Search, Sorting, and Urban Agglomeration

Christopher H. Wheeler, Tulane University

Studies have suggested that urban agglomeration enhances productivity by facilitating the firm-worker matching process. This article develops a model that formalizes this notion and demonstrates that, when firm capital and worker skill are complementary in production, urban agglomeration will tend to generate more efficient, yet segregated matches. As a result, not only will local market size be positively associated with average productivity, it will also generate greater between-skill-group wage inequality and a higher expected return to skill acquisition. Recent data from the counties and metropolitan areas of the United States is consistent with each of these implications.

I. Introduction

For over a century, it has been observed that the concentration of economic activity in urban areas generates sizable returns to producers in the form of greater productivity. To be sure, agents located in densely populated markets are commonly believed to take advantage of positive externalities, such as those associated with knowledge spillovers across firms both within and between industries (e.g., Marshall 1916; Jacobs 1969; Lucas 1988; Glaeser, Kallal, Scheinkman, and Shleifer 1992); the presence of a more extensive division of labor (e.g., Abdel-Rahman 1988; Abdel-Rahman and Fujita 1990; Becker and Murphy 1992); or increasing returns

I would like to thank Buz Brock, John Jones, Dan McMillen, Bill Oakland, and Stefani Smith for their comments and suggestions. All errors are, of course, my own.
owing to firm-level economies of scale (e.g., Mills 1967; Dixit 1973; Krugman 1991).

Yet, while these particular explanations for “agglomeration economies” have received considerable attention, an additional explanation—that cities offer significant returns in the form of lower search costs—has not. Intuitively, the idea is straightforward: by allowing firms and workers to conduct searches with greater ease, spatial concentration permits agents to search more extensively and, thus, generate more productive matches. Indeed, Glaeser (1994, p. 19) observes: “In a one-company town, individuals who are imperfectly matched to that company have nowhere else to go. These workers will stay at the company (or leave the town), being underproductive. In a city, the workers could easily move to a competitor and improve the quality of their match. Thus, we expect that cities will have better matches between workers and employers and that human capital will be utilized more effectively than in rural counterparts.” A similar argument is made by Henderson (1986, p. 48), who suggests that industrial concentration generates “labor market economies where industry size reduces search costs for firms looking for workers with specific training relevant to that industry.”

This lack of attention is somewhat surprising. The idea that cities facilitate search is certainly consistent with a host of casual empirical evidence. Scott (1988), for instance, finds that rates of employment turnover tend to be higher in large, urban labor markets than in smaller ones, suggesting that moving between jobs is less costly in thicker markets. A similar finding is established by Alperovich (1993), who documents a significantly negative relationship between population and the duration of unemployment in a sample of urban areas in Israel.

This article develops a model that formalizes the link between search costs and firm-worker matching and shows that three of the model’s direct implications are supported empirically. In particular, assuming that firm capital and worker skill are complementary in production, the theory indicates that larger local markets, by generating more productive (albeit stratified) matches, will simultaneously exhibit greater output per worker, wage inequality, and expected returns to a worker’s skill.

Empirical evidence from the counties and cities of the United States is consistent with each of these implications. Numerous studies (e.g., Ciccone and Hall 1996; Glaeser and Mare, in press) have shown that measures of average productivity, including wages and output per worker, rise substantially with local market size.

1 A recent survey by Quigley (1998) lists only one paper, Helsley and Strange (1990), as having modeled search costs and matching in cities.

Moreover, recent data from a sample of 286 U.S. metropolitan areas suggests that the wages of highly educated workers rise significantly in relation to those of their less educated counterparts as resident population increases. Estimation of a simple wage regression indicates that, on average, a doubling of metropolitan area population corresponds to a 4% increase in the hourly wages of workers with 16 or more years of schooling. For workers with fewer than 9 years of education, by contrast, the estimated correlation between population and wage earnings is negligible.

In addition, there is a strong, positive association between county and city-level population and education. Two recent papers, Glaeser (1999) and Glaeser and Mare (in press), have found that urban populations are significantly more educated than rural ones. Among U.S. counties, the pattern is quite similar. Using data from 1990, I estimate that a doubling of population corresponds to a 2 percentage point increase in the proportion of the resident population with a bachelor’s degree or higher. Although not conclusive, such results are at least consistent with the notion that highly educated individuals are drawn to urban labor markets where the returns to their human capital are the highest.

As such, the contribution of this article is twofold. First, it provides evidence that urban agglomeration enhances productivity by facilitating labor market search. Second, and more important, it demonstrates that differences in the ability of firms and workers to identify and establish productive matches may help to rationalize differences in a number of additional outcomes—patterns of inequality, expected returns to skill, and, thus, distributions of worker skill or human capital—observed across local geographic markets.

The rest of the article is organized as follows. Section II describes the model in which firms and workers must match themselves into pairs in an environment with costly search. Each firm and worker is characterized by an exogenously determined level of skill/capital, or “quality,” and qualities are assumed to be complementary in production. The matching process itself takes the form of firms searching over and hiring from the pool of available labor. Since the process is costly, firms use cutoff rules, whereby each matches with the first worker it encounters who satisfies its minimum quality requirement.

To capture the notion that large, urban areas facilitate matching, search costs are specified as a decreasing function of the size of the market. As a result, markets of different sizes tend to exhibit differing degrees of sorting, all else equal. In particular, larger markets permit the most productive firms to set higher reservation qualities, which, in turn, induces more stratified firm-worker matches: the most productive workers are hired by the most productive firms, leaving low-end agents to match among themselves.

