

A Spatial Knowledge Economy*

Donald R. Davis[†]

Jonathan I. Dingel[‡]

Columbia University and NBER

Columbia University

February 20, 2013

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. The 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)

*We thank Pol Antras, Kerem Coşar, Arnaud Costinot, Gilles Duranton, Jessie Handbury, Walker Hanlon, Sam Kortum, Corinne Low, Ben Marx, Joan Monras, Suresh Naidu, Kentaro Nakajima, Stephen Redding, Holger Sieg, 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, Conference on Urban and Regional Economics, Empirical Investigations in International Trade, NBER ITI meeting, NYU, Urban Economics Association annual meeting, Princeton IES Summer Workshop, Spatial Economic Research Centre annual conference, and University of Toronto for helpful comments on various drafts. We thank Paul Piveteau and Yuxiao Huang for research assistance. We are grateful to Enrico Moretti and Stuart Rosenthal for sharing their housing-price measures with us. Dingel gratefully acknowledges financial support from the Institute for Humane Studies and the Program for Economic Research at Columbia University.

[†]drdavis@columbia.edu

[‡]jid2106@columbia.edu

1 Introduction

In modern economies driven by innovation and ideas, local economic outcomes increasingly depend on local idea generation. The spatial distribution of the knowledge economy thus has implications for productivity, prices, and inequality (Moretti, 2012). Economists have long understood that cities play an important role in producing knowledge. Marshall (1890) wrote that in cities the “mysteries of the trade become no mysteries; but are as it were in the air.” One interpretation of this statement is that learning in a city is a special case of the external economies described by Henderson (1974).¹ But the freely-available-spillovers interpretation makes idea exchange a “black box,” which some have criticized as evanescent in empirical terms and close to assuming the conclusion in theoretical terms (Fujita, Krugman, and Venables 1999, p.4; Fujita and Thisse 2002, p.129). To better understand the spatial knowledge economy and its consequences, we need models in which idea exchange is an explicit economic decision.

This paper introduces a model in which costly exchange of ideas is the agglomeration force driving a variety of spatial phenomena. The essence of the model is that skilled individuals devote time to exchanging ideas with each other in order to raise their productivity, and cities with more numerous and more able conversation partners are better idea-exchange environments. High-ability individuals benefit most from these conversations, so they locate in larger cities, accessing more valuable idea exchanges at the cost of higher local prices. Individuals of lower abilities are employed in every city producing non-tradables.

Formally, we describe a population of heterogeneous individuals with a continuum of abilities in a perfectly competitive economic environment. Individuals may produce two goods, tradables and non-tradables, and those of higher abilities have comparative advantage in producing tradables. Tradables producers 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 idea exchange. 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. In equilibrium, larger cities exhibit better idea-exchange opportunities because they are populated by higher-ability individuals who devote more time to exchanging ideas. Larger cities have higher non-tradables prices that compensate the lower-ability individuals producing them for their higher costs of living.

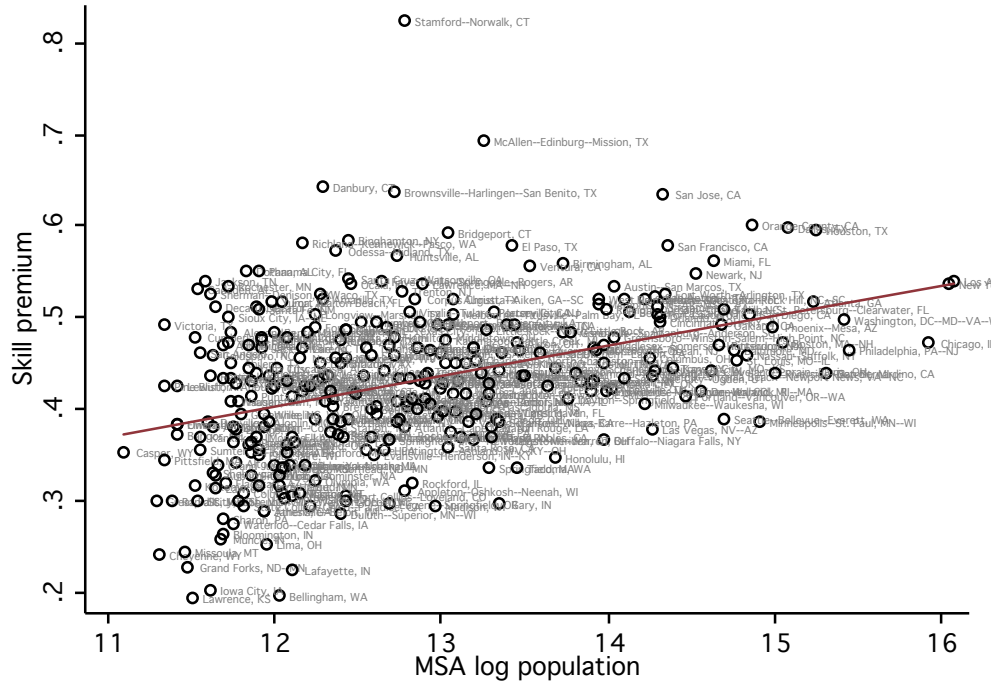
¹Abdel-Rahman and Anas (2004, p.2300): “One way to interpret this black-box model [of Marshallian externalities] is that the productivity of each worker is enhanced by the innovative ideas freely contributed by the labor force working in close proximity.”

Our model matches a broad set of facts from the empirical literature. First, cities exhibit substantial heterogeneity in size (Gabaix, 1999). While our model has symmetric fundamentals, idea-driven agglomeration 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; De la Roca and Puga, 2012), this sorting is incomplete and individuals of many skill types are present in every city. Our approach yields this imperfect sorting, since there is sorting amongst tradables producers but not amongst non-tradables producers.

Our emphasis on labor heterogeneity naturally yields predictions about spatial variation in wage inequality. Since higher-ability tradables producers locate in larger cities in order to raise their productivity by exchanging ideas and non-tradables productivity does not vary across locations, the relative productivity of tradables producers is increasing in city size. This causes relative wages to increase with city size, since the productivity gap is only partially offset by higher non-tradables prices. Figure 1 depicts spatial variation in the skilled wage premium, a relative price that captures important dimensions of wage inequality. The scatterplot shows substantial cross-city variation in skill premia, measured as differences in average log weekly wages between college graduates and high school graduates, and that a large share of this variation 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. Our model is the first spatial-equilibrium theory to predict this spatial pattern of skill premia.

Our theory’s contrasting implications for the spatial choices of skilled and unskilled workers also implies observable differences in their migratory behavior. We develop a simple dynamic extension of our model that describes costly migration in the limit as all individuals’ migration costs converge to zero, that is, as we converge to spatial equilibrium. Since tradables producers are spatially sorted, skilled individuals prefer to move to the city whose idea-exchange environment matches their ability level. Individuals producing non-tradables are equally productive in all cities, so they have weaker economic incentives to migrate. As a result, skilled tradables producers migrate more frequently and farther than unskilled non-tradables producers. This matches the prominent empirical contrast in migration rates

Figure 1: Skill premia and metropolitan populations, 2000



NOTE: The skill premium in a (primary) metropolitan statistical area 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. See appendix C for a detailed description of the data and estimation.

Table 1: Educational attainment and migration

	High school degree	Bachelor's degree
Different residence than five years prior	42%	48%
Different metropolitan area different residence	21%	33%
Average distance (km) different residence	204	365
Standard error	(0.8)	(1.4)
Average distance (km) different metropolitan area	777	979
Standard error	(3.1)	(3.5)

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.

across educational levels. Table 1 demonstrates that prime working age US-born individuals who change residences are more than 50% 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 gradu-

ate is about 80% farther than that of a high school graduate.² Comparing only those who *change* metropolitan areas, college graduates move more than 25% farther than high school graduates.

This theory is 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 a spatial-equilibrium explanation 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. Our approach is sufficiently flexible that it can be adapted to address a variety of questions about the spatial organization of economic activity within and between 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.³ 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.

2 Related literature

Our theory is related to a wide variety of previous work, which we describe in terms of idea exchange, skill premia, and migratory patterns.

2.1 Idea exchange

The role of cities in facilitating idea exchange has been noted by economists since at least Marshall (1890). The relative importance of idea exchange is presumably much greater in modern advanced economies. Glaeser and Gottlieb (2009, p. 983) suggest that “modern cities are far more dependent on the role that density can play in speeding the flow of ideas” than lowering the cost of trading goods. Similarly, Krugman (2011, pp. 5-6) writes “How can you de-emphasize technology and information spillovers in a world in which everyone’s prime

²In this calculation, we assign a distance of zero to residence changes within the same public-use microdata area. See appendix C for details.

³Recent 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.

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.” This emphasis accords well with empirical evidence suggesting that larger cities reward cognitive and people skills rather than motor skills or physical strength (Bacolod, Blum, and Strange, 2009; Michaels, Rauch, and Redding, 2013). Studies also suggest that knowledge exchanges and communication skills are more common and more valuable in larger cities (Charlot and Duranton, 2004).

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, although no money changes hands, the knowledge acquired in these exchanges is not really free. The opportunity cost is time not devoted to other productive activities. Second, since much knowledge is tacit, requiring face-to-face communication, we treat cities as the loci of learning communities.⁴ 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 idea-exchange opportunities is quite natural. Finally, we assume idea exchange 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 individuals’ spatial choices when knowledge spillovers are exogenous externalities (Henderson, 1974; Black, 1999; Lucas, 2001). Another has focused on choices of learning activities within a single location of exogenous population (Jovanovic and Rob, 1989; Helsley and Strange, 2004; Berliant, Reed III, and Wang, 2006; Berliant and Fujita, 2008; Lucas and Moll, 2011). 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.⁵ Our characterization of idea exchanges is simple compared to those

⁴This 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), Gaspar and Glaeser (1998), Audretsch and Feldman (2004), Storper and Venables (2004), and Arzaghi and Henderson (2008).

⁵Glaeser (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.

presented in the second strand of literature, but this allows us to tractably model endogenous exchanges of ideas in a system of cities.

