A Spatial Knowledge Economy*

Donald R. Davis†
Columbia University and NBER

Jonathan I. Dingel‡
Columbia University

July 19, 2012

Abstract

Leading empiricists and theorists of cities have recently argued that the generation and exchange of ideas must play a more central role in the analysis of cities. This paper develops the first system of cities model with costly idea exchange as the agglomeration force. Our model replicates a broad set of established facts about the cross section of cities. It provides the first spatial equilibrium theory of why skill premia are higher in larger cities, how variation in these premia emerges from symmetric fundamentals, and why skilled workers have higher migration rates than unskilled workers when both are fully mobile. (JEL: J24, J61, R01)

*This is a modest revision of NBER Working Paper No. 18188. We thank Pol Antras, Arnaud Costinot, Jessie Handbury, Walker Hanlon, Sam Kortum, Corinne Low, Ben Marx, Joan Monras, Suresh Naidu, Daniel Sturm, Eric Verhoogen, Reed Walker, David Weinstein, and seminar participants at the CESifo conference on heterogeneous firms in international trade, Columbia applied micro and international trade colloquia, NYU, Princeton IES Summer Workshop, Spatial Economic Research Centre annual conference, and University of Toronto for helpful comments on various drafts. We thank Paul Piveteau for research assistance. We are grateful to Enrico Moretti and Stuart Rosenthal for sharing their housing-price measures with us. Dingel acknowledges financial support from the Program for Economic Research at Columbia University.

†`drdavis@columbia.edu`

‡`jid2106@columbia.edu`

1 Introduction

Cities differ markedly. They differ in size, of course. But a large city is much more than the summation of many small towns. Larger cities have more educated populations and higher productivity, wages, housing prices, and inequality. These differences across cities are not external facts of nature. They are the result of hundreds of millions of individual decisions, each made in light of different cities offering different jobs, associates, earnings, and costs of living. What links these individual decisions to the aggregate outcomes we observe in the cross section of cities? That is the question we address in this paper.

In the last couple of decades, theorists have focused on the role of cities as loci for the exchange of goods as the agglomeration force in the cross section of cities. This is the "new economic geography" launched by Krugman (1991). Recently, however, important voices have argued that the exchange of ideas as an agglomeration force needs to take a more central role in the discussion. Notably, Krugman (2011, pp. 5-6) writes

> How can you de-emphasize technology and information spillovers in a world in which everyone's prime examples of localization are Silicon Valley and Wall Street?... The New Economic Geography style, its focus on tangible forces, seems less and less applicable to the actual location patterns of advanced economies.

Similarly, Glaeser and Gottlieb (2009, p. 983) write

> Some manufacturing firms cluster to reduce the costs of moving goods, but this force no longer appears to be important in driving urban success. Instead, modern cities are far more dependent on the role that density can play in speeding the flow of ideas.

This emphasis accords well with empirical evidence suggesting that wages are higher in larger cities for those with occupations emphasizing cognitive and people skills rather than motor skills and physical strength (Bacolod, Blum, and Strange, 2009). Studies also suggest that knowledge exchanges and communication skills are more common and more valuable in larger cities (Charlot and Duranton, 2004).

Economists have long understood that cities provide an opportunity to learn from others. Marshall (1890) wrote that in cities the "mysteries of the trade become no mysteries; but are as it were in the air." A seminal formalization of this idea treats learning in a city as a pure local externality (Henderson, 1974). But the influence of research on idea exchange as an agglomeration force has been limited by a "black box" critique. The difficulty is that if ideas are a pure externality, costlessly available to all in the city, then they are both evanescent in empirical terms and close to assuming your conclusion in theoretical terms (Fujita, Krugman, and Venables, 1999, p.4). To advance, we need models of idea exchange that, like the new economic geography, provide explicit microeconomic foundations.

We consider idea exchange among heterogeneous workers. A number of recent contributions have sought to explain differences in outcomes for skilled and unskilled workers across cities by appealing to exogenous differences in fundamental characteristics of those cities.[1] We instead follow the example of the new economic geography: spatial heterogeneity across cities emerges from perfectly symmetric fundamentals.[2]

Spatial heterogeneity is the outcome of individuals' locational choices. When individuals choose their locations freely and optimally, a system of cities is in "spatial equilibrium."[3] Glaeser and Gottlieb (2009) refer to spatial equilibrium as "the field's central theoretical tool." Similarly, Moretti (2011) stresses the importance of spatial equilibrium as a necessary condition for thinking about long-term spatial patterns. We note this because there is a counter tradition that uses observed differences in movement between skilled and unskilled workers as reason to assume that unskilled workers are immobile. In models of long-run spatial outcomes, differential movement should be a result, not an assumption.

In this paper, we develop the first system of cities model in which costly exchange of ideas is the agglomeration force. Our model is consistent with a broad set of established facts about the cross section of cities. It provides the first spatial-equilibrium account of why skill premia rise with city sizes. Our model also provides the first spatial-equilibrium account of how variation in such skill premia may arise from symmetric fundamentals. We provide the first explanation of why skilled workers move more than unskilled workers when both are mobile. Our approach is sufficiently flexible that it can be adapted to address a variety of questions about the spatial organization of activity within and between cities.

[1] For example, Glaeser (2008) and Beaudry, Doms, and Lewis (2010) model skill-segmented housing markets and skill-biased housing supplies to explain spatial variation in skilled wage premia. Gyourko, Mayer, and Sinai (2006) and Eeckhout, Pinheiro, and Schmidheiny (2010) model exogenous differences in housing supply elasticities and city-level productivities, respectively.

[2] We do not reject the idea that so-called "first nature" fundamental differences across locations have influenced and continue to influence population patterns. For example, Glaeser (2005) traces how the geographic advantages of the obscure Dutch trading outpost of New Amsterdam helped it become the colossus of New York City. But these are not the proximate forces that, for example, led Google to recently buy one of the city's largest buildings. Much more likely is that Manhattan provides Google with valuable opportunities to interact with others.

[3] Abdel-Rahman and Anas (2004) survey the literature on systems of cities.

Understanding the sources of differences in the cross section of cities is of considerable importance in its own right (Glaeser, 2008; Glaeser and Gottlieb, 2009). This importance is amplified by the fact that many fields of economics also use the cross section of cities and regions as a laboratory for testing theories beyond the traditional bounds of urban and regional economics.[4] A clearer understanding of the forces shaping key economic patterns in the cross section of cities will provide a stronger foundation for studies making use of this variation.

1.1 Idea exchange

Our model of idea exchange is in the spirit of Lucas (1988). He wrote

> Most of what we know we learn from other people. We pay tuition to a few of these teachers... but most of it we get for free, and often in ways that are mutual – without a distinction between student and teacher. (p.38)

We develop this in several respects. First, we make explicit that the knowledge acquired in these exchanges is not really free. The opportunity cost is time not devoted to other productive activities. In our model, agents choose their time allocation optimally. Second, since much knowledge is tacit, requiring face-to-face communication, we treat cities as the loci of learning communities.[5] Third, we use a continuous distribution of heterogeneous labor. Because what one has to offer other learners and what one can learn oneself varies across these individuals, spatial sorting of learners into distinct cities with distinct learning opportunities is quite natural. Finally, learning depends not only on the average ability of learners in one's community but also the mass of learners (cf. Glaeser 1999). A solitary genius is not enough.

Our approach unites two strands of literature on the exchange of ideas. One has focused on spatial choices of learning opportunities when knowledge spillovers are exogenous and freely available within a city (Henderson, 1974; Black, 1999). Another has focused on choices of learning activities within a single location of exogenous population (Helsley and Strange,

[4]Recent examples include Albouy (2009) on federal taxation of nominal income, Autor and Dorn (2012) on the polarization of jobs, Beaudry, Doms, and Lewis (2010) on the introduction of computers as a technological revolution, and Nakamura and Steinsson (2011) on fiscal stimulus in a monetary union.

[5]This is in line with Lucas's observation: "What can people be paying Manhattan or downtown Chicago rents for, if not for being near other people?" (Lucas, 1988, p.39) For more on how proximity facilitates knowledge transmission, see Jaffe, Trajtenberg, and Henderson (1993), Audretsch and Feldman (2004), and Arzaghi and Henderson (2008).

2004; Berliant, Reed III, and Wang, 2006; Berliant and Fujita, 2008; Lucas and Moll, 2011).[6] In our model, locational choices shape knowledge exchanges because learning opportunities are heterogeneous and depend upon the time-allocation decisions of the learners in each location. Our characterization of idea exchanges is simple compared to those presented in the second strand of literature, but this allows us to tractably model endogenous exchanges of ideas in a system of cities.

An issue that is unavoidable when considering endogenous idea exchange is how one will treat labor heterogeneity. One possibility is to work with purely homogeneous labor, so that all exchange is purely horizontal. A second possibility is to work with two classes of labor, skilled and unskilled. We take heterogeneity to its limit and consider a continuum of labor types. Common experience tells us that even PhDs from elite universities are highly heterogeneous in their knowledge and skills. This holds *a fortiori* when we consider the very wide range of labor in the economy as a whole. Moreover, we will show that this labor heterogeneity is of considerable analytic convenience in making sense of important features of the cross-city data.

1.2 Idea exchange and the cross section of cities

We embed our process of idea exchange in a perfectly competitive economic environment. Inter-city trade costs are zero or infinite. Cities are sites where producers interact in order to acquire productivity-increasing ideas. Our model features people who are heterogeneous in a single dimension. In the core model, there are two produced goods, tradables and non-tradables. Tradables production makes use of the underlying heterogeneity of individuals; non-tradables production does not. By comparative advantage, as in Roy (1951), high-ability individuals sort into the tradables sector. In the tradables sector, individuals can divide their time between directly producing the homogeneous tradable good and raising their productivity by exchanging ideas with others in their city who also devote time to learning. All tradables producers find attractions in large, high-ability cities where learning opportunities are greatest. However, congestion leads to high prices for housing and non-tradable services. A tradable producer's productivity gains from idea exchanges are supermodular in own ability and a city's learning opportunities, so tradables producers sort across cities.

[6] Glaeser (1999) is an important precursor to our approach. His model specifies two locations, a city and a rural hinterland. In contrast to our approach, the fundamental difference between the two locations is exogenous, since learning is possible only in the city.

Larger cities are populated by higher-ability individuals who, in equilibrium, devote more time to exchanging ideas. Non-tradables are produced in every city by the least able agents who are exactly compensated for cities' price differences in housing and non-tradables.

Our model matches a broad set of facts from the empirical literature. First, cities exhibit substantial heterogeneity in size, as required by the literature on the city-size distribution (Gabaix, 1999). While our model has symmetric fundamentals, it generically yields asymmetric outcomes. Second, these size differences are accompanied by differences in wages, housing prices, and productivity (Glaeser, 2008). Our model's agglomeration and congestion forces link these components together in equilibrium so that larger cities are more expensive and more productive. Third, while there is evidence that a meaningful share of spatial wage variation is attributable to spatial sorting of heterogeneous workers (Combes, Duranton, and Gobillon, 2008; Gibbons, Overman, and Pelkonen, 2010; De la Roca, 2012), this sorting is incomplete and individuals of many skill types are present in every city. The Roy-model component of our approach yields this imperfect sorting, since there is sorting within tradables producers but not within non-tradables producers. Fourth, people are highly mobile in advanced economies and respond to spatial arbitrage opportunities (Borjas, Bronars, and Trejo, 1992; Dahl, 2002; Notowidigdo, 2011). Our model follows the spatial-equilibrium tradition in assuming zero mobility costs.

Our emphasis on labor heterogeneity naturally yields predictions about spatial variation in wage inequality. Workers in the skilled tradables sector can raise their productivity by exchanging ideas. In equilibrium, larger cities offer more valuable idea-exchange environments, so higher-ability tradables producers locate there and benefit more from idea exchanges. Our focus on the within-group heterogeneity of skilled workers matches findings that attending a higher quality college is particularly associated with higher wages in larger cities (Bacolod, Blum, and Strange, 2009) and that larger cities exhibit greater within-group wage dispersion (Baum-Snow and Pavan, 2011). Making idea exchange among skilled tradables producers the agglomeration force also links cities' population sizes and skill premia. The spatial sorting of skilled tradables producers yields a positive premium-population relationship.

Theoretically linking together cities, ideas, and skill premia is non-trivial. Unlike temporal differences in wage premia, spatial differences in wage premia are disciplined by a no-arbitrage condition. As Glaeser (2008, p.85) notes, when people are mobile, differences in productivity "tend to show up exclusively in changes in quantities of skilled people, not in different returns to skilled people across space." The canonical spatial-equilibrium model, in which there are two homogeneous skill groups and preferences are homothetic, predicts that skill premia are spatially invariant (Black, Kolesnikova, and Taylor, 2009).

In short, spatial theory lags behind the empirical evidence. Glaeser, Resseger, and Tobio (2009, p.639) state that "we are much more confident that differences in the returns to skill can explain a significant amount of income inequality across metropolitan areas than we are in explaining why areas have such different returns to human capital." We provide an explanation by modeling cities, heterogeneous skills, and idea exchanges.

We are not aware of a prior spatial-equilibrium model that links skill premia to cities' sizes. Nor are we aware of a prior spatial-equilibrium model that generates spatial variation in skill premia from symmetric fundamentals. Previous system of cities models amended the canonical model by introducing spatial variation in fundamentals, namely skill-segmented housing markets and skill-biased housing supplies, in order to explain spatial variation in skill premia (Glaeser, 2008; Beaudry, Doms, and Lewis, 2010). These neoclassical models did not relate skill premia to city sizes.

