship so that the population elasticity of industry employment is rising with the skill intensity of the industry. The apparel industry is an outlier, with low average education and a high population elasticity of employment. This may reflect the share of apparel industry employees who are less-skilled foreign-born individuals, consistent with our previous discussion of skills.[52] Testing the hypothesis that $\beta_\sigma \geq \beta_{\sigma'} \iff \sigma \geq \sigma'$ for the 21 manufacturing industries involves 210 ($= 21 \times 20/2$) comparisons of these estimated elasticities.[53] This hypothesis is rejected in 26 comparisons, so the elasticity implication holds true for manufacturing industries about

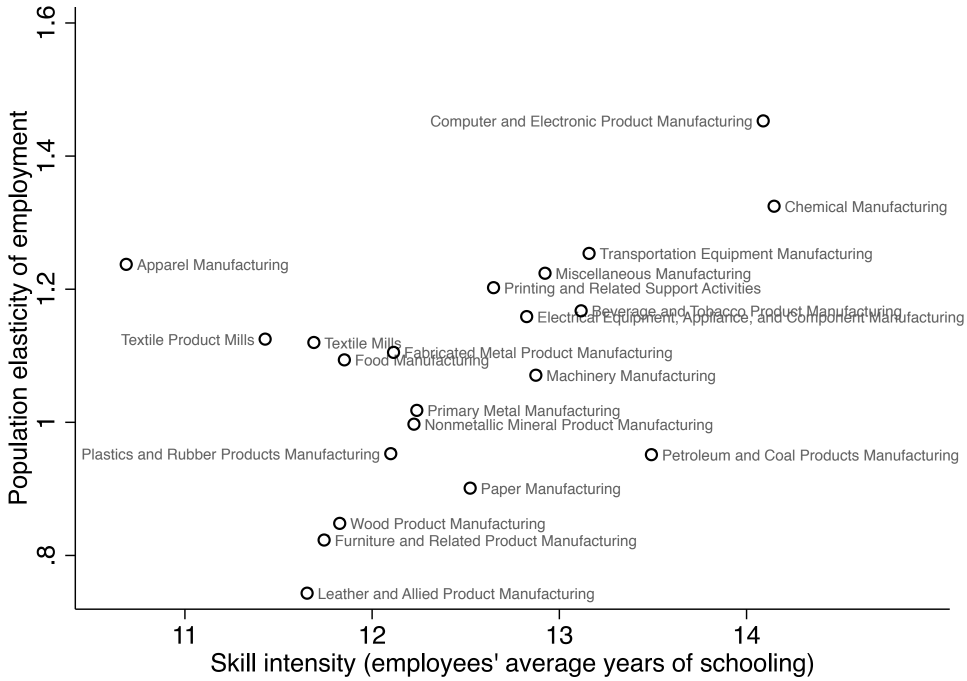
[50]We focus on manufacturing industries since we believe they have the lowest trade costs, but we have found broadly similar results when using all industries.

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

[52]Fifty-seven percent of apparel-manufacturing employees are foreign-born. Fifty-eight percent of foreign-born and 15% of US-born apparel-manufacturing employees have less than a high-school degree.

[53]The elasticity estimates appear in Appendix Table 11.

Figure 11: Industries' population elasticities and skill intensities



87 percent of the time.[54] This success rate is higher than the corresponding statistic for occupational elasticities.