We focus on the exchange of ideas between rather than within firms. Idea exchange within firms is surely of great significance but it does not motivate firms to locate in cities, since intra-firm idea exchange may occur in geographic isolation. Our model describes inter-firm interactions because these are the idea exchanges that can provide a foundation for urban agglomeration.⁶

2.2 Skill premia

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. Wheeler (2001) reports that returns to schooling rise with city size.⁷

In this paper, we focus on the college wage premium, a measure of inequality that has received much attention (Acemoglu and Autor, 2011). The first column of Table 2 shows a strong positive correlation between premia and population sizes, whether we measure skill premia controlling for individuals' observable demographic characteristics or not. The second through fourth columns show that this relationship is robust to controlling for two other city characteristics that prior work has linked to cities' skill premia, the fraction of the population possessing a college degree (Glaeser, 2008; Glaeser, Resseger, and Tobio, 2009; Beaudry, Doms, and Lewis, 2010) and housing prices (Black, Kolesnikova, and Taylor, 2009). 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.⁸

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

⁶Recent research supports the idea that the close physical proximity afforded by clustering in cities facilitates inter-firm communication. Allen, Raz, and Gloor (2010) describe inter-firm communication amongst individual scientists at biotech firms in the Boston area. They show that geographic proximity and firm size are both positively associated with inter-firm communication on the extensive and intensive margins. Inoue, Nakajima, and Saito (2012) describe inter-firm collaboration on Japanese patent applications. They find that inter-firm collaboration is more geographically concentrated than intra-firm collaboration and that the importance of geographic proximity has not weakened over the last two decades.

⁷Wheeler (2001) presents a theory linking inequality to market size in a single location. His approach does not address cross-city comparisons and spatial-equilibrium concerns.

⁸Regressions 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 2: Skill premia and metropolitan characteristics, 2000

<i>Skill premia</i>				
log population	0.033** (0.0038)	0.030** (0.0057)	0.037** (0.0043)	0.027** (0.0053)
log rent		0.032 (0.037)		0.12** (0.038)
log college ratio			-0.035 (0.020)	-0.075** (0.018)
R ²	0.161	0.165	0.178	0.213
<i>Composition-adjusted skill premia</i>				
log population	0.029** (0.0032)	0.030** (0.0050)	0.031** (0.0037)	0.029** (0.0048)
log rent		-0.013 (0.035)		0.029 (0.036)
log college ratio			-0.027 (0.017)	-0.037* (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.

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). 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 in a setting with spatially symmetric fundamentals and fully mobile individuals. Prior theories of spatial variation in wage inequality have relied upon introducing heterogeneity in spatial fundamentals or immobility of some factors. 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.⁹ We instead follow the example of the new economic geography: spatial heterogeneity across cities emerges from spatially symmetric fundamentals.¹⁰ Recent theoretical work by Behrens, Duranton, and Robert-Nicoud (2012) and Behrens and Robert-Nicoud (2011) has also analyzed cross-city inequality differences in a setting with spatially symmetric fundamentals and a continuum of heterogeneous individuals. However, their explanations of spatial heterogeneity in inequality are based on individuals not relocating in response to idiosyncratic shocks.¹¹ In our model, individuals freely choose their optimal locations. Our emphasis on mobility reflects the idea that spatial equilibrium is “the field’s central theoretical tool” (Glaeser and Gottlieb, 2009) and necessary for thinking about long-run spatial patterns (Moretti, 2011).

2.3 Migration

Our analysis of migration also emphasizes the mobility of individuals. It is well known that the skilled migrate more frequently than the unskilled (Greenwood, 1997; Molloy, Smith, and Wozniak, 2011). How should this contrast be incorporated 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.¹² 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,

⁹For 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. These neoclassical models do not relate skill premia to city sizes. Gyourko, Mayer, and Sinai (2006) and Eeckhout, Pinheiro, and Schmidheiny (2010) model exogenous differences in housing supply elasticities and city-level productivities, respectively.

¹⁰We 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 employees valuable opportunities to interact with others.

¹¹All workers entering a city have identical abilities in these models. Upon choosing a city, workers draw random “luck” that determines their productivities, so all within-city inequality is attributable to this stochastic process. Individuals may not relocate and take their “luck” with them. 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.”

¹²See, for example, Fujita and Thisse (2002, p.270), Tabuchi and Thisse (2002), Borck, Pflüger, and Wrede (2010), and Autor and Dorn (2012). Helpman (1998) shows how the results of Krugman (1991) are altered by modeling the centrifugal force as housing supplies rather than immobile “peasants.”

location has to be a choice, not an assumption.¹³ 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.

Our model’s explanation of differential migration rates is that more skilled individuals have greater incentives to migrate. This accords with the fact that people are highly mobile in advanced economies and respond to spatial arbitrage opportunities (Borjas, Bronars, and Trejo, 1992; Dahl, 2002; Notowidigdo, 2011) and advances from theories in which unskilled individuals are immobile. 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. 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). More recently, Wildasin (2000) suggested that more educated workers with more specialized skills find fewer potential employers in each city and thus search for employment across a larger number of cities. In his model, skilled workers change cities in response to city-industry-level productivity shocks, and unskilled workers, who are geographically immobile by assumption, change industries. Because all individuals are mobile in our theory, our model does not rely on a failure of arbitrage to make sense of spatial welfare heterogeneity.¹⁴

3 A spatial knowledge economy

The economy consists of a continuum of individuals of mass L , whose heterogeneous abilities are indexed by z and distributed with density $\mu(z)$ on connected support on $\mathbb{R}_+$. There are a number of homogeneous sites that may be populated cities. The number of cities and their populations are endogenously determined.

¹³Assuming immobility precludes other explanations for lack of movement, a point underscored in Notowidigdo (2011). These explanations have important policy implications. Citing differences in migration rates, Moretti (2012) proposes “mobility vouchers” to subsidize migration of the unemployed from high-unemployment locations to low-unemployment locations. These vouchers may be welfare-improving if the less skilled move less due to shortsightedness or credit constraints. They may be welfare-reducing if the lower migration rates of the less skilled reflect optimal choices given weak incentives to migrate.

¹⁴We recognize that short-run responses to economic shocks may be highly localized due to movement 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.

3.1 Preferences and production

Individuals consume three goods: tradables, non-tradable services, and (non-tradable) housing. Services and housing are strict necessities; after consuming $\bar{n}$ 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.¹⁵ The indirect utility function, therefore, for a consumer with income y facing prices p_c in city c is

$$V(p_c, y) = y - p_{n,c}\bar{n} - p_{h,c} \quad (1)$$

Individuals are perfectly mobile across cities and jobs, so their locational and occupational choices maximize $V(p_c, y)$.

An individual can produce tradables (t) or non-tradables (n). Non-tradables can be produced at a uniform level of productivity by all individuals. Tradables, by contrast, make use of the underlying heterogeneity. We choose units of output so that an individual's productivity in non-tradable services is unity. An individual'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 . An individual working in sector σ earns income equal to the value of her output, which is

$$y = \begin{cases} p_{n,c} & \text{if } \sigma = n \\ \tilde{z}(z, Z_c) & \text{if } \sigma = t \end{cases} \quad (2)$$

Tradables productivity depends both on an individual's ability and participation in idea exchanges. Tradables producers can acquire knowledge to increase their productivity.¹⁶ 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. Exchanging ideas is an economic decision, because time spent interacting ($1 - \beta$) trades off with time spent producing output directly (β). Production depends on own ability (z), time spent producing (β), time spent exchanging ideas ($1 - \beta$), the productivity benefits of learning (A), and local

¹⁵This specification is obviously quite special. The perfectly inelastic demand for housing and services is not meant to be realistic. The preferences are non-homothetic, but none of the results we emphasize in the paper rely on this feature – indeed non-homotheticity works against finding a positive premia-population relationship (cf. Black, Kolesnikova, and Taylor 2009). See footnote 27. These preferences have been used previously in the related work of Glaeser, Gyourko, and Saks (2006) and Moretti (2011). We use this specification for reasons of tractability.

¹⁶To focus on spatial issues, our static theory abstracts from knowledge accumulation by assuming that production may involve exchanging ideas to solve problems that are *iid*. Lucas and Moll (2011) study knowledge accumulation while abstracting from the spatial dimension.

learning opportunities (Z_c). The tradables output of an agent of ability z is

$$\tilde{z}(z, Z_c) = \max_{\beta \in [0,1]} \beta z(1 + (1 - \beta)AZ_c z) \quad (3)$$

A is a parameter common to all locations that indexes the scope for productivity gains from interactions. When A is higher, conversations with other tradables producers 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 moment of time to idea exchange in a city is the probability of encountering another individual during that moment times the expected ability of the individual encountered. Since idea exchanges are instantaneous and individuals devote an interval of time to idea exchange, every individual devoting time to exchanging ideas realizes the expected gains from these exchanges.

The probability of encountering someone during each moment of 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 a continuous, increasing function, with $m(0) = 0$ and $m(\infty) = 1$. This embodies the idea of Glaeser (1999) 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.¹⁷ In our setting the population of individuals available for such encounters is determined endogenously by tradable producers' time-allocation choices.

The average ability of the individuals encountered in these matches is $\bar{z}_c$. This is a weighted average of the abilities of producers participating in idea exchanges in which the weights are the time each type of individual devotes to interactions. Conditional on meeting another learner, the scope for gains from interactions, and one's own ability, conversations with more able individuals are more valuable.