Because agent qualities are complementary in production, greater strat-
iffication leads to higher overall output and, thus, output per worker. Yet it does so by increasing the payoffs of the most skilled workers relative to those of the least skilled. Not only does this generate greater between-skill-group wage inequality, but it also creates a larger expected payoff associated with moving from the bottom end of the skill distribution to the top. Hence, average productivity, inequality, and the expected return to skill acquisition all increase with local market size.

Although the model does not deal with unemployment and job turnover, a brief discussion of these two issues is given in Section III. Section IV then presents the empirical evidence on productivity, wage inequality, and returns to skill for samples of counties and metropolitan areas in the United States. Section V concludes.

II. Theory

Consider an economy consisting of two types of agents, firms and workers, of which there are \( n \) each. Every agent is characterized by an exogenously determined, strictly positive scalar measure of quality (i.e., skill, productive capital), which is denoted by \( q \) for workers, \( k \) for firms. Each agent is assumed to supply one unit of input into production, hence \( q_i \) represents the effective labor supplied by worker \( i \), \( k_j \) the effective capital supplied by firm \( j \). Let individual workers and firms be indexed such that \( q_1 > q_2 > \cdots > q_n \) and \( k_1 > k_2 > \cdots > k_n \), and the distribution of worker qualities be known by all firms.

Production involves the matching of workers and firms into pairs or “production coalitions,” each of which generates output given by a Cobb-Douglas function of agent inputs. Thus, worker \( i \), when paired with firm \( j \), generates

\[
Y_{ij} = q_i k_j^{1-\alpha}, \quad 0 < \alpha < 1.
\]  

(1)

Firm and worker payoffs are assumed to be given by, respectively, \((1 - \alpha)\) and \(\alpha\) shares of the production generated. Hence, when matched with firm \( j \), worker \( i \) receives a wage equal to \(\alpha Y_{ij} \). Note that, given this specification of agent payoffs, higher quality partners are more desirable since any worker’s (firm’s) payoff strictly increases in the quality of the firm (worker) to which he or she is matched. Because production requires inputs from both firms and workers, any unmatched agent receives a payoff of zero.

The assignment of workers to firms is accomplished by means of search, which is conducted by firms only. The search process itself takes the form of random draws from a pool of available workers. After having received a random draw gratis, a firm may choose to match with that worker or continue searching by taking additional draws, each of which costs \( c(n) \), from the available supply of labor. For simplicity, I assume that draws
are conducted with replacement so that there is equal probability that a
firm will select any worker, including one that it has just rejected.

To capture the idea that thick markets facilitate search for the appro-
priate type of labor, $c(.)$ is assumed to decrease in the number of workers
in the market. Intuitively, we can imagine that paying a fixed search cost,
$C$ (for advertising, interviewers, etc.), will allow a firm to see a certain
number of workers who arrive at rate $\lambda(n)$ over some time period. Since
urban density very likely increases the extent to which individuals interact
with one another by increasing the frequency of face-to-face contacts
(e.g., Glaeser 1999), firms in cities may experience a more rapid arrival
rate of potential workers for any job opening. As a result, paying a par-
ticular cost will allow firms in cities to see a larger number of workers,
and, on a per-worker basis, the search cost, $c(n) = C/\lambda(n)$, will decrease
in $n$.

The matching process is performed in a sequential manner so that the
firm with the highest quality searches and makes a match first. After
having seen which worker has been removed from the supply, the second
highest acts, and so on. At every stage, all firms know which workers
have been hired so that the exact distribution of available labor is always
known. This matching procedure is particularly convenient in that it en-
sures that when a firm selects a worker to hire, the worker accepts the
match because he or she realizes that it is the best job available. Conse-
quently, the configuration that results from the process is one in which
all matches are mutually agreeable.

Given this setup, let us first consider the case in which search costs are
zero, $c(.) = 0$. In such an instance, perfectly positive assortative matching
(PPAM) results: firm $j$ is matched with worker $j$, for all $j$. Since there are
no costs associated with search, the most productive firm searches until
it finds the highest quality worker. The second most productive firm then
searches until it finds the worker with the next highest quality, and so
on. Hence, the matching process described here is consistent with the
results of Becker (1973) and Kremer (1993), who find that frictionless
environments generate complete sorting when inputs are complements
and production coalitions take on a unique size.

Suppose now that search costs are strictly positive, $c(.) > 0$. The first
problem that we must consider is that of firm 1, which takes draws from
the entire supply of workers. If firm 1 has found a particular worker, $i$,
then it can either match with him and receive a payoff of $(1 - \alpha)Y_i =

3 Based on the specification of payoffs, higher quality firms are more desirable
employers for all workers. Hence, even in a case where all firms search simulta-
nceusly, firm 1 should still be able to select any worker it wants. Similarly, after
firm 1 has made its choice, firm 2 should be able to select any of the remaining
workers that it wants. In this sense, the matching process can be thought of as
sequential.
(1 - \(\alpha\))q^\(\alpha\)k^\(\alpha\)_, or pay \(c(n)\) to take another draw. Formally, the Bellman equation describing firm 1’s optimal behavior can be written as follows:

\[ V(q_1) = \max [(1 - \alpha)q^\alpha k_1^\alpha, \frac{1}{n} \sum_{j=1}^{n} V(q_j) - c(n)]. \]  

(2)

**Proposition 1.** The solution to the maximization condition given by (2) is a cutoff rule, in which firm 1 will have a reservation quality, \(q^*\).