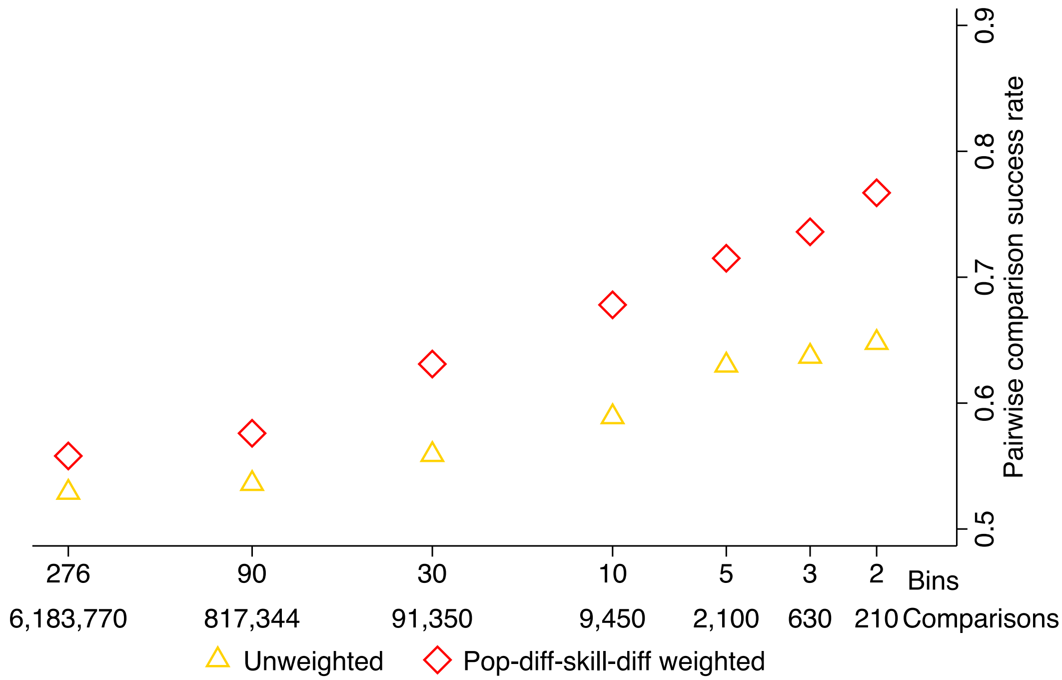
The pairwise comparisons results for industries appear in Figure 12 and Appendix Table 12. When we do this for 276 individual cities and 21 industries, we have a total of more than 6 million pairwise comparisons, of which just over half are correct.[55] Weighting this by skill and population differences raises this to 56 percent, again low compared to our results for pairwise comparisons of skills. We can maintain the weighting and consider this for cities grouped by size into 30, 5, or 2 groups. The corresponding proportion of successes rises respectively to 63, 72, and 77 percent. These are modestly low relative to the prior results on occupations and even more so relative to the results on skills. Nonetheless, they do show that there is systematic variation across cities of different sizes in the composition of manufacturing.[56] Note that prior work contrasting large and medium-size cities, Henderson (1997), is analogous to our comparisons of two or three groups of cities ordered by population.

[54]If we restrict the data to uncensored observations, which reduces the sample considerably, this hypothesis is rejected in 32 comparisons, for an 85 percent success rate. See appendix D for a discussion of censoring in County Business Patterns data.

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

[56]These results are not driven solely by the largest metropolitan areas; excluding the ten largest cities from pairwise comparisons of occupations and industries yields similar success rates.

Figure 12: Pairwise comparisons of 21 manufacturing industries



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.

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 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, with cross-city variation in size reflecting cross-industry variation in the strength of agglomeration economies. Our theory, by contrast, focuses on urbanization economies and allows that large cities may be the largest site of economic activity for all sectors (Corollary 3). Our empirical exercise asks what weight should be placed on the predictions flowing from each of these archetypes.

In the data, larger cities tend to have larger sectoral employment in all activities. This tendency is clear from the population elasticities plotted in Figures 9 and 11, as they are all strongly positive. Amongst the 21 3-digit NAICS manufacturing industries, the prediction that $c > c' \Rightarrow f(\sigma, c) \geq f(\sigma, c')$ is true in 77 percent of 796,950 cases. Sixteen manufacturing industries attain their maximal size in the three largest of 276 metropolitan areas (New York, Los Angeles, and Chicago), and all but one do so in the ten largest cities.[57] The exception is textile mills, which employ the most people in the Greenville-Spartanburg-Anderson metropolitan area, the 52nd largest city. The analogous results for occupational categories show an even stronger tendency for larger cities to have higher employment levels in all occupations.[58]

These findings are more consistent with urbanization economies than localization mechanisms at the city level. While particular examples such as South Carolina's concentration of textile mills are consistent with localization economies, the typical manufacturing industry exhibits 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.

