# LEARNING IN CITIES

Edward L. Glaeser

Working Paper **6271**

NBER WORKING PAPER SERIES

# LEARNING IN CITIES

Edward L. Glaeser

Working Paper 6271
http://www.nber.org/papers/w6271

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1997

Helpful comments were provided by José Scheinkman. An earlier version of this model appeared in joint work with David Maré who also greatly enhanced my understanding of this topic. Financial assistance was provided by the Foundation and the National Science Foundation through a grant to the National Bureau of Economic Research. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Learning in Cities
Edward L. Glaeser
NBER Working Paper No. 6271
November 1997
JEL Nos. J24, O15, R10

## ABSTRACT

Alfred Marshall argues that industrial agglomerations exist in part because individuals can learn skills from each other when they live and work in close proximity to one another. An increasing amount of evidence suggests that the informational role of cities is a primary reason for their continued existence. This paper formalizes Marshall's theory in a model where individuals acquire skills by interacting with one another, and dense urban areas increase the speed of interactions. The model predicts that cities will have a higher mean and higher variance of skills. Cities will attract young people who are not too risk averse and who benefit most from learning (e.g. more patient people). Older, more skilled workers will stay in cities only if they can internalize some of the benefits that their presence creates for young people. The level of urbanization will rise when the demand for skills rises, when the ability to learn by imitation rises, or when the level of health in the economy rises. Empirical evidence on urban wages supports the learning view of cities and a variety of other implications of the theory are corroborated empirically.

Edward L. Glaeser
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@kuznets.fas.harvard.edu

"Great are the advantages which people following the same skilled trade get from near neighborhood to one another. The mysteries of the trade become no mystery: but are as it were, in the air....."

Alfred Marshall, *Principles of Economics*, 1890

## I. Introduction

Alfred Marshall (1890) argues that since intellectual flows between individuals depreciate over space, dense concentrations facilitate the flow of ideas. Jacobs (1968) more broadly claims that cities are intellectual furnaces where new ideas are formed. Lucas (1988) trumpets Jacobs' world view and suggests that the human capital spillovers found in cities are intrinsic to the creation of new ideas which underpin economic growth. A variety of recent papers have confirmed that there is some basic truth to these claims.

Jaffe, Trajtenberg and Henderson (1991) use patent data to document the extent to which distance limits the flow of ideas. Feldman and Audretsch (1997) show that intellectual innovations are strongly concentrated in urban areas. Rauch (1993) documents that workers' wages (and presumably productivity) are higher in those cities where there are more skilled workers, where, presumably, human capital spillovers are greater. Glaeser et al. (1995) shows that high human capital cities have grown in both population and income per capita faster in the post-war period, possibly showing an increased demand for human capital spillovers.[1] The recent example of Silicon Valley has further seems to imply that spatial agglomeration helps to create cutting edge technology (Saxenian, 1993).

As impressive as the role of cities in generating new innovations may be, the primary informational role of cities may not be in creating cutting edge technologies, but rather in creating learning opportunities for everyday people. Dense urban agglomerations provide a faster rate of contact between individuals, and each new contact provides an opportunity for learning. Alfred Marshall's quotation which heads this paper pertains to young apprentices learning

[1]Glaeser et al. (1992) and Henderson et al. (1995) disagree about which urban milieu is most suitable for idea creation (diversity or specialization), but both argue locational density spurs the flow of ideas.

commonly known mechanical skills in industrial centers, not just to cutting edge innovations.

Glaeser and Maré (1994, discussed at length later in the paper) present evidence on the nature of the substantial urban wage premium, where the urban wage premium is defined as the earnings gap between workers in metropolitan areas surrounding large cities and workers in non-metropolitan areas or in metropolitan areas without a large city. We find that rural-urban migrants do not receive the urban wage premium immediately upon arrival in a city. Instead, over time rural-urban migrants experience unusually fast wage growth. Urban-rural migrants do not experience wage declines when they leave the city. Furthermore, the gap between urban and non-urban workers is larger among older workers, who have presumably lived in the city longer. These facts are strongly compatible with a model where learning is faster in cities, and workers then take the skills that they learn in dense urban areas wherever they go. The fact are less compatible with urban productivity models that explain agglomeration through lower transportation costs which would predict an immediate positive wage impact for rural-urban migrants and an immediate negative impact for urban-rural migrants.

The model in this paper focuses on faster human capital accumulation in cities as a result of learning through imitation. The number of contacts per period rises with density and so does the speed of learning. Of course, cities may provide an opportunity for learning beyond the ability to copy the more skilled. While rural areas may contain only a narrow range of experiences, urban areas may contain a plethora of experiences which build skills. Cities may facilitate the division of labor which enables individuals to specialize in a particular range of skills and therefore become more productive more quickly (as in Becker and Murphy, 1992). Finally, the scale economies afforded by urban areas may allow better schools to be built in big cities and therefore facilitate formal education.[2]

[2]The connection of formal schooling and cities is complex and differs substantially across the globe. Anglo-Saxon countries have tended to locate major universities outside of central cities (e.g Oxford, Cambridge, Harvard, and Princeton). European university and their Latin American descendants have been much more likely to be located in major urban areas. These differences are presumably due to political factors.

