Available online at www.sciencedirect.com
ScienceDirect
Journal of Urban Economics 61 (2007) 112–128
JOURNAL OF Urban Economics
www.elsevier.com/locate/jue

# Cities, matching and the productivity gains of agglomeration

Fredrik Andersson [a], Simon Burgess [b], Julia I. Lane [c,*]

[a] Cornell University, Cornell Institute for Social and Economic Research, 391 Pine Tree Road, Ithaca, NY 14850, USA
[b] University of Bristol, CMPO and CEPR, Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK
[c] NORC/University of Chicago, 55 E Monroe St, Suite 4800, Chicago, IL 60603, USA

Received 7 November 2005; revised 9 June 2006

Available online 12 September 2006

**Abstract**

The striking geographical concentration of economic activities suggests that there are substantial benefits to agglomeration. However, the nature of those benefits remains unclear. In this paper we take advantage of a new data set to quantify the role of one of the main contenders—the matching of workers and jobs. We show that thicker urban labor markets are associated with more assortative matching in terms of worker and firm quality. When we estimate establishment-level production functions we also find evidence of complementarities between worker and firm quality. Putting together the production and matching relationships, we show that production complementarity and assortative matching is an important source of the urban productivity premium.


*JEL classification:* R23; R12; J24

*Keywords:* Urban productivity; Matching; Agglomeration

## 1. Introduction

Cities are home to 75% of Americans, yet occupy less than 2% of the land area of the lower 48 states.[1] This striking geographical concentration of economic activities, evident in both de-

* Corresponding author.
*E-mail address:* Lane-Julia@norc.uchicago.edu (J.I. Lane).

[1] These facts are from Rosenthal and Strange [14].

0094-1190/$ – see front matter 
doi:10.1016/j.jue.2006.06.005

veloped and developing countries, suggests that there are substantial benefits to agglomeration Yet it has been difficult to econometrically identify the major sources of these benefits, primarily due to data deficiencies (Rosenthal and Strange [14]). The contribution of this paper is to use a unique data set to examine the role of one potential source: the improved matching between firms and workers made possible by dense urban labor markets.

The underlying idea is straightforward. Suppose that both workers and firms differ in quality. If production is characterized by complementarity between worker and firm quality, productivity will be higher when workers and firms are assortatively matched. Urban areas will be more productive than rural if they are characterized by a greater degree of assortative matching, which would arise if dense markets have lower search frictions.

In order to empirically test this idea, we need to quantify the strength of two relationships: the presence of complementarity between worker and firm quality in production, and the degree of spatial variation in assortative matching. Our data, which are universal and longitudinal in both firms and workers, are uniquely suited to address the issue. We can test the first relationship because we have direct measures of "worker quality" (or the labor-market value of human capital that is independent of the identify of the employer) and "firm quality" (or the firm-specific wage mark-up) and we can use these to estimate the complementarity between the two in firm level production functions. We can test the second relationship because we also have information on the spatial coordinates of each worker and each firm. These data allow us to characterize the joint distribution of worker and firm quality and to describe how it varies over different labor market densities.

Our data show that there is a significant urban productivity premium. The raw average productivity differential between firms located in counties with employment per square mile in the upper decile and those located in counties with employment per square mile below the median is between 0.09 and 0.18 log points across the states in our sample, in favor of the urban firms. These raw productivity differentials cannot be accounted for by differences in industry structure between urban and rural areas—in fact the urban productivity premium is larger within industry. We show that the two conditions for matching to matter are met: there is complementarity in production, and workers and firms are more assortatively matched in dense labor markets. Putting the matching and the production function results together, we calculate how important the effect of location is for firm productivity and show that labor market matching is an important source of the urban productivity premium.

The theoretical background for this idea originates from the assortative matching observed in labor and marriage markets [6]. This argument was extended to a production framework by Kremer and Maskin [12], and Shimer and Smith [15] provided general results on the existence and characterization of equilibrium in a context with search frictions, and established restrictions on the production function that ensure positive assortative matching (PAM).[2]

The analysis in this paper derives most directly from work by Burdett and Coles [7] and Delacroix [9]. Burdett and Coles [7] set out a model with heterogeneity on both sides of the market, Nash bargained utilities, and an exogenous arrival rate of offers. They show that five types of pure strategy equilibria[3] will occur for different specifications of the joint production function. In particular sufficient complementarity in production yields PAM (the 'elite' equilibrium in their description). The most important result of their work from the point of view of this paper is that

[2] These are supermodularity of the production function, but also of its log first- and cross-derivatives (see p. 344).