[57] We find similar results for 4-digit NAICS manufacturing industries, though there is considerably more censoring at this more disaggregated level of observation. Larger cities have larger employment levels in 80 percent of comparisons. Sixty percent of 4-digit manufacturing industries attain their maximal size in the five largest cities, and nearly 80 percent do so in the fifty largest cities.

[58] The $c > c' \Rightarrow f(\sigma, c) \geq f(\sigma, c')$ prediction holds true in 88 percent of occupational comparisons, and 19 of the 22 occupations attain their maximal size in the largest metropolitan area, New York.

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 more precise 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 equal-sized categories of educational attainment are quite strong. Even disaggregated to nine educational categories, the cross-city distribution of US-born individuals is well described by our theory.

Empirically, we find that larger cities specialize relatively in skill-intensive activities. More skill-intensive occupations and industries tend to have higher population elasticities of employment. In making pairwise comparisons, our model does better in describing the pattern of occupational employment than industrial employment. This is consistent with a recent emphasis in the literature on workers performing tasks. Our results demonstrate that metropolitan skill distributions shape the comparative advantage of cities. Consistent with our approach based on urbanization economies, larger cities tend to have larger absolute employment in all industries.

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. 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.

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.

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., 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.

Brueckner, J. K. (1987): "The structure of urban equilibria: A unified treatment of the muth-mills model," in *Handbook of Regional and Urban Economics*, ed. by E. S. Mills, vol. 2, chap. 20, pp. 821–845. Elsevier.

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.

——— (2014): "Beyond Ricardo: Assignment Models in International Trade," mimeo.

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. (2012): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," mimeo.

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.

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.

Eeckhout, 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.

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.

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.

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.

——— (2012): "Coagglomeration and the Scale and Composition of Clusters," UC Berkeley mimeo.

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.

Henderson, V. (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.

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.

Lucas, R. E., and E. Rossi-Hansberg (2002): "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445–1476.

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. (1975): "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica*, 43(3), 455–68.

——— (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.

Trefler, D. (1995): "The Case of the Missing Trade and Other Mysteries," *American Economic Review*, 85(5), 1029–46.

von Thünen, J. H. (1826): *Der Isolierte Staat in Beziehung auf Landschaft und Nationalökonomie.*

A Appendix: 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

$$q(c, \tau, \sigma; \omega) = A(c)H(\omega, \sigma) \tag{13}$$

$$U(c, \tau, \sigma; \omega) = T(\tau)\left[A(c)H(\omega, \sigma)p(\sigma) - r(c, \tau)\right] \tag{14}$$

where $T(\tau)$ determines the value of the individual's disposable income after paying his or her locational price.[59] 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.

[59] Recall that the final good is the numeraire.

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 $K : \Gamma \to \Omega$, $\bar{\gamma}$, 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 that satisfies Costinot's Definition 4 (see footnote 23), so Corollary 2 holds true. As a result, the predictions about cities' population, sectors, and productivities described in sections 3.5 and 3.6 are unaltered by interpreting $T(\tau)$ as describing consumption benefits rather than production benefits.

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 $r(c,\bar{\tau}(c)) = 0$. □

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

$$A(c)T(\tau')G(\omega') - r(c, \tau') \geq A(c)T(\tau)G(\omega') - r(c, \tau)$$
$$A(c)T(\tau)G(\omega) - r(c, \tau) \geq A(c)T(\tau')G(\omega) - r(c, \tau')$$

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 $\{\tau_1, \ldots, \tau_j\} \equiv \mathcal{T}_0^j \subset \mathcal{T}_0$ for which $\bar{\omega}(\tau_i) - \underline{\omega}(\tau_i) \geq (\bar{\omega} - \underline{\omega})/j$ for $i = 1, \ldots, j$. 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

$$f(\omega, c, \tau) = f(\omega)\delta^{Dirac}[1 - \mathbf{1}_{\Omega(\tau)}] \quad \text{for } \mu\text{-almost all } \omega \in \Omega(\tau), \tag{15}$$

where δ^{Dirac} is a Dirac delta function. Combining equations (9) and (15) with $\mu[\Omega(\tau)] > 0$ yields $S'(\tau) = +\infty$, which contradicts our assumptions about $S(\tau)$.