Proofs of all propositions appear in the appendix. Firm 1 simply matches with the first worker it encounters with quality \(q \geq q^*\). If confronted with a lower quality worker, it pays \(c(n)\) and takes another draw. Simple manipulation can be used to show that firm 1’s reservation quality satisfies:

\[ \min q \text{ subject to } \sum_{q \geq q^*} [(1 - \alpha)q^\alpha k_1^\alpha - (1 - \alpha)q^\alpha k_1^\alpha] \leq nc(n), \]  

(3)

where the summation is over all workers with quality greater than or equal to \(q\).

From (3), we see that a lower value of \(c(n)\), all else constant, will tend to increase firm 1’s cutoff, which is quite intuitive. After all, lower search costs will allow firm 1 to be more selective. Note, however, that the right-hand side of (3) need not decline in population size \(n\). On one hand, \(c(n)\) declines with \(n\), representing lower search costs in thick markets, but this term is multiplied by \(n\) itself, representing an increase in the complexity of search (or “congestion”) with market size.

By imposing the restriction that the benefits of size outpace its costs, we can derive the following three properties of firm 1’s reservation quality.

**Proposition 2.** Let firm 1 have initial cutoff quality \(q^*\), which satisfies equation (3) given some set of initial conditions governing agent qualities and the per worker search cost; and let \(n_1c(n_1) < n_1c(n_2)\) for \(n_1 > n_2\). Then, (i) an increase in market size, \(n\), (ii) an increase in firm 1’s quality, \(k_1\), and/or (iii) an increase in any of the qualities of the workers, \(q_i\), for whom \(q_i > q^*\), will have a nondecreasing impact on firm 1’s cutoff quality.

Due to the fact that we are dealing with discrete values of \(k\) and \(q\), firm 1’s reservation quality need not strictly decrease in search costs, nor strictly increase in its own capital and the quality levels of the best workers. However, given a large enough change, this result indicates that firm 1 will choose to search for a more capable worker the larger is the labor market, the greater is its own stock of productive capital, and the more capable are the top workers. Note that this last property is similar to that of Kremer and Maskin’s (1996) model, which implies that greater skill
dispersion is associated with more extensive segregation of workers by skill across production coalitions.

Let us now move on to the second firm in the rank-ordering. Once firm 1 has made its decision and hired a worker, firm 2 enters the market and undergoes the same search process. The problem facing firm 2 is the same as the one derived for firm 1, with one important difference: the supply of labor now contains \((n - 1)\) workers. By assumption, this does not influence the per worker search cost that the firm faces since the size of the population has not changed. However, it does influence firm 2’s cutoff quality, \(q\), which satisfies:

\[
\min q \text{ subject to } \sum_{q \geq q} [(1 - \omega)q_1 - (1 - \omega)q_1] \leq (n - 1)c(n). \tag{4}
\]

Note that the summation is over all available workers of quality \(q\) or higher.

Naturally, firm 2’s reservation quality will be affected by the worker that firm 1 hires since that choice will influence the distribution of qualities firm 2 faces. Thus, given a set of workers and firms with a particular per worker search cost, the quality satisfying (4) will tend to vary across matching trials as long as firm 1’s cutoff is less than \(q_1\).

After firm 2 has found a satisfactory worker, we continue to move down the distribution of firms, deriving a series of optimal search rules, each of which involves the selection of a reservation quality satisfying a condition analogous to (3) and (4) above. Again, as long as firms are willing to match with more than one worker, the sequence of cutoff qualities will vary from trial to trial depending on which workers are actually drawn and hired. The matching strategy of the lowest quality firm, of course, is simply to hire the only remaining worker in the labor supply.

From this basic setup, we can now formalize the relationship between market size and the extent to which firm-worker matches are stratified by quality, a natural metric for which is the covariance between \(q\) and \(k\) across all \(n\) production coalitions. That is, defining a matching configuration by a bijective mapping \(i()\) which assigns to firm, \(j\), the worker given by \(i(j)\), let the degree of stratification exhibited by matching \(i()\) be quantified by

\[
\text{cov}(q_{i(j)}, k_j) = \frac{1}{n} \sum_{j=1}^{n} q_{i(j)}k_j - \bar{q}\bar{k}, \tag{5a}
\]

where the bars over \(q\) and \(k\) represent means. Then, we have the following result.

**Proposition 3.** An increase in market size, \(n\), will, on average, induce a greater degree of stratification of workers and firms by quality if \(c(n)\) decreases rapidly enough in \(n\).
Larger markets allow the most productive firms to conduct more extensive searches because the cost of doing so is smaller. The end result is greater sorting: higher quality firms tend to be matched with higher quality workers, leaving lower quality firms to match with lower quality workers.

Now consider the impact of an increase in market size, and thus a lower per worker search cost, on the following three market characteristics: average productivity, wage inequality, and the expected returns to skill. Average productivity in this setup is just the average level of output per production coalition and, therefore, can be written as follows:

$$\frac{1}{n} \sum_{i=1}^{n} q_{ij} k_{ij}^{1-\alpha}.$$ (5b)

To quantify inequality and the expected returns to skill, let us first introduce the following notation. Let $w_{i,1}, w_{i,2}, \ldots, w_{i,n}$ represent the wages earned under the matching configuration $i(\cdot)$, by workers with qualities $q_1, q_2, \ldots, q_n$. I define inequality as

$$\frac{w_{i,1}}{w_{i,n}},$$ (5c)

which is merely the ratio of the wage of the most skilled worker to that of the least skilled worker. Although (5c) is not necessarily the ratio of the highest wage received to the lowest, it does indicate the degree of spread between the wage of the worker with the greatest potential wage earnings to that of the worker with the least. As search costs drop to zero, of course, (5c) will quantify the ratio of the highest wage to the lowest.