Throughout this paper, individuals become skilled through random contact with more skilled neighbors. The probability of learning is a function of the fraction of skilled individuals in the community and the density of the community--individuals living in denser areas are assumed to have more contacts per period. This model yields a distribution of skills that is a function of city size and suggests that there will be a much greater variance of utility levels and a higher average level of skills in big cities than in the countryside. The model also predicts that the size of large cities will rise when skills become more important, when it is easier to learn through imitation, or when life expectancy improves.

An unappealing technical assumption in the first model is that individuals choose locations initially and then do not move over their lives. This feature makes it impossible for us to understand the implications of the theory for the demographic composition of cities. The second model relaxes this assumption and also allows industrial specialization of cities but simplifies other aspects. The model predicts that younger, skilled individuals tend to be particularly represented in cities, because they value the learning that cities provide. This prediction appears to be true empirically.

Industrial diversity increases the productivity of older workers because there is less competition between these more diverse producers, but it also decreases the level of learning since I assume that younger workers can only learn from the skilled members of their own industry. This model predicts that cities that are full of young people or growing will tend to be specialized (supporting the result of Henderson et al. 1995, and contradicting the results of Glaeser et al., 1992). Diversification will tend to be lower when the returns to skills are higher or when imitation is more feasible.

The model offers some predictions about the future of cities. Like Gaspar and Glaeser (1997), it predicts that the future of the city may be much more secure than many believe. The increase in the importance of skills (see e.g. Murphy and Welch, 1992) suggests that if cities facilitate skill acquisition they will continue to be important far into the future. However, since the skill acquisition discussed in this paper works best in concentrated environments, it may be that the highly concentrated Silicon Valley-type environments packed with skilled workers in a particular industry may be the model for the future and may replace more diverse metropolises. If Jacobs (1968) and Glaeser et al. (1992) are right and diversity generates the most important ideas, then even though these concentrated cities may be best for skill accumulation, they may be less efficient at engendering long-term intellectual fertility.

## II. Learning by Imitation in Cities

In this model, skill accumulation occurs when individuals imitate and learn from their more skilled neighbors. While there are many other sources of education, including formal schools or explicit training programs, this model argues that dense cities are special and offer a unique environment for informal learning from social and business contacts. The model is written to capture skill acquisition after formal training, and it is meant include to the acquisition of any valuable ideas or knowledge or contacts as part of this learning process (i.e. an idea about where to open a new restaurant would be a form of skill acquisition).

*The Learning Technology*

All agents begin life with skill level zero. Their skill level is denoted S and must be an integer. Individuals increase their skill levels when they meet more skilled persons whom they imitate. In each period individuals produce goods, and the value of this production is a function w(S) of their skill level. In a meeting between two unequally skilled individual, with probability c, the less skilled individual learns from the more skilled individual and increases his skill level by one. Each meeting occurs with a random member of the city population.

The number of meetings per period is D, which is an increasing function of the number of individuals working in the central business district (CBD). Many different microstructures can justify this assumption. For example, if individuals use space within the CBD when they engage in production, and they make contacts when other people share their space, the number of random contacts will be higher when more people are using the same space. Of course, there will also be congestion cost which I will explicitly model later. Individuals become skilled immediately after learning from a more skilled individual (i.e. if there are three meetings per period and an individual gains a skill level after the first meeting, then this individual will be more skilled for 2/3 of the period). If $\Pi(j,t)$

denotes the share of the population with skill level j or higher at time t, then the probability that a person with skill level j-1 will increase his skill level in any period is $1-(1-c\Pi(j,t))^{D}$. The rate of skill accumulation is a function of the rate of imitation (c), the rate of new meetings (D-- which is itself a function of city density) and the skill distribution of the city.

Only a fraction δ of the population (chosen randomly among all skill and age levels) survives each period, so every 1/Dth period, $1-\delta^{1/D}$ agents die. In a stationary equilibrium, new agents (with zero skills) are born so that the overall population of the city is constant. If π(J, t) denotes the share of the population at exactly skill level J at time t, then the evolution of Π(J, t) is governed by the following equation:

(1) $$\Pi\left(J,t+\frac{1}{D}\right)=\delta^{\frac{1}{D}}\Pi(J,t)+\delta^{\frac{1}{D}}c\Pi(J,t)\pi(J-1,t).$$

The quantity of individuals with skills of J or higher at time t+1/D is a combination of the individuals with skills of J or higher at time t who have survived, plus the individuals who had skill level J-1 exactly at time t who moved to skill level J between time t and time t+1/D.[3] The movement into the skill level J and the higher group equals π(J-1,t) (the proportion of the population that might move up into the group) times Π(J, t) (the proportion of the encounters that are with agents who are skilled enough to be imitated) times c (the proportion of the encounters that lead to imitation).