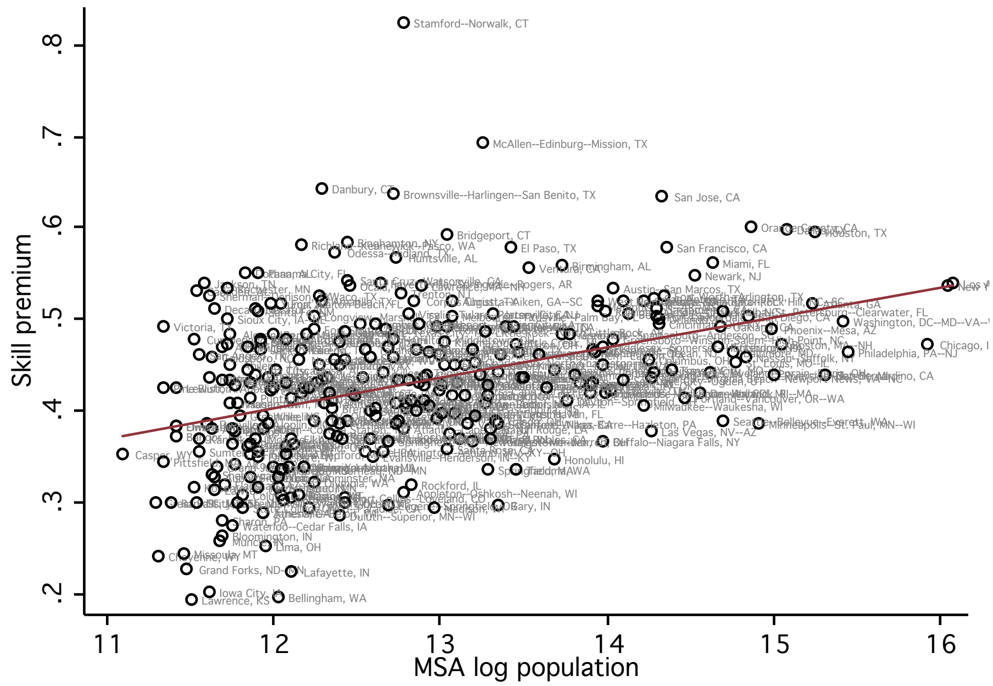
Our model's results about city size and wage inequality are related to recent theoretical work by Behrens, Duranton, and Robert-Nicoud (2010) and Behrens and Robert-Nicoud (2011). These authors also focus on labor heterogeneity by using a continuum of abilities. Their work differs in two important respects. First, they model agglomeration driven by the exchange of goods. This emphasis potentially complements our study of idea exchange. Second, their explanations of cross-city inequality differences stem from assuming that laborers make irreversible one-time locational choices.[7] Our model provides the first spatial-equilibrium explanation of these phenomena.

Wage inequality and city size are strongly linked in the data. Glaeser, Resseger, and Tobio (2009) and Behrens and Robert-Nicoud (2011) report that larger cities exhibit higher Gini coefficients; Baum-Snow and Pavan (2011) show that they have greater overall variance in nominal wages. In this paper, we focus on the skilled wage premium, a relative price that captures important dimensions of wage inequality. Figure 1 demonstrates that skill premia, measured as differences in average log weekly wages between college graduates and high school graduates, are higher in more populous metropolitan areas.[8] The scatterplot shows substantial cross-city variation in skill premia and that a large share of this variation

[7] All workers entering a city have identical abilities in these models. Upon choosing a city, workers randomly draw their productivity levels. Behrens and Robert-Nicoud (2011) note that allowing for mobility in their model "would imply that a city's equilibrium income distribution is independent of its size."

[8] Appendix Figure A1 of Baum-Snow and Pavan (2011) also appears to suggest this relationship.

Figure 1: Skill premia and metropolitan populations, 2000



NOTE: The skill premium is the difference in average log weekly wages between full-time, full-year employees whose highest educational attainment is a bachelor's degree and those whose is a high school degree in a (primary) metropolitan statistical area. See appendix C for a detailed description of the data and estimation.

is explained by cities' sizes. College wage premia range from about 47% in metropolitan areas with 100,000 residents to about 71% in places with 10 million residents.

Prior work on spatial variation in skill premia has studied how skill premia correlate with other city characteristics, such as the fraction of the population possessing a college degree (Glaeser, 2008; Glaeser, Resseger, and Tobio, 2009; Beaudry, Doms, and Lewis, 2010) or housing prices (Black, Kolesnikova, and Taylor, 2009). Table 1 shows that the positive premium-population relationship is robust to controlling for these other characteristics. Furthermore, the relationship does not depend on whether we measure skill premia controlling for individuals' observable characteristics or not. The positive correlation between cities' population sizes and skill premia is a robust, persistent, first-order feature of the data that requires a spatial-equilibrium explanation.[9]

[9]Regressions for 1990 and 2007 also demonstrate a strongly positive premium-population relationship. See appendix C.2. This spatial pattern does not appear to be a temporary or disequilibrium phenomenon.

Table 1: Skill premia and metropolitan characteristics, 2000

Skill premia				
log population	0.033**	0.030**	0.037**	0.027**
	(0.0038)	(0.0057)	(0.0043)	(0.0053)
log rent		0.032		0.12**
		(0.037)		(0.038)
log college ratio			-0.035	-0.075**
			(0.020)	(0.018)
R^2	0.161	0.165	0.178	0.213
Composition-adjusted skill premia				
log population	0.029**	0.030**	0.031**	0.029**
	(0.0032)	(0.0050)	(0.0037)	(0.0048)
log rent		-0.013		0.029
		(0.035)		(0.036)
log college ratio			-0.027	-0.037*
			(0.017)	(0.016)
R^2	0.161	0.162	0.176	0.179
Observations	325	325	325	325

Robust standard errors in parentheses
** p<0.01, * p<0.05

NOTE: Each column reports two OLS regressions. In the upper panel, the dependent variable is a metropolitan area's skill premium, measured as the difference in average log weekly wages between college and high school graduates. The lower panel uses composition-adjusted skill premia. See appendix C for a detailed description of the data and estimation.

1.3 Spatial equilibrium and skill patterns of migration

Our aim in this paper is to understand the spatial choices of skilled and unskilled workers as well as the observable, heterogeneous consequences of these choices. One prominent contrast between skilled and unskilled workers is that the skilled migrate more frequently than the unskilled (Greenwood, 1997; Molloy, Smith, and Wozniak, 2011). Table 2 demonstrates that prime working age US-born individuals who change residences are nearly 70% more likely to change metropolitan areas if they hold a bachelor's degree rather than just a high school degree. Moreover, bachelor's degree holders move farther when they change residences. The typical move of a college graduate is about 80% greater than that of a high school graduate.[10] Even if we compare only those who *change* metropolitan areas, college graduates move more than 25% farther than high school graduates.

[10] In this calculation, we assign a distance of zero to residence changes within the same public-use microdata area. See appendix C for details.

Table 2: Educational attainment and migration

	High school degree	Bachelor's degree
Different residence than five years prior	42%	48%
Different metropolitan area | different residence	19%	32%
Average distance (km) | different residence	204	365
Standard error	(0.9)	(1.4)
Average distance (km) | different metropolitan area	771	977
Standard error	(3.3)	(3.7)

NOTE: The sample is made up of US-born individuals ages 30–55 residing in metropolitan areas in the 2000 Census public-use microdata whose highest educational attainment is a bachelor's degree or a high school degree. See appendix C for details.

How shall we incorporate this contrast in movement of the skilled and unskilled into our thinking about spatial patterns of activity across cities? One answer is embodied in the Krugman (1991) core-periphery model, which translates the observation of differential *movement* into an assumption of differential *mobility*. This has been extremely influential in subsequent work and so deserves careful attention.[11] This has two key shortcomings. The first is that if the fundamental problem that one wants to address is the spatial pattern of economic activity, location has to be a choice, not an assumption.[12] Second, since many of these models assume that labor is homogeneous within a broad class, this also has important consequences for welfare. In particular, perfectly mobile skilled workers receive the same utility everywhere. Perfectly immobile unskilled workers receive utility that varies by location, but only because they are *assumed* unable to move.

We develop a simple dynamic extension of our model that considers costly migration in the limit as those common costs for skilled and unskilled workers converge to zero, that is, as we converge to full spatial equilibrium. We believe this extension provides important advances on the prior literature. The greater rate of movement of skilled than unskilled, as well as the greater average distance of moves by the skilled, is a result rather than an assumption. Moreover, because we can explain the facts in spatial equilibrium, our model does not rely on a failure of arbitrage to make sense of spatial welfare heterogeneity.[13]

[11] See, for example, Tabuchi and Thisse (2002) and Borck, Pflüger, and Wrede (2010). Autor and Dorn (2012) also make this assumption. Helpman (1998) shows how the results of Krugman (1991) are altered by modeling the centrifugal force as housing supplies rather than immobile "peasants."

[12] Assuming immobility precludes other explanations for lack of movement, a point underscored in Notowidigdo (2011).

[13] We recognize that short-run responses to economic shocks may be highly localized due to movement

2 A spatial knowledge economy

This section develops a simple spatial knowledge economy, explores the basic model's relation to important empirical regularities in the cross section of cities, and then extends it to a simple dynamic model of migration and outsourcing to explore differential movement of the skilled and unskilled.

2.1 Consumption

Individuals consume three goods: tradables, non-tradable services, and (non-tradable) housing. Services and housing are necessities; after consuming $\bar{s}$ units of non-tradable services and one unit of housing, consumers spend all of their remaining income on tradables, which we use as the numeraire.[14] The indirect utility function, therefore, for a consumer with income I facing prices p in city c is

$$V(p, I) = I_c - p_{s,c}\bar{s} - p_{h,c} \tag{1}$$

Consumers are perfectly mobile across cities and jobs, so their locational and occupational choices maximize $V(p, I)$.

2.2 Production

2.2.1 Housing and non-tradable services

Every system of cities model must have both agglomeration and congestion forces. Since our contribution will focus on the force for agglomeration, we model the congestion force in the most stripped-down way possible. Alonso (1964), Mills (1967), and Muth (1969) developed a simple model of the internal structure of the city in which residents commute from home to a central business district. We follow Behrens, Duranton, and Robert-Nicoud (2010) and introduce this in a standard form.[15] Each location is endowed with housing sites that serve

costs (Autor, Dorn, and Hanson, 2011). Still, we believe spatial equilibrium is the right starting point for an analysis of long-run spatial patterns, which may be stable across decades or longer. Moreover, one cannot measure the speed at which spatial arbitrage occurs without the baseline provided by a model in which such arbitrage is costless.

[14] This specification, in which consumers demand a fixed quantity of non-tradables, is also found in Glaeser, Gyourko, and Saks (2006) and Moretti (2011). We use it for analytical convenience; it is not crucial to our results. See also footnote 23.

[15] See appendix section A.1 for details.

as residences and that do not require any labor input. We will refer to $p_{h,c}$ as the consumer price of housing in city c, but the reader should keep in mind that this incorporates both land rents and commuting costs and is invariant across locations within a city. This yields a simple increasing relation between housing prices, $p_{h,c}$, and a city's population, L_c, of the form $p_{h,c} = \theta L_c^{\gamma}$, with $\theta, \gamma > 0$.

There is a mass L of workers of heterogeneous ability, indexed by z and distributed with density $\mu(z)$. They choose to produce non-tradables or tradables. Non-tradables can be produced at a uniform level of productivity by anyone employed in that sector. Tradables, by contrast, make use of the underlying heterogeneity. A person's productivity in tradables is $\tilde{z}(z, Z_c)$, which is increasing in z and depends on the learning opportunities available through interacting with others working in the tradable sector in that city, governed by Z_c (discussed below).

By comparative advantage, low-z people will specialize in producing non-tradables, which make no use of the underlying heterogeneity, while high-z people will specialize in tradables. Denote the marginal worker indifferent between the two sectors as z_m.

We choose units of output so that an individual's productivity in non-tradable services is unity. Since productivity in non-tradable services is independent of individual ability, the total output of services in a city is equal to the mass of agents working in services, $L_{s,c}$. The income of a non-tradables producer in city c is therefore $p_{s,c}$.

2.2.2 Idea exchange and tradables productivity

Tradables producers can acquire knowledge to increase their productivity. They do this by spending time interacting with other tradables producers in their city. Each person has one unit of time that they divide between interacting and producing. Production depends on own ability (z), time spent producing (β), time spent exchanging ideas ($1-\beta$), the productivity benefits of learning (A), and local learning opportunities (Z_c). Exchanging ideas is an economic decision, because time spent interacting ($1-\beta$) trades off with time spent producing output directly (β). The tradables output of an agent of ability z is

$$\tilde{z}(z, Z_c) = \max_{\beta \in [0,1]} \beta z(1 + (1-\beta) A Z_c z) \tag{2}$$

A is a parameter common to all locations that indexes the scope for productivity gains from interactions. When A is higher, conversations with other agents raise productivity more. Knowledge has both horizontal and vertical differentiation. Horizontal differentiation implies that producers can learn something from anyone. Vertical differentiation means that they learn more from more able counterparts.

Local learning opportunities Z_c are the result of a random-matching process in which producers devoting time to idea exchanges encounter other producers doing likewise. The expected value of devoting a unit of time to idea exchange in a city is the probability of encountering another individual times the expected ability of the individual encountered.

The probability of encountering a person during time spent seeking idea exchanges is $m(M_c)$, where M_c is the total time devoted to learning by producers in the city. $m(\cdot)$ is an increasing function, with $m(0) = 0$ and $m(\infty) = 1$. Like Glaeser (1999), we assume that face-to-face interactions occur with greater frequency in denser places, so that random matches occur more often in the central business districts of larger cities. In our setting the population of agents available for such encounters is determined endogenously by tradable producers' time-allocation choices.