Similarly, I define the expected returns to skill as

$$\frac{w_{i,1} - w_{i,n}}{q_1 - q_n},$$ (5d)

which represents the average change in a worker’s wage per unit of skill, $q$, as one moves from the bottom end of the skill distribution to the top. Using these measures, (5b)–(5d), I first state the following result.

**Proposition 4.** Perfectly positive assortative matching (PPAM) maximizes average productivity (5b), inequality (5c), and the expected returns to skill (5d).

Essentially, this statement establishes that, as we move from a situation in which there are search costs that are sufficiently high to generate less than a completely stratified configuration to one in which PPAM always results, we will, on average, see an increase in productivity, inequality, and the expected returns to skill.
More generally, however, we can describe the variation of these three characteristics with market size as follows.

**Proposition 5.** An increase in market size, \( n \), will, on average, generate greater average productivity, inequality, and expected returns to skill if \( c(n) \) decreases rapidly enough in \( n \).

Hence, as long as search frictions decrease sufficiently rapidly as markets become larger, the model indicates that, on average, we should see increases in each of these three characteristics with size in a cross section of markets.

**III. Extensions**

A. Unemployment

Although the model presented above is one of full employment, the idea that search frictions systematically influence equilibrium rates of unemployment has been widely researched (for surveys, see Mortensen 1986; Pissarides 1990). In the most basic of these types of models, workers draw wage offers from some underlying distribution until one exceeding a reservation level is found. When this happens, the worker leaves the pool of unemployed labor.

Commonly, reservation wages vary inversely with a search cost parameter, which acts much as the term, \( c(n) \), in the model presented above. The fact that some studies have documented a positive connection between city size and unemployment (e.g., Vipond [1974] in the case of men) may, therefore, be explained by larger local markets possessing fewer search frictions, which serves to entice workers or firms to search more extensively.

It is interesting that reservation wage models also suggest that workers will hold out for a higher wage the greater is the variance of the distribution from which wage offers are drawn. Intuitively, this happens because the probability of drawing a particularly high wage increases with the dispersion of the distribution. Unemployment, therefore, should rise with wage variance.

In the model presented above, we can derive a similar, positive association between unemployment and wage dispersion as follows. If firms in cities choose to search for higher quality workers because search costs are lower, unemployment may be higher in urban areas because firms tend to hold out for better workers. Yet, as matches are made, there will be a more dispersed distribution of wages, induced by more extensive sorting. Essentially, the complementarity in the wage function serves to increase the “average spread” in distribution of worker payoffs as matches become increasingly assortative. Hence, just as in many reservation wage models, higher unemployment will be positively associated with greater
wage dispersion. However, unlike in those models, it is not caused by a higher variance in the distribution of wages.

Of course, two additional factors related to search costs may serve to place downward pressure on rates of unemployment. First, while firms may be more selective in denser markets, the matching process itself may occur at a faster rate. This result is suggested by Alperovich’s (1993) evidence. Second, lower search costs may make on-the-job searches for new positions easier. Instead of searching for a new job by quitting an old position first, urban dwellers may remain in a particular position until a new job is found. This might help to explain why the unemployment–city size relationship, which is sometimes found to be positive, is not particularly strong.

B. Job Turnover

Jovanovic (1979) suggests that, because workers have imperfect information about how productive they will be in a particular position, they move from job to job in an attempt to find an optimal match. Topel and Ward (1992) suggest that this type of occupational churning is a particularly important aspect of the growth in the earnings of young workers, who have relatively little information about what they do best.

Because the extent to which workers move from one position to another very likely depends on the cost of doing so, we should expect to see concentrations of young workers, who seem to be the primary beneficiaries of the search process, in cities where search frictions are smaller. This implication is certainly consistent with Glaeser’s (1999) finding that urban populations tend to be significantly younger than rural ones. In particular, he reports that, among the three largest metropolitan areas in the United States—New York, Chicago, and Los Angeles—the share of the resident population under the age of 35 is nearly 4 percentage points higher than the share residing in nonmetropolitan areas.

IV. Empirical Evidence

This section presents empirical evidence on the relationship between population and three characteristics—average productivity, wage inequality, and returns to skill—for various samples of local markets in the United States.

A. Average Productivity

Numerous studies have established a strong, positive connection between various measures of average productivity and the extent to which economic activity is geographically clustered. Glaeser and Mare (in press), for example, find evidence of a substantial urban wage premium, even after conditioning on measures of a worker’s human capital. Using data
from the Panel Study of Income Dynamics, they estimate the city premium to be nearly 27% after controlling for experience, education, and job characteristics. Results from the analysis of data from the National Longitudinal Survey of Youth (NLSY) show the premium to be nearly 16% after conditioning on experience, education, job characteristics, and the Armed Forces Qualification Test (AFQT) score. Furthermore, their results show that workers who move to cities from rural locations do not experience immediate wage gains (level effects) but, instead, wage increases in the 2–6 years after moving. This finding clearly suggests that large markets are characterized by mechanisms that actually enhance worker productivity in some manner.

Additional evidence is documented by Ciccone and Hall (1996), whose examination of the states of the United States shows a direct link between a state’s level of average productivity and the degree to which employment is geographically concentrated in its counties. After having accounted for statewide levels of average education and measures of public capital, they find that a doubling of employment density corresponds to a 6% increase in gross state product per worker.

As mentioned in the introduction, since this connection between urban agglomeration and productivity has been widely documented, numerous explanations have been suggested. However, unlike theories that appeal to positive spillovers, the division of labor, or firm-level scale economies, the sorting model presented above also makes predictions about inequality and skill distributions. I consider these issues next.