Thus, the value of local learning opportunities Z_c reflects both a scale effect and an average ability effect. Consider city c with a population described by $\mu(z, c)$, where $\frac{\mu(z, c)}{\mu(z)}$ is

¹⁷This scale effect is not necessary for most of our results, but it is needed to explain why the best idea-exchange locations are large cities rather than gated communities in which high- z individuals exclude less able tradables producers. Most empirical evidence on matching processes describes job search, which is distinct from idea exchange in numerous dimensions. Early job-search studies, while noisy, were often interpreted to suggest constant returns (Pissarides and Petrongolo, 2001). More recent studies have found results more favorable to increasing returns to scale (Petrongolo and Pissarides, 2006; Di Addario, 2011; Bleakley and Lin, 2012).

the share of z -ability individuals in c . When individuals 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:

$$\begin{aligned} Z_c &= m(M_c) \bar{z}_c \\ M_c &= L \int_{z:\sigma(z)=t} (1 - \beta_{z,c}) \mu(z, c) dz \\ \bar{z}_c &= \begin{cases} \int_{z:\sigma(z)=t} \frac{(1 - \beta_{z,c})z}{\int_{z \geq z_m} (1 - \beta_{z,c}) \mu(z, c) dz} \mu(z, c) dz & \text{if } M_c > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (4)$$

This agglomeration mechanism trades off with a very simple congestion force. Each individual in a city of population L_c pays a net urban cost of

$$p_{h,c} = \theta L_c^\gamma \quad (5)$$

in units of the numeraire, with $\theta, \gamma > 0$. Behrens, Duranton, and Robert-Nicoud (2012) provide microeconomic foundations for this functional form, which they derive from a standard model of the internal structure of a monocentric city in which commuting costs increase with population size as governed by the technological parameters θ and γ .¹⁸ 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.

3.2 Individual behavior

Individuals choose their locations, occupations, and time allocations optimally. Since individuals are perfectly mobile, two individuals with the same ability z will obtain the same utility in equilibrium wherever they are located.

Occupational choices are governed by comparative advantage. High-ability individuals produce tradables since labor heterogeneity matters in that sector. The proof of lemma 1 is given in appendix section A.2.

Lemma 1 (Comparative advantage). *There is an ability level z_m such that individuals of greater ability produce tradables and individuals of lesser ability produce non-tradables.*

¹⁸See appendix section A.1 for details.

$$\sigma(z) = \begin{cases} t & \text{if } z \geq z_m \\ n & \text{if } z \leq z_m \end{cases}.$$

For a tradables producer of ability z in city c with learning opportunities Z_c , the optimal time spent exchanging ideas and the resulting output are

$$1 - \beta_{z,c} = \begin{cases} \frac{1}{2} \frac{AZ_c z - 1}{AZ_c z} & \text{if } AZ_c z \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad \tilde{z}(z, Z_c) = \begin{cases} \frac{1}{4AZ_c} (AZ_c z + 1)^2 & \text{if } AZ_c z \geq 1 \\ z & \text{otherwise} \end{cases} \quad (6)$$

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 idea exchange by a tradables producer is increasing in the scope for productivity gains from such exchange, the time devoted to idea exchange by others, the average ability of other learners in that location, and the producer's own ability.

There is a complementarity between a tradable producer's own ability and the value of a city's idea-exchange environment. In particular, $\tilde{z}(z, Z_c)$ is supermodular.

Lemma 2 (Supermodularity of $\tilde{z}(z, Z_c)$). *$\tilde{z}(z, Z_c)$ is supermodular in its arguments. If $AZ_c z \geq 1$, $\tilde{z}(z, Z_c)$ is strictly supermodular.*

Proof. The twice-differentiable function $\tilde{z}(z, Z_c)$ is supermodular if and only if $\frac{\partial^2}{\partial z \partial Z_c} \tilde{z}(z, Z_c) \geq 0$ (Topkis, 1998).

$$\frac{\partial^2}{\partial z \partial Z_c} \tilde{z}(z, Z_c) = \begin{cases} \frac{Az}{2} & \text{if } AZ_c z \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

□

3.3 Equilibrium

An equilibrium for a population L with ability distribution $\mu(z)$ in a set of locations $\{c\}$ is a set of prices $\{p_{h,c}, p_{n,c}\}$ and populations $\mu(z, c)$ such that workers maximize (1) by their choices of c, σ , and β and markets clear.¹⁹ Markets clear when equations (4) and (5) and the

¹⁹In this exposition, we define equilibrium where each member of the set $\{c\}$ is populated, $L_c > 0$. In appendix section A.3, 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.

following conditions hold:

$$\mu(z) = \sum_c \mu(z, c) \quad \forall z \quad (7)$$

$$L_c = L \int \mu(z, c) dz \quad \forall c \quad (8)$$

$$L_{n,c} = L \int_{z:\sigma(z)=n} \mu(z, c) dz = \bar{n} L_c \quad \forall c \quad (9)$$

The equilibrium value of local idea exchanges Z_c in equation (4) 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 .²⁰ Equation (5) defines the market-clearing housing price in each city. Equations (7) and (8) are adding-up constraints for worker types and city populations. Equation (9) equalizes demand and supply of non-tradable services within each location. The tradables market clears by Walras' Law.

4 The cross section of cities in equilibrium

4.1 Equilibrium occupations and prices

By lemma 1 and equations (7) through (9), the ability level of the individual indifferent between producing tradables and non-tradables, z_m , is given by

$$\int_0^{z_m} \mu(z) dz = \bar{n} .$$

Individuals of ability $z \geq z_m$ produce tradables and those of ability $z \leq z_m$ produce non-tradables.

Since $\tilde{z}(z, Z_c)$ is supermodular in z and Z_c , there is spatial sorting of tradables producers in which higher-ability tradables producers locate in cities with better idea-exchange environments.

Lemma 3 (Spatial sorting of tradables producers). *For $z'' > z' > z_m$, if $\mu(z'', c'') > 0$ and $\mu(z', c') > 0$ then $Z_{c''} \geq Z_{c'}$.*

Proof. $\mu(z'', c'') > 0 \Rightarrow \tilde{z}(z'', Z_{c''}) - \bar{n} p_{n,c''} - p_{h,c''} \geq \tilde{z}(z'', Z_c) - \bar{n} p_{n,c} - p_{h,c} \quad \forall c$

$\mu(z', c') > 0 \Rightarrow \tilde{z}(z', Z_{c'}) - \bar{n} p_{n,c'} - p_{h,c'} \geq \tilde{z}(z', Z_c) - \bar{n} p_{n,c} - p_{h,c} \quad \forall c$

²⁰The fixed point $Z_c = 0$ always exists. A strictly positive fixed point always exists when matching is sufficiently successful at low scale ($m(M_c)$ is close enough to one for low values of M_c).

Therefore $\tilde{z}(z'', Z_{c''}) + \tilde{z}(z', Z_{c'}) \geq \tilde{z}(z'', Z_{c'}) + \tilde{z}(z', Z_{c''})$. Since $\tilde{z}$ is supermodular, it must be that $Z_{c''} \geq Z_{c'}$. $\square$

As a result, individuals of ability z_m producing tradables will be located in the city with the lowest value of Z_c . Label cities in order of the value of their idea exchanges so that $Z_1 = \min_c \{Z_c\}$. Indifference between producing tradables and non-tradables implies that $p_{n,1}$ satisfies

$$p_{n,1} = \tilde{z}(z_m, Z_1) \quad (10)$$

There is a population of non-tradables producers located in each city. In spatial equilibrium, each of these individuals obtains the same utility, so equation (1) implies that spatial differences in non-tradables prices exactly compensate for spatial differences in housing prices.

$$(1 - \bar{n})p_{n,c} - p_{h,c} = (1 - \bar{n})p_{n,c'} - p_{h,c'} \quad \forall c, c' \quad (11)$$

All equilibria exhibit this pattern of occupations and prices. We now distinguish between equilibria based on whether cities vary in size.

4.2 Equilibrium systems of cities

There are two classes of equilibria for this economy: equilibria in which all cities have the same population sizes and equilibria with heterogeneous cities. The latter are the empirically relevant class, since there is considerable and systematic variation in cities' populations. We analyze the properties of equilibria with heterogeneous cities after describing why systems of equal-sized cities are only relevant if the gains from idea exchange are too small to cause agglomeration. When idea exchange is sufficiently rewarding, a system of heterogeneous cities emerges.²¹

4.2.1 Systems of equal-sized cities

Systems of equal-sized cities are possible equilibria because our model has spatially symmetric fundamentals. In such a system, local prices $p_{h,c}$ and $p_{n,c}$ are the same in all cities, as given

²¹We provide sufficient conditions for the existence of a stable equilibrium with two heterogeneous cities in appendix section A.4.

by equations (5) and (11). Such a system may have zero, one, or multiple cities in which idea exchanges take place. We discuss each case in turn.

In a system of cities in which idea exchange occurs nowhere, no tradables producer devotes time to idea exchange because no other tradables producer does, and $Z_c = 0 \forall c$. While the no-idea-exchange equilibrium will not be the focus of our discussion, it does illustrate an important aspect of the economic mechanisms. It underscores the fact that ideas here are not manna from heaven but the outcome of a costly allocation of time by those acquiring knowledge. The no-idea-exchange equilibria are not of theoretical interest because they are not robust to coordination when there are potential gains from idea exchange. There are potential gains when the scope for gains (A), the local population size, and the abilities of tradables producers are sufficiently high that there is a value of Z_c satisfying equation (4) for the relevant idea-exchange participants. Since exchanging ideas is a Pareto improvement (it raises productivity for all participants without lowering the productivity of any other individual), communication or coordination among (a sufficiently large set of) tradables producers would facilitate its choice.

A system of cities of identical population size in which idea exchange occurs in a single location is only an equilibrium when the potential gains from idea exchange are too low to support agglomeration. Denote the single city in which idea exchange occurs by C such that $Z_C > 0$. Since all cities have the same prices, all individuals who might gain from idea exchange, those of ability $z > \frac{1}{AZ_C}$, live in city C and participate in idea exchange. This arrangement is only an equilibrium when the scope for gains from idea exchange and the abilities of tradables producers are so low that all those possibly gaining from idea exchange can live in a single city.

In a system of cities of identical population size in which idea exchange occurs in multiple cities, the value of idea exchange must be identical across locations. If cities differed in the value of their idea-exchange opportunities while exhibiting the same prices, tradables producers would move to the city with a higher value of Z_c , making it more populous. Unless the gains from idea exchange are too small relative to congestion costs, such a system of identical cities is not locally stable, because the movement of some high-ability tradables producers from one city to the other would improve the latter's idea-exchange environment, thereby drawing in more tradables producers.²²

Thus, a system of equal-sized cities is only a stable equilibrium configuration if the potential gains from idea exchange are so low as to prevent agglomeration. When broad

²²See appendix section A.5 for our definition of local stability and the relevant argument.

participation in idea exchange occurs, a system of heterogeneous cities emerges.