Step 5 means there is a function $N : \mathcal{T} \to \Omega$ 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 $d\tau \to 0$, 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 $d\gamma \to 0$, we obtain $\frac{\partial r_\Gamma(\gamma)}{\partial \gamma} = G(K(\gamma))$. Integrating from $\underline{\gamma}$ to γ and using the boundary condition $r_\Gamma(\gamma) = 0$ yields $r_\Gamma(\gamma) = \int_{\underline{\gamma}}^{\gamma} G(K(x))dx$. □

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 $d\omega \to 0$ yields the population of ω in c. Using the definition of $s(\gamma, c)$ yields the desired expression. □

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 $d\sigma \to 0$ yields the population employed in σ in c. Using the definition of $s(\gamma, c)$ yields the desired expression.

□

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

Lemma 8. *Let $f(z) : \mathbb{R} \to \mathbb{R}^{++}$ and $g(x, y) : \mathbb{R}^2 \to \mathbb{R}^{++}$ 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

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

$\square$

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 $\frac{\gamma}{A(c)}$ is submodular and log-modular in (γ, c). Therefore, by lemma 8, $s(\gamma, c)$ is log-supermodular if and only if $V(z)$ has a decreasing elasticity. $\square$

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

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

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: Empirical Tests

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 $\mathcal{C}'$ ($\inf_{c \in \mathcal{C}} L(c) > \sup_{c' \in \mathcal{C}'} L(c')$), then

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

Suppose that the world is noisy. Consider the following form for $\ln f(\nu, c)$, 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

$$\begin{aligned} &\Pr\left(\sum_{c \in \mathcal{C}} \ln f(\nu, c) + \sum_{c' \in \mathcal{C}'} \ln f(\nu', c') \geq \sum_{c \in \mathcal{C}} \ln f(\nu', c) + \sum_{c' \in \mathcal{C}'} \ln f(\nu, c')\right) \\ &= \Pr\left(\sum_{c \in \mathcal{C}} \epsilon_{\nu',c} - \epsilon_{\nu,c} + \sum_{c' \in \mathcal{C}'} \epsilon_{\nu,c'} - \epsilon_{\nu',c'} \leq (\beta_\nu - \beta_{\nu'})(\sum_{c \in \mathcal{C}} \ln L_c - \sum_{c' \in \mathcal{C}'} \ln L_{c'})\right). \end{aligned}$$

To illustrate the properties of this probability, consider the special case in which the error term is normally distributed, $\epsilon_{\nu,c} \sim \mathcal{N}(0, \sigma^2)$, so that the sum of the error terms is also normally distributed.[60] Then this probability $\mathcal{P}$ is

$$\mathcal{P} = \Omega\left(\frac{\beta_\nu - \beta_{\nu'}}{\sqrt{2\sigma^2}} \frac{\sum_{c \in \mathcal{C}} \ln L_c - \sum_{c' \in \mathcal{C}'} \ln L_{c'}}{\sqrt{\sum_{c \in \mathcal{C}} 1 + \sum_{c' \in \mathcal{C}'} 1}}\right),$$