### B. Wage Inequality

The model of Section II suggests that, as local market size increases, workers at the top end of the skill distribution should experience increases in their wages relative to those of the workers at the low end of the distribution. To test this idea, I use data from the 1980 1% metro sample of the Integrated Public Use Microdata Series (Ruggles and Sobek 1997), which reports over 2 million person records from the U.S. Census of Population and Housing, to estimate the following statistical model:

\[
\ln(w_{ij}) = \alpha + \beta x_{ij} + \gamma z_j + \epsilon_{ij},
\]

where \(w_{ij}\) is the log hourly wage of worker \(i\) who works in metropolitan area \(j\), \(x_{ij}\) is a vector of personal covariates, \(z_j\) is a vector of characteristics of metropolitan area \(j\), and \(\epsilon_{ij}\) is an individual-specific residual.

The variables that appear in the estimation of (6) are reported in table 1. The personal covariates, \(x_{ij}\), include a number of basic demographic characteristics (e.g., education, experience, gender) as well as five occupational indicators. Among the city-level variables, \(z_j\), are average years of education and average years of experience among individuals in the
Table 1
Estimates of Equation (6)

<table>
<thead>
<tr>
<th></th>
<th>All Individuals</th>
<th>16+ Years of Schooling</th>
<th>13–15 Years of Schooling</th>
<th>9–12 Years of Schooling</th>
<th>0–8 Years of Schooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1 (0.06)</td>
<td>-0.36 (.14)</td>
<td>-0.75 (.13)</td>
<td>-1.2 (.08)</td>
<td>-1.04 (.24)</td>
</tr>
<tr>
<td>Education</td>
<td>0.05 (.0005)</td>
<td>0.025 (.002)</td>
<td>0.04 (.003)</td>
<td>0.07 (.002)</td>
<td>0.03 (.002)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.04 (.0004)</td>
<td>0.05 (.001)</td>
<td>0.045 (.001)</td>
<td>0.04 (.006)</td>
<td>0.04 (.002)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.006 (.00008)</td>
<td>-0.001 (.00002)</td>
<td>-0.0008 (.00002)</td>
<td>-0.0006 (.00001)</td>
<td>-0.0005 (.00003)</td>
</tr>
<tr>
<td>Disabled</td>
<td>-0.13 (.006)</td>
<td>-0.12 (.002)</td>
<td>-0.12 (.001)</td>
<td>-0.13 (.008)</td>
<td>-0.17 (.002)</td>
</tr>
<tr>
<td>Married</td>
<td>0.18 (.004)</td>
<td>0.18 (.008)</td>
<td>0.16 (.008)</td>
<td>0.16 (.006)</td>
<td>0.19 (.002)</td>
</tr>
<tr>
<td>In school</td>
<td>-0.07 (.004)</td>
<td>-0.07 (.009)</td>
<td>-0.06 (.006)</td>
<td>-0.07 (.006)</td>
<td>-0.03 (.005)</td>
</tr>
<tr>
<td>Female</td>
<td>0.05 (.008)</td>
<td>0.14 (.02)</td>
<td>0.01 (.02)</td>
<td>0.04 (.01)</td>
<td>0.18 (.02)</td>
</tr>
<tr>
<td>Race (white = 1)</td>
<td>0.11 (.005)</td>
<td>0.13 (.01)</td>
<td>0.12 (.01)</td>
<td>0.11 (.007)</td>
<td>0.06 (.005)</td>
</tr>
<tr>
<td>Female*experience</td>
<td>-0.15 (.006)</td>
<td>-0.15 (.002)</td>
<td>-0.01 (.001)</td>
<td>-0.015 (.001)</td>
<td>-0.02 (.003)</td>
</tr>
<tr>
<td>Female*experience squared</td>
<td>0.002 (.0001)</td>
<td>0.002 (.0004)</td>
<td>0.0008 (.0003)</td>
<td>0.0002 (.0002)</td>
<td>0.0003 (.0005)</td>
</tr>
<tr>
<td>Category</td>
<td>Coefficient 1</td>
<td>Coefficient 2</td>
<td>Coefficient 3</td>
<td>Coefficient 4</td>
<td>Coefficient 5</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Female*race</td>
<td>-.14</td>
<td>-.18</td>
<td>-.13</td>
<td>-.14</td>
<td>-.09</td>
</tr>
<tr>
<td></td>
<td>(.007)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Female*married</td>
<td>-.17</td>
<td>-.16</td>
<td>-.15</td>
<td>-.18</td>
<td>-.15</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.008)</td>
<td>(.025)</td>
</tr>
<tr>
<td>Female*number of children</td>
<td>-.03</td>
<td>-.05</td>
<td>-.05</td>
<td>-.03</td>
<td>-.02</td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.005)</td>
<td>(.004)</td>
<td>(.002)</td>
<td>(.007)</td>
</tr>
<tr>
<td>Professional or managerial</td>
<td>.44</td>
<td>.27</td>
<td>.42</td>
<td>.45</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.02)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Technical or sales</td>
<td>.28</td>
<td>.1</td>
<td>.28</td>
<td>.35</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.02)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Service</td>
<td>.08</td>
<td>-.12</td>
<td>.13</td>
<td>.13</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.03)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Craft</td>
<td>.36</td>
<td>.11</td>
<td>.36</td>
<td>.43</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.02)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Operator or laborer</td>
<td>.3</td>
<td>-.08</td>
<td>.26</td>
<td>.35</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.03)</td>
<td>(.02)</td>
<td>(.01)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Average MSA education</td>
<td>.06</td>
<td>.05</td>
<td>.04</td>
<td>.06</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.01)</td>
<td>(.008)</td>
<td>(.006)</td>
<td>(.02)</td>
</tr>
<tr>
<td>Average MSA experience</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.008</td>
<td>.009</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.003)</td>
<td>(.003)</td>
<td>(.002)</td>
<td>(.006)</td>
</tr>
<tr>
<td>Log of resident population</td>
<td>.027</td>
<td>.04</td>
<td>.03</td>
<td>.02</td>
<td>.00005</td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.002)</td>
<td>(.002)</td>
<td>(.001)</td>
<td>(.005)</td>
</tr>
<tr>
<td>Observations</td>
<td>312,625</td>
<td>64,180</td>
<td>64,453</td>
<td>162,521</td>
<td>21,471</td>
</tr>
</tbody>
</table>