The expected ability of the individual encountered is $\bar{z}_c$, the weighted average of the abilities of producers participating in idea exchanges. The weights are the time agents devote to interactions.[16] Conditional on meeting another learner, the scope for gains from interactions, and one's own ability, conversations with more talented agents are more productive.

Thus, the value of local learning opportunities Z_c reflects both a scale effect and an average ability effect. Consider city c with population ability distribution $\mu(z, c)$. When agents of ability z in city c devote $1 - \beta_{z,c}$ of their time to exchanging ideas, the value of idea exchange in city c is described by the following:

$$\begin{aligned}
Z_c &= m(M_c)\bar{z}_c \\
M_c &= L \int_{z \geq z_m} (1 - \beta_{z,c})\mu(z, c)\mathrm{d}z \\
\bar{z}_c &= \int_{z \geq z_m} \frac{(1 - \beta_{z,c})z}{\int_{z \geq z_m}(1 - \beta_{z,c})\mu(z, c)\mathrm{d}z}\mu(z, c)\mathrm{d}z
\end{aligned} \tag{3}$$

This characterization of idea exchanges as mutually beneficial meetings in which each party is both student and teacher follows Lucas (1988). The matching process that yields exchanges

[16]The expression for $\bar{z}_c$ in equation (3) is not well defined when $\beta_{z,c} = 1$ for all agents. We will define $\bar{z}_c = 0$ for the case in which no one invests in learning. This is not an average, of course, but it seems an appropriate definition that reflects the absence of opportunities to learn from others. The particular (finite) value assigned to $\bar{z}_c$ when $\beta_{z,c} = 1$ for all agents is immaterial, since $m(0) = 0$ and therefore $Z_c = 0$.

means that a city's population size and average ability both matter.

For an individual worker, the optimal time spent interacting is

$$1-\beta_{z,c}=\begin{cases}\frac{1}{2}\frac{AZ_cz-1}{AZ_cz} & \text{if } AZ_cz\geq 1\\ 0 & \text{otherwise}\end{cases}.$$

Conditional on the population in city c, which is described by the ability distribution $\mu(z,c)$, the equilibrium value of local idea exchanges Z_c is a fixed point defined by $Z_c = m(M_c)\bar{z}_c$, since individual choices of $\beta_{z,c}$, which determine M_c and $\bar{z}_c$, depend on the city-level Z_c.

Of course, there is also an equilibrium in which $Z_c = 0$, since no individual will allocate time to interacting with others when there are no others with whom to interact. While the no-learning equilibrium will not be the focus of our discussion, it does illustrate an important aspect of the economic mechanisms. It underscores the fact that learning here is not manna from heaven but the outcome of a costly allocation of time by those acquiring knowledge. Thus, larger cities are better learning environments because, in equilibrium, they offer a higher frequency of face-to-face interactions with a more talented population of partners, as we show below.

An individual allocates her time in order to maximize her income, so she solves the maximization problem described in equation (2). The tradable output of type z in city c with learning opportunities Z_c is

$$\tilde{z}(z,Z_c)=\begin{cases}\frac{1}{4AZ_c}\left(AZ_cz+1\right)^2 & \text{if } AZ_cz\geq 1\\ z & \text{otherwise}\end{cases}. \tag{4}$$

We have a few key conclusions. Tradables producers choose to engage other producers in encounters from which they both learn. This learning takes time away from direct production but maximizes their total output by raising their productivity. Time devoted to learning by a tradables producer is increasing in the time devoted to idea exchange by others, the scope for productivity gains from idea exchange, the average quality of other learners in that location, and the producer's own ability. Given this knowledge economy, we now characterize the patterns of economic outcomes in spatial equilibrium.

2.3 Equilibrium

This section develops the conditions for equilibrium in our spatial knowledge economy. Consumers optimally choose their city, occupation, and consumption. Tradables producers optimally allocate their time between direct production and idea exchange. Prices clear markets and the individual locational choices must be consistent with aggregate population measures. There are three types of equilibria: equilibria without idea exchange, equilibria with symmetric cities, and equilibria with heterogeneous cities. The latter are stable, match many empirical findings in the systems of cities literature, and will be our primary object of interest.

An equilibrium for a population L with talent distribution $\mu(z)$ in a set of locations $\{c\}$ is a set of prices $\{p_{h,c}, p_{s,c}\}$ and locational choices $\mu(z,c)$ such that workers optimize and markets clear.[17] Define the set of cities in which agents of ability z are found by $C(z) = \{c : \mu(z,c) > 0\}$. We can then write our equilibrium conditions as equations (5) through (14).

Equations (5) and (6) are adding-up constraints for worker types and city populations.

$$\mu(z) = \sum_c \mu(z,c) \quad \forall z \tag{5}$$

$$L_c = L \int \mu(z,c)\mathrm{d}z \quad \forall c \tag{6}$$

Equation (7) defines the land-market-clearing housing price within each city.

$$p_{h,c} = \theta L_c^{\gamma} \quad \forall c \tag{7}$$

Equation (8) equalizes demand and supply of non-tradable services within each location.

$$L_{s,c} = L \int_{z \leq z_m} \mu(z,c)\mathrm{d}z = \bar{s} L_c \quad \forall c \tag{8}$$

The tradables market clears by Walras' Law.

Equation (9) characterizes the value of potential idea exchanges in each city, Z_c, which depends on scale (M_c) and average ability ($\bar{z}_c$). Equation (10) characterizes the latter, the

[17] In this exposition, we define equilibrium where each member of the set $\{c\}$ is populated, $L_c > 0$. In appendix section A.2, we describe how the number of populated locations is endogenously determined when there are many potential city locations, not all of which must be populated.

time-weighted average ability of learners in each city.

$$Z_c = m(M_c)\bar{z}_c = m\Big(L\int_{z\geq z_m}(1-\beta_{z,c})\mu(z,c)\mathrm{d}z)\Big)\bar{z}_c \quad \forall c \tag{9}$$

$$\bar{z}_c = \begin{cases} \int_{z\geq z_m}\frac{(1-\beta_{z,c})z}{\int_{z\geq z_m}(1-\beta_{z,c})\mu(z,c)\mathrm{d}z}\mu(z,c)\mathrm{d}z & \text{if } M_c > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall c \tag{10}$$

Equations (11) through (14) describe agents' optimal choices. Equation (11) says that tradables producers allocate their time optimally between directly producing and exchanging ideas. Equation (12) says that agents choose their occupations optimally so that the marginal producer is indifferent between the two sectors.

$$\beta_{z,c} = \arg\max_{\beta\in[0,1]} \beta z(1+(1-\beta)AZ_c z) \quad \forall z \geq z_m \; \forall c \tag{11}$$

$$\tilde{z}(z_m, Z_c) = p_{s,c} \quad \forall c \in C(z_m) \tag{12}$$

Equations (13) and (14) describe the prices consistent with spatial equilibrium. Equation (13) says that non-tradables producers' expenditure on tradables, which is their net income after purchasing non-tradable services and housing, is equal across locations. Equation (14) means that tradables producers are located in their most-preferred place.

$$(1-\bar{s})p_{s,c} - p_{h,c} = (1-\bar{s})p_{s,c'} - p_{h,c'} \quad \forall c, c' \tag{13}$$

$$C(z) = \arg\max_{c} \tilde{z}(z, Z_c) - \bar{s}p_{s,c} - p_{h,c} \quad \forall z \geq z_m \tag{14}$$

There are three classes of equilibria that satisfy equations (5) through (14): no-learning equilibria in which all cities have identical aggregate characteristics; learning equilibria in which some or all cities have identical aggregate characteristics; and learning equilibria with heterogeneous cities.

In no-learning equilibria, no tradables producer devotes time to idea exchange because no other tradables producer does, and $Z_c = 0 \; \forall c$. Since $\tilde{z}(z, 0)$ does not vary across locations, all cities in which tradables are produced must have same prices for housing and non-tradables to satisfy the spatial no-arbitrage conditions (13) and (14). By equation (7), therefore, all populated cities are the same size.

The no-learning equilibrium is not of interest for two reasons. First, it is empirically irrelevant. There is considerable and systematic variation in cities' populations. Second, it is not a stable equilibrium. Since exchanging ideas is a Pareto improvement (it raises productivity for all learners without lowering the productivity of any other agent), communication or coordination among (a sufficiently large set of) tradables producers would facilitate its choice.

The second type of equilibria are those in which learning occurs and some cities' aggregate characteristics are identical. Suppose that $L_c = L_{c'}$. Then, by equations (7) and (13), housing prices and non-tradables prices are equal in these locations. Equation (14) requires that $Z_c = Z_{c'}$. These cities are therefore identical in their populations, prices, and learning opportunities.

Learning equilibria with symmetric cities are not of interest for the two reasons the no-learning equilibrium is not. First, they are empirically irrelevant. Second, they are not stable. When two cities' learning environments differ at all, higher-ability tradables producers are drawn to the better learning environment. Thus, their movement reinforces initial differences in learning opportunities and moves the system of cities towards the asymmetric equilibrium. See appendix section A.3 for details.

Finally, there are equilibria with heterogeneous cities, our object of interest. Equilibria with heterogeneous cities exhibit cross-city patterns that can be established independent of the number of cities that arise.[18] Equations (5) through (14) jointly imply that larger cities have higher housing prices, higher non-tradables prices, exhibit better learning opportunities, and are populated by more talented tradables producers.

To understand this logic, suppose that Z_c varies across cities. Housing and service prices $(p_{h,c} + p_{s,c}\bar{s})$ must be higher in locations with higher Z_c, lest all tradables producers prefer those locations, in accordance with equation (14) and violating equation (8). Equation (13) requires that locations with greater Z_c and therefore greater $p_{h,c} + p_{s,c}\bar{s}$ have higher $p_{s,c}$ so that non-tradables producers earn higher nominal incomes in locations with higher prices. If $p_{h,c}$ were not higher in cities with higher $p_{s,c}$, then these locations would attract all non-tradables producers, so cities with better learning opportunities (Z_c) have both higher $p_{s,c}$ and higher $p_{h,c}$, which means that they are more populous, by equation (7).

Because $\tilde{z}(z, Z_c)$ is supermodular in z and Z_c, higher-z tradables producers gain more from locating in high-Z_c locations with higher prices. As a result, tradables producers sort across cities in equilibria with heterogeneous cities. This sorting according to ability supports

[18]Since these patterns characterize all equilibria with heterogeneous cities, we do not address issues of uniqueness.

equilibrium differences in Z_c.[19]

If there are n locations with positive population and we label the cities by their size so that $L_1 < L_2 < \cdots < L_n$, then the stable-equilibria correspondence $C(z)$ is increasing and lower hemicontinuous for all $z > z_m$. It is single-valued for all $z > z_m$ except for $n-1$ "boundary values" of z, who are indifferent between cities c and $c+1$ because the benefit from $Z_{c+1} > Z_c$ is offset by the higher prices in $c+1$.

We provide sufficient conditions for the existence of this equilibrium when $n = 2$ in appendix section A.4. In short, the existence of an equilibrium with two heterogeneous cities requires that congestion costs are sufficiently strong so that not everyone will locate in a single city in equilibrium and that the potential productivity gains from idea exchanges are sufficiently high that all tradables producers in the larger city will spend time learning in equilibrium.

These equilibria with heterogeneous cities, our object of interest, are robust to perturbation.[20] They match the fundamental facts that cities differ in size and these size differences are accompanied by differences in wages, housing prices, and productivity (Glaeser, 2008). Empirically, larger cities exhibit higher nominal wages in industries that produce tradable goods, which means that productivity is higher in these locations (Moretti, 2011). Our model of why larger cities generate more productivity-increasing idea exchanges is a microfounded explanation of these phenomena. Having matched these well-established facts, we now describe the novel empirical implication that skill premia will be higher in larger cities.

2.4 The spatial pattern of skill premia

When cities are heterogeneous, equations (5) through (14) jointly imply that larger cities have higher housing prices, higher non-tradables prices, exhibit better learning opportunities, and are populated by more talented tradables producers. They also imply that skill premia are higher in larger cities. Appendix section A.5 formally derives this prediction for a two-city equilibrium when ability is distributed Pareto or uniform. This section uses numerical examples to illustrate the economic mechanisms and logic of the novel prediction.

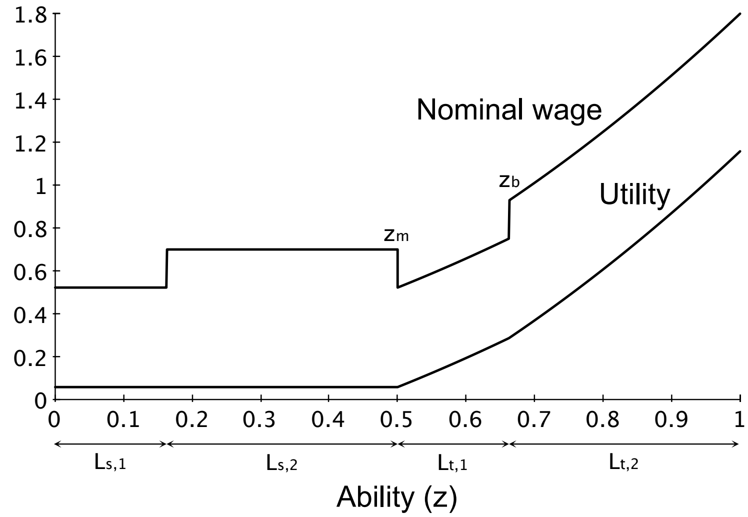
Figure 2 shows the nominal wage and utility outcomes for a particular parameterization

[19] Any microfoundations for Z_c in which cities with a larger mass of more-talented tradables producers exhibit a higher endogenous value of Z_c will support a sorting outcome. Supermodularity of $\tilde{z}(z, Z_c)$ is sufficient for sorting among tradables producers, since the prices they face do not vary with z.