4.2.2 Systems of heterogeneous cities

Equilibria with heterogeneous cities exhibit cross-city patterns that can be established independent of the number of cities that arise.²³ Proposition 1 states that larger cities exhibit higher housing prices, higher non-tradables prices, better idea-exchange opportunities, and more able populations of tradables producers. Its proof is in appendix section A.6.

Proposition 1 (Heterogeneous cities' characteristics). *In equilibrium, if $L_c > L_{c'}$ then $p_{h,c} > p_{h,c'}$, $p_{n,c} > p_{n,c'}$, $Z_c > Z_{c'}$, and $\mu(z, c)\mu(z', c') \geq \mu(z, c')\mu(z', c) = 0$ for any $z > z' > z_m$.*

The mechanics of Proposition 1 are straightforward. Larger cities have higher housing prices due to congestion, so non-tradables producers require higher wages in these locations. Larger cities attract tradables producers because the benefits of more valuable idea exchanges offset their higher housing and non-tradables prices. More able tradables producers benefit more from participating in better idea exchanges, so there is spatial sorting of tradable producers. This spatial sorting supports equilibrium differences in idea-exchange environments because these high-ability individuals are better idea-exchange partners.²⁴

Equilibria with heterogeneous cities 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 empirical implication that skill premia will be higher in larger cities.

4.3 Skill premia with heterogeneous cities

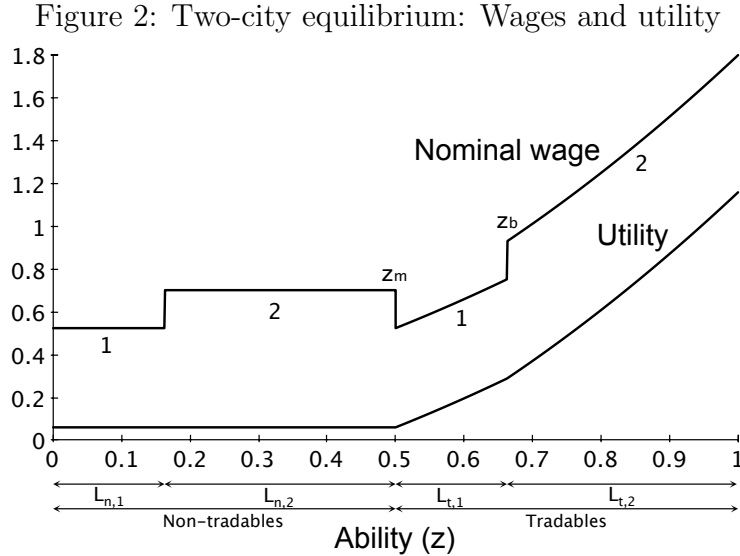
When cities are heterogeneous, we find that observed skill premia are higher in more populous cities. We first use numerical examples to illustrate the economic mechanisms driving this novel implication. Proposition 2 then formally states this prediction for a two-city equi-

²³Since these patterns characterize all equilibria with heterogeneous cities, we do not address issues of uniqueness.

²⁴Any microfoundations for Z_c in which cities with a larger mass of higher-ability tradables producers exhibit a higher endogenous value of Z_c will support a sorting outcome.

librium when ability is distributed Pareto and provides sufficient conditions for the two-city, uniform-distribution case.

Figure 2 shows the nominal wage and utility outcomes for a particular parameterization of our model in a two-city equilibrium.²⁵ City 2 is more populous than city 1. Worker ability, indexed by z , appears on the horizontal axis. We assume here that ability is uniformly distributed, and this means that the width of the interval is proportional to city population. 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.²⁶ 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.



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 their utility constant across cities.

Tradables producers' wages are higher in larger cities for three reasons. First, there is a composition 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

²⁵See appendix section B for details of this parameterization.

²⁶See appendix section A.4 for the formal definition of this $\mu(z, c)$.

for tradables producers. Third, there is a compensation effect. Producers who are indifferent between two cities must have a wage gap that exactly matches the gap in non-tradables and housing prices between those cities.

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 (common) non-tradables wage $p_{n,c}$.

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

In equilibria with heterogeneous cities, the cross-city pattern of skill premia depends upon the composition, learning, and compensation effects. The composition and learning effects yield higher nominal incomes for tradables producers in larger cities. These raise tradables producers' wages relative to non-tradables producers' wages in larger cities and therefore generate a positive premium-population relationship. The compensation effect that reflects the differences in local prices between cities makes the nominal wages of both tradables and non-tradables producers higher in larger cities. Since higher-ability individuals 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 .²⁷ This pushes towards a negative premium-population relationship. When the composition and learning effects dominate this implication of the compensation effect, the skill premium is higher in the larger city.

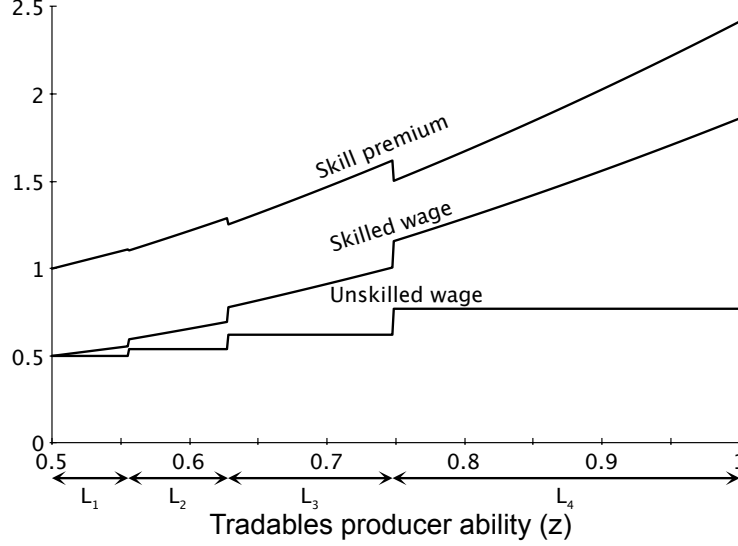
Figure 3 illustrates the pattern of wage premia for a four-city example.²⁸ It compares the incomes of tradables and non-tradables producers by placing the wage schedules on a common horizontal axis. The ratio of the wage schedules gives the skill premium of each tradables producer relative to the non-tradables 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 composition and learning effects that raise the skill premium, due to the differences in inframarginal tradables producers' abilities and the differences in the productivity gains arising from idea exchanges,

²⁷This 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.

²⁸See appendix section B for the parameter values underlying this example. Interval widths are proportionate to city populations.

are greater than the compensation effect that lowers the skill premium. Here larger cities exhibit higher observed skill premia.²⁹

Figure 3: Four-city equilibrium: Skill premia



The equilibrium pattern of skill premia depends on the distribution of abilities, $\mu(z)$. Proposition 2 states that, in a two-city equilibrium, the skill premium is higher in the more populous city when ability is Pareto distributed. It provides sufficient conditions for this inequality to hold for the uniform distribution. The proof is in appendix section A.7.

Proposition 2 (Skill premia in two-city equilibria). *In an equilibrium in which $L_2 > L_1$, $\tilde{z}(z_m, Z_1) = p_{n,1}$, and z_b is the value such that $\tilde{z}(z_b, Z_1) - p_{n,1}\bar{n} - p_{h,1} = \tilde{z}(z_b, Z_2) - p_{n,2}\bar{n} - p_{h,2}$:*

- (a) *If $\mu(z) = \frac{kb^k}{z^{k+1}}$ for $z \geq b$ with $k > 1$, then $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$*
- (b) *If $z \sim U(\mathfrak{z}, \hat{z})$ and $\hat{z}z_m > z_b^2$, then $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$*

To study the robustness of this prediction for more than two cities, we use numerical optimization to search for parameter values minimizing the correlation between city sizes and skill premia for equilibria with two to ten cities. For the uniform distribution, the premia-size correlation is minimized by letting $\bar{n} \rightarrow 1$ so that the mass of inframarginal tradables producers shrinks to zero and the relative influence of the compensation effect is

²⁹The assumption of homogeneous non-tradables productivity is not crucial to this result. With heterogeneous non-tradables productivity, the preferences in equation (1) would induce spatial sorting amongst non-tradables producers. This would introduce another composition effect and weaken the compensation effect. It can be shown that an analogue of Proposition 2 holds so long as labor heterogeneity is sufficiently more important in the tradables sector.

maximized. For the Pareto distribution, the correlation is minimized by letting $k \rightarrow \infty$ so that variation in tradables producers' abilities is minimized. For all parameter values yielding equilibria with heterogeneous cities, the observed skill premia are strictly increasing in city population. The prediction that skill premia are higher in larger cities appears to be a robust feature of our model.

Our model does not yield closed-form comparative statics for the response of skill premia to parameter values, but a numerical example demonstrates that it can yield results consistent with the literature. Work in labor economics has emphasized skill-biased technical change as one reason for growth in the (nationwide) skill premium (Acemoglu and Autor, 2011), and work in urban economics has suggested that increasingly inelastic housing supplies may contribute to spatial inequality (Ganong and Shoag, 2012). We interpret the former as an increase in A and the latter as an increase in γ . The equilibrium with four heterogeneous cities presented in Figure 3 exhibits an economy-wide average skill premium of about 60%, and its city-size distribution yields a power law exponent of about 1.3, which are plausible magnitudes. Locally, an increase in the scope for gains from idea exchange, A , is associated with a higher average skill premium and greater geographic concentration of population in the largest city. An increase in the population elasticity of congestion costs, γ , is associated with lower population in the largest city and thus a higher power law exponent. Since the population pattern is more stable than housing prices and wages over time, we consider a simultaneous increase in A and γ that leaves the four cities' populations largely unchanged. This shift causes an increase in the economy-wide average skill premium and a stronger correlation between cities' skill premia and population sizes. Thus, the parameterization depicted in Figure 3 exhibits comparative statics that are qualitatively consistent with mechanisms that have been suggested as important to explaining wage and spatial inequality.