In the stationary distributions where Π(J, t) and π(J, t) are independent of t (we denote these stationary values as Π(J) and π(J) respectively), equation (1) reduces to $\delta^{\frac{-1}{D}}=1+c\pi(J-1)$ or $\pi(J-1)=\theta/c$ where θ is defined as $\left(\delta^{-1/D}-1\right)$ and is required to lies between zero and one, for all J except for the highest level of J, which we denote as $\bar{J}$.[4] The distribution of skills across the community will be uniform. If the share of the population in any skill level j' is greater or less than $\theta/c$, then the share or the population at skill level j'+1 will either grow or shrink

[3]Looking at 1/D time intervals simplifies matters because it allows the exclusive consideration of skill jumps of at most one level.

[4]As $1>\delta$, it follows that $\delta^{-1/D}>1$. If $\delta^{-1/D}>2$, then all the people have skill level zero. This case occurs only when the density is sufficiently low or the death rate is sufficiently high.

in the next period and the equilibrium will not be stationary.[5] The cumulative distribution of individuals who have skill level J-1 or lower is J times $\theta / c$, which implies that Π(J) must equal one minus J (the number of lower skill levels) times $\theta / c$, or $\Pi(J) = 1 - J\theta / c$.

There are several different stationary distributions, depending on the highest level of skill in the community. Any Z≥0 can be a stationary distribution since no agents of level Z can increase their skill level without higher level agents to imitate. This multiplicity stresses that communities can get caught in low skill traps where no one becomes skilled because there is no one who is skilled to imitate. A simple stability rule can eliminate all but one stationary equilibrium--the only stable equilibrium is the one in which $\bar{J}$ is defined so that $\theta / c \geq 1 - \bar{J}\theta / c \geq 0$ or $c / \theta \geq \bar{J} \geq c / \theta - 1$.[6]

The number of skill levels in a city (and the average skill level in that city and the variance of skills in that city) increases with the probability of imitations (c), the speed of interactions (D) and decreases with the probability of death (δ). As a result, skills will be higher and more varying in high density cities. Industries that depend on learning through contacts (i.e. with high levels of c) should sort into cities to take advantage of the more interactive environment. Figure 1 illustrates these comparative statics by comparing the spread out, high mean distribution of a large city (parameters chosen are c=1, δ=.8 and D=2) with the more compressed, low mean distribution of a small city (parameters chosen are c=1, δ=.8 and D=1).

---

[5]In order to achieve a Pareto distribution which matched the data on wage distributions more accurately, we could make the probability of imitation decline with the skill level of the imitator (as suggested by José Scheinkman). Alternatively a Pareto distribution can be generated if higher skilled agents have more interactions per period than lower skilled agents.

[6]In any candidate stationary equilibrium where $\tilde{\pi} > \theta / c$ individuals have the highest level (denoted J'), we can perturb this equilibrium and transfer ε agents to skill level (J'+1). In 1/D periods there will be $\delta^{\frac{1}{D}}(1 + c\tilde{\pi})\varepsilon > \varepsilon$ individuals with skill level J'+1, so the perturbation generates a greater perturbation and eventually skill level J'+1 will remain as the highest skill level. As such, unless there are fewer than $\theta / c$ individuals with the highest skill level, then the stationary distribution will not be stable.

*Lifetime Earnings*

Individuals choose workplace location at the beginning of their lifetime, and this location is fixed for their lives. The locational decisions will maximize expected lifetime earnings minus lifetime commuting cost. The value of production each period is assumed to be separable into a city-specific component $w_i$ which is independent of skills and a skill-related component $a(J)=\alpha J$, which is independent of location a $\alpha$ reflects the returns to skill.

For any worker at skill level $J < \bar{J}$ at time t, the expected skill level earnings conditional upon surviving at the period t+1/D equals current skill level plus $c\Pi(J+1)$, or:

(2) $$E\left(J\left(t+\frac{1}{D}\right)|J(t)\right) = J(t) + c\left(1 - \frac{J(t)+1}{c}\theta\right) = (1-\theta)J(t) + (c-\theta)$$

where $E\left(J\left(t+\frac{1}{D}\right)|J(t)\right)$ denotes the expectation of skill as of time t+1/D for someone who has skill level J at time t. While this equation holds for $J < \bar{J}$ absolutely, it will not hold for $J = \bar{J}$ unless $\theta / c = \pi(\bar{J})$ exactly, which I assume.[7] Minor calculations and assuming that agents are paid at the beginning of each period implies that expected discounted lifetime earnings are:[8]

(3) $$Earnings = \sum_{t=0}^{\infty} \beta^t \delta^t \left( w_i + \frac{\alpha(c-\theta)\left(1-(1-\theta)^{Dt}\right)}{\theta} \right) = \frac{w_i + \dfrac{\alpha\beta\delta(c-\theta)\left(1-(1-\theta)^D\right)}{\theta\left(1-\beta\delta(1-\theta)^D\right)}}{(1-\beta\delta)}.$$

Lifetime earnings will be higher when c is higher or when $\alpha$ is higher or when D is higher.