where $\Omega(\cdot)$ denotes the cumulative distribution function of $\mathcal{N}(0,1)$. The probability of obtaining the inequality depends on the difference in population size $\left(\frac{\sum_{c \in \mathcal{C}} \ln L_c - \sum_{c' \in \mathcal{C}'} \ln L_{c'}}{\sqrt{\sum_{c \in \mathcal{C}} 1 + \sum_{c' \in \mathcal{C}'} 1}}\right)$, the difference in population elasticities $(\beta_\nu - \beta_{\nu'})$, and the noisiness (σ^2) of the relationship. When the deterministic function is log-supermodular ($c > c' \Rightarrow L_c \geq L_{c'}; \nu > \nu' \Rightarrow \beta_\nu \geq \beta_{\nu'}$), $\mathcal{P} \to 1$ as $\sigma^2 \to 0$ (and $\mathcal{P} \to 1/2$ as $\sigma \to \infty$). When the function is log-modular, $\mathcal{P} \to 1/2$ as

[60] This specification assumes that the errors are homoscedastic. Empirically, we estimate that the residuals' variance does not increase with population size; if anything, there is a slight negative relationship. Even in the heteroscedastic case in which the variance increases with population size, with $\epsilon_{\nu,c} \sim \mathcal{N}(0, \sigma^2 \ln L_c)$ and $\mathcal{P} = \Omega\left(\frac{\beta_\nu - \beta_{\nu'}}{\sqrt{2\sigma^2}} \frac{\sum_{c \in \mathcal{C}} \ln L_c - \sum_{c' \in \mathcal{C}'} \ln L_{c'}}{\sqrt{\sum_{c \in \mathcal{C}} \ln L_c + \sum_{c' \in \mathcal{C}'} \ln L_{c'}}}\right)$, the probability $\mathcal{P} = \Omega\left(\frac{\beta_\nu - \beta_{\nu'}}{\sqrt{2\sigma^2}} \cdot \sqrt{n_{\mathcal{C}}} \cdot \frac{\ln \bar{L}_{\mathcal{C}} - \ln \bar{L}_{\mathcal{C}'}}{\sqrt{\ln \bar{L}_{\mathcal{C}} + \ln \bar{L}_{\mathcal{C}'}}}\right)$ exhibits the same desirable properties as the homoscedastic case.

$\sigma^2 \to 0$, and when the function is log-submodular, $\mathcal{P} \to 0$ as $\sigma^2 \to 0$.

This probability increases with aggregation by summation. Denote the number of elements in $\mathcal{C}$ and $\mathcal{C}'$ by $n_{\mathcal{C}}$ and $n_{\mathcal{C}'}$, respectively, and define the average $\ln \bar{L}_{\mathcal{C}'} \equiv \frac{1}{n_{\mathcal{C}'}} \sum_{c' \in \mathcal{C}'} \ln L_{c'}$. If we aggregate into bins with equal numbers of cities so that $n_{\mathcal{C}} = n_{\mathcal{C}'}$, the probability of obtaining the inequality simplifies to

$$\mathcal{P} = \Omega \left(\frac{\beta_\nu - \beta_{\nu'}}{\sqrt{2\sigma^2}} \cdot \sqrt{n_{\mathcal{C}}} \cdot (\ln \bar{L}_{\mathcal{C}} - \ln \bar{L}_{\mathcal{C}'}) \right),$$

which is increasing in the number of cities in the "bin".

Thus, 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. The figures in section 1 suggest that these are reasonable approximations. They also show noise, such that $\sigma^2 \gg 0$, so we should not expect the pairwise comparison test to have a 100% success rate.

D 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 Burea of Labor Statistics website.

Geography: We use (consolidated) metropolitan statistical areas as defined by the OMB as our 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 253 consolidated MSAs for the regressions in Table 5.

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 includes all full-time, full-year prime-age workers, defined as individuals 25 to 55 years of age who reported working at least 35 hour per week and 40 weeks in the previous year. Using the "educd" variable from IPUMS, we construct nine levels of educational attainment: less than high school, high school dropout, high school graduate, some college, associate's degree, bachelor's degree, master's degree, professional degree, and doctorate. There is at least one observation in every educational category in every metropolitan area.