[3] They note that mixed strategy equilibria can occur but they ignore them.

as the offer rate increases (as search frictions decline), the market equilibrium tends to the elite outcome.[4]

Delacroix's [9] implementation is similar. His assumption that the production function is supermodular rules out a number of potential equilibria and implies that high quality agents are better off matched with other high quality agents; without it there is no particular reason to expect an equilibrium with PAM. Delacroix's approach yields an assortatively matched equilibrium and a pooled equilibrium. Delacroix uses a series of simulation exercises to show that the PAM equilibrium is more likely as the exogenous offer arrival rate increases. This result is the theoretical basis for saying that dense urban labor markets lead to more sorted matching and therefore, given the nature of the production function, higher productivity.

The empirical analysis of the contribution of labor market matching to agglomeration is thin. Indeed, Rosenthal and Strange [14] acknowledge that there is little direct econometric evidence on the importance of any of the three main sources they identify. Of the work that has been done, Moretti [13] surveys the evidence on human capital externalities and productivity spillovers in cities. Ellison and Glaeser [10] and Rosenthal and Strange [14] show that some proxies for labor market pooling explain the regional degree of spatial correlation quite well. Baumgartner [5] shows that the division of labor is finer in big cities, which suggests a more efficient labor market. The paper closest to ours is Combes, Duranton and Gobillon [8] who use a French panel data set to establish that individual skills account for a large fraction of existing spatial wage disparities with strong evidence of spatial sorting by skills. They also find that interaction effects are mostly driven by the local density of employment.

The rest of the paper is organized as follows. The next section describes the empirical approach. Section 3 provides a description of the data, and Section 4 presents the results. The final section offers some conclusions.

## 2. Empirical approach

There are two hypotheses to investigate. First, we examine whether there is complementarity between firm and worker quality in production. We test for this feature by setting up a simple firm level production function in which output depends both on firm and worker quality, and estimate the degree of complementarity between these two inputs. However, a finding of complementarity by itself is not enough to generate an urban productivity premium, since if workers and firms were similarly paired up in all markets, the productivity impact of complementarity would be the same everywhere. Thus, the second part of the empirical exercise is an examination of whether the result that PAM is more likely when matching is easier is empirically valid: whether the degree of PAM is increasing in labor market density.

### *2.1. Production complementarities*

We estimate a production function in which output is a function of worker and firm quality, together with the interaction of the two, and include dummies for firm size classes and 2-digit industry. We estimate:

$$\ln y_j = a_0 + a_1\psi_j + a_2\bar{\theta}_{i\in j} + \beta(\psi_j.\bar{\theta}_{i\in j}) + \boldsymbol{\pi}.\mathbf{Z}_j + \upsilon_j \tag{1}$$

[4] pp. F325, F326.

where $y_j$ is the revenue per head of establishment $j$, $\psi_j$ is the estimated wage mark-up of firm $j$, and $\bar{\theta}_{i\in j}$ is the estimated average market value of the human capital of workers working for firm $j$. Firm size and industry dummies are in **Z**. The key coefficient is $\beta$, the interaction between worker and firm quality, which captures the complementarity in production between these inputs in the data. If the parameter $\beta$ is positive, a high $\psi$ firm will be more productive with high $\theta$ workers, and this complementarity in production between worker and firm quality would provide the incentive for assortative matching, as in Burdett and Coles [7] and Delacroix [9]. But this incentive is general; to generate a differential effect between urban and rural areas, the degree of assortative matching in equilibrium must be different.

### 2.2. *Assortative matching across space*

The hypothesis to be tested is whether matching is facilitated in dense urban labor markets (Delacroix [9]). In order to examine this, we characterize the matching equilibrium by the joint density function of worker quality $\theta$ and firm mark-up, $\psi$, and test whether the joint density function varies with labor market density, $\delta$. If there is positive assortative matching, there should be a positive correlation between $\theta$ and $\psi$. The factor generating differential positive assortative matching (PAM) between areas is the assumption that dense urban labor markets generate higher offer arrival rates and hence we can test whether the degree of PAM is higher in urban markets than rural ones.