5 Outsourcing and migration in spatial equilibrium

In this section, we extend our model to explain key facts about migration, notably that skilled workers move more often and farther than unskilled workers. The challenge is to explain the differential *movement* of the 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 tradables production component carried out by the unskilled, “back-office tasks”, which can be carried out locally or outsourced. Producers in larger cities, where unskilled nominal wages are higher, outsource back-office tasks. Outsourcing makes larger cities exhibit a higher skilled to unskilled ratio and will enrich our model of migration.³⁰

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 moving only seek the nearest city with notional excess demand for their labor. Differential mobility and differential movement are not the same. Our theory delivers the latter without requiring the former. The empirical observation of differential migration rates can be accounted for in a spatial-equilibrium framework.

5.1 Outsourcing

To this point, there has been a single, competitively produced homogeneous tradable good, so this notionally tradable good has not in fact been traded. We now introduce a richer model of tradables production. 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.³¹

We thus amend tradables production to require a second, back-office task.³² The back-

³⁰If 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.

³¹Duranton 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.

³²The *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,

office task requires $l < 1 - \bar{n}$ 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 $l \cdot p_{n,c}$ when producing output in city c .³³

Tradables firms may pay a fixed cost f_b to establish a back-office center in a location other than where the firm is headquartered. The net benefit to a tradables producer located in city c of outsourcing this task to city c' is $l(p_{n,c} - p_{n,c'}) - f_b$. Thus, tradables producers in a large, high- z city will outsource back-office tasks to the smaller, lower- z city if the gap in back-office prices is sufficiently large. Such outsourcing raises the relative skill level of larger cities by shifting unskilled back-office 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 a back-office assignment function $\rho(c, c')$, which describes the fraction of back-office tasks for firms headquartered in c that are performed in c' .³⁴ The equilibrium conditions for tradables production require that tradables producers take back-office costs into account when choosing their headquarters location, the back-office location assignments minimize back-office costs, and labor markets clear.

$$y = \begin{cases} p_{n,c} & \text{if } \sigma = n \\ \tilde{z}(z, Z_c) - l \sum_{c'} \rho(c, c') p_{n,c'} - \sum_{c' \neq c} \rho(c, c') f_b & \text{if } \sigma = t \end{cases} \quad (2')$$

$$\{\rho(c, c')\} \in \arg \min_{\{\rho(c, x)\}} l \sum_x \rho(c, x) p_{n,x} - \sum_{x \neq c} \rho(c, x) f_b \quad \text{s.t.} \quad \sum_x \rho(c, x) = 1 \quad (12)$$

$$L_{n,c} = L \int_{z \leq z_m} \mu(z, c) dz = \bar{n} L_c + l \sum_{c'} \rho(c', c) L \int_{z \geq z_m} \mu(z, c') dz \quad (9')$$

The rest of the equilibrium conditions in section 3 are unchanged.

In the absence of outsourcing, the skilled share of each city's population was $1 - \bar{n}$ 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 weakly increasing in

“Financial Giants Moving Jobs Off Wall Street,” 1 July 2012).

³³ Assuming that each tradables producer requires l units of back-office tasks regardless of $\tilde{z}(z, Z_c)$ is the simplest conceivable process.

³⁴ The optimal choice of back-office 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.

city size, matching the empirical tendency. The strength of this relationship depends on the fragmentation cost f_b and the relative magnitudes of l and $\bar{n}$.

Many believe that fragmentation costs have fallen substantially in recent decades (Duranton and Puga, 2005). Our model predicts that this will trigger outsourcing of back-office tasks and generate a positive correlation between cities' skilled population share and total population.

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 skilled workers move more frequently than unskilled workers in the data. In this section we develop a very simple stationary dynamic version of our static model in which all individuals are 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 back-office tasks 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 ability distribution time-invariant, so the dynamic equilibrium is stationary. 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 people migrate? Consider first those born in city 2, the relatively "skilled city." Newborns with ability $z \in (z_b, \infty)$ will stay in the skilled city. Newborns with ability $z \in (z_m, z_b)$ will migrate to the unskilled city. Some individuals with ability $z \leq z_m$ (newborn or not) have reason to migrate to the unskilled city because the larger city's outsourcing-induced lower unskilled share ($\frac{L_{n,2}}{L_2} < \frac{L_{n,1}}{L_1}$) means that the fraction of newborns with ability $z \leq z_m$ there exceeds the equilibrium fraction. In city 1, only those with ability $z \in (z_b, \infty)$ have reason to migrate to city 2. There is therefore 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 flow of the skilled matching the gross flow 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 ability, 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 a multi-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 city's equilibrium ability distribution and the abilities of the $\delta\lambda L_c$ newborns whose characteristics are orthogonal to those of their parents and drawn from the economy-wide ability distribution. 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 multi-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.8 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. Skilled workers are employed in tasks that are sensitive to differences in ability and depend on local interactions, which motivates more spatially extensive searches. Less skilled workers are employed in tasks that exhibit less spatial heterogeneity in real returns, so they search across cities less.³⁵ Both the frequency and distance of moves reflect this.

6 Conclusion

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. In our model, individuals allocate their time according to the expected gains from exchanging ideas in their city. The gains from idea exchange are greater in places where conversation partners are more numerous and of higher ability. 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.

³⁵By assuming zero heterogeneity in non-tradables productivity, we have obtained the result that each tradables producer has a uniquely optimal location and all non-tradables producers are indifferent amongst all populated places. The logic of our migration results would still apply when non-tradables producers prefer some cities to others, provided greater heterogeneity in economic outcomes across locations for the skilled than the unskilled and some costs of migration. Migration costs diminish migration more amongst those with smaller spatial differences in economic outcomes. CEOs search for jobs in a national market; janitors do not.

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. These few assumptions cause cities to vary in size, and those in larger cities exchange ideas more frequently with more people whose average ability is higher. Larger cities exhibit 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.
- ALLEN, T. J., O. RAZ, AND P. GLOOR (2010): “Does geographic clustering still benefit high tech new ventures? The case of the Cambridge/Boston biotech cluster,” in *Innovation in Complex Social Systems*, ed. by P. Ahrweiler. Routledge.
- 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.
- BACOD, 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 (2012): “Productive cities: Sorting, selection and agglomeration,” mimeo.
- 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.
- BLEAKLEY, H., AND J. LIN (2012): “Thick-market effects and churning in the labor market: Evidence from US cities,” *Journal of Urban Economics*, 72(2), 87–103.
- 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.
- CHEN, Y., AND S. S. ROSENTHAL (2008): “Local amenities and life-cycle migration: Do people move for jobs or fun?,” *Journal of Urban Economics*, 64(3), 519–537.
- 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.
- DE LA ROCA, J., AND D. PUGA (2012): “Learning by working in big cities,” mimeo.
- DI ADDARIO, S. (2011): “Job search in thick markets,” *Journal of Urban Economics*, 69(3), 303–318.
- 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.
- ECKHOUT, 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 (2002): *Economics of Agglomeration*. Cambridge University 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.
- GANONG, P., AND D. SHOAG (2012): “Why Has Regional Convergence in the U.S. Stopped?,” HKS Working Paper RWP12-028.
- GASPAR, J., AND E. L. GLAESER (1998): “Information Technology and the Future of Cities,” *Journal of Urban Economics*, 43(1), 136–156.
- GIBBONS, S., H. G. OVERMAN, AND P. PELKONEN (2010): “Wage Disparities in Britain: People or Place?,” SERC Discussion Paper 0060.
- 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.
- (2012): “Coagglomeration and the Scale and Composition of Clusters,” UC Berkeley mimeo.
- HENDERSON, J. V. (1974): “The Sizes and Types of Cities,” *American Economic Review*, 64(4), 640–56.
- INOUE, H., K. NAKAJIMA, AND Y. U. SAITO (2012): “Geographical Concentration of Inter-Organizational Collaborations,” mimeo.
- 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.
- JOVANOVIC, B., AND R. ROB (1989): “The Growth and Diffusion of Knowledge,” *Review of Economic Studies*, 56(4), 569–82.
- 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.
- (2001): “Externalities and Cities,” *Review of Economic Dynamics*, 4(2), 245–274.
- 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.
- MICHAELS, G., F. RAUCH, AND S. J. REDDING (2013): “Task Specialization in U.S. Cities from 1880-2000,” NBER Working Paper 18715.
- 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.
- (2012): *The New Geography of Jobs*. Houghton Mifflin Harcourt.
- NAKAMURA, E., AND J. STEINSSON (2011): “Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions,” NBER Working Paper 17391.
- NOTOWIDIGDO, M. J. (2011): “The Incidence of Local Labor Demand Shocks,” NBER Working Paper 17167.
- PETRONGOLO, B., AND C. PISSARIDES (2006): “Scale Effects in Markets with Search,” *Economic Journal*, 116(508), 21–44.
- PISSARIDES, C. A., AND B. PETRONGOLO (2001): “Looking into the Black Box: A Survey of the Matching Function,” *Journal of Economic Literature*, 39(2), 390–431.
- 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.
- STORPER, M., AND A. J. VENABLES (2004): “Buzz: face-to-face contact and the urban economy,” *Journal of Economic Geography*, 4(4), 351–370.
- TABUCHI, T., AND J.-F. THISSE (2002): “Taste heterogeneity, labor mobility and economic geography,” *Journal of Development Economics*, 69(1), 155–177.
- TOPKIS, D. M. (1998): *Supermodularity and Complementarity*. Princeton University Press.
- WHEELER, C. H. (2001): “Search, Sorting, and Urban Agglomeration,” *Journal of Labor Economics*, 19(4), 879–99.
- WILDASIN, D. E. (2000): “Labor-Market Integration, Investment in Risky Human Capital, and Fiscal Competition,” *American Economic Review*, 90(1), 73–95.