---

7 Since the failure of this equation to hold exactly has little economic content and simply reflects an integer problem, I will proceed assuming $\theta / c = \pi(\bar{J})$ which implies that (3) holds for all values of J.

8 Iterating on equation (3) reveals that:
$E\left(J\left(t+\frac{q}{D}\right)|J(t)\right) = (1-\theta)^q J(t) + (c-\theta)\left(1-(1-\theta)^q\right)/\theta$ which in turn implies that $E(J(t)|J(0)) = (c-\theta)\left(1-(1-\theta)^{Dt}\right)/\theta$.

Equation (3) implies that lifetime earnings always rise with city size. However, I now introduce congestion to put a natural limit on city size. Individuals must commute to the CBD from their homes and commuting costs are k per period per distance from the city. If city size is N, housing lots are of fixed area A, the value of land at the fringe is $\underline{R}$ and the city extends along a single dimension (the classic long-narrow city), then the edge of the city will be AN/2 distant from the center city. The rent function which makes consumers indifferent between different housing locations will be $\underline{R} + kAN/2 - k^*W$, where W denotes the distance from the city center. Total living costs in this city (housing plus commuting) are, therefore, $\underline{R} + kAN/2$ or $(\underline{R} + kAN/2)/(1-\beta\delta)$ in expected value over the life cycle.

Individuals will make a permanent decision about location at the beginning of their lifetimes in order to maximize the expected sum of lifetime earnings minus living costs. Figure 2 illustrates a possible situation generated from a simulation of the model where total welfare in the city is initially rising with city population as density increases the rate of learning. However, eventually diminishing returns to learning set in, and the negative effect of population on congestion outweighs the positive effect on learning. The second line in Figure 2 illustrates a hypothetical reservation utility curve (lifetime utility in the hinterland) and the city will only be in equilibrium at the points at which utility in the city equals reservation utility. The figure suggests that there are three equilibrium outcomes, one in which there are no people at all in the area, and two where the city actually has some population. The middle equilibrium is unstable because a perturbation that increases its size will also increase the incentives to live in the city so that the perturbation will be reinforced by other migration.

To generate comparative statics on the level of urbanization, we will assume that individuals either choose to live in a large city or in an edge city/hinterland. In this edge city there are no commuting costs, wages are fixed at $\underline{w}$ and no learning from neighbors. The equilibrium condition for city size will be:

$$(4) \qquad (w_i - \underline{w}) + \alpha\beta\delta \frac{(c-\theta)\left(1-(1-\theta)^D\right)}{\theta\left(1-\beta\delta(1-\theta)^D\right)} = \frac{kAN}{2}$$

which implies comparative statics for city size or the degree of urbanization as long as the city is stable in the sense that increases to city population lower the relative returns to living in the city. Unsurprisingly, city size rises with $w_i$ and declines with $\underline{w}$, k or A. As the importance of accumulating human capital ($\alpha$) rises, or as the ability to absorb information from one's neighbors (c) rises, the value of being in a large city also rises. Perhaps, the dramatic connection of urbanization and development occurs in part because development is generally associated with increases in the returns to skill.

City size will also increase as the discount factor rises because living in a city is a form of human capital investment that offer delayed benefits (more skills) for greater immediate costs. An increase in the survival rate will also increase city size under most circumstances. Health increases the value of urban living not only because it acts to increase the value of future periods (in the same as greater patience) but also because increased health means that there is a greater number of more skilled individuals around from whom to learn. Since the value of the city lies in facilitating learning from its older, skilled inhabitants, when there are more of these individuals the value of the city grows.

## III. Industrial Diversity and Specialization

This model simplifies some aspects of the previous model to allow me to investigate intra-lifetime migration and the role of industrial diversity and specialization. Individuals now live for only two periods in an overlapping generations model, and there are only two skill levels. Individuals no longer need to stay in the location that they enter early in life. The discount rate, probability of death and urban congestion cost notation remains the same.

The probability of becoming skilled is dependent upon meeting skilled individuals, but now the meeting must be in one's own industry. If we let S denote the share of skilled individuals in the city who are assumed to be equally spread across "I" industries, then the probability that a given encounter is with a person who is both skilled and in one's own industry is S/I. Again, the probability of learning is c times this quantity, and the number of meetings in the first period is D. Thus, the probability of becoming skilled equals $1-(1-cS/I)^D$. I denote the benefit from becoming skilled as V. The expected total second period earnings benefits for the young from living in a city is $\beta\delta V\left(1-(1-cS/I)^D\right)$. The benefits from a successful meeting are shared between the skilled and the unskilled worker who met; the young learner must pay 1-σ percent of his benefits to the older person who taught him these skills. This internalization presumably comes primarily through firms which enable unskilled workers to compensate skilled workers for their input by accepting lower wages. Thus, the total expected benefits received by the young from a particular city are σ times $\sigma\beta V\left(1-(1-cS/I)^{D(N)}\right)$.