We assume that a simple linear relationship underlies the joint density function. This implementation has the interpretation that the expected worker quality for a firm with mark-up $\Psi$ in a labor market with density $\delta$ can be written as:

$$E(\theta \mid \Psi, \delta) = \mathrm{b}_0 + \mathrm{b}_1\Psi + \mathrm{b}_2\delta + \mathrm{b}_3\Psi.\delta. \tag{2}$$

If $\mathrm{b}_3$ is positive, then as density increases so does the correlation between $\theta$ and $\psi$.[5] Thus the empirical specification is to regress an individual worker's quality on her matched firm mark-up, local density, and an interaction of quality and density.

### 2.3. *Calibration of the productivity effect*

In order to examine the importance of agglomeration, we estimate a firm's productivity across areas with different densities. We take the production function used above (Eq. (1)), and substitute in for the expected worker quality from the representation of the matching outcome in (2):

$$\mathrm{E}(\ln y_j \mid \Psi_j) = a_0 + a_1\Psi_j + a_2\mathrm{E}(\theta \mid \Psi_j; \delta) + \beta\big(\Psi_j.\mathrm{E}(\theta \mid \Psi_j; \delta)\big). \tag{3}$$

The effect of location on a given firm's expected productivity can be calculated by computing Eq. (3) for given values of $\Psi$ and at locations with different densities, $\delta$.

## 3. Data

Our data provide a unique opportunity to test the two empirical relationships: the presence of complementarity between worker and firm quality in production, and the degree of spatial

[5] To be clear, this relationship is not to be interpreted causally, but rather as a summary of the joint density of worker and firm quality, and its dependence on density.

variation in assortative matching. We briefly describe the general characteristics of the data set, then the characteristics that enable us to test the two relationships.

The new database that enables us to match workers with past and present employers has been assembled at the Longitudinal Employer-Household Dynamics Program at the US Census Bureau (Abowd, Haltiwanger and Lane [3]). This database consists of quarterly records of the employment and earnings of almost all individuals from the unemployment insurance (UI) systems of a number of US states in the 1990s—the UI number provides the key link between workers and firms. This data set has been extensively described elsewhere (Haltiwanger, Lane and Spletzer, [11]), but it is worth noting a number of advantages. Earnings are reasonably accurately reported, since there are financial penalties for misreporting. The data are current, and the data set is extremely large. The UI records have also been matched to internal administrative records at the Census Bureau that contain information on date of birth, place of birth, race, and gender for all workers, thus providing limited demographic information. One limitation of the data is that there are no direct data links between workers and establishments, but only between workers and firms. Thus, for about 30% of the workforce—whose employing firm consists of more than one establishment—we cannot tell with certainty in which particular establishment a worker is employed. Thus, probabilistic links are used to impute a place of work for workers who work for multi-unit businesses.

In this study we use data on workers and their employers in 2001 for two large states—California and Florida. There are 98,874,945 annual observations for California and 35,756,341 observations for Florida on 22,524,212 individuals in California and 10,274,254 individuals in Florida. They worked at a total of 1,314,847 firms in California and 592,086 in Florida during the period 1992 to 1999. The estimation sample consists of some 26 million individuals employed in either California or Florida during at least one quarter in 2001 and for which estimated firm and worker effect are available. These individuals are distributed across roughly 1.2 million unique employers. Because of differences in coverage and data collection period productivity data are available only for a subset of all employers in our sample, which reduces the sample size in the firm-level analysis to about half of all employers.

### 3.1. Production complementarities

In order to test for the presence of complementarity between worker and firm quality in production we first need estimates of worker and firm quality at the firm level, and then need firm level productivity measures.