Note.—Ordinary least squares estimates. Dependent variable is log hourly wage. Eight census division dummies are also included in the estimation of each specification. Standard errors are reported in parentheses. MSA = metropolitan statistical area.
sample, eight census division dummies, and the logarithm of resident population. After dropping all observations for which either the metropolitan area of work or some of the individual covariates are not reported, I arrive at a total sample size of 312,625 individuals covering 286 metropolitan areas.

Ordinary least squares estimates of (6) using all 312,625 observations appear in the first column of table 1. From the results, we can see that there is a significantly positive premium associated with the resident population of an individual’s metropolitan area of work, which, in light of the studies cited in the previous subsection, is not surprising. In this particular sample, a doubling of population corresponds to a 2.7% increase in a worker’s hourly wage on average.

To determine whether this effect is consistent across workers of different educational groupings and, thus, provide a simple test of the sorting theory, I break the sample into four categories by educational attainment—16 or more years of schooling, 13–15 years, 9–12 years, and 0–8 years—and estimate (6) for each. The results appear in the second through fifth columns of table 1.

Looking at the estimated coefficients on the logarithm of resident population, we see that the local market size premium seems to increase with years of schooling completed. In fact, across the four groups, there is a monotonic rise in the estimated relationship between population and wages. For workers at the top end of the educational distribution, a doubling of resident population is associated with a 4% increase in hourly wages, on average. Workers with 13–15 years of education and those with 9–12 years exhibit smaller gains: 3% and 2%, respectively. Among workers at the bottom end of the distribution, by contrast, the estimated partial correlation is not significantly different from zero.

Although not conclusive, this evidence is certainly consistent with the predictions of the sorting model. Again, if larger markets generate more stratified matches, such areas will tend to allocate high-skill workers to more productive employers, leaving low-skill workers in less productive positions. The overall effect will be a boost in the wages of the former relative to those of the latter.

C. Returns to Skill and Education Levels

A third implication of the theory is that larger local markets will generate greater expected returns to skill in the sense that the average gain in wages per unit of skill across all workers will be higher. We should therefore expect to see higher-skill workers in larger markets, where they will be sorted more efficiently.

If we measure a worker’s skill by the level of education that he or she has attained, the evidence is consistent with this implication too. The
Search, Sorting, and Urban Agglomeration

Table 2
Education and Population in U.S. Counties

<table>
<thead>
<tr>
<th>Education Measure</th>
<th>Estimated Slope</th>
<th>SE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage with no high school</td>
<td>−.02 (.001)</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Percentage with high school diploma</td>
<td>−.01 (.0009)</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Percentage with some college or associate’s degree</td>
<td>.008 (.001)</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Percentage with bachelor’s degree or higher</td>
<td>.02 (.0009)</td>
<td>.2</td>
<td></td>
</tr>
</tbody>
</table>

Note.—Ordinary least squares estimates. Independent variable is logarithm of 1990 population. Heteroskedasticity-consistent standard errors in parentheses.

results of the previous subsection suggest that the wage differential between workers at the top of the educational distribution and those at the bottom is greater in larger markets. Hence, the average slope of the wage function with respect to educational attainment should be higher in these areas.

Moreover, the evidence on educational distributions across local markets of different sizes indicates that urban populations in the United States are significantly more educated than rural ones. Glaeser (1999) finds that, in 1990, only 13.4% of the populations of nonmetropolitan areas were college graduates, on average. By contrast, the corresponding figure for the three most populous cities (New York, Chicago, and Los Angeles) was 23.6%; for all other metropolitan areas it was 22.5%. Glaeser and Mare (in press) find that the average city dweller possesses one more year of schooling than the average nonmetropolitan area resident.

United States county-level data from the 1990 Census of Population and Housing shows similar results. Using the 3,111 counties of the contiguous United States, I find a significantly negative relationship between overall resident population and the percentages of their adult populations possessing low levels of schooling. Table 2 reports the estimated slope from a regression of the proportion of a local market’s adult population with less than a high school education on the logarithm of total resident population. The estimate shows that, on average, a doubling of county-level population is associated with a decrease of 0.02 in the proportion of residents with only 0–8 years of schooling. There is also a significant decline in the proportion of county populations with only a high school degree as population rises in the cross section.

Not surprisingly, then, the proportion of resident populations possessing relatively high levels of schooling rises significantly with population. The estimated coefficients from the regressions of the proportion of the population with a bachelor’s degree or higher as well as the proportion with some college or an associate’s degree on the logarithm of total population are both positive. In particular, the results indicate that
a doubling of population is associated with an increase of roughly 2 percentage points, on average, in the proportion of the population with 16 or more years of education.