The marginal product of the unskilled is assumed to equal $\underline{W}$ in the hinterland, and the returns to the skilled are $W_S$ in the hinterland (since some skilled and unskilled older workers always remain in the hinterland, the returns to becoming skilled are determined by the returns to skill in the hinterland and the value of becoming skilled, V, must be $W_S - \underline{W}$). The productivity of the unskilled (old and young) is assumed to be independent of industry because the unskilled are using a linear technology to produce traded goods where prices are set on the world market. However, the marginal product of the skilled in the city is diminishing with the number of people in the particular industry in the city and increasing in the overall size of the city (holding the size of the industry constant). Thus, wages for the skilled equal W(S/I,N).

The equilibrium size of a large city is found by calculating this size so that utility levels are equal between the city and the hinterland. In this case, both old and young workers must be indifferent between the two locations. Therefore, for the young:

(5) $$w_i - \underline{w} + \sigma\beta\delta V\left[1-(1-cS/I)^D\right] = kAN/2,$$

and for the old

(6) $$w\left(\frac{S}{I},N\right)-w_S+(1-\sigma)\beta\delta V\frac{(1-S)(1-(1-cS/I)^D)}{S}=\frac{kAN}{2}$$

for the older skilled workers. If unskilled young workers are indifferent between the big city and the hinterland, then all of the unskilled old workers will prefer the hinterland. The voluntary selection of the old unskilled to the suburbs proves to be quite valuable socially since added unskilled workers living in the big city creates a negative externality by congesting the learning process. High moving costs might lead to some unskilled older workers remaining in the city in practice and such workers could significantly damage the value of the city as a learning center.

These two indifference conditions provide a system of equations with solutions for S and N. To satisfy intuitive stability criteria, we assume that both returns to the skilled and unskilled in the city are declining in city size (near the equilibrium), returns to the skilled are declining in the share of the population that is skilled and returns to the unskilled are declining with the share of the population that is unskilled. Standard differentiation then provides many of the same comparative statics as the previous section. City size is rising with $\beta$, $\delta$, c, and V; city size is declining with $\underline{w}$, k and A. An increase in industrial diversification can either increase of decrease city size depending on whether the negative effect of diversification on skill accumulation is outweighed by the positive effect of diversification on the earnings of the skilled, which then attracts more skilled workers and indirectly more young workers.

As $\sigma$ rises, the share of the population that is skilled falls because the skilled receive fewer of the returns from learning. This effect can imply that the young unskilled workers would prefer to receive a lower share of their gains from learning, because accepting a lower share could attract more skilled workers, as long as the young could commit collectively to this taking this lower share.

As long as $\sigma$ is close enough to one, then increases in $\beta$, $\delta$, V or c will increase the share of young unskilled workers in the city. The more that the young value the future, either because their survival level is greater or because their rate of time discount is higher, the greater we should expect young people to live in the city, and as the returns to skill increase, we should also expect an added inflow of the young into the city. We should also expect to see young individuals who are particularly healthy, patient or able to learn or who have a greater ability to utilize new skills to select into the city (a subset of young people with levels of these variables that are greater than the marginal young person would all choose to live in the city). As a result, we should expect to see cities not only filled with the young, but also particularly filled with the more skilled young.

*A Developer's City*

A developer who is choosing both the size of the city and the skill composition of the city to maximize total social surplus subject to the constraint that workers receive their reservation utilities would, of course, design a different city since externalities abound in this model. In this case, I assume that the developer's must deliver utility equal to that which individuals would receive the hinterland and that $w_i = \underline{w}$, so that the developer has access to the same technology as is available elsewhere. Developer's are assumed to develop a city in order to maximize the total earnings from that city which would equal the expected benefits of all workers from that city minus their opportunity costs, which implies that a developer will maximize:

$$(7) \qquad N\left[S\left(w\left(\frac{S}{I},N\right)-w_S\right)+(1-S)\beta\delta V(1-(1-cS/I)^D)-\frac{kAN}{2}\right]$$

which yields:

$$(8) \quad w\left(\frac{S}{I},N\right)-w_S+\frac{S}{I}w_1\left(\frac{S}{I},N\right)=\beta\delta V(1-(1-cS/I)^D)-(1-S)\frac{\beta\delta VDc}{I}(1-cS/I)^{D-1},$$

as the first order condition for the choice of skill level in this city. The left hand side shows the marginal return from including an additional skilled person and the right hand side shows the advantage of added less skilled persons. This equation can be compared with:

$$(9) \qquad w\left(\frac{S}{I},N\right)-w_S=\beta\delta V(1-(1-cS/I)^D)\left(\sigma+\frac{(1-\sigma)(1-S)}{S}\right)$$

which is the comparable equality that holds for the open city described by equations (5) and (6). However, it is unclear whether the developer will create a more or less skilled city than the free market, even when $\sigma$=1. The developer internalizes two externalities that work against each other. Comparing the left hand sides of (8) and (9) illustrates that the developer internalizes the fact that increasing the number of skilled persons reduces the productivity of each skilled person. Comparing the right hand sides of (8) and (9) illustrates that the developer internalizes the negative effect of learning created by expanding the number of unskilled persons in the city. Since the developer will internalize both an added cost of increasing the skill level of the city and an added benefit of increasing the skill level, it is unclear whether the developer will create a more or less skilled city. Of course, if we assume that the skilled workers do not congest each other very much (which seems possible), then the developer will create a more skilled city.