[20] Applying the dynamic extension presented in appendix section A.3 shows that a small perturbation to the asymmetric equilibrium will yield sorting that converges back to the same asymmetric equilibrium.

of our model in a two-city equilibrium.[21] Worker ability, indexed by z, appears on the horizontal axis. We assume here that ability is uniformly distributed. Since the spatial allocation of non-tradables producers ($z < z_m$) is indeterminate due to indifference, we order them by ability only for ease of illustration.[22] Tradables producers ($z > z_m$) are sorted according to ability because this maximizes their utility. z_b is the ability of the tradables producer who is indifferent between the two cities. Since ability is uniformly distributed, the width of the interval is proportional to city population.

Figure 2: Two-city equilibrium: Wages and utility



The nominal wages of both tradables and non-tradables producers are higher in larger cities. This matches the well-established empirical literature on the urban wage premium (Glaeser and Maré, 2001; Glaeser and Gottlieb, 2009). For non-tradables producers, higher nominal wages in larger cities may be thought of as compensation for higher housing prices that keeps real wages constant across cities.

Tradables producers' wages are higher in larger cities for three reasons. First, there is a compositional effect. Since there is spatial sorting among tradables producers, those in larger cities have higher innate abilities that generate higher incomes in any location. Second, there is a learning effect. Since larger cities provide more valuable learning opportunities, idea exchanges in larger cities yield larger productivity gains and thus higher nominal incomes for tradables producers. Third, there is a compensation effect. Producers who are indifferent at the margin between two cities must have a wage gap that exactly matches the gap in non-

[21]See appendix section B for details of this parameterization.

[22]See appendix section A.4 for the formal definition of this $\mu(z, c)$.

tradables and housing prices between those cities. Among the skilled tradables producers, this is a measure zero set that defines the boundary ability level z_b. Among the unskilled non-tradables producers, the compensation effect applies to all individuals, because their homogeneity of productivity makes them all indifferent across cities. Since higher-ability agents earn higher incomes, the nominal wage difference between the two cities is a larger proportion of the non-tradables producers' incomes than that of the marginal tradables producer z_b.[23]

What do these outcomes imply for the spatial pattern of skill premia? We define a city's observed skill premium as its average tradables wage divided by its (average) non-tradables wage $p_{s,c}$.

$$\frac{\bar{w}_c}{p_{s,c}} = \frac{\frac{\int_{z \geq z_m} \tilde{z}_c(z)\mu(z,c)dz}{\int_{z \geq z_m} \mu(z,c)dz}}{p_{s,c}}$$

In equilibria with heterogeneous cities, the cross-city pattern of skill premia depends upon the compositional, learning, and compensation effects. The compositional and learning effects yield higher nominal incomes for tradables producers in the larger city and affect all tradables producers. Each of these effects raises the skill premium of the larger city relative to the smaller city. The compensation effect lowers the skill premium in the larger city relative to the smaller city. When the compositional and learning effects dominate the compensation effect, the skill premium is higher in the larger city.

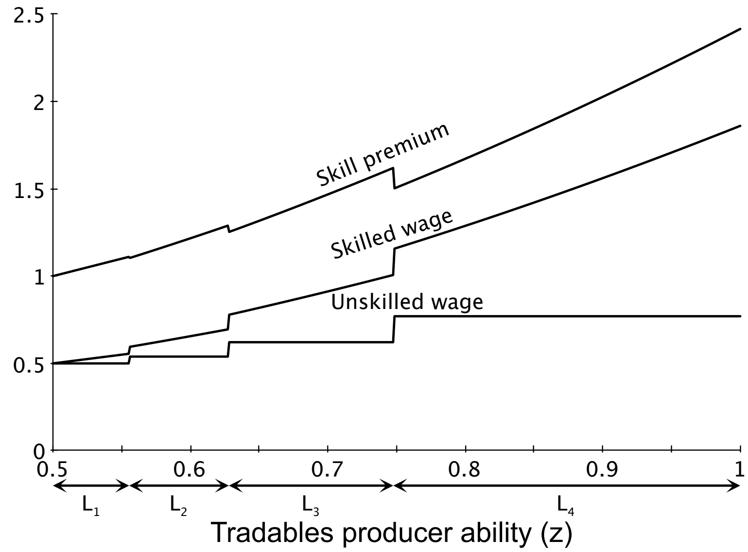
Figure 3 illustrates the pattern of wage premia for a four-city example.[24] It compares the incomes of tradables and non-tradables producers by placing the wage schedules on a common horizontal axis. The ratio of the wage schedules gives the skill premium of each tradables producer relative to the non-tradables producers in the same location. The observed skill premium is the average of these observations in each location. The skill premia curve steps down at the boundaries where tradables producers are indifferent between two locations, due to the compensation effect. The figure illustrates how the compositional and learning effects that raise the skill premium, due to the differences in inframarginal tradables

[23]This compensation effect, which stems from non-homothetic preferences in which lower-income individuals spend a larger fraction of their budget on non-tradables, is the basis for the prediction of Black, Kolesnikova, and Taylor (2009) that skill premia will be lower in cities with higher housing prices. It cannot explain why skill premia are higher in larger cities, since larger cities generally have higher housing prices.

[24]See appendix section B for the parameter values underlying this example. Interval widths are proportionate to city populations.

producers' abilities and the differences in the productivity gains arising from idea exchanges, are greater than the compensation effect that lowers the skill premium. Here larger cities exhibit higher skill premia.

Figure 3: Four-city equilibrium: Skill premia



We can state the condition formally for a two-city asymmetric equilibrium. The skill premium is higher in the larger city when:

$$\frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)dz}{\int_{z_b}^{\infty} \mu(z)dz}}{p_{s,2}} = \frac{\bar{w}_2}{p_{s,2}} > \frac{\bar{w}_1}{p_{s,1}} = \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}}{p_{s,1}}$$

The equilibrium pattern of skill premia depends on the distribution of abilities, $\mu(z)$. In appendix section A.5, we show that this condition for skill premia to increase with city size holds true in the two-city case for the Pareto distribution and provide sufficient conditions for this inequality to hold for the uniform distribution.

To study the pattern of cross-city wage premia predicted by our model with an arbitrary number of cities, we used numerical optimization to search the parameter values minimizing the correlation between city sizes and skill premia when $z \sim U(0,1)$ for equilibria with more than two cities. The numerical results suggest that the premia-size correlation is minimized by letting $\bar{s} \to 1$ so that the mass of inframarginal tradables producers shrinks to zero and the relative influence of the compensation effect is maximized. We did not find a set of parameter values yielding an equilibrium in which the observed skill premia were not strictly increasing in city population. The prediction that skill premia are higher in larger cities appears to be a robust feature of our model.

2.5 Outsourcing and migration in spatial equilibrium

In this section, we develop a model that explains key facts developed in section 1.3, notably that skilled workers move more often and farther than unskilled workers. The challenge is to explain the differential *movement* of skilled and unskilled although they are both perfectly *mobile*. We do this in two steps.

The first step brings our model closer to an important feature of the data. Thus far, we have abstracted from the fact that larger cities tend to have a higher ratio of skilled to unskilled workers. We address this by introducing an additional task carried out by the unskilled, assembly of final tradable output, which can be carried out locally or outsourced. Producers in larger cities, where unskilled nominal wages are higher, outsource assembly tasks. Outsourcing makes larger cities exhibit a higher skilled to unskilled ratio and will enrich our model of migration.[25]

With this model of outsourcing in hand, our second step is to introduce a formal model of migration of skilled and unskilled workers. In this model, the skilled will move both more often and a greater distance on average than the unskilled even though both are perfectly mobile. The intuition is simple. Skilled workers have a most-preferred city that best rewards their skill, so they choose to move there. This gives rise to long-distance moves for many of the skilled and simultaneous outflows and inflows of skilled workers from the same city. The unskilled receive the same utility in all cities, so those who move only need seek the nearest city with notional excess demand for their labor. Differential mobility and differential movement are not the same. The empirical observation of differential migration rates can be accounted for in a spatial-equilibrium framework.

2.5.1 Outsourcing

To this point, we have production of a single, competitively produced homogeneous tradable good, so that this notionally tradable good is not in fact traded. We now introduce a richer

[25] If we take the limit as outsourcing goes to zero, our simple dynamic migration model has *only* migration of the skilled. Obviously this implies that the skilled would move more than the unskilled. With outsourcing and the empirically relevant cross-city heterogeneity in skill composition, members of both labor groups migrate. Thus our result that the skilled move more often is shown to hold in this more realistic setting.

model of tradables production that amends this in interesting ways. Heretofore, tradables producers have been unable to fragment their production process across locations because each producer is self-employed in her residential location. Self-employment also collapses the distinction between a worker and a firm. Assuming that all of a firm's activities take place in a single location is plausible when all elements of production depend upon face-to-face information exchange. But new communication technologies increasingly facilitate the separation of knowledge-intensive headquarters activities from some production-plant-level activities.[26]

We thus amend tradables production to add a second task, assembly of the output.[27] Assembly requires $l < 1 - \bar{s}$ units of homogeneous labor of the same type that produces non-tradable services. Each tradables producer $z \geq z_m$ must therefore incur the additional production cost $l \cdot p_{s,c}$ when assembling output in city c.[28]

Tradables firms may pay a fixed cost f_a to establish an assembly plant using homogeneous labor in a location other than where the firm is headquartered. The net benefit to a tradables producer located in city c of outsourcing assembly to city c' is $l(p_{s,c} - p_{s,c'}) - f_a$. Thus, tradables producers in a large, high-z city will outsource assembly to the smaller, lower-z city if the gap in assembly prices is sufficiently large. Such outsourcing raises the relative skill level of larger cities by shifting unskilled assembly activities to smaller cities and attracting more tradables producers to larger cities to benefit from idea exchange. Thus, larger cities become sites of human-capital-intensive activities that are home to more skilled populations.

To formalize this, we define an assembly assignment function $\rho(c, c')$, which describes the fraction of assembly tasks for firms headquartered in c that are performed in c'.[29] The equilibrium conditions for tradables production require that the assembly location assignments minimize assembly costs, labor markets clear, and that tradables producers take assembly

[26] Duranton and Puga (2005) study the fragmentation of production in a model with homogeneous workers and exogenously assigned occupations. Fujita and Thisse (2006) study how trade costs and communication costs determine this fragmentation of production in a setting with homogeneous firms and two worker types.

[27] "Assembly" should be interpreted broadly here. The *New York Times* recently reported on the ongoing process of Wall Street financial firms "nearshoring" some less skilled back-office jobs, a process consistent with our approach (Nelson D. Schwartz, "Financial Giants Moving Jobs Off Wall Street," 1 July 2012).

[28] In assuming that each tradables producer requires l units for assembly regardless of $\tilde{z}(z, Z_c)$, we are assuming that the greater revenues accruing to higher-z producers are due to selling higher-quality products rather than greater quantities. This is the simplest conceivable assembly process.

[29] The optimal choice of assembly location is orthogonal to producer ability, so $\rho(c, c')$ is independent of z. While the location of production is discretely chosen by each tradables producer, there is a continuum of them, so ρ may be a fraction.

costs into account when choosing their headquarters location.

$$\{\rho(c,c')\} \in \arg\min_{\{\rho(c,x)\}} l\sum_x \rho(c,x)p_{s,x} - \sum_{x\neq c}\rho(c,x)f_a \qquad \text{s.t.} \sum_x \rho(c,x) = 1 \tag{15}$$

$$L_{s,c} = L\int_{z\leq z_m}\mu(z,c)\mathrm{d}z = \bar{s}L_c + l\sum_{c'}\rho(c',c)L\int_{z\geq z_m}\mu(z,c')\mathrm{d}z \tag{8'}$$

$$C(z) = \arg\max_c \tilde{z}(z,Z_c) - \bar{s}p_{s,c} - p_{h,c} - l\sum_{c'}\rho(c,c')p_{s,c'} - \sum_{c'\neq c}\rho(c,c')f_a \quad \forall z\geq z_m \tag{14'}$$

The rest of the equilibrium conditions are unchanged.

In the absence of outsourcing, the skilled share of each city's population was $1-\bar{s}$ and therefore independent of city size. When tradables production incorporates a task that may be outsourced and does not benefit from physical proximity to better learning opportunities, larger cities with higher local prices will outsource those tasks to smaller cities with lower local prices. This means that the skilled population share is increasing in city size, matching the empirical tendency. The strength of this relationship depends on the fragmentation cost f_a and the relative magnitudes of l and $\bar{s}$.

Many believe that fragmentation costs have fallen substantially in recent decades (Duranton and Puga, 2005). Our model predicts that this will trigger outsourcing of assembly tasks and generate a positive correlation between cities' skilled population share and total population. Since larger cities exhibit higher nominal skill premia in our model, outsourcing generates a positive correlation between skill premia and skill shares.

2.5.2 Migration in the outsourcing model

As noted earlier, influential models in the spatial literature assume that skilled workers are mobile while unskilled workers are not, justifying this on the basis that empirically skilled workers move more frequently than unskilled workers. Thus the challenge we take up in this section is to develop a very simple dynamic extension in which all agents are (essentially) perfectly mobile but skilled workers move more frequently than unskilled workers. With modest additional assumptions, we can develop an additional prediction. Not only will skilled workers move more frequently than unskilled workers, but they will also typically move a greater distance, a prediction that also finds support in the data.