We use estimates of worker and firm fixed effects that have been generated by LEHD program staff using econometric techniques developed by Abowd, Lengermann and McKinney [1]. The estimated model for the log of annualized earnings of worker $i$ in firm $j$ in period $t$ is:

$$\ln w_{ijt} = c + x_{it}\beta + \theta_i + \psi_j + \varepsilon_{ijt} \tag{4}$$

where $\theta_i$ is the worker effect, $x_{it}$ is a set of time-varying personal characteristics, $\psi_j$ is the firm mark-up, and $\varepsilon_{ijt}$ is the statistical residual. The personal characteristics are a quartic in labor force experience and a set of work history dummies (to capture censored employment spells), all interacted with gender. The individual fixed effect is a measure of human capital, the market value of the portable component of an individual's skill set, which we refer to as worker quality. It includes observable factors (gender, years of education) and unobservable factors such as innate ability, non-cognitive skills, family background, and educational quality. Symmetrically, the firm fixed effect is a summary measure of the wage premium (or discount) that each firm

Table 1
Summary of estimated wage components

| Component | Standard deviation | Correlation with | | | | |
|---|---|---|---|---|---|---|
| | | $y$ | $\theta$ | $\psi$ | $x\beta$ | $\varepsilon$ |
| Log real annual wage rate ($y$) | 0.901 | 1.000 | 0.475 | 0.503 | 0.231 | 0.397 |
| Person effect ($\theta$) | 0.876 | 0.475 | 1.000 | 0.065 | −0.631 | 0.000 |
| Firm effect ($\psi$) | 0.353 | 0.503 | 0.065 | 1.000 | 0.064 | 0.000 |
| Time-varying personal characteristics ($x\beta$) | 0.741 | 0.231 | −0.631 | 0.064 | 1.000 | 0.000 |
| Residual ($\varepsilon$) | 0.361 | 0.397 | 0.000 | 0.000 | 0.000 | 1.000 |

*Note*. Based on 134,631,286 annual observations from 1992 to 1999 for 32,798,466 persons and 1,906,933 firms in California and Florida.
*Source*. Data from the LEHD Program Employment Dynamics Estimates Database.

pays to observationally equivalent workers, which we refer to as the firm mark-up. This mark-up can reflect a variety of different factors such as the organizational structure, the degree of rent-sharing, the capital intensity, or the degree of unionization at a firm (see Andersson, Holzer and Lane [4], for a non-technical description). Although it is impossible to separate out how much of the firm wage mark-up is due to each of these factors, it does capture the key elements of the firm's production and personnel decisions.

Table 1 shows the correlation between the different wage components in California and Florida over the period 1992 to 1999. The first thing to note is the explanatory power of this decomposition. The correlation between the residual and the wage measure is 0.397, which translates into an $R^2$ of 84%. The second thing to note is the importance of firm effects. The simple pairwise correlation of the estimated firm effect and earnings is 0.503. This number is substantially higher than the correlation between the effects of observable personal characteristics and earnings and comparable to the correlation between the effects of unobservable person characteristics and earnings. Finally, note that firm and worker effects are virtually uncorrelated. We show below that there is an important element of positive assortative matching once the spatial dimension of data is incorporated.

The other key measure is the productivity of the establishment. The data from the Economic Census in 1997 provide measures of sales at the establishment level, which, together with employment, is used to create a proxy for productivity—sales (or revenue) per worker, similar to the measure used by Haltiwanger, Lane and Spletzer [11]. Although clearly the preferred productivity measure would be value-added per hour, Haltiwanger, Lane and Spletzer point out that there is a close correspondence both conceptually and in terms of measurement between this measure of gross output at the establishment level and the industry-level measures published by BLS. The standard BLS measure of labor productivity at the detailed industry level is output per hour.

### 3.2. *Capturing spatial variation*

In order to capture the degree of spatial variation in assortative matching, it is necessary to have detailed geographic information, which is also present in this data set. The physical location of each establishment is geocoded to the latitude and longitude level, as is the place of residence of each worker (from 1999 on). This information is available on a longitudinal, annual basis (geocoded businesses are available all years and residences have been geocoded in 1999, 2000 and 2001). The data set allows us to describe the geographical distribution of workers and employers as well as commuting and mobility patterns.