A Theory

A.1 Internal urban structure

To introduce congestion costs, we follow Behrens, Duranton, and Robert-Nicoud (2012) and adopt a standard, highly stylized model of cities' internal structure.³⁶ 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.

Individuals choose a residential location x to minimize the sum of land rent and commuting cost, $r(x) + \tau x^\gamma$. In equilibrium, individuals 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

$$TLR = \int_{-\frac{L}{2}}^{\frac{L}{2}} r(x) dx = 2 \int_0^{\frac{L}{2}} r(x) dx = 2\tau \left(\left(\frac{L}{2}\right)^{\gamma+1} - \frac{1}{\gamma+1} \left(\frac{L}{2}\right)^{\gamma+1} \right) = \frac{2\tau\gamma}{\gamma+1} \left(\frac{L}{2}\right)^{\gamma+1}$$

The city's total commuting cost is

$$TCC = 2 \int_0^{\frac{L}{2}} \tau x^\gamma dx = \frac{2\tau}{\gamma+1} \left(\frac{L}{2}\right)^{\gamma+1} \equiv \theta L^{\gamma+1}$$

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 $p_{h,c} = \theta L_c^\gamma$.

³⁶There is nothing original in this urban structure. We use notation identical to, and taken from, Behrens, Duranton, and Robert-Nicoud (2012).

A.2 Comparative advantage

There is an ability level z_m such that individuals of greater ability produce tradables and individuals of lesser ability produce non-tradables.

$$\sigma(z) = \begin{cases} t & \text{if } z \geq z_m \\ n & \text{if } z \leq z_m \end{cases}$$

Proof. First, we can identify an ability level dividing tradables and non-tradables producers in each city. Consider city c with price $p_{n,c} \geq 0$ and idea-exchange opportunities Z_c . Since tradables output $\tilde{z}(z, Z_c)$ is strictly increasing in z and not bounded from above, there is a unique value $z_{m,c}$ such that $p_{n,c} = \tilde{z}(z_{m,c}, Z_c)$. Individuals of ability $z < z_{m,c}$ produce non-tradables and individuals of ability $z > z_{m,c}$ produce tradables in city c .

Second, there is an ability level dividing tradables and non-tradables producers across all locations, which we denote z_m . Individuals of ability $z \leq z_m$ produce non-tradables and individuals of ability $z \geq z_m$ produce tradables. Suppose not. If there is not an ability level dividing tradables and non-tradables production across all locations, there are abilities z', z'' such that, without loss of generality, $z' < z''$ and z' produces tradables in city c' and z'' produces non-tradables in city c'' . The former's choice means $\tilde{z}(z', Z_{c'}) - p_{n,c'}\bar{n} - p_{h,c'} \geq (1 - \bar{n})p_{n,c''} - p_{h,c''}$. The latter's choice means $(1 - \bar{n})p_{n,c''} - p_{h,c''} \geq \tilde{z}(z'', Z_{c'}) - p_{n,c'}\bar{n} - p_{h,c'}$. Together, these imply $\tilde{z}(z', Z_{c'}) \geq \tilde{z}(z'', Z_{c'})$, contrary to the fact that $\tilde{z}(z, Z_c)$ is strictly increasing in z . $\square$

A.3 The number of cities

In section 3.3, 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 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 ($p_{n,c} = \frac{p_{h,c}}{1-\bar{n}} \forall c : L_c > 0$), non-tradables producers living in cities obtain strictly positive utility. Therefore the entire

population lives in a finite number of cities. For a given set of parameters, there may exist multiple equilibria that have distinct numbers of heterogeneous cities. We see no theoretical reason to believe that the equilibrium number of populated cities must 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.

A.4 Existence of equilibrium with two heterogeneous cities

Here we characterize three sufficient conditions for $\{L, \mu(z), \bar{n}, \theta, \gamma, A, m(\cdot)\}$ such that there exists a two-city equilibrium in which $L_1 < L_2$. The first is that idea exchange creates potential gains from agglomeration. The second is that congestion costs prevent the entire population from living in a single city. The third is that it is feasible for the entire population to live in two cities.

To help define the three conditions, let $Z_c(x, y)$ denote the maximum value of Z_c satisfying equations (4) and (6) when the population of tradables producers in city c is all individuals with abilities in the $[x, y]$ interval. Formally, the maximum value of Z_c satisfying those equations when $\mu(z, c) = \mu(z) \forall z \in [x, y]$ and $\mu(z, c) = 0 \forall z \in [z_m, x) \cup (y, \infty)$ where z_m is given by $\bar{n} = \int_0^{z_m} \mu(z) dz$. The *agglomeration condition* is that $AZ_c(z, \infty)z > 1$ where z is the median tradables producer, identified by $\frac{1-\bar{n}}{2} = \int_{z_m}^z \mu(z) dz$. This condition says that technology $(A, m(\cdot), \bar{n})$ and population $(L$ and $\mu(z))$ are such that the median tradables producer and every individual of greater ability would find idea exchange with one another profitable if they all colocated. In other words, there are potential gains from agglomeration via idea exchange. The *congestion condition* is that the congestion costs of locating the economy's entire population in a single city exceed the gains from idea exchange for the lowest-ability tradables producer, $\frac{\theta}{1-\bar{n}}L^\gamma > \tilde{z}(z_m, Z_c(z_m, \infty)) - z_m$. The *feasibility condition* is that the least-able tradables producer generates enough output to cover the congestion costs associated with two cities, $z_m \geq \frac{\theta}{1-\bar{n}} \left(\frac{L}{2}\right)^\gamma$.

We now characterize the economy in terms of L_1 and define an equation $\Omega(L_1)$ that equals zero when the economy is in equilibrium. Choose a value $L_1 \leq \frac{1}{2}L$, which implies $L_2 = L - L_1$. Define values z_b and $z_{b,n}$ that respectively denote the highest-ability tradables and non-tradables producers in city 1 by

$$(1 - \bar{n})L_1 = L \int_{z_m}^{z_b} \mu(z) dz \quad \bar{n}L_1 = L \int_0^{z_{b,n}} \mu(z) dz.$$

Because the support of $\mu(z)$ is connected, z_b is continuous in L_1 . The locational assignments

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

satisfy equations (7), (8), and (9). These assignments imply values for $p_{h,1}, p_{h,2}, p_{n,1}, p_{n,2}, Z_1, Z_2$, and $\beta_{z,c}$ via equations (4), (5), (6), (10), and (11), where we select the maximal values of Z_1 and Z_2 satisfying those equations. The feasibility condition ensures these assignments are possible for all L_1 .

This is a spatial equilibrium if z_b is indifferent between the two cities. Utility in the smaller city minus utility in the larger city for the marginal tradables producer, z_b , is

$$\tilde{z}(z_b, Z_1(z_m, z_b)) - \bar{n}p_{n,1} - p_{h,1} - (\tilde{z}(z_b, Z_2(z_b, \infty)) - \bar{n}p_{n,2} - p_{h,2})$$

Using equations (5) and (11) and rearranging terms, we call this difference $\Omega(L_1)$.

$$\Omega(L_1) \equiv \frac{\theta}{1-\bar{n}}(L_2^\gamma - L_1^\gamma) - \tilde{z}(z_b, Z_2(z_b, \infty)) + \tilde{z}(z_b, Z_1(z_m, z_b))$$

Ω can be written solely as a function of L_1 because all the other variables are given by L_1 via $z_{b,n}$ and z_b through the locational assignments and other equilibrium conditions.

$\Omega(L_1) = 0$ is an equilibrium. $\Omega(\frac{L}{2}) < 0$ since equal-sized cities have equal prices and the agglomeration condition ensures that $Z_2 > Z_1$ at $L_1 = \frac{1}{2}L$. $\lim_{L_1 \rightarrow 0} \Omega(L_1) > 0$ due to the congestion condition. If $\Omega(L_1)$ is appropriately continuous, then there is an intermediate value $L_1 \in (0, \frac{L}{2})$ satisfying $\Omega(L_1) = 0$. We now show that $\Omega(L_1)$ is continuous almost everywhere and that any discontinuity point is an upward jump and therefore not a problem.

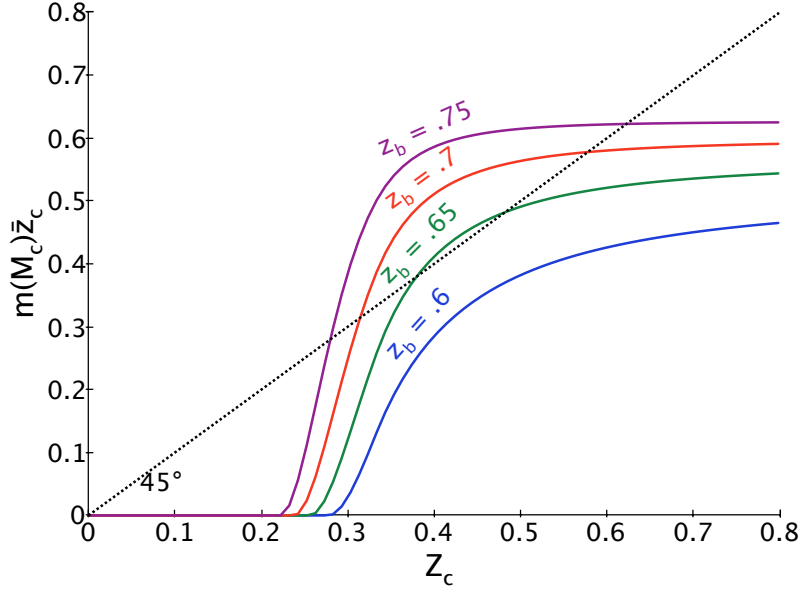
The first term, $\frac{\theta}{1-\bar{n}}(L_2^\gamma - L_1^\gamma)$, is obviously continuous in L_1 .