Consider first a 2-city system with spatial sorting and outsourcing of unskilled assembly as described above, in which city 2 is larger and more skilled. Assume that each period a fraction δ of the population in each city simultaneously gives birth to a succeeding child and dies. There is no saving, accumulation of capital, or other intertemporal economic interaction, and the total population size is time-invariant. A fraction $1-\lambda$ of the newborns inherit the same z as their parent and a fraction λ of the newborns have their type distributed according to $\mu(z)$. This makes the aggregate talent distribution time-invariant and so the full equilibrium is unchanged. We assume there are positive but arbitrarily small costs of movement, so that gross migration is the minimum necessary to achieve the equilibrium population allocation.

Will agents migrate? Consider first those born in city 2, the relatively "skilled city." Newborns with talent $z \in (z_b, 1)$ will stay in the skilled city. Newborns with talent $z \in (z_m, z_b)$ will migrate to the unskilled city. Some agents with talent $z \leq z_m$ (newborn or not) have reason to migrate to the unskilled city because the larger city's outsourcing-induced lower unskilled share ($\frac{L_{s,2}}{L_2} < \frac{L_{s,1}}{L_1}$) means that the fraction of newborns with talent $z \leq z_m$ there exceeds the equilibrium fraction. Conversely, there will be net migration of tradables producers to the skilled city from the unskilled city.

Gross skilled ($z \geq z_m$) migration exceeds net skilled migration, which equals gross unskilled ($z \leq z_m$) migration. This therefore provides an endogenous, economic reason for the greater movement of more-educated workers. There are two-way flows of skilled workers and a one-way flow of unskilled workers, with the net flows of the skilled matching the gross flows of the unskilled. Provided that less than half the population is skilled, this matches the empirical regularity that more skilled workers move more frequently as a consequence of the equilibrium allocation of talent, rather than an assumption that less-skilled workers are immobile. It also matches empirical work suggesting that movement reflects differential returns to skills (Borjas, Bronars, and Trejo, 1992; Dahl, 2002).

This insight generalizes to an n-city setting, and the proof is simple. The economy-wide skill distribution is invariant across time, so any initial equilibrium is also an equilibrium in later periods. We focus on this equilibrium. For each city, there will be a mismatch between the $\delta\lambda L_c$ newborns each period whose characteristics are orthogonal to those of their parents and their parents' characteristics. This difference represents the net migration offer of city c to all other cities in the system. Note that newborns whose z determines they will work as skilled workers in tradables have (except for a measure zero set) a unique city to which they must move, while as of yet we have not determined the exact patterns of flows of the unskilled, although we consider the case of arbitrarily small costs of migration to rule out cross-hauling of unskilled migrants.

It is convenient to define two groups of cities. Let C_X be the set of cities that are exporters of unskilled migrants and C_M be the set of cities that are importers of unskilled migrants (gross and net being the same due to the absence of cross-hauling of the unskilled). Since each individual city has zero change in total population, that is also true of any partition of the set of cities. Thus exports of unskilled migrants from C_X to C_M must be exactly matched by net imports of skilled migrants in the reverse direction.

Note that all exports of the unskilled must move from C_X to C_M as a matter of definition. But these cannot be the only exports of migrants from C_X to C_M; there are also skilled workers unique to cities in C_M who travel that direction. Thus exports of workers from C_X to C_M are comprised of all the unskilled who move plus some skilled. The volume of exports the reverse direction, by balanced migration, must equal this sum. All of these are skilled. Thus we already have that the majority of migrants between cities in C_X and C_M are skilled. We need to add in the migrants *among* the cities of C_X and C_M, respectively. All of these are skilled as well, since the arbitrarily small migration costs prevent cross-hauling of unskilled migrants. Hence, we can claim, *a fortiori*, that the skilled will be the majority of migrants. So long as the skilled are less than half the labor force, this suffices to show that the *fraction* of migrants is higher among the skilled than unskilled, the first fact that we wanted to explain.

Moreover, the n-city framework also allows us to make a novel prediction – not only will the skilled move more often but they will typically move a greater distance. Again, the logic is simple. With arbitrarily small positive trade costs, the skilled move to their most preferred city. Movements of the unskilled can be considered the solution to a linear programming problem that minimizes the total distance moved of the unskilled while matching net offers of unskilled by cities (cf. Dorfman, Samuelson, and Solow 1958). Appendix section A.6 formalizes the result that skilled individuals will migrate greater distances on average.

The model of migration we have developed here is surely special in a number of dimensions. This notwithstanding, we believe that there is a deeper logic at work here that is consistent with our story of spatial equilibrium.[30] Skilled workers find employment in tasks that make considerable use of heterogeneity, which motivates more spatially extensive

[30] An older, empirical literature on differential migration by education suggested broader geographic labor markets for higher skilled workers without explaining their economic foundations (Long, 1973; Frey, 1979; Frey and Liaw, 2005).

searches. Less skilled workers find employment in tasks that make little use of heterogeneity, which induces less spatial searching. Both the frequency and distance of moves reflect this.

3 Conclusion

The productivity of modern cities depends crucially on their role as loci for idea exchange. Workers in knowledge-intensive industries choose their locations based on the number and skill of the workers with whom they may interact. But there is no prior spatial theory in which individuals choose both their locations and investments in exchanging ideas.

This paper presents the first system of cities model in which costly idea exchange is the agglomeration force. Our emphasis on the costly and optimal allocation of effort to idea exchange is designed to overcome the "black box" critique that has inhibited research in this crucial area. Individuals allocate their time according to the expected gains from exchanging ideas in their city. Everyone would like to be where learning opportunities are greatest, but the best learners are those most able to take advantage of these opportunities and so most willing to pay for them. The broad desire to access the best learning opportunities induces differences in city sizes and housing prices. The higher willingness of the most skilled to pay these prices induces sorting among learners. In larger cities, they exchange ideas more frequently with more people whose average ability is higher.

Our approach is quite parsimonious, but it yields a rich set of spatial patterns. Labor is the sole factor of production and is heterogeneous in a single dimension. There are two goods, tradables and non-tradables. Housing acts as a simple dispersion force. Idea exchanges are local and depend on the scale and average ability of learners. These few assumptions cause cities to vary in size and larger cities to have better learning environments, as well as higher wages, productivity, housing prices, and skill premia – all prominent features in the data. Extended to outsourcing and cross-city migration, our model also provides a simple account of the differential movement of the skilled and unskilled that does not rely on making one of these immobile.

A distinguishing feature of our contribution is that our theory explains spatially heterogeneous outcomes as emergent results of economic processes. Previous explanations for empirical regularities such as differential migration rates or spatial variation in skill premia relied on assumed asymmetries in individuals' freedom to move or cities' fundamental characteristics. We hope that our theory illuminates another path for long-run models of the spatial distribution of economic activity in a world in which cities are defined by the skills and ideas of those who choose to live in them.

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.

Albouy, D. (2009): "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117(4), 635–667.

Alonso, W. (1964): *Location and land use: Toward a general theory of land rent*, Publication of the Joint Center for Urban Studies. Harvard University Press.

Arzaghi, M., and J. V. Henderson (2008): "Networking off Madison Avenue," *Review of Economic Studies*, 75(4), 1011–1038.

Audretsch, D. B., and M. P. Feldman (2004): "Knowledge spillovers and the geography of innovation," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4, chap. 61, pp. 2713–2739. Elsevier.

Autor, D. H., and D. Dorn (2012): "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market," MIT working paper.

Autor, D. H., D. Dorn, and G. Hanson (2011): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," MIT working paper.

Bacolod, M., B. S. Blum, and W. C. Strange (2009): "Skills in the city," *Journal of Urban Economics*, 65(2), 136–153.

Baum-Snow, N., and R. Pavan (2011): "Inequality and City Size," mimeo.

Beaudry, P., M. Doms, and E. Lewis (2010): "Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 118(5), 988 – 1036.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2010): "Productive cities: Sorting, selection and agglomeration," CEPR Discussion Paper 7922, CEPR.

Behrens, K., and F. Robert-Nicoud (2011): "Survival of the Fittest in Cities: Urbanisation, Agglomeration, and Inequality," mimeo.

Berliant, M., and M. Fujita (2008): "Knowledge Creation As A Square Dance On The Hilbert Cube," *International Economic Review*, 49(4), 1251–1295.

Berliant, M., R. R. Reed III, and P. Wang (2006): "Knowledge exchange, matching, and agglomeration," *Journal of Urban Economics*, 60(1), 69–95.

Black, D. (1999): "Local knowledge spillovers and inequality," mimeo.

Black, D., N. Kolesnikova, and L. Taylor (2009): "Earnings Functions When Wages and Prices Vary by Location," *Journal of Labor Economics*, 27(1), 21–47.

Borck, R., M. Pflüger, and M. Wrede (2010): "A simple theory of industry location and residence choice," *Journal of Economic Geography*, 10(6), 913–940.

Borjas, G. J., S. G. Bronars, and S. J. Trejo (1992): "Self-selection and internal migration in the United States," *Journal of Urban Economics*, 32(2), 159–185.

Charlot, S., and G. Duranton (2004): "Communication externalities in cities," *Journal of Urban Economics*, 56(3), 581–613.

Combes, P.-P., G. Duranton, and L. Gobillon (2008): "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, 63(2), 723–742.

Dahl, G. B. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367–2420.

De la Roca, J. (2012): "Selection in initial and return migration: Evidence from moves across Spanish cities," mimeo.

Dorfman, R., P. Samuelson, and R. Solow (1958): *Linear programming and economic analysis*. McGraw-Hill.

Duranton, G., and D. Puga (2005): "From sectoral to functional urban specialisation," *Journal of Urban Economics*, 57, 343–370.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2010): "Spatial Sorting: Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled," CEPR Discussion Paper 8151.

Frey, W. H. (1979): "The Changing Impact of White Migration on the Population Compositions of Origin and Destination Metropolitan Areas," *Demography*, 16(2), pp. 219–237.

Frey, W. H., and K.-L. Liaw (2005): "Migration within the United States: Role of Race-Ethnicity," *Brookings-Wharton Papers on Urban Affairs*, pp. 207–262.

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press.

Fujita, M., and J.-F. Thisse (2006): "Globalization And The Evolution Of The Supply Chain: Who Gains And Who Loses?," *International Economic Review*, 47(3), 811–836.

Gabaix, X. (1999): "Zipf's Law For Cities: An Explanation," *The Quarterly Journal of Economics*, 114(3), 739–767.

Gibbons, S., H. G. Overman, and P. Pelkonen (2010): "Wage Disparities in Britain: People or Place?," SERC Discussion Papers 0060, Spatial Economics Research Centre, LSE.

Glaeser, E. L. (1999): "Learning in Cities," *Journal of Urban Economics*, 46(2), 254–277.

——— (2005): "Urban colossus: Why is New York America's largest city?," *Economic Policy Review*, 11(2), 7–24.

——— (2008): *Cities, Agglomeration, and Spatial Equilibrium*, The Lindahl Lectures. Oxford University Press.

Glaeser, E. L., and J. D. Gottlieb (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983–1028.

Glaeser, E. L., J. Gyourko, and R. E. Saks (2006): "Urban growth and housing supply," *Journal of Economic Geography*, 6(1), 71–89.

Glaeser, E. L., and D. C. Maré (2001): "Cities and Skills," *Journal of Labor Economics*, 19(2), 316–42.

Glaeser, E. L., M. Resseger, and K. Tobio (2009): "Inequality In Cities," *Journal of Regional Science*, 49(4), 617–646.

Greenwood, M. J. (1997): "Internal migration in developed countries," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig, and O. Stark, vol. 1, chap. 12, pp. 647–720. Elsevier.

Gyourko, J., C. Mayer, and T. Sinai (2006): "Superstar Cities," NBER Working Paper 12355, National Bureau of Economic Research.

Helpman, E. (1998): "The size of regions," in *Topics in Public Economics*, ed. by I. Z. David Pines, Efraim Sadka, pp. 33–54. Cambridge University Press.

Helsley, R. W., and W. C. Strange (2004): "Knowledge barter in cities," *Journal of Urban Economics*, 56(2), 327–345.

Henderson, J. V. (1974): "The Sizes and Types of Cities," *American Economic Review*, 64(4), 640–56.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 108(3), 577–98.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483–99.

——— (2011): "The New Economic Geography, Now Middle-aged," *Regional Studies*, 45(1), 1–7.

Long, L. H. (1973): "Migration differentials by education and occupation: Trends and variations," *Demography*, 10, 243–258.

Lucas, R. E. (1988): "On the mechanics of economic development," *Journal of Monetary Economics*, 22(1), 3–42.

Lucas, R. E., and B. Moll (2011): "Knowledge Growth and the Allocation of Time," NBER Working Paper 17495, National Bureau of Economic Research.

Marshall, A. (1890): *Principles of Economics*. MacMillan and Co.

Mills, E. (1967): "An aggregative model of resource allocation in a metropolitan area," *The American Economic Review*, 57(2), 197–210.

Molloy, R., C. L. Smith, and A. Wozniak (2011): "Internal Migration in the United States," *Journal of Economic Perspectives*, 25(3), 173–96.

Moretti, E. (2011): "Local Labor Markets," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4, chap. 14, pp. 1237–1313. Elsevier.

Muth, R. F. (1969): *Cities and Housing*. University of Chicago Press.