A major question is determining the appropriate unit of geography for analysis. Unfortunately, there is no consensus about the definition of a local labor market area. The Employment and Training Administration of the Department of Labor defines Workforce Investment Areas for administrative purposes, that are typically either counties or agglomerations of counties The Office of Management and Budget uses commuting patterns to define Metropolitan Statistical Areas, Micropolitan Statistical Areas, Combined Statistical Areas and Metropolitan Divisions that are used by the US Census Bureau, and the Bureau of Labor Statistics. The US Department of Transportation uses traffic analysis zones (TAZs) that are defined by "state and/or local transportation officials for tabulating traffic-related data- especially journey-to-work and place-of-work statistics. A TAZ usually consists of one or more census blocks, block groups, or census tracts." The US Department of Agriculture defines local labor market areas differently again (see http://www.ers.usda.gov/Briefing/Rurality/LMACZ/). Although most academic researchers work with the broadest geographic definitions of Metropolitan areas (see, e.g. Wheaton and Lewis [16]), it is not a satisfactory aggregation from our perspective, since this unit is too large and heterogeneous for our purposes. Two well defined geographic units are counties and tracts. There are advantages and disadvantages associated with each of these measures to measure local labor market areas. Counties may also be too large, since they often cover large areas containing both urban and rural parts. Census tracts, on the other hand, are relatively small areas of between 1500 and 8000 individuals, and while they are not designed to be a local labor market, they are chosen to be relatively homogeneous in terms of population characteristics, economic status, and living conditions. Thus, tract-based density measures will pick up some of the within-county variation in density. However, the small size of tracts is not unproblematic either, since in urban areas tracts cover a small physical area by construction. In a number of tracts, then, either the population per square mile is high—if it is in the residential areas of the city—or the employment per square mile is high—if it is in the commercial districts of the city— but the two measures are not necessarily highly correlated. To check whether our results are sensitive to the level of geographical aggregation, we estimate our results using four different measures of density: employment and population per square mile at the level of both county and of tract. For brevity, however, we present results based on employment per square mile as our density measure, at county level.

## 4. Results

### 4.1. Productivity

As a preliminary measure, we confirm the existence of a raw urban productivity premium in our data. We calculate simple correlations of employment density with productivity and wages. The results in Table 2 show a significant positive correlation in both states between employment density and productivity. The Table also shows that the dispersion of productivity is positively correlated with density, which follows from the matching approach. The results in Table 3 are similar, showing a significant positive correlation between employment density and wages.

We estimate the production function introduced above, Eq. (1). The key coefficient is $\beta$, the interaction between worker and firm quality, which captures the complementarity in production between these inputs in the data. The results of this analysis are reported in Table 4. The coefficients on worker and firm quality are quantitatively and statistically significant as expected. Firms employing workers with more valuable general skills are more productive. The critical part for our purposes is the degree of complementarity. In both California and Florida the interaction

Table 2
Correlations between productivity, productivity dispersion and density

| | California | Florida |
|---|---|---|
| Corr(P,E) | 0.489** | 0.726** |
| Corr(PD,E) | 0.709** | 0.275* |
| *N* | 58 | 67 |

*Notes*. Corr(P,E) is the correlation between the mean of log of labor productivity and log of employment per square mile across the counties within each state. PD is the standard deviation of log of labor productivity across firms within the geographical unit.
* Significant at 5%.
** Significant at 1%.

Table 3
Correlations between wages, wage dispersion and density

| | California | Florida |
|---|---|---|
| Corr(W,E) | 0.525** | 0.748** |
| Corr(WD,E) | 0.559** | 0.659** |
| Corr(W9010,E) | 0.247 | 0.504** |
| *N* | 58 | 67 |

*Notes*. Corr(W,E) is the correlation between the mean of log of annualized earnings and log of employment per square mile across the counties within each state. PD is the standard deviation of log of annualized earnings across all workers within the geographical unit.
* Significant at 5%.
** Significant at 1%.

Table 4
Firm-level productivity regressions

| | California | Florida |
|---|---|---|
| (1) Mean worker quality ($\bar{\theta}$) | 0.430 | 0.346 |
| | (163.46)** | (82.46)** |
| (2) Firm mark up ($\psi$) | 0.638 | 0.568 |
| | (246.76)** | (129.79)** |
| Interaction term between (1) and (2) | 0.028 | 0.059 |
| | (14.92)** | (15.92)** |
| Constant | 4.828 | 4.739 |
| | (686.22)** | (468.25)** |
| Observations | 400,770 | 152,367 |
| $R$-squared | 0.29 | 0.24 |

*Notes*. Unit is a firm. The dependent variable is the log of labor productivity. In addition the specification includes controls for size and industry of firm (not reported in table).
* Significant at 5%.
** Significant at 1%.

term is statistically significant and positive. We return to discussing its quantitative significance below.