The second term is continuous in L_1 , provided that the agglomeration condition holds. Since β_z 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 (4) and (6) is where the function $m(M_c)\bar{z}_c$ intersects the 45-degree line. The agglomeration condition means that such an intersection $Z_2 = m(M_2)\bar{z}_2$ exists for all values $L_1 \in (0, \frac{1}{2}L)$. Since $\beta_{z,2}$ is continuous in Z_2 , and M_2 and $\bar{z}_2$ are continuous in Z_2 and z_b , $m(M_2)\bar{z}_2$ is continuous in Z_2 and z_b . This means that $Z_2(z_b, \infty)$ is a continuous function of L_1 . $\tilde{z}(z, Z_c)$ is continuous in its arguments. Thus, $\tilde{z}(z_b, Z_2(z_b, \infty))$ is continuous in L_1 .

$Z_1(z_m, z_b)$ 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 (4) and (6) because $m(M_c)$ is continuous and $m(0) = 0$. When L_1 becomes sufficiently large that there is a value of Z_1 satisfying equations (4) and (6), 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 $Z_1(z_m, z_b) > 0$ for some $L_1 \in (0, \frac{1}{2}L)$, then $Z_1(z_m, z_b)$ and Ω discontinuously increases at one value of L_1 and are continuous everywhere else in $(0, \frac{1}{2}L)$.

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



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

Since $\lim_{L_1 \rightarrow 0} \Omega(L_1) > 0$, $\Omega(\frac{L}{2}) < 0$, and Ω increases at any point at which Ω is not continuous in L_1 , there exists a value of L_1 such that $\Omega(L_1) = 0$. This is an equilibrium with heterogeneous cities. Since $\Omega(L_1)$ crosses zero from above, it is a stable equilibrium, as will be defined in appendix section A.5.

A.5 Stability of equilibria

In this section, we consider stability of equilibria. First, we describe why a traditional definition of stability is inapplicable to our model. The traditional definition assumes that potential movement plays no role in economic outcomes when individuals are at their equilibrium locations, while our model's equilibrium outcomes depend on potential movement through a spatial no-arbitrage condition. Second, we define a notion of local stability and show that stable equilibria can have equal-sized cities only if agglomeration is weak relative to congestion.

A traditional definition of stability in spatial theory considers perturbations that re-allocate a small, arbitrary mass of individuals away from their equilibrium locations. If individuals would obtain greater utility in their initial locations than in their arbitrarily assigned locations, then the equilibrium is said to be stable. This concept requires calculating each individual's utility in a location given an arbitrary population allocation. Doing so is straightforward in models in which goods and labor markets clear city-by-city, so that an individual's utility in a location can be written solely as a function of the population in that location, as in Henderson (1974), Behrens, Duranton, and Robert-Nicoud (2012), and Helsley and Strange (2012). It is also feasible in models in which the goods and labor markets clear for any arbitrary population allocation through inter-city trade, as in Krugman (1991). In these settings, shutting down labor mobility does not change any economic outcomes if individuals are at their equilibrium locations.

By contrast, in our model the potential movement of individuals is crucial to determining equilibrium outcomes. Non-tradables prices are linked across cities in equilibrium by a no-arbitrage condition, equation (11). Consider solving for equilibrium with an arbitrary population allocation and not allowing individuals to move. Clearing the goods and labor markets would require $p_{n,c} = \tilde{z}(z_{m,c}, Z_c)$ in each city, where $z_{m,c}$ is defined by $\int_0^{z_{m,c}} \mu(z, c) dz = \bar{n} \int_0^\infty \mu(z, c) dz$ for the arbitrary $\mu(z, c)$. Therefore, the prices and utilities obtained when clearing markets conditional on an arbitrary population allocation would not equal the equilibrium prices and utilities even when evaluated at the equilibrium population allocation. This stems from the inseparability of labor-market outcomes and labor mobility. So we need a distinct approach to evaluating stability.

We define a class of perturbations that maintains spatial equilibrium amongst non-tradables producers so that stability can be assessed in terms of tradables producers' incentives. Starting from an equilibrium allocation $\mu^*(z, c)$, we consider perturbations in which a small mass of tradables producers and a mass of non-tradables producers whose net supply

equals the tradables producers' demand for non-tradables move from one city to another. The equilibrium allocation is stable if the tradables producers who moved would obtain higher utility in their equilibrium city than in their new location.

Definition 1 (Perturbations). *A perturbation of size ϵ is a measure $d\mu(z, c)$ satisfying*

- $L \sum_c \int |d\mu(z, c)| dz = 2\epsilon$ [we move a mass of ϵ]
- $\sum_c d\mu(z, c) = 0 \ \forall z$ [the aggregate population of any z is unchanged]
- $\{c : d\mu(z, c) > 0\}$ is a singleton and $\{c : d\mu(z, c) < 0\}$ is a singleton [we move them in one direction from a single city to another]
- $(1 - \bar{n}) \int_0^{z_m} |d\mu(z, c)| dz = \bar{n} \int_{z_m}^\infty |d\mu(z, c)| dz$ [we move non-tradables producers to satisfy the tradables producers' demand]

Definition 2 (Local stability). $\mu^*(z, c)$ is locally stable if there exists an $\bar{\epsilon} > 0$ such that

$$\tilde{z}(z, Z'_{c_1}) - \frac{\theta}{1 - \bar{n}} L'_{c_1}{}^\gamma \geq \tilde{z}(z, Z'_{c_2}) - \frac{\theta}{1 - \bar{n}} L'_{c_2}{}^\gamma \ \forall z, c_1, c_2 : z > z_m \ \& \ d\mu(z, c_1) < 0 \ \& \ d\mu(z, c_2) > 0$$

for all population allocations $\mu'(z, c) = \mu^*(z, c) + d\mu(z, c)$ in which $d\mu$ is a perturbation of size $\epsilon \leq \bar{\epsilon}$, where Z'_c , and L'_c denote the values of these variables when the population allocation is μ' , individuals maximize (1) by their choices of σ and β , markets clear, and prices satisfy equations (10) and (11).

Proposition 3 (Instability of symmetric cities). *If agglomeration benefits are sufficiently strong relative to congestion costs, two cities of the same population size with positive idea exchange cannot coexist in a locally stable equilibrium.*

Proof. Suppose $L_1 = L_2$ and $Z_1 = Z_2 > 0$. Without loss of generality, consider perturbations of size $\epsilon \leq \bar{\epsilon}$ moving individuals from city 1 to city 2. Since $\tilde{z}(z, Z_c)$ is supermodular, the highest-ability producers have the most to gain from a move and it is sufficient to consider perturbations of size ϵ in which all tradables producers in the range $[z^*(\epsilon), \infty]$ move from city 1 to city 2; these are perturbations $d\mu$ that satisfy $L \int_{z^*(\epsilon)}^\infty \mu(z, 1) dz = (1 - \bar{n})\epsilon$ and $d\mu(z, 1) = -\mu(z, 1) \ \forall z \geq z^*(\epsilon)$. Since an interval of the highest-ability tradables producers, accompanied by the appropriate mass of non-tradables producers, moves from city 1 to city 2, $Z'_2 > Z'_1$ and $L'_2 > L'_1$ with $L'_2 = L_1 + \epsilon$ and $L'_1 = L_1 - \epsilon$. Denote $\hat{z} = \sup\{z : \mu(z, 1) > 0\}$. The equilibrium is stable with respect to this perturbation only if

$$\tilde{z}(\hat{z}, Z'_2) - \tilde{z}(\hat{z}, Z'_1) \leq \frac{\theta}{1 - \bar{n}} ((L_1 + \epsilon)^\gamma - (L_1 - \epsilon)^\gamma)$$

This inequality is violated whenever A and $\hat{z}$ are sufficiently high relative to γ , since $\tilde{z}(\hat{z}, Z'_2) - \tilde{z}(\hat{z}, Z'_1)$ is unboundedly increasing in both A and $\hat{z}$. $\square$

Proposition 3 implies that a system of equally sized cities would be a stable equilibrium only when the gains from idea exchange are small relative to congestion costs. Empirically, this does not seem the relevant case. Theoretically, if z has unbounded support, then there always exists a perturbation in which $\hat{z}$ is arbitrarily large, so the left side of the inequality is arbitrarily large, and systems of equal-sized cities are not stable.

Proposition 4 (Stability of two heterogeneous cities). *If the agglomeration, congestion, and feasibility conditions defined in A.4 hold, there exists a locally stable equilibrium with two heterogeneous cities.*

Proof. Appendix section A.4 shows that these three conditions are sufficient for the existence of an equilibrium with two cities in which $L_1 < L_2$ and $\Omega(L_1)$ crosses zero from above. Amongst tradables producers in city 1, those with the most to gain by moving to city 2 are those of the highest ability. Amongst tradables producers in city 2, those with the most to gain by moving to city 1 are those of the lowest ability. It is therefore sufficient to consider perturbations that are changes in z_b and consummate changes in $z_{b,n}$ as defined in appendix section A.4. Since $\Omega(L_1)$ crosses zero from above, this equilibrium is stable. $\square$

A.6 Heterogeneous cities' characteristics

Proof of Proposition 1.

- Equation (5) says that $L_c > L_{c'} \iff p_{h,c} > p_{h,c'}$.
- Equation (11) says that $p_{h,c} > p_{h,c'} \iff p_{n,c} > p_{n,c'}$.
- If $p_{h,c} > p_{h,c'}$ and $p_{n,c} > p_{n,c'}$, then $Z_c > Z_{c'}$. Suppose not. Then $\tilde{z}(z, Z_c) - \bar{n}p_{n,c} - p_{h,c} < \tilde{z}(z, Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'} \forall z > z_m$ and $\mu(z, c) = 0 \forall z > z_m$. Then $L_c = 0$ by equations (8) and (9), contrary to the premise that $L_c > L_{c'}$.
- If $z > z' > z_m$ and $L_c > L_{c'}$, then $\mu(z, c')\mu(z', c) = 0$. Suppose not. If $\mu(z, c') > 0$ and $\mu(z', c) > 0$, then $\tilde{z}(z, Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'} \geq \tilde{z}(z, Z_c) - \bar{n}p_{n,c} - p_{h,c}$ and $\tilde{z}(z', Z_c) - \bar{n}p_{n,c} - p_{h,c} \geq \tilde{z}(z', Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'}$. Therefore, $\tilde{z}(z, Z_{c'}) + \tilde{z}(z', Z_c) \geq \tilde{z}(z, Z_c) + \tilde{z}(z', Z_{c'})$. By strict supermodularity of $\tilde{z}(z, Z_c)$, this is false. Therefore $\mu(z, c')\mu(z', c) = 0$.