Nakamura, E., and J. Steinsson (2011): "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions," NBER Working Paper 17391, National Bureau of Economic Research.

Notowidigdo, M. J. (2011): "The Incidence of Local Labor Demand Shocks," Working Paper 17167, National Bureau of Economic Research.

Roy, A. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, pp. 135–146.

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.

Tabuchi, T., and J.-F. Thisse (2002): "Taste heterogeneity, labor mobility and economic geography," *Journal of Development Economics*, 69(1), 155–177.

A Theory

A.1 Internal urban structure

To introduce congestion costs, we follow Behrens, Duranton, and Robert-Nicoud (2010) and adopt a standard, highly stylized model of cities' internal structure.[31] City residences of unit size are located on a line and center around a single point where economic activities occur, called the central business district (CBD). Residents commute to the CBD at a cost that is denoted in units of the numeraire. The cost of commuting from a distance x is τx^{γ} and independent of the resident's income and occupation.

Agents choose a residential location x to minimize the sum of land rent and commuting cost, $r(x) + \tau x^{\gamma}$. In equilibrium, agents are indifferent across residential locations. In a city with population mass L, the rents fulfilling this indifference condition are $r(x) = r\left(\frac{L}{2}\right) + \tau\left(\frac{L}{2}\right)^{\gamma} - \tau x^{\gamma}$ for $0 \leq x \leq \frac{L}{2}$. Normalizing rents at the edge to zero yields $r(x) = \tau\left(\frac{L}{2}\right)^{\gamma} - \tau x^{\gamma}$.

The city's total land rent is

$$TLR = \int_{\frac{-L}{2}}^{\frac{L}{2}} r(x)\mathrm{d}x = 2\int_{0}^{\frac{L}{2}} r(x)\mathrm{d}x = 2\tau\left(\left(\frac{L}{2}\right)^{\gamma+1} - \frac{1}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1}\right) = \frac{2\tau\gamma}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1}$$

The city's total commuting cost is

$$TCC = 2\int_{0}^{\frac{L}{2}} \tau x^{\gamma}\mathrm{d}x = \frac{2\tau}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1} \equiv \theta L^{\gamma+1}$$

The city's total land rents are lump-sum redistributed equally to all city residents. Since they each receive $\frac{TLR}{L}$, every resident pays the average commuting cost, $\frac{TCC}{L} = \theta L^{\gamma}$, as her net urban cost. Since this urban cost is proportionate to the average land rent, we say the "consumer price of housing" in city c is $p_{h,c} = \theta L_c^{\gamma}$.

A.2 The number of cities

In section 2.3, we defined equilibrium for a set of locations $\{c\}$ in which each member of the set is populated, $L_c > 0$. Here we describe how the equilibrium number of cities is determined when there are an arbitrary number of potential city sites, some of which are

[31] There is nothing original in this urban structure. We use notation identical to, and taken from, Behrens, Duranton, and Robert-Nicoud (2010).

unpopulated in equilibrium.

Consider a potential city site that is unoccupied. The modern technologies employed require specialization, so individuals cannot divide their time between producing tradables and non-tradables. Since non-tradables are a necessity, an individual living in isolation will produce only non-tradables. Thus, an individual moving to an empty location would engage in subsistence production of non-tradables, consume free housing, and obtain utility of zero.

Unless all non-tradables producers consume zero tradables ($p_{s,c} = \frac{p_{h,c}}{1-\bar{s}}\ \forall c : L_c > 0$), non-tradables producers living in cities obtain strictly positive utility. Therefore the entire population lives in a finite number of cities. Accommodating an arbitrary set of locations $\{c\}$ that includes uninhabited places ($L_c = 0$) only requires modifying one equation. When there are potentially empty locations, the spatial-equilibrium indifference condition 13 only applies to occupied locations. The equilibrium condition is thus

$$(1-\bar{s})p_{s,c} - p_{h,c} = (1-\bar{s})p_{s,c'} - p_{h,c'} \quad \forall c, c' : L_c > 0, L_{c'} > 0 \tag{13'}$$

There may be multiple equilibria satisfying equations (5) through (12), (13'), and (14) that have different numbers of cities. We see no theoretical reason to believe that the equilibrium number of populated cities should be unique for a given set of parameters. The qualitative, cross-city predictions of the model do not depend upon the equilibrium number of cities. The particulars of our numerical examples do, of course.

When there are tradables producers who do not spend time learning, it is welfare-maximizing for the population of non-learning tradables producers and a corresponding fraction of the non-tradables producers to reside in every potential location, since this minimizes congestion costs and there are no agglomeration benefits for non-learners. Particular assumptions about city developers, *à la* Henderson (1974), would ensure that this outcome would occur in equilibrium, but we do not consider these issues to be crucial to the topics we explore in this paper.

A.3 Stability of equilibria

In this section, we describe the instability of equilibria in which there are symmetric cities with the same population size and learning opportunities. For simplicity, consider a two-city symmetric equilibrium in which initially $L_1 = L_2$ and $Z_1 = Z_2 > 0$. Since sorting among tradables producers distinguishes equilibria with symmetric cities from equilibria with heterogeneous cities and since in all equilibria non-tradables producers are indifferent across all locations, we simplify the discussion by assuming that $\frac{\bar{s}}{1-\bar{s}}$ individuals of ability $z < z_m$ move locations whenever individuals of ability $z > z_m$ move locations. We now introduce a simple dynamic process for the locational choices of tradables producers that demonstrates that equilibria with symmetric cities are unstable.

Suppose that there is a shock to the symmetric equilibrium such that $L_1 \neq L_2$ and $Z_1 \neq Z_2$ but we are in the neighborhood of $L_1 = L_2$ and $Z_1 = Z_2$. Without loss of generality, let $Z_2 > Z_1$. By the supermodularity of $\tilde{z}(z, Z_c)$, the net benefit of locating in city 2 relative to city 1 ($\tilde{z}(z, Z_2) - \tilde{z}(z, Z_1) - \frac{\theta}{1-\bar{s}}(L_2^\gamma - L_1^\gamma)$) is increasing in z.

Individuals move according to the myopic net benefits of changing locations. The residents of each city have the opportunity to move, and we alternate between the two cities *ad infinitum*. Without loss of generality, individuals located in city 1 consider moving, followed by individuals located in city 2, followed by individuals now located in city 1, and so forth. We now assume that locational changes are ordered according to the net benefits of changing locations. That is, the tradables producers who have the most to gain from moving move first. Tradables producers take full account of the changes in cities' economic characteristics induced by those who have moved before them and are completely myopic with respect to future changes.

Suppose that the highest ability tradables producer in city 1 has a positive net benefit of moving to city 2.[32] By supermodularity of $\tilde{z}(z, Z_c)$, this producer has the most to gain by relocating to city 2. Since we start from a symmetric equilibrium, this move raises Z_2 and L_2 and lowers Z_1 and L_1.[33] The ordering of the net benefit of locating in city 2 relative to city 1 is unchanged by these outcomes. If the net benefit is positive for the (remaining) highest ability tradables producer located in city 1, that producer relocates. This process continues until the net benefit for the highest ability tradables producer is zero.[34] At this point, all the individuals located in city 1 wish to remain in city 1. We know that $L_2 > L_1$ because individuals have moved from city 1 to city 2, and we know that $Z_2 > Z_1$ because the net benefit is zero for the last mover.

Next, individuals in city 2 have the opportunity to relocate to city 1. If there is a tradables

[32]If not, then no individual in city 1 moves. In the next step, individuals in city 2 have the opportunity to move.

[33]Since we are in the neighborhood of $Z_1 = Z_2$ and $Z_c = m(M_c)\bar{z}_c$ where $\bar{z}_c$ is a weighted average of producers' abilities and $m(\cdot) \leq 1$, the highest ability producer in city 1 must have ability $z > Z_1 \approx Z_2$.

[34]Such a producer exists so long as not everyone wishes to locate in a single city. We provide sufficient conditions on parameters such that a two-city asymmetric equilibrium exists in section A.4.

producer in city 2 with ability z lower than the ability of the highest-z tradables producer in city 1, then the lowest-z tradables producer in city 2 has a positive net benefit of relocating to city 1. This movement is followed by subsequent movements until we reach a tradables producer in city 2 who is indifferent between the two locations. We know that we reach the indifferent producer while $L_2 > L_1$ because $Z_2 > Z_1$. At this point, all the individuals located in city 2 wish to remain in city 2.

Next, individuals in city 1 have the opportunity to relocate. If there is a tradables producer in city 1 with ability z greater than the ability of the lowest-z tradables producer in city 2, then the highest-z tradables producer in city 1 has a positive net benefit of relocating to city 2. This process continues *ad infinitum* until the ability of the lowest-z tradables producers in city 2 is equal to the ability of the highest-z tradables producer in city 1. At that point, no tradables producers wish to move and we have obtained an equilibrium with heterogeneous cities.

A.4 Existence of two-city asymmetric equilibrium

Here we characterize sufficient conditions for parameter values such that there exists a two-city equilibrium in which $L_1 < L_2$. Since the two cities differ in size, any equilibrium will exhibit sorting among tradables producers $z \geq z_m$. The allocation of non-tradables producers $z < z_m$ is both indeterminate and inessential. z_m is given by $\bar{s} = \int_0^{z_m} \mu(z)\mathrm{d}z$.

We start by guessing $L_1 \leq \frac{1}{2}L$, which implies $L_2 = L - L_1$. Define the values z_b and $z_{b,s}$ by

$$(1-\bar{s})L_1 = L\int_{z_m}^{z_b} \mu(z)\mathrm{d}z \qquad \bar{s}L_1 = L\int_0^{z_{b,s}} \mu(z)\mathrm{d}z$$

The locational assignments

$$\mu(z,1) = \begin{cases} \mu(z) & 0 \leq \; z < z_{b,s} \\ 0 & z_{b,s} \leq \; z < z_m \\ \mu(z) & z_m \leq \; z < z_b \\ 0 & z_b \leq \; z \end{cases} \qquad \mu(z,2) = \begin{cases} 0 & 0 \leq \; z < z_{b,s} \\ \mu(z) & z_{b,s} \leq \; z < z_m \\ 0 & z_m \leq \; z < z_b \\ \mu(z) & z_b \leq \; z \end{cases}$$

satisfy equations (5), (6), and (8). We then suppose equations (7) and (9) through (13) hold true, where $M_c > 0$ if there is a value of $M_c > 0$ satisfying those equations.

If all tradables producers are in their optimal location, so that equation (14) is satisfied, then the value of L_1 that we guessed supports an asymmetric equilibrium. To check whether this holds, we define an expression $\Omega(L_1)$ that is utility in the smaller city minus utility in the larger city for the marginal tradables producer, z_b.

$$\Omega(L_1) \equiv \frac{\theta}{1-\bar{s}}\left(L_2^{\gamma} - L_1^{\gamma}\right) - \left(\tilde{z}(z_b, Z_c(z_b, 1)) - \tilde{z}(z_b, Z_c(z_m, z_b))\right)$$

where, in an abuse of notation, $Z_c(x, y)$ is the maximum value of Z_c satisfying equation (10) when x and y are the lower and upper limits of integration and $\mu(z, c) = \mu(z)$. Ω can be written solely as a function of L_1 because all the other variables in a sorting equilibrium are given by L_1 via $z_{b,s}$ and z_b through the locational assignments and other equilibrium conditions. The marginal tradables producer is indifferent between the two locations when $\Omega(L_1) = 0$, and all the inframarginal tradables producers are in their optimal locations by the supermodularity of $\tilde{z}$ in z and Z_c.

In an asymmetric equilibrium, learning must occur in the larger city and all tradables producers located in the larger city must participate in learning. Otherwise, they would raise their utility by locating in the smaller city with lower local prices. A sufficient condition for full-participation learning to occur in the larger city in equilibrium is to assume a value of A and functional form $m(\cdot)$ such that full-participation learning occurs in the larger city for all potential values of z_b. That is, let A be sufficiently large and $m(\cdot)$ approach one sufficiently quickly that $\exists Z_2 > 0$ satisfying equations (16) through (19) for all $L_1 : 0 < L_1 < \frac{1}{2}L$.

$$Z_2 = m(M_2)\bar{z}_2 \tag{16}$$

$$M_2 = L\int_{z_b}^{\infty}(1-\beta_{z,2})\mu(z)\mathrm{d}z \tag{17}$$

$$\bar{z}_2 = \int_{z_b}^{\infty}\frac{(1-\beta_{z,2})\mu(z)}{\int_{z_b}^{\infty}(1-\beta_{z,2})\mu(z)\mathrm{d}z}\mathrm{d}z \tag{18}$$

$$\beta_{z,2} = \arg\max_{\beta\in[0,1]} \beta z(1 + (1-\beta)AZ_2 z) \tag{19}$$

$\Omega\left(\frac{L}{2}\right) < 0$, since the cities are equally sized, equalizing housing and non-tradable services prices, but they differ in Z_c, with $Z_2 > Z_1$.

We require $\lim_{L_1\to 0}\Omega(L_1) > 0$, so that the entire population does not live in a single city in equilibrium. This requires $\frac{\theta}{1-\bar{s}}L^{\gamma} > \tilde{z}(z_m, Z_c(z_m, 1)) - \tilde{z}(z_m, Z_c(z_m, z_m)) = \tilde{z}(z_m, Z_c(z_m, 1)) - z_m$. In words, provided that congestion costs are sufficiently strong relative to idea-exchange benefits, the second location will not be empty.