We carry out two robustness checks. First, productivity at the establishment level might depend on the distribution of workforce quality, rather than simply the mean. The fact that we have universe data means that we can look at the impact of the whole distribution of worker quality in the firm. We calculate the quality of workers at the 25th, 50th, and 75th percentiles of the human capital distribution at each establishment, and include these as independent variables in the regression (following the approach taken in Abowd et al. [2]). The results in Table 5 also support the finding of complementarity in production, though only rather weakly in Florida. We revert to using mean worker quality for the rest of the paper.

Second, we check whether differences in industrial structure are driving these findings of a significant complementarity between worker and firm quality. We run the analysis separately by major industry, and report the results in Table 6. The California results indicate that the interaction is significantly positive in 4 out of 7 industries, and significantly negative in none. In Florida,

Table 5
Firm-level productivity regressions

| | California | Florida |
|---|---|---|
| (1) 25th %tile of firm's worker quality distribution | 0.060 | 0.102 |
| | (32.67)** | (6.04)** |
| (2) 50th %tile of firm's worker quality distribution | 0.142 | 0.063 |
| | (35.07)** | (16.14)** |
| (3) 75th %tile of firm's worker quality distribution | 0.164 | 0.111 |
| | (8.05)** | (16.53)** |
| (4) Firm mark up ($\psi$) | 0.631 | 0.571 |
| | (188.57)** | (95.02)** |
| Interaction between (1) and (4) | −0.016 | 0.017 |
| | (2.79)** | (1.90) |
| Interaction between (2) and (4) | 0.035 | 0.026 |
| | (3.29)** | (1.63) |
| Interaction between (3) and (4) | 0.024 | 0.024 |
| | (3.88)** | (2.50)* |
| Constant | 4.832 | 4.739 |
| | (670.70)** | (457.08)** |
| Observations | 396020 | 150756 |
| $R$-squared | 0.29 | 0.24 |

*Notes*. Unit is a firm. In addition the specification includes controls for size and industry of firm (not reported in table).
* Significant at 5%.
** Significant at 1%.

Table 6
Interaction terms by industry

| | California | Florida |
|---|---|---|
| Construction | −0.003 | 0.065** |
| Manufacturing | 0.012 | 0.022** |
| Transportation & Utilities | 0.065** | 0.093** |
| Wholesale Trade | 0.007 | 0.014 |
| Retail Trade | 0.148** | 0.108** |
| FIRE | 0.033** | 0.046** |
| Services | 0.044** | 0.047** |

*Notes*. Unit is a firm. Dependent variable is the log of labor productivity. Coefficient on the interaction between mean worker quality and firm mark up from firm-level productivity regressions. In addition the specification by industry includes controls for mean worker quality, firm mark up, firm size and a constant (not reported in table).
* Significant at 5%.
** Significant at 1%.

there is a positive effect in all but wholesale trade. Comparing across industries, the interaction is greater in services than in manufacturing, largest in retail trade, transportation and utilities, and educational services (in Florida).

These results support the existence of widespread and strong complementarity in productivity. This complementarity provides the incentive for firms and workers to match assortatively.

### *4.2. Matching across space*

We start with a county level analysis and compute the empirical correlation between a firm's mark-up and its mean worker quality across all the firms in each county. This correlation is plotted for all counties against their employment density in Fig. 1 (California) and Fig. 2 (Florida). The plots reveal a clear positive relationship between the two and regression results confirm the visual impression. Table 7 provides county level regressions of the matching correlation on density. The dependent variable is the matching correlation, and we also control for the mean level of worker and firm mark-up. The regressions show a positive and statistically significant effect of (employment) density on the matching correlation in both California and Florida. Of the other variables, worker quality seems to matter in California, but not in Florida.