Sectoral skill intensity: Using the same sample of full-time, full-year prime-age workers, 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 3-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 96 3-digit NAICS industries, of which 21 are manufacturing industries. 75 of these industries, including all 21 manufacturing industries, appear in both the Census of Population microdata and the County Business Patterns data. 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 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 390 (C)MSA-manufacturing-industry pairs, out of $5796 = 21 \times 276$, in which there are zero establishments. The County Business Patterns data omit self-employed individuals and employees of private households, railroads, agriculture production, the postal service, and public administrations. See the CBP methodology webpage for details.

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

E Appendix: Tables

Table 6: Pairwise comparisons of three skill groups

Bins	Birthplace	Weights	College vs some college	College vs HS or less	Some college vs HS or less	Total comparisons	Average
2	All	Unweighted	1	1	1	3	1
2	All	Population difference	1	1	1	3	1
2	US-born	Unweighted	1	1	1	3	1
2	US-born	Population difference	1	1	1	3	1
3	All	Unweighted	1	1	1	9	1
3	All	Population difference	1	1	1	9	1
3	US-born	Unweighted	1	1	1	9	1
3	US-born	Population difference	1	1	1	9	1
5	All	Unweighted	1	1	.900	30	.967
5	All	Population difference	1	1	.955	30	.985
5	US-born	Unweighted	1	1	.900	30	.967
5	US-born	Population difference	1	1	.955	30	.985
10	All	Unweighted	.844	.844	.711	135	.800
10	All	Population difference	.927	.944	.852	135	.908
10	US-born	Unweighted	.844	.889	.756	135	.830
10	US-born	Population difference	.927	.956	.870	135	.918
30	All	Unweighted	.768	.726	.632	1305	.709
30	All	Population difference	.887	.853	.716	1305	.819
30	US-born	Unweighted	.782	.784	.694	1305	.753
30	US-born	Population difference	.893	.898	.812	1305	.868
90	All	Unweighted	.684	.667	.58	12,015	.644
90	All	Population difference	.804	.779	.627	12,015	.737
90	US-born	Unweighted	.679	.693	.639	12,015	.670
90	US-born	Population difference	.799	.809	.727	12,015	.778
270	All	Unweighted	.629	.616	.556	108,945	.600
270	All	Population difference	.717	.695	.588	108,945	.667
270	US-born	Unweighted	.624	.635	.589	108,945	.616
270	US-born	Population difference	.712	.726	.647	108,945	.695

Table 7: Pairwise comparisons of nine skill groups with one city per bin

Unweighted comparisons								
	LHS	HSD	HS	CD	AA	BA	MA	Pro
HSD	.423							
HS	.399	.413						
CD	.428	.486	.587					
AA	.43	.483	.571	.483				
BA	.476	.555	.644	.619	.602			
MA	.484	.558	.643	.614	.615	.528		
Pro	.484	.57	.645	.617	.604	.524	.499	
PhD	.49	.548	.598	.576	.577	.521	.501	.511
Population-difference weighted comparisons of US born								
	LHS	HSD	HS	CD	AA	BA	MA	Pro
HSD	.568							
HS	.488	.435						
CD	.583	.569	.649					
AA	.552	.53	.616	.453				
BA	.644	.65	.738	.695	.682			
MA	.648	.651	.738	.686	.695	.544		
Pro	.654	.654	.73	.676	.676	.533	.493	
PhD	.611	.605	.651	.605	.617	.502	.476	.497

Table 8: Pairwise comparisons of nine skill groups

Bins	Weights	Total comparisons	All individuals success rate	US-born success rate
2	Unweighted	36	.583	.778
2	Population difference	36	.583	.778
2	Pop diff x edu shares	36	.766	.860
3	Unweighted	108	.574	.713
3	Population difference	108	.609	.771
3	Pop diff x edu shares	108	.773	.847
5	Unweighted	360	.581	.714
5	Population difference	360	.622	.768
5	Pop diff x edu shares	360	.768	.837
10	Unweighted	1620	.575	.662
10	Population difference	1620	.620	.733
10	Pop diff x edu shares	1620	.742	.800
30	Unweighted	15,660	.567	.616
30	Population difference	15,660	.610	.691
30	Pop diff x edu shares	15,660	.719	.763
90	Unweighted	144,180	.547	.582
90	Population difference	144,180	.585	.647
90	Pop diff x edu shares	144,180	.666	.698
270	Unweighted	1,307,340	.536	.559
270	Population difference	1,307,340	.564	.605
270	Pop diff x edu shares	1,307,340	.614	.636