Taking the derivative of (7) with respect to N (and recalling that D=D(N)) yields:

$$\begin{aligned} (10)\qquad & S\left(w\left(\frac{S}{I},N\right)-w_S+Nw_2\left(\frac{S}{I},N\right)\right)+ \\ & (1-S)\beta\delta V\left(1-\left(1-\frac{cS}{I}\right)^D\left(1+ND'(N)\ln\left(1-\frac{cS}{I}\right)\right)\right)=kAN \end{aligned}$$

The corresponding free market equation is:

$$(11)\qquad S\left(w\left(\frac{S}{I},N\right)-w_S\right)+(1-S)\beta\delta V\left[1-(1-cS/I)^D\right]=kAN/2$$

The developer internalizes both positive and negative effects of city size which may lead the city to be either larger or smaller than in the free market equilibrium. Comparing the left hand sides of (10) and (11) illustrates that the developer incorporates the positive effect of added population on the returns to the skilled and the positive effect of density on learning. Comparing the right hand sides of the two equations shows that the developer also internalizes the negative effect of added population on congestion costs. Overall, we can also not tell whether the city will be bigger or smaller under a profit-maximizing developer or under the free market.

*The Determinants of Industrial Structure*

I now return to assuming the free market equilibrium and assume that industrial structure is itself an outcome of economic forces. The simplest means of modeling industrial structure is to assume that industries are formed by the older generation, and this formation acts to maximize the welfare of the average member of the older generation taking city size and skill composition of the city as fixed.

In this case the number of industries will be chosen so that:

(12) $$-w_1\left(\frac{S}{I},N\right)=(1-\sigma)c\beta\delta VD\frac{S}{1-S}\left(1-\frac{cS}{I}\right)^{D-1},$$

assuming that second order conditions also hold. Industrial diversity will fall when the old internalize more of the learning gains of the young, when the young are more patient or have a greater life expectancy or value skills more.

The connection between industrial diversity and city size depends on two different effects of city size. First, city size may either increase or decrease the importance of spreading the skilled across multiple industries to eliminate congestion in their static productivity. Second, city size increases the number of interactions which increases the value of industrial diversity as long as $cS/I$ is sufficiently small. Therefore, as long as it is less valuable to spread individuals into multiple industries in large cities, and as long as $cS/I$ is small, then a larger city will have more industries. In general, greater values of S also lead to a greater level of industrial diversity as well, so more skilled cities will have a greater range of industrial diversity.

Finally, this model where the benefits of industrial diversity display themselves in greater static gains to the skilled, but also act to reduce the learning benefits to the young has a very specific model of the learning process. Alternatively, we could assume that some of the most interesting ideas are transmitted across fields, but these transmissions are harder to internalize. In this case, the optimal city will be much more diverse than the equilibrium city, and we will see that diverse cities are particularly skill laden and will possibly grow faster if they begin away from the equilibrium city size (as in Glaeser et al., 1992).

## IV. Empirical Implications and Discussion

This section discusses some empirical evidence that relates to the implications of the model.

*The Evidence Presented by Glaeser and Maré (1994)*

Much of the motivation for theoretically examining these issues comes out of previous empirical work done with David Maré. This work on wage patterns suggests that cities exist to facilitate learning. First, we find the urban wage premium (which is defined as the premium for living in a metropolitan area surrounding a city of greater than 1 million inhabitants) is quite sizable (34 percent) when compared to non-metropolitan areas and is even large (10 percent) when compared to metropolitan areas not surrounding a large city. These differences remain robust when we control for race, demographics, industry, education and occupation controls. While we do not have good enough urban price data to know for sure, it appears that this wage difference compensates workers for higher prices, but the puzzle still remains: why are firms willing to pay workers more in cities? Neoclassical theory maintains that this willingness indicates a greater marginal productivity of labor in cities.

When we examine workers wage movements over time using the Panel Study of Income Dynamics, we find that workers who move to the city experience no immediate wage gains, despite the predictions of most agglomeration theories that their marginal product of labor should have risen immediately upon entering the urban milieu. Instead, workers who have moved to cities experience faster wage growth over the next five years and eventually appear to be earning wages that are close to those of long term residents. Workers who leave cities experience no visible wage declines, despite the predictions of most agglomeration theories, but just as the model in this paper predicts.

Finally, across numerous data sets the difference between workers in metropolitan areas surrounding big cities and other workers grows over the lifecycle. While there are other theories that could potentially explain this increasing gap, the change would certainly be predicted by the model presented here, where older workers have learned more in cities. The data presented by Glaeser and Maré (1994) has various interpretations, but it does appear to be quite suggestive of a model where the primary urban agglomeration force comes from faster learning in cities.