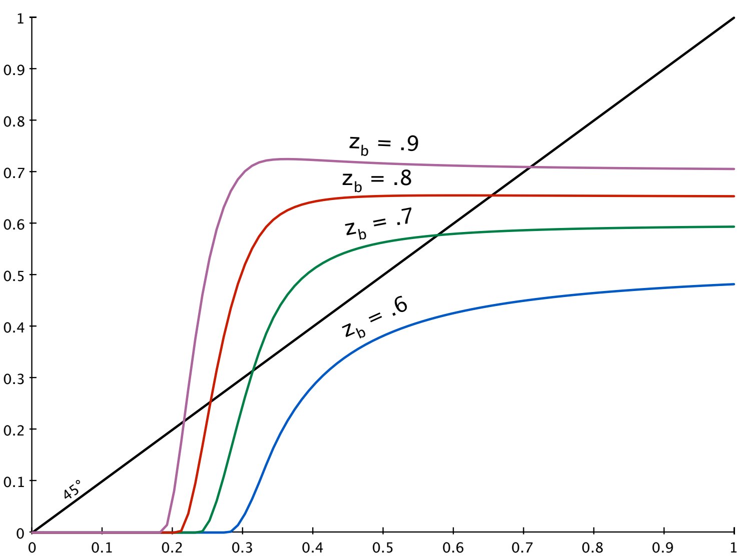
With $\lim_{L_1 \to 0} \Omega(L_1) > 0$ and $\Omega\left(\frac{L}{2}\right) < 0$, continuity delivers the existence of a L_1 such that $\Omega(L_1) = 0$. We now describe where Ω is continuous and why its discontinuities are not a problem.

$$\Omega(L_1) \equiv \frac{\theta}{1-\bar{s}}\left(L_2^\gamma - L_1^\gamma\right) - \left(\tilde{z}(z_b, Z_c(z_b, 1)) - \tilde{z}(z_b, Z_c(z_m, z_b))\right)$$

$\frac{\theta}{1-\bar{s}}\left(L_2^\gamma - L_1^\gamma\right)$ is obviously continuous in L_1.

We assume that $\mu(z)$ is continuous in z. Since $\beta_{z,c}$ is a function of Z_c, and M_c and $\bar{z}_c$ are functions of $\beta_{z,c}$, the equilibrium value of Z_c satisfying equations (9) through (11) is where the function $m(M_c)\bar{z}_c$ intersects the 45-degree line. Since $\beta_{z,c}$ is continuous in Z_c, and M_c and $\bar{z}_c$ are continuous in Z_c and z_b, $m(M_c)\bar{z}_c$ is continuous in Z_c and z_b. This means that $Z_2(z_b, 1)$ is a continuous function of L_1. $\tilde{z}(z, Z_c)$ is continuous in its arguments. Thus, $\tilde{z}(z_b, Z_c(z_b, 1))$ is continuous in L_1.

Figure 4: Finding the fixed points of $Z_1(z_m, z_b)$



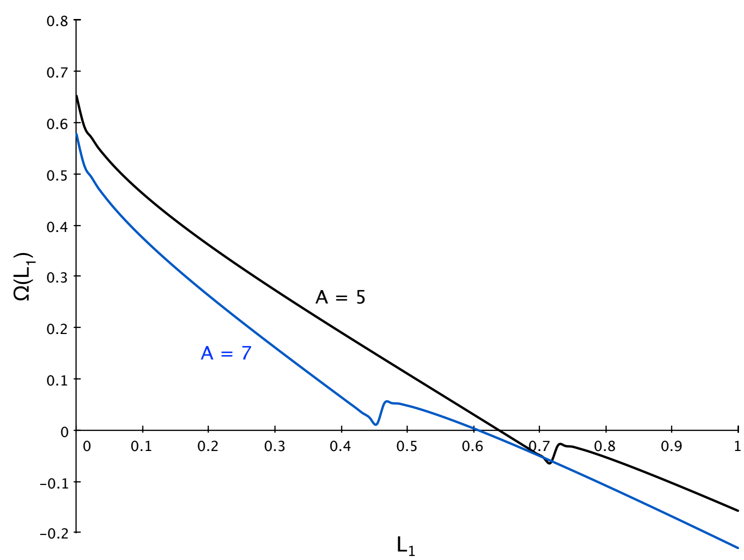
NOTE: $z \sim U(0,1), A = 6, z_m = .5, m(M_c) = \frac{\exp(30M_c)-1}{\exp(30M_c)}, L = 2$

$Z_1(z_m, z_b)$ is (weakly) increasing in L_1. $Z_1(z_m, z_b)$ is not continuous in L_1. For sufficiently small values of L_1, there is no value of $Z_1 > 0$ satisfying equations (9) through (11). The smaller size and lower abilities of the smaller city's population rule out an equilibrium with idea exchange. When L_1 becomes sufficiently large that there is a value of Z_1 satisfying equations (9) through (11), there is a discontinuous increase in $Z_1(z_m, z_b)$ at this point because the maximum value of Z_1 given the population jumps from zero to a positive number. This causes a discontinuous increase in Ω at this value of L_1. $Z_1(z_m, z_b)$ is continuous in L_1 for greater values of L_1 by the continuity of $\beta_{c,z}$, M_c, and $\bar{z}_c$ in z_b, Z_1, and L_1. An example of how these fixed points vary with z_b (which is determined by L_1) is illustrated in Figure 4.

If $Z_1(z_m, z_b) = 0 \; \forall L_1 \in (0, \frac{1}{2}L)$, then Z_1 is continuous in L_1 and Ω is continuous in L_1. If $Z_1 > 0$ for some $L_1 \in (0, \frac{1}{2}L)$, then $Z_1(z_m, z_b)$ and Ω discontinuously increase at one value of L_1 and are continuous everywhere else in $(0, \frac{1}{2}L)$.

Since $\lim_{L_1 \to 0} \Omega(L_1) > 0$, $\Omega\left(\frac{L}{2}\right) < 0$, and Ω increases at any point at which Ω is not continuous in L_1, there exists a value of L_1 such that $\Omega(L_1) = 0$. Two examples of this are illustrated in Figure 5.

Figure 5: $\exists L_1 : \Omega(L_1) = 0$



NOTE: $z \sim U(0,1), \bar{s} = .5, \theta = .25, \gamma = .5, m(M_c) = \frac{\exp(30M_c)-1}{\exp(30M_c)}, L = 2$

In summary, the existence of an equilibrium with two heterogeneous cities requires that congestion costs are sufficiently strong so that not everyone will locate in a single city in equilibrium and that the potential productivity gains from idea exchanges are sufficiently high that everyone in the larger city will spend time learning in equilibrium. We have formalized sufficient conditions for these outcomes as θ and $\bar{s}$ taking values such that $\frac{\theta}{1-\bar{s}}L^{\gamma} > \tilde{z}(z_m, Z_c(z_m, 1)) - z_m$ and A be sufficiently high and $m(\cdot)$ approaching one sufficiently quickly that $\exists \; Z_2 > 0$ satisfying equations (16) through (19) for all $L_1 : 0 < L_1 < \frac{1}{2}L$.

A.5 Skill premia in a two-city equilibrium

A.5.1 Pareto distribution

In an asymmetric two-city equilibrium with z distributed Pareto, $\mu(z) = \frac{kb^k}{z^{k+1}}$ with $k > 1$, the skill premium in the larger city is higher when

$$\frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\infty} \mu(z)\mathrm{d}z}}{p_{s,2}} > \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}}{p_{s,1}} \iff \frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\infty} \mu(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} > \frac{p_{s,2}}{p_{s,1}}$$

We now show that this condition is always true. In steps:

1. Because $\tilde{z}_2(z)$ is increasing in z, for any $\bar{z} \geq z_b$ the following inequality holds:

$$\frac{\frac{\int_{z_b}^{\bar{z}} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\bar{z}} \mu(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} < \frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\infty} \mu(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}}$$

2. Define a change of variables by $f(z) = \left(z_b^{-k} + z^{-k} - z_m^{-k}\right)^{\frac{-1}{k}}$ and $\bar{z} = \left(2z_b^{-k} - z_m^{-k}\right)^{\frac{-1}{k}}$ such that $\int_{z_b}^{\bar{z}} \tilde{z}_2(z)\mu(z)\mathrm{d}z = \int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)\mathrm{d}z$. By construction $\mu(z) = \mu(f(z))f'(z)$.
3. $\frac{\tilde{z}_2(f(z_m))}{\tilde{z}_1(z_m)} = \frac{\tilde{z}_2(z_b)}{p_{s,1}} > \frac{p_{s,2}}{p_{s,1}}$ because $\tilde{z}_2(z_b) > p_{s,2}$.
4. $\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}$ is increasing in z, so $\tilde{z}_2(f(z)) > \frac{p_{s,2}}{p_{s,1}}\tilde{z}_1(z)\ \forall z \in (z_m, z_b)$.
5. Multiplying by $\mu(z)$and integrating yields $\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)\mathrm{d}z > \frac{p_{s,2}}{p_{s,1}} \int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z$. Thus $\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z} > \frac{p_{s,2}}{p_{s,1}}$ and $\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z} = \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(f(z))f'(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} > \frac{p_{s,2}}{p_{s,1}}$
6. Therefore,

$$\frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\infty} \mu(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} > \frac{\frac{\int_{z_b}^{\bar{z}} \tilde{z}_2(z)\mu(z)\mathrm{d}z}{\int_{z_b}^{\bar{z}} \mu(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} = \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(f(z))f'(z)\mathrm{d}z}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)\mathrm{d}z}{\int_{z_m}^{z_b} \mu(z)\mathrm{d}z}} > \frac{p_{s,2}}{p_{s,1}}$$

Only the fourth step ($\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}$ is increasing in z) requires further elaboration.

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}z}\Big(\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}\Big) &= \frac{\tilde{z}_1(z)\tilde{z}_2'(f(z))f'(z) - \tilde{z}_2(f(z))\tilde{z}_1'(z)}{\tilde{z}_1(z)^2} \\
\tilde{z}_c'(z) &= \frac{1}{2}(AZ_c z + 1) \\
\tilde{z}_c(z) &= \frac{1}{AZ_c}(\tilde{z}_c'(z))^2 \\
\frac{\mathrm{d}}{\mathrm{d}z}\Big(\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}\Big) &= \frac{1}{A\tilde{z}_1(z)^2}\Big(\tilde{z}_2'(f(z))\tilde{z}_1'(z)\Big)\Big(\frac{\tilde{z}_1'(z)}{Z_1}f'(z) - \frac{\tilde{z}_2'(f(z))}{Z_2}\Big) \\
&= \frac{1}{A\tilde{z}_1(z)^2}\Big(\tilde{z}_2'(f(z))\tilde{z}_1'(z)\Big)\Big(\underbrace{\frac{f'(z)}{Z_1} - \frac{1}{Z_2}}_{>0} + \underbrace{A(f'(z)z - f(z))}_{>0}\Big)
\end{aligned}
$$

Those inequalities are true because

$$
\begin{aligned}
f(z) &= \left(z_b^{-k} + z^{-k} - z_m^{-k}\right)^{\frac{-1}{k}} \\
f'(z) &= \left(z_b^{-k} + z^{-k} - z_m^{-k}\right)^{\frac{-1-k}{k}} z^{-1-k} \\
&= \left(1 + \frac{z^{-k} - z_m^{-k}}{z_b^{-k}}\right)^{\frac{-1-k}{k}} \left(\frac{z_b}{z}\right)^{k+1} > 1 \\
f'(z)z - f(z) &= f(z)\left(\frac{z_m^{-k} - z_b^{-k}}{z_b^{-k} - z_m^{-k} + z^{-k}}\right) > 0
\end{aligned}
$$

A.5.2 Uniform distribution

In an asymmetric two-city equilibrium with $z \sim U(\mathfrak{z}, \hat{z})$, the skill premium in the larger city is higher when

$$
\frac{\frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\frac{1}{\hat{z}-\mathfrak{z}}\mathrm{d}z}{\int_{z_b}^{\hat{z}} \frac{1}{\hat{z}-\mathfrak{z}}\mathrm{d}z}}{p_{s,2}} > \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\frac{1}{\hat{z}-\mathfrak{z}}\mathrm{d}z}{\int_{z_m}^{z_b} \frac{1}{\hat{z}-\mathfrak{z}}\mathrm{d}z}}{p_{s,1}} \iff \frac{z_b - z_m}{\hat{z} - z_b}\frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mathrm{d}z}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mathrm{d}z} > \frac{p_{s,2}}{p_{s,1}}
$$

A sufficient condition for this to be true in equilibrium is $z_m > z_b^2$. In steps:

1. By change of variable, $\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mathrm{d}z = \int_{z_m}^{z_b} \tilde{z}_2(f(z))f'(z)\mathrm{d}z$, where $f(z) = z_b + \frac{\hat{z}-z_b}{z_b-z_m}(z - z_m)$. Therefore $\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mathrm{d}z = \frac{1}{f'(z)}\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mathrm{d}z = \frac{z_b-z_m}{1-z_b}\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mathrm{d}z$.

2. $\frac{\tilde{z}_2(f(z_m))}{\tilde{z}_1(z_m)} = \frac{\tilde{z}_2(z_b)}{p_{s,1}} > \frac{p_{s,2}}{p_{s,1}}$

3. If $\hat{z}z_m > z_b^2$, then $\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}$ is increasing in z, so $\tilde{z}_2(f(z)) > \frac{p_{s,2}}{p_{s,1}}\tilde{z}_1(z)\ \forall z \in (z_m, z_b)$.