Table 9: Occupational employment population elasticities

β_{σ_1} Farming, Fishing, and Forestry Occupations	0.803	$\beta_{\sigma_{12}}$ Sales and Related Occupations	1.037
× log population	(0.048)	× log population	(0.010)
β_{σ_2} Building and Grounds Cleaning and Maintenance	1.039	$\beta_{\sigma_{13}}$ Management occupations	1.082
× log population	(0.011)	× log population	(0.015)
β_{σ_3} Food Preparation and Serving Occupations	0.985	$\beta_{\sigma_{14}}$ Arts, Design, Entertainment, Sports, and Media	1.158
× log population	(0.011)	× log population	(0.019)
β_{σ_4} Construction and Extraction Occupations	1.037	$\beta_{\sigma_{15}}$ Business and Financial Operations Occupations	1.204
× log population	(0.014)	× log population	(0.018)
β_{σ_5} Production Occupations	1.045	$\beta_{\sigma_{16}}$ Architecture and Engineering Occupations	1.209
× log population	(0.025)	× log population	(0.026)
β_{σ_6} Transportation and Material Moving Occupations	1.061	$\beta_{\sigma_{17}}$ Computer and Mathematical Occupations	1.395
× log population	(0.014)	× log population	(0.034)
β_{σ_7} Installation, Maintenance, and Repair Workers	1.015	$\beta_{\sigma_{18}}$ Healthcare Practitioners and Technical Occupations	1.001
× log population	(0.011)	× log population	(0.014)
β_{σ_8} Healthcare Support Occupations	0.980	$\beta_{\sigma_{19}}$ Community and Social Services Occupations	0.986
× log population	(0.013)	× log population	(0.020)
β_{σ_9} Personal Care and Service Occupations	1.065	$\beta_{\sigma_{20}}$ Education, Training, and Library Occupations	1.011
× log population	(0.017)	× log population	(0.017)
$\beta_{\sigma_{10}}$ Office and Administrative Support Occupations	1.081	$\beta_{\sigma_{21}}$ Life, Physical, and Social Science Occupations	1.170
× log population	(0.010)	× log population	(0.030)
$\beta_{\sigma_{11}}$ Protective Service Occupations	1.123	$\beta_{\sigma_{22}}$ Legal Occupations	1.200
× log population	(0.014)	× log population	(0.022)
Observations		5943	
R-squared		0.931	
Occupation fixed effects		Yes	

Standard errors, clustered by MSA, in parentheses

Table 10: Pairwise comparisons of occupations

Bins	Weights	Comparisons	Success rate
2	Unweighted	231	.714
2	Population difference	231	.714
2	Population difference × skill difference	231	.775
3	Unweighted	693	.688
3	Population difference	693	.694
3	Population difference × skill difference	693	.736
5	Unweighted	2,310	.684
5	Population difference	2,310	.71
5	Population difference × skill difference	2,310	.756
10	Unweighted	10,395	.653
10	Population difference	10,395	.689
10	Population difference × skill difference	10,395	.735
30	Unweighted	100,485	.599
30	Population difference	100,485	.628
30	Population difference × skill difference	100,485	.662
90	Unweighted	925,155	.564
90	Population difference	925,155	.582
90	Population difference × skill difference	925,155	.606
276	Unweighted	8,073,382	.543
276	Population difference	8,073,382	.571
276	Population difference × skill difference	8,073,382	.598

Note: The number of cities per "bin" may differ by one, due to the integer constraint.