$\square$

A.7 Skill premia in a two-city equilibrium

Denote $\tilde{z}_c(z) \equiv \tilde{z}(z, Z_c)$.

A.7.1 Pareto distribution

In an equilibrium with $L_2 > L_1$ in which z is distributed Pareto, $\mu(z) = \frac{kb^k}{z^{k+1}}$ for $z \geq b$ with $k > 1$, the skill premium in the larger city is higher when

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

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

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

$$\frac{\frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mu(z)dz}{\int_{z_b}^{\hat{z}} \mu(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}} < \frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)dz}{\int_{z_b}^{\infty} \mu(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}}$$

2. Define a change of variables by $f(z) = (z_b^{-k} + z^{-k} - z_m^{-k})^{\frac{-1}{k}}$ and $\hat{z} = (2z_b^{-k} - z_m^{-k})^{\frac{-1}{k}}$ such that $\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mu(z)dz = \int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)dz$. By construction $\mu(z) = \mu(f(z))f'(z)$.

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

4. $\frac{\tilde{z}_2(f(z))}{\tilde{z}_1(z)}$ is increasing in z , so $\tilde{z}_2(f(z)) > \frac{p_{n,2}}{p_{n,1}} \tilde{z}_1(z) \forall z \in (z_m, z_b)$.

5. Multiplying by $\mu(z)$ and integrating yields $\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)dz > \frac{p_{n,2}}{p_{n,1}} \int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz$.

$$\text{Thus } \frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)dz}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz} > \frac{p_{n,2}}{p_{n,1}} \text{ and } \frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(z)dz}{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz} = \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)dz}{\int_{z_m}^{z_b} \mu(f(z))f'(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}} > \frac{p_{n,2}}{p_{n,1}}$$

6. Therefore,

$$\frac{\frac{\int_{z_b}^{\infty} \tilde{z}_2(z)\mu(z)dz}{\int_{z_b}^{\infty} \mu(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}} > \frac{\frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z)\mu(z)dz}{\int_{z_b}^{\hat{z}} \mu(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}} = \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_2(f(z))\mu(f(z))f'(z)dz}{\int_{z_m}^{z_b} \mu(f(z))f'(z)dz}}{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z)\mu(z)dz}{\int_{z_m}^{z_b} \mu(z)dz}} > \frac{p_{n,2}}{p_{n,1}}$$

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

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

Those inequalities are true because

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

A.7.2 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

$$\frac{\frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z) \frac{1}{\hat{z}-z} dz}{\int_{z_b}^{\hat{z}} \frac{1}{\hat{z}-z} dz}}{p_{n,2}} > \frac{\frac{\int_{z_m}^{z_b} \tilde{z}_1(z) \frac{1}{\hat{z}-z} dz}{\int_{z_m}^{z_b} \frac{1}{\hat{z}-z} dz}}{p_{n,1}} \iff \frac{z_b - z_m}{\hat{z} - z_b} \frac{\int_{z_b}^{\hat{z}} \tilde{z}_2(z) dz}{\int_{z_m}^{z_b} \tilde{z}_1(z) dz} > \frac{p_{n,2}}{p_{n,1}}$$

A sufficient condition for this to be true in equilibrium is $\hat{z}z_m > z_b^2$. In five steps:

1. By change of variable, $\int_{z_b}^{\hat{z}} \tilde{z}_2(z) dz = \int_{z_m}^{z_b} \tilde{z}_2(f(z)) f'(z) dz$, where $f(z) = z_b + \frac{\hat{z}-z_b}{z_b-z_m}(z - z_m)$. Therefore $\int_{z_m}^{z_b} \tilde{z}_2(f(z)) dz = \frac{1}{f'(z)} \int_{z_b}^{\hat{z}} \tilde{z}_2(z) dz = \frac{z_b-z_m}{\hat{z}-z_b} \int_{z_b}^{\hat{z}} \tilde{z}_2(z) dz$.
2. $\frac{\tilde{z}_2(f(z_m))}{\tilde{z}_1(z_m)} = \frac{\tilde{z}_2(z_b)}{p_{n,1}} > \frac{p_{n,2}}{p_{n,1}}$
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 $\tilde{z}_2(f(z)) > \frac{p_{n,2}}{p_{n,1}} \tilde{z}_1(z) \forall z \in (z_m, z_b)$.

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

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

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

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

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

$\hat{z}z_m > z_b^2$ 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 composition and learning effects. As reported in section 4.3, 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.8 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 back-office 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 $\frac{\int_{z_m}^{z_{b,1}} \mu(z) dz}{\int_{z_m}^{\infty} \mu(z) dz}$ 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, $\frac{L \int_{z_m}^{\infty} \mu(z, c) dz}{L_c}$. The gross migratory flows of unskilled workers and tradables producers of ability $z_m \leq z \leq z_{b,1}$ solve

$$\begin{aligned} \min_{\{x_{c,c'}\}} \sum_{c \in C_u} \sum_{c' \in C_s} x_{c',c} d(c', c) \text{ subject to } \sum_{c' \in C_s} x_{c',c} &= \frac{1}{N_u} \sum_{c' \in C_s} L_{c'} \delta \lambda \quad \forall c \\ \min_{\{y_{c,c'}\}} \sum_{c \in C_u} \sum_{c' \in C_s} y_{c',c} d(c', c) \text{ subject to } \sum_{c' \in C_s} y_{c',c} &= \frac{1}{N_u} \sum_{c' \in C_s} L_{c'} \delta \lambda \int_{z_m}^{z_{b,1}} \mu(z) dz \quad \forall c \end{aligned}$$

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

The average distance migrated by unskilled individuals to city c is

$$\sum_{c' \in C_s} \frac{N_u x_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda} d(c', c)$$

The average distance migrated by unskilled individuals is

$$\sum_{c \in C_u} \frac{1}{N_u} \sum_{c' \in C_s} \frac{N_u x_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda} d(c', c)$$

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

$$\sum_{c' \in C_s} \frac{N_u y_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda \int_{z_m}^{z_{b,1}} \mu(z) dz} d(c', c)$$

The average distance migrated by skilled individuals is

$$\frac{\int_{z_m}^{z_{b,1}} \mu(z, c) dz}{\int_{z_m}^{\infty} \mu(z) dz} \sum_{c \in C_u} \frac{1}{N_u} \sum_{c' \in C_s} \frac{N_u y_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda \int_{z_m}^{z_{b,1}} \mu(z) dz} d(c', c) + \frac{\int_{z_{b,1}}^{\infty} \mu(z, c) dz}{\int_{z_m}^{\infty} \mu(z) dz} \sum_c \sum_{c'} w_c w_{c'} d(c', c)$$

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^* ,

$$\begin{aligned} \frac{\int_{z_m}^{z_{b,1}} \mu(z, c) dz}{\int_{z_m}^{\infty} \mu(z) dz} \sum_{c \in C_u} \sum_{c' \in C_s} \frac{y_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda \int_{z_m}^{z_{b,1}} \mu(z) dz} d(c', c) + \frac{\int_{z_{b,1}}^{\infty} \mu(z, c) dz}{\int_{z_m}^{\infty} \mu(z) dz} \sum_c \sum_{c'} w_c w_{c'} d(c', c) \geq \\ \sum_{c \in C_u} \sum_{c' \in C_s} \frac{x_{c',c}^*}{\sum_{c'' \in C_s} L_{c''} \delta \lambda l} d(c', c) \end{aligned}$$

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 $A, \bar{n}, \theta, \gamma$, 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 back-office task or outsourcing ($l = 0$), and we do not address life-cycle migration.

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

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 observations 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 week. 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.

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

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.

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 MSAs. 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 2 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. We identify residence changes over the five-year span using the “migrate5d” variable. We identify metropolitan changes by comparing the current and prior MSAs for individuals who lived in a MSA identified by the “migplac5” and “migpuma” variables. 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, dummies for white, Hispanic, and black demographics, and the college dummy interacted with the male and demographic dummies. 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 2.

Table 3: Skill premia and metropolitan populations

	1990 PMSA	1990 CMSA	2000 PMSA	2000 CMSA	2005-7 MSA
Skill premia	0.015** (0.0038)	0.014** (0.0039)	0.033** (0.0038)	0.029** (0.0036)	0.040** (0.0038)
Composition-adjusted skill premia	0.013** (0.0030)	0.013** (0.0031)	0.029** (0.0032)	0.025** (0.0030)	0.028** (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.

In Table 2, we used average gross monthly rent as our housing price measure. In Table 4, we use a quality-adjusted annual rent from Chen and Rosenthal (2008) that includes both owner-occupied housing and rental properties. The number of observations differs from that in Table 2 because Chen and Rosenthal do not report quality-adjusted rent values for every PMSA in 2000, but the results are very similar.

Table 4: Skill premia and metropolitan characteristics, 2000

<i>Skill premia</i>				
log population	0.031** (0.0040)	0.030** (0.0053)	0.033** (0.0046)	0.031** (0.0054)
log quality-adjusted rent		0.0046 (0.028)		0.026 (0.027)
log college ratio			-0.021 (0.021)	-0.033 (0.018)
R ²	0.139	0.139	0.145	0.150
<i>Composition-adjusted skill premia</i>				
log population	0.027** (0.0034)	0.030** (0.0047)	0.029** (0.0040)	0.030** (0.0048)
log quality-adjusted rent		-0.019 (0.026)		-0.015 (0.025)
log college ratio			-0.014 (0.018)	-0.0069 (0.015)
R ²	0.145	0.150	0.149	0.151
Observations	297	297	297	297

Robust standard errors in parentheses

** p<0.01, * p<0.05