*The Distribution of Populations across City Sizes*

Two of the most basic predictions of the model are that cities will attract young persons and repel older, less skilled individuals. Table 1 examines this prediction by looking at the distribution of more or less skilled individuals of different age categories across three different groups of U.S. cities. I have grouped the United States population between the ages of 18 and 65 excluding those in school into three urban groups: individuals living in the New York, Chicago and Los Angeles metropolitan areas, individuals living in all other metropolitan areas and individuals living outside of metropolitan areas. Within each urban size category, I examine the distribution of age groups (under 35, 35 to 50 and 50 and up) for the total population and for college graduates.

The larger cities appear to attract more skilled persons, especially relative to non-metropolitan areas. The striking difference in the first row which illustrates that individuals in the three largest metropolitan areas are 10.2 percent more likely to be college graduates. This fact supports the notion that cities attract individuals who care about skills.

There appears to be a sizable gap in the distribution of young persons between the largest metropolitan areas and non-metropolitan regions, especially among college educated persons in the labor force below the age of 35. In particular, the share of the three largest MSAs that is less than 35 is 3.9 percentage points higher than the share of non-metropolitan regions. Among college graduates, the share of the three largest MSA is 8.4 percentage points higher. The prevalence towards youth appears to be quite strong, especially among those younger persons who are likely to be willing to invest in city-related learning.

*The Future of the City*

The models presented in this paper directly predict the determinants of urban concentration. Therefore, they may help us predict the changes in urban structure that are forecast by current changes in the overall economy.

First, the model predicts that the rise in returns to skill that has been reported by a plethora of authors (e.g. Murphy and Katz, 1992) will surely act to increase the demand for cities. If cities act as forges of human capital, then heightened demand for human capital will increase the demand for cities as well. Of course, the model predicts that if face-to-face learning becomes obsolete then cities will as well, but as long as we believe that this fundamental technology will remain in action, then cities should actually be strengthened.

However, the model also emphasizes that urban structure will change in other ways as the demand for skills rises. If there is increased demand for learning from others, then industrial diversity will decline as cities become focused centers of intellectual transmission. Silicon Valley may indeed become the model of the city in the next century.

Finally, greater importance placed on face-to-face learning will also create an increase in the demand for high skill cities. Since contact with less skilled individuals eliminates the advantages of urban density, cities may be designed to minimize contacts between the more skilled and those less skilled who are able to pay the housing costs of a high skilled city. Of course, this will mean that those less skilled who are less willing or able to pay will learn less just as learning becomes more important. If this is so, then the increasing role of information in the economy will also lead to heightened segregation by skill level at the city and neighborhood level.

## V. Conclusion

Cities exist in part to facilitate learning between individuals who come into contact with one another. This view of the urban form generates a variety of testable hypotheses, only a few of which were discussed in this paper. Glaeser and Maré (1994) confirm that cities have the patterns of wage growth that we might expect if cities primarily facilitate learning. The distribution of age and skill groups between urban and non-urban areas also corresponds with this viewpoint. The important informational role of the city means that if we are entering an informational age, then the city will not die but flourish.

## References


Becker, G. and K. Murphy (1992) "The Division of Labor, Coordination Costs and Knowledge," *Quarterly Journal of Economics* CVII: 1137-1160.

Chinitz, B. (1961) "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review* 71: 279-289.

Gaspar, J. and E. Glaeser (1997) "Information Technology and the Future of Cities" *Journal of Urban Economics*, forthcoming.

Glaeser, E. and D. Maré (1994) "Cities and Skills," Hoover Institution Working Paper E-94-11.

Glaeser, E., Kallal, H., Scheinkman, J. and A. Shleifer (1992) "Growth in Cities," *Journal of Political Economy* 100: 1126-1152.

Henderson, Kuncuro and Turner (1995) "Industrial Development in Cities" *Journal of Political Economy* 103 (5): 1067-1090.

Jacobs, J. (1968) *The Economy of Cities,*, New York: Vintage Books.

Jaffe, A., Trajtenberg, M. and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics.*