Table 11: Industrial employment population elasticities

β_{σ_1} Apparel Manufacturing	1.237	1.024	$\beta_{\sigma_{11}}$ Nonmetallic Mineral Product Manufacturing	1.018	0.955
× log population	(0.070)	(0.148)	× log population	(0.036)	(0.042)
β_{σ_2} Textile Product Mills	1.125	0.905	$\beta_{\sigma_{12}}$ Paper Manufacturing	0.901	0.539
× log population	(0.056)	(0.135)	× log population	(0.063)	(0.104)
β_{σ_3} Leather and Allied Product Manufacturing	0.743	0.147	$\beta_{\sigma_{13}}$ Printing and Related Support Activities	1.202	1.122
× log population	(0.099)	(0.284)	× log population	(0.036)	(0.047)
β_{σ_4} Furniture and Related Product Manufacturing	1.120	1.000	$\beta_{\sigma_{14}}$ Electrical Equipment, Appliance & Component	1.159	0.813
× log population	(0.050)	(0.076)	× log population	(0.074)	(0.111)
β_{σ_5} Textile Mills	0.823	0.352	$\beta_{\sigma_{15}}$ Machinery Manufacturing	1.071	0.960
× log population	(0.105)	(0.208)	× log population	(0.055)	(0.069)
β_{σ_6} Wood Product Manufacturing	0.848	0.608	$\beta_{\sigma_{16}}$ Miscellaneous Manufacturing	1.224	1.208
× log population	(0.055)	(0.085)	× log population	(0.044)	(0.059)
β_{σ_7} Fabricated Metal Product Manufacturing	1.094	1.036	$\beta_{\sigma_{17}}$ Beverage and Tobacco Product Manufacturing	1.168	1.010
× log population	(0.048)	(0.050)	× log population	(0.065)	(0.147)
β_{σ_8} Food Manufacturing	0.953	0.864	$\beta_{\sigma_{18}}$ Transportation Equipment Manufacturing	1.254	0.940
× log population	(0.050)	(0.067)	× log population	(0.075)	(0.101)
β_{σ_9} Plastics and Rubber Products Manufacturing	1.105	0.975	$\beta_{\sigma_{19}}$ Petroleum and Coal Products Manufacturing	0.951	0.393
× log population	(0.056)	(0.070)	× log population	(0.074)	(0.308)
$\beta_{\sigma_{10}}$ Primary Metal Manufacturing	0.997	0.449	$\beta_{\sigma_{20}}$ Computer and Electronic Product Manufacturing	1.453	1.254
× log population	(0.078)	(0.107)	× log population	(0.075)	(0.108)
			$\beta_{\sigma_{21}}$ Chemical Manufacturing	1.325	0.992
			× log population	(0.065)	(0.098)
Observations	5406	2130	Observations	5406	2130
R-squared	0.564	0.541	R-squared	0.564	0.541
Industry fixed effects	Yes	Yes	Industry fixed effects	Yes	Yes
Only uncensored observations		Yes	Only uncensored observations		Yes

Standard errors, clustered by MSA, in parentheses

Table 12: Pairwise comparisons of manufacturing industries

Bins	Weights	Comparisons	Success rate
2	Unweighted	210	.648
2	Population difference	210	.648
2	Population difference × skill difference	210	.767
3	Unweighted	630	.637
3	Population difference	630	.64
3	Population difference × skill difference	630	.736
5	Unweighted	2100	.63
5	Population difference	2100	.629
5	Population difference × skill difference	2100	.715
10	Unweighted	9450	.589
10	Population difference	9450	.604
10	Population difference × skill difference	9450	.678
30	Unweighted	91,350	.559
30	Population difference	91,350	.577
30	Population difference × skill difference	91,350	.631
90	Unweighted	817,344	.536
90	Population difference	817,344	.545
90	Population difference × skill difference	817,344	.576
276	Unweighted	6,183,770	.529
276	Population difference	6,183,770	.538
276	Population difference × skill difference	6,183,770	.558

Note: The number of cities per "bin" may differ by one, due to the integer constraint.