Although the model presented in Section II neither endogenizes the education process nor examines the location decisions of agents with varying degrees of education across geographic markets, these results still lend support to the idea that larger local markets facilitate search and recruitment. Because larger markets allocate high quality (say, highly educated) workers more efficiently, urban areas will encourage a larger fraction of their residents to invest in schooling than smaller markets, all else equal. Recall that the expected return to an additional unit of quality (say, a year of schooling) was, on average, higher in larger markets given an equal distribution of worker skills and firm capital.

Suppose, for example, that all workers have the option to enhance their qualities by purchasing education. Doing so increases a worker's skill from $q_i$ to $\mu q_i$, where $\mu > 1$, although a fixed cost, $E$, must be paid. Given $E$ and the expected return to an additional unit of quality, $\beta$, an agent with quality $q_i$ will choose to invest in education only if

$$\beta(\mu - 1)q_i \geq E.$$ 

If $\mu$ and $E$ are constant across markets, urban areas will be able to support a larger proportion of their populations investing in education than rural ones, provided the increase in the number of workers purchasing education does not depress the expected return, $\beta$, too much (if at all).

V. Concluding Remarks

In a recent survey of assignment models, Sattinger (1993) notes that the means by which heterogeneous workers are allocated to heterogeneous jobs has significant implications for an economy's aggregate output and distribution of earnings, whether conditional on worker characteristics or not. This article has constructed a model to explore the hypothesis that urban agglomeration reduces the costs associated with assigning workers to jobs. Assuming that firm capital and worker skill are complementary in production, increases in market size will tend to induce greater sorting: high-capital firms will hire high-skill workers, leaving low-end firms to match with low-end workers.

Three implications of the model—that larger local markets will exhibit greater output per worker, wage inequality, and expected returns to skill—are all consistent with recent data on U.S. counties and metropolitan areas. The evidence indicates significantly positive connections between county- and city-level population, on the one hand, and average productivity, the wages of highly educated workers relative to their less ed-
ucated counterparts, and the extent to which resident populations are educated on the other.

As such, this article demonstrates that search frictions, which previous studies have identified as a potentially important element in the creation of agglomeration economies, also provide a basis for the rationalization of several additional labor market outcomes across local geographic areas. In particular, the ability of markets to match workers and firms efficiently will directly affect the extent of inequality, the magnitude of an individual’s expected return to skill and, thus, the resulting stocks of human capital held by the resident population.

Appendix

Proof of Proposition 1

This result is standard in the search literature. The second term in the brackets of equation (2) is just a constant, representing the expected value of the value function given one more round of search. The first term is strictly increasing in \( q_i \), the quality of the worker under consideration. Since the firm can do no better than receiving \((1 - \alpha)Y_{i1}\), \( V(q_i) = (1 - \alpha)Y_{i1} \). At the same time, since the firm can do no worse than receiving \((1 - \alpha)Y_{i0}\), firm 1 will choose to search if it encounters worker \( n \), as long as the search cost is not prohibitively high. If it is not, there will exist some critical \( q_i > q^* \) such that the firm will reject \( q_i \), \((j \geq 0)\), but accept \( q_{i-}(j \geq 0)\). If search costs are too high to justify searching even after selecting the lowest quality worker, firm 1 will merely have a reservation quality equal to \( q_{n1} \), which corresponds to random matching. Q.E.D.

Proof of Proposition 2

All follow from (3). First, note that the left-hand side strictly decreases in the cutoff quality \( q^* \). Then, (i) an increase in population strictly decreases the right-hand side, implying that the new cutoff quality must be at least as large as \( q^* \). (ii) An increase in firm 1’s quality, \( k_{i1} \), will increase the left-hand side of (3) if \( q^0 \neq q_1 \) and leave it unchanged if \( q^0 = q_1 \). The cutoff level of quality, therefore, cannot fall if (3) is to be satisfied. (iii) Similarly, a rise in the productive capabilities of the workers whose qualities are greater than \( q^* \) will strictly increase the value of the left-hand side of the equation if \( q^0 \neq q_1 \). Hence, to maintain equation (3), firm 1’s cutoff cannot decrease. Q.E.D.

Proof of Proposition 3

The previous proposition already establishes that an increase in \( n \) can induce firm 1 to set a higher reservation quality. In such a case, because there is equal probability of drawing any worker, the expected quality of the worker hired by firm 1 will increase in \( n \). While this rise in \( n \) will, on average, reduce the qualities of the available workers firm 2 may select, it will also decrease the per worker search cost, which appears on the right-hand side of (4). A sufficiently large drop in this search cost places upward
pressure on firm 2’s cutoff quality and thus increases the expected quality of its worker. In general, we can specify a cost function such that an increase in \( n \) leads the top \( T \) firms (for some \( T < n \)) to set higher cutoffs and, thus, match with higher quality workers on average, leaving firms \( T + 1, T + 2, \ldots, n \) to match with lower quality workers, on average. Given \( \{k_i\}_{i=1}^n \), the average degree of stratification will depend directly on the term \( \sum_{i=1}^n E(q_{ij})k_i \), where \( E(q_{ij}) \) represents the expected quality of the worker assigned to firm \( j \) taken across all possible bijective mappings \( i(\cdot) \).