Lucas, R. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics.*

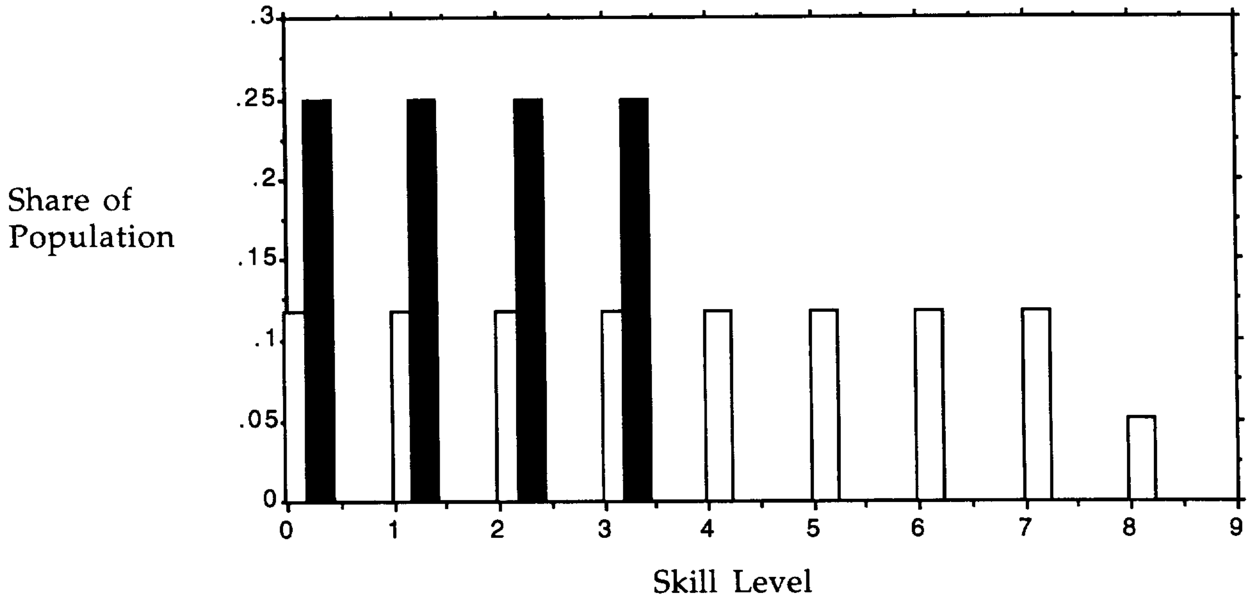
Marshall, A. (1890) *Principles of Economics*, Macmillan.

Murphy, K. and F. Welch (1992) "The Structure of Wages" *Quarterly Journal of Economics* CVII: 285-326.

Rauch, J. (1993) "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics* 34: 3-33.

Saxenian, A. (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Cambridge: Harvard University Press.

## Figure 1: Skill Distributions by City Density

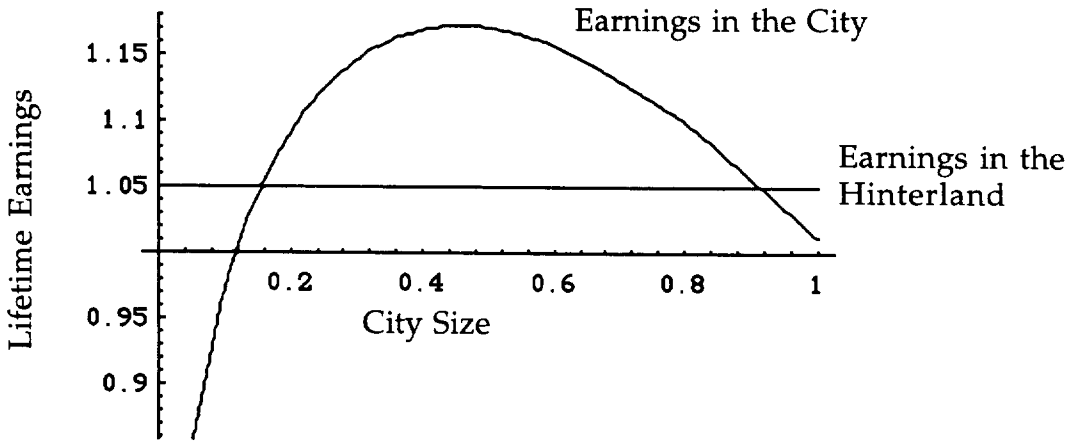



These skill distributions are based on the model described in the text, with parameter values are c=1 and delta=.8.

The gray bars represent the skill distribution of a "Small City" with D=1.

The clear bars represent the skill distribution of a "Big City" with D=2.

## Figure 2: Earnings and City Size




This graph represents result from simulations of the model using the same parameter values as in Figure 1, and also assuming that alpha=1, beta=1, A=1, D=.5+the square root of N and k=2.

**Table 1: The Distribution of Population across Metropolitan Areas**

| | New York, Chicago and Los Angeles | | All Other Metropolitan Areas | | Non-Metropolitan Areas | |
|---|---|---|---|---|---|---|
| | Total Sample | College Graduate (23.6% of area total) | Total Sample | College Graduate (22.5% of area total) | Total Sample | College Graduate (13.4% of area total) |
| Share that is Age 18-34 | 41.0 | 38.7 | 39.7 | 35.2 | 37.1 | 29.3 |
| Share that is Age 35-49 | 34.8 | 41.4 | 35.6 | 44.6 | 35.3 | 47.4 |
| Share that is Age 50-65 | 24.2 | 20.0 | 24.4 | 20.2 | 27.5 | 23.2 |

Note: The source of this information is the 1990 U.S. Census 1% Public Use Micro Sample. Percentages may not add to 100 because of rounding.