4. Integrating, $\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mathrm{d}z > \frac{p_{s,2}}{p_{s,1}} \int_{z_m}^{z_b} \tilde{z}_1(z)\mathrm{d}z$

5. Therefore, $\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mathrm{d}z}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mathrm{d}z} = \frac{z_b - z_m}{\hat{z} - z_b} \frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mathrm{d}z}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mathrm{d}z} > \frac{p_{s,2}}{p_{s,1}}$. The skill premium is higher in the larger city.

$\hat{z} z_m > z_b^2$ is sufficient for the third step because

$$\frac{\mathrm{d}}{\mathrm{d}z}\Big(\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}\Big) = \frac{1}{A\tilde{z}_1(z)^2}\Big(\tilde{z}_2'(f(z))\tilde{z}_1'(z)\Big)\Big(\underbrace{\frac{f'(z)}{Z_1} - \frac{1}{Z_2}}_{>0} + A(f'(z)z - f(z))\Big)$$

$$f'(z) = \frac{\hat{z} - z_b}{z_b - z_m} > 1$$

$$\hat{z} z_m > z_b^2 \Rightarrow f'(z)z - f(z) > 0 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}z}\Big(\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}\Big) > 0$$

$\hat{z} z_m > z_b^2$ is far from necessary. In fact, when it fails is when z_b is relatively large, which means that the two cities are relatively similar in size. But this similarity in size causes a similarity in housing prices, which diminishes the compensation effect relative to the compositional and learning effects. We have not found a set of parameter values yielding a two-city equilibrium in which the skill premium is lower in the larger city.

A.6 Migration and distance

Here we characterize migration flows for a special case of the outsourcing model and show that they imply that the average migration of non-tradables producers will be shorter than that of tradables producers. Suppose that there are n cities in equilibrium, with n_s "skilled cities" outsourcing their assembly activities to n_u "unskilled cities", such that $n_s + n_u = n$. Denote the set of skilled cities by C_s and the set of unskilled cities by C_u. We denote gross migration flows of the unskilled from city c to c' by $x_{c,c'}$ and gross migration flows of the skilled by $y_{c,c'}$. The cost of migrating from c to c' is arbitrarily small and proportionate to the distance between the cities, $d(c, c') = d(c', c)$.

Denote the lowest ability tradables producers in the skilled cities by $z_{b,1}$. With arbitrarily small migration costs, newborn tradables producers of ability $z \geq z_{b,1}$ whose ability lies outside the skill interval of their birthplace migrate to their unique destination. Tradables producers of ability $z_m \leq z \leq z_{b,1}$ born in skilled cities migrate to the unskilled cities in order to support the steady-state population levels while minimizing migration costs. Some workers who do not produce tradables migrate from skilled cities to unskilled cities in order to support the steady-state population levels while minimizing migration costs.

If the bilateral distances between cities are orthogonal to their population characteristics and $n_u > 1$, then the expected migratory distance of tradables producers ($z \geq z_m$) exceeds the expected migratory distance of unskilled workers ($z \leq z_m$). Gross migratory flows of the unskilled are arranged so as to minimize migration costs, while only a fraction $\frac{\int_{z_m}^{z_{b,1}} \mu(z)\mathrm{d}z}{\int_{z_m}^{\infty} \mu(z)\mathrm{d}z}$ of gross flows of tradables producers are arranged to minimize migration costs.

By optimal choices of outsourcing destinations, unskilled cities exhibit identical prices and total population. Suppose that they also have identical ratios of tradables producer population to total population, $\frac{L\int_{z_m}^{\infty} \mu(z,c)\mathrm{d}z}{L_c}$. The gross migratory flows of unskilled workers and tradables producers of ability $z_m \leq z \leq z_{b,1}$ solve

$$\min_{\{x_{c,c'}\}} \sum_{c\in C_u} \sum_{c'\in C_s} x_{c',c} d(c',c) \text{ subject to } \sum_{c'\in C_s} x_{c',c} = \frac{1}{n_u} \sum_{c'\in C_s} L_{c'}\delta\lambda l \quad \forall c$$
$$\min_{\{y_{c,c'}\}} \sum_{c\in C_u} \sum_{c'\in C_s} y_{c',c} d(c',c) \text{ subject to } \sum_{c'\in C_s} y_{c',c} = \frac{1}{n_u} \sum_{c'\in C_s} L_{c'}\delta\lambda \int_{z_m}^{z_{b,1}} \mu(z)\mathrm{d}z \quad \forall c$$

Denote the optimal solutions x^* and y^*. Due to linearity, the optimal solutions are proportionate to each other. Denote the fraction $w_c = \frac{\int_{z_{b,1}}^{\infty} \mu(z,c)\mathrm{d}z}{\int_{z_{b,1}}^{\infty} \mu(z)\mathrm{d}z}$.

The average distance migrated by unskilled individuals to city c is

$$\sum_{c'\in C_s} \frac{n_u x^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda l} d(c',c)$$

The average distance migrated by unskilled individuals is

$$\sum_{c\in C_u} \frac{1}{n_u} \sum_{c'\in C_s} \frac{n_u x^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda l} d(c',c)$$

The average distance migrated by skilled individuals to city $c \in C_u$ is

$$\sum_{c'\in C_s} \frac{n_u y^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda \int_{z_m}^{z_{b,1}} \mu(z)\mathrm{d}z} d(c',c)$$

The average distance migrated by skilled individuals is

$$\frac{\int_{z_m}^{z_{b,1}} \mu(z,c)\mathrm{d}z}{\int_{z_m}^{\infty} \mu(z)\mathrm{d}z} \sum_{c\in C_u} \frac{1}{n_u} \sum_{c'\in C_s} \frac{n_u y^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda \int_{z_m}^{z_{b,1}} \mu(z)\mathrm{d}z} d(c',c) + \frac{\int_{z_{b,1}}^{\infty} \mu(z,c)\mathrm{d}z}{\int_{z_m}^{\infty} \mu(z)\mathrm{d}z} \sum_{c} \sum_{c'} w_c w_{c'} d(c',c)$$

If the bilateral distances between cities are orthogonal to their other characteristics, then by the optimality of x^* the following inequality holds:

$$\sum_{c\in C_u} \sum_{c'\in C_s} \frac{x^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda l} d(c',c) \leq \sum_{c} \sum_{c'} w_c w_{c'} d(c',c)$$

Then, because x^* is proportionate to y^*,

$$\frac{\int_{z_m}^{z_{b,1}} \mu(z,c)\mathrm{d}z}{\int_{z_m}^{\infty} \mu(z)\mathrm{d}z} \sum_{c\in C_u} \sum_{c'\in C_s} \frac{y^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda \int_{z_m}^{z_{b,1}} \mu(z)\mathrm{d}z} d(c',c) + \frac{\int_{z_{b,1}}^{\infty} \mu(z,c)\mathrm{d}z}{\int_{z_m}^{\infty} \mu(z)\mathrm{d}z} \sum_{c} \sum_{c'} w_c w_{c'} d(c',c) \geq \sum_{c\in C_u} \sum_{c'\in C_s} \frac{x^*_{c',c}}{\sum_{c''\in C_s} L_{c''}\delta\lambda l} d(c',c)$$

The expected average distance migrated by a skilled individual is greater than the expected average distance migrated by an unskilled individual.

B Parameterization

Parameterizing the model means picking a function $m(\cdot)$, a distribution $\mu(z)$, and values for $A, \bar{s}, \theta, \gamma$, and L. In the parameterizations we present in this paper, we use $m(M_c) = \frac{\exp(\nu M_c)-1}{\exp(\nu M_c)}$ with $\nu = 30$. We choose $\mu(z) = 1$ when $0 \leq z \leq 1$, so that $z \sim U(0,1)$. There is no assembly or outsourcing ($l = 0$), and we do not address life-cycle migration.

To produce the two-city wage schedule in Figure 2, we chose $A = 6, \bar{s} = .5, \theta = .25, \gamma = .5, L = 2$. To produce the four-city wage schedule in Figure 3, we chose $A = 6, \bar{s} = .5, \theta = .3, \gamma = .3, L = 4$.

C Data and estimates

C.1 Data description

Data sources: Our population data are from the US Census website (1990, 2000, 2007). Our data on individuals' wages, education, demographics, and housing costs come from public-use samples of the decennial US Census and the annual American Community Survey made available by IPUMS-USA (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek, 2010). We use the 1990 5%, and 2000 5% Census samples and the 2005-2007 American Community Survey 3-year sample. We use the 2005-2007 ACS data because ACS data from 2008 onwards only report weeks worked in intervals.

Wages: We exclude observation missing the age, education, or wage income variables. We study individuals who report their highest educational attainment as a high-school diploma or GED or a bachelor's degree and are between ages 25 and 55. We study full-time, full-year employees, defined as individuals who work at least 40 weeks during the year and usually work at least 35 hours per work. We obtain weekly and hourly wages by dividing salary and wage income by weeks worked during the year and weeks worked times usual hours per week. Following Acemoglu and Autor (2011), we exclude observations reporting an hourly wage below \$1.675 per hour in 1982 dollars, using the GDP PCE deflator. We define potential work experience as age minus 18 for high-school graduates and age minus 22 for individuals with a bachelor's degree. We weight observations by the "person weight" variable provided by IPUMS.

Housing: To calculate the average housing price in a metropolitan statistical area, we use all observations in which the household pays rent for their dwelling that has two or three bedrooms. We do not restrict the sample by any labor-market outcomes. We drop observations that lack a kitchen or phone. We calculate the average gross monthly rent for each metropolitan area using the "household weight" variable provided by IPUMS.

College ratio: Following Beaudry, Doms, and Lewis (2010), we define the "college ratio" as the number of employed individuals in the MSA possessing a bachelor's degree or higher educational attainment plus one half the number of individuals with some college relative to the number of employed individuals in the MSA with educational attainment less than college plus one half the number of individuals with some college. We weight observations by the "person weight" variable provided by IPUMS.

Note that both income and rent observations are top-coded in IPUMS data.

Geography: We map the public-use microdata areas (PUMAs) to metropolitan statistical areas (MSAs) using the "'MABLE Geocorr90, Geocorr2K, and Geocorr2010" geographic correspondence engines from the Missouri Census Data Center. For 1990 and 2000, we consider both primary metropolitan statistical areas (PMSAs) and consolidated metropolitan statistical areas (CMSAs). The 2005-2007 geographies are defined by core-based statistical areas (CBSAs). In some sparsely populated areas, only a fraction of a PUMA's population belongs to a MSA. We include PUMAs that have more than 50% of their population in a metropolitan area. Figure 1 and Table 1 describe PMSAs in 2000.

Migration: We study individuals in the 2000 Census public-use microdata who are born in the United States, 30 to 55 years of age, whose highest educational attainment is a high school degree or a bachelor's degree, and who currently live in a metropolitan area as identified by the "metaread" IPUMS variable. We identify residence changes over the five-year span using the "migrate5d" variable. We identify metropolitan changes by comparing the "migmet5" and "metaread" variables for individuals who lived in an identified metropolitan area five years earlier. We calculate distances between public-use microdata areas using the latitude and longitude coordinates of the PUMAs' centroids, calculated from US Census cartographic boundary files. We assign residences changes that do not change PUMAs a distance of zero.

C.2 Empirical estimates

Our empirical approach is to estimate cities' college wage premia and then study spatial variation in those premia. Our first-stage estimates of cities' skill premia are obtained by comparing the average log weekly wages of full-time, full-year employees whose highest educational attainment is a bachelor's degree to those whose highest educational attainment is a high school degree.

Our first specification uses the difference in average log weekly wages y in city c without any individual controls as the first-stage estimator. The dummy variable college_i indicates that individual i is a college graduate. Expectations are estimated by their sample analogues.

$$\text{premium}_c = \mathbb{E}(y_{ic}|\text{college}_i = 1) - \mathbb{E}(y_{ic}|\text{college}_i = 0)$$

Our second approach uses a first-stage Mincer regression to estimate cities' college wage premia after controlling for experience, sex, and race. The first-stage equation describing variation in the log weekly wage y of individual i in city c is

$$y_i = \gamma X_i + \alpha_c + \rho_c \text{college}_i + \epsilon_i$$

X_i is a vector containing years of potential work experience, potential experience squared, a dummy variable for males, and dummies for white, Hispanic, and black demographics. The estimated skill premium in each city, $\hat{\rho}_c$, is the dependent variable used in the second-stage regression. We refer to these estimates as "composition-adjusted skill premia."

One may be inclined to think that the estimators that control for individual characteristics are more informative. But if differences in demographics or experience are correlated with differences in ability, controlling for spatial variation in skill premia attributable to spatial variation in these factors removes a dimension of the data potentially explained by our model. To the degree that individuals' observable characteristics reflect differences in their abilities, the unadjusted estimates of cities' skill premia are more informative for comparing our model's predictions to empirical outcomes.

Table 3 shows the correlation between estimated skill premia and population sizes for various years and geographies. These coefficients are akin to those appearing in the first column of Table 1.

Table 3: Skill premia and metropolitan populations

	1990 PMSA	1990 CMSA	2000 PMSA	2000 CMSA	2005-7 CBSA
Skill premia	0.015**	0.014**	0.033**	0.029**	0.040**
	(0.0038)	(0.0039)	(0.0038)	(0.0036)	(0.0038)
Composition-adjusted skill premia	0.013**	0.013**	0.029**	0.025**	0.028**
	(0.0030)	(0.0031)	(0.0032)	(0.0030)	(0.0033)
Observations	322	271	325	270	353

Robust standard errors in parentheses
** p<0.01, * p<0.05

NOTE: Each cell reports the coefficient and standard error from an OLS regression of the estimated college wage premia on log population (and a constant). The sample is full-time, full-year employees whose highest educational attainment is a bachelor's degree or a high-school degree.