To prove that this sum increases when \( E(q_{ij}) \) rises for \( j = 1, \ldots, T \) but decreases for all \( j > T \), note first that the term \( \sum_{i=1}^n E(q_{ij})k_i \) is unaffected by the probabilities with which each matching configuration occurs. In particular, given that there are a total of \( m = n! \) matching configurations: \( i_1(\cdot), i_2(\cdot), \ldots, i_m(\cdot) \),

\[
\sum_{j=1}^n E(q_{ij}) = q_{i_0(1)}\text{Prob}(i_1(\cdot)) + q_{i_0(2)}\text{Prob}(i_2(\cdot)) + \cdots \\
+ q_{i_0(1)}\text{Prob}(i_m(\cdot))) + q_{i_0(2)}\text{Prob}(i_i(\cdot)) \\
+ q_{i_0(3)}\text{Prob}(i_j(\cdot)) + \cdots + q_{i_0(t)}\text{Prob}(i_m(\cdot)) \\
+ \cdots + q_{i_0(1)}\text{Prob}(i_m(\cdot)) + q_{i_0(2)}\text{Prob}(i_2(\cdot)) + \cdots + q_{i_0(t)}\text{Prob}(i_m(\cdot))
\]

where \( \text{Prob}(i_p(\cdot)) \) is the probability with which matching configuration \( i_p(\cdot) \) occurs. This sum can be rewritten as

\[
\sum_{j=1}^n E(q_{ij}) = \text{Prob}(i_1(\cdot))q_{i_1(1)} + q_{i_1(2)} + \cdots + q_{i_1(t)} \\
+ \text{Prob}(i_2(\cdot))q_{i_2(1)} + q_{i_2(2)} + \cdots + q_{i_2(t)} + \cdots \\
+ \text{Prob}(i_m(\cdot))q_{i_m(1)} + q_{i_m(2)} + \cdots + q_{i_m(t)}.
\]

Because each mapping \( i(\cdot) \) is bijective and \( \sum_{i=1}^m \text{Prob}(i_p(\cdot)) = 1 \), then \( \sum_{i=1}^n E(q_{ij}) = \sum_{i=1}^n q_{ij} \) regardless of the probabilities governing the \( m \) matching configurations.

As a result, any increase in \( \sum_{j=1}^T E(q_{ij}) \), which I denote by \( \sum_{j=1}^{T-1} \Delta E(q_{ij}) \), will equal the decrease in \( \sum_{j=T+1}^{T+m} E(q_{ij}) \), which I denote by \( -\sum_{j=T}^{T+m-1} \Delta E(q_{ij}). \) Now, notice that the maximum amount by which \( \sum_{i=1}^n E(q_{ij})k_i \) may decrease is given by \( -\sum_{j=T}^{T+m-1} \Delta E(q_{ij})k_{T+1} \) and that the minimum amount by which \( \sum_{j=1}^n E(q_{ij})k_i \) may increase is given by \( \sum_{j=1}^T \Delta E(q_{ij})k_T \). Consequently, the change in the term \( \sum_{j=1}^n E(q_{ij})k_i \) must be at least as large as
\[ k_T \sum_{j=1}^{T} \Delta E(q_{i(j)}) + k_{T+1} \sum_{j=T+1}^{n} \Delta E(q_{i(j)}) = (k_T - k_{T+1}) \sum_{j=1}^{T} \Delta E(q_{i(j)}) > 0 \]

since \( k_T > k_{T+1} \). Hence, the change in the expected covariance with this drop in search costs must be positive. Q.E.D.

**Proof of Proposition 4**

Average productivity depends on the term \( \sum_i q_i^a k_i^{1-a} \). Since the function, \( f(q,k) = q^a k^{1-a} \) with \( 0 < \alpha < 1 \), is supermodular, PPAM maximizes this quantity (a proof is given by Becker [1973], pp. 841–42). PPAM generates measures of inequality and returns to skill given by, respectively

\[ \frac{q_i^a k_i^{1-a}}{q_n^a k_n^{1-a}} \]

and

\[ \frac{\alpha(q_i^a k_i^{1-a} - q_n^a k_n^{1-a})}{q_1 - q_n}. \]

Given any distribution of agent qualities, \( \{q_i\}_{i=1}^n \) and \( \{k_i\}_{i=1}^n \), the maximum and minimum wages possible are, respectively, \( \alpha q_1^a k_1^{1-a} \) and \( \alpha q_n^a k_n^{1-a} \). Q.E.D.

**Proof of Proposition 5**

Expected average productivity is given by \( \frac{1}{n} \sum_{i=1}^{n} E(q_i^a) k_i^{1-a} \). Following the proof of proposition 3, a sufficiently large decline in \( c(n) \) will increase this term by increasing the expected value of the quality of the worker assigned to the top \( T \) firms (for some \( T < n \), hence \( E(q_{i(T)}) \) for \( j = 1, 2, \ldots, T \), but lowering \( E(q_{i(j)}) \) for \( j > T \). Since \( \sum_{i=1}^{n} E(q_{i(j)}) = \sum_{j=1}^{n} q_i^a \) regardless of the probabilities with which each matching configuration occurs, the increase in \( \sum_{i=1}^{T} E(q_{i,j}) \) will equal the decrease in \( \sum_{j=T+1}^{n} E(q_{i,j}) \). Thus, as in proposition 3, since \( k_T > k_{T+1} \), the increase in \( \sum_{i=1}^{T} E(q_{i,j}) k_i^{1-a} \) must be greater than the decrease in \( \sum_{j=T+1}^{n} E(q_{i,j}) k_j^{1-a} \), which causes expected average productivity to rise.

Additionally, since an increase in \( n \) will have a nondecreasing impact on the cutoff qualities of the firms at some top end of the firm quality distribution, the expected quality of the firm to which worker 1 will be assigned will increase. Hence, worker 1’s wage will rise in expectation. By the same logic, an increase in \( n \) will tend to make matching with the bottom worker less attractive, thus pushing down the expected quality of the firm to which he or she will be assigned. In expectation, the difference between \( w_{i,j} \) and \( w_{u,j} \) will increase, causing (5c) and (5d) to rise on average. Q.E.D.
References