We can tie down the relationship between PAM and density more precisely by using all the individual data. We estimate the simple linear characterization of the joint density function, given in Eq. (2). The key issue for this analysis is that as density increases, the correlation between $\theta$ and $\psi$ increases, so we expect the interaction between them in (2) to be positive. An individual worker's quality is regressed on her matched firm mark-up, local density, and an interaction of quality and density. The results in Table 8 show very clearly that the interaction term is positive and significant for both states: 0.020 in California, and 0.045 in Florida.

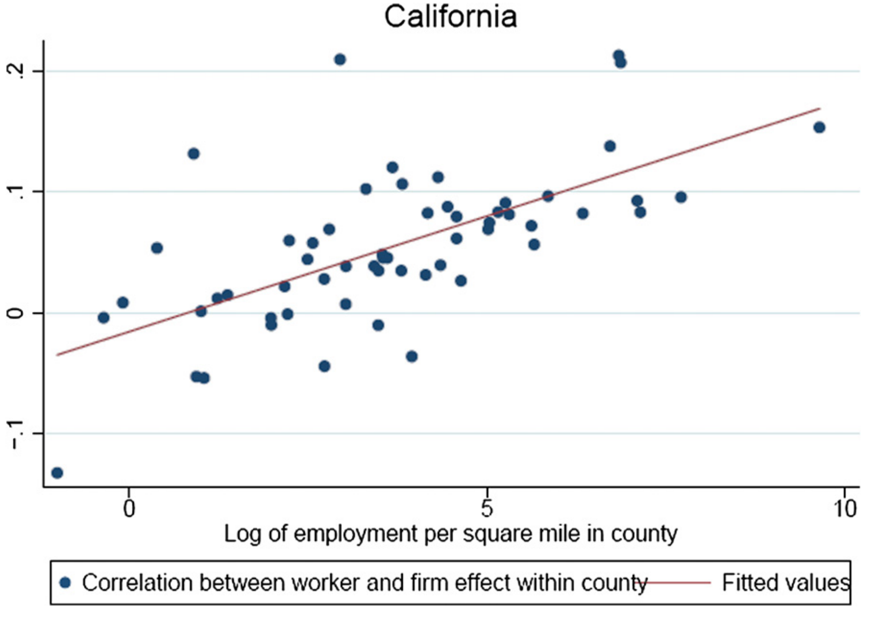


Fig. 1. Matching and Density in California.

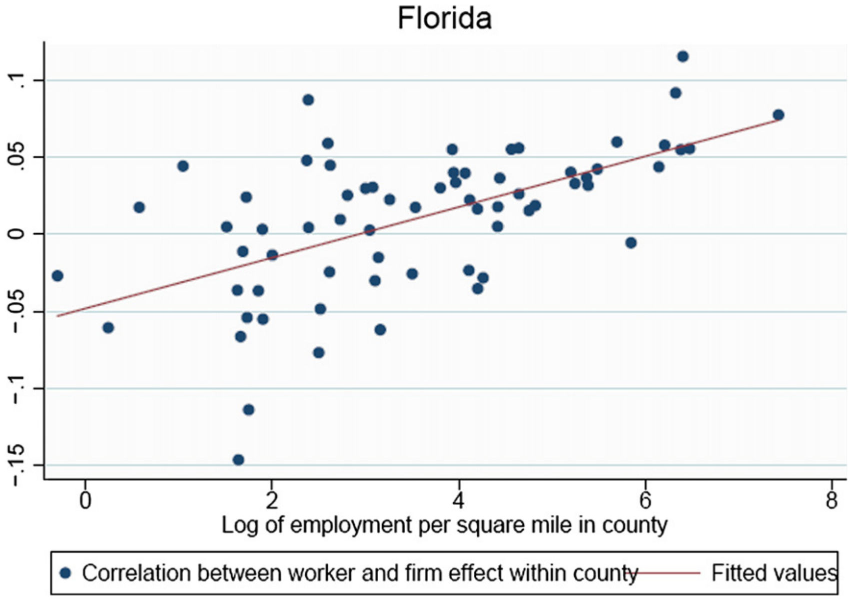


Fig. 2. Matching and Density in Florida.

Table 7
Matching regressions

| | California | Florida |
|---|---|---|
| County mean worker quality | −0.322 | −0.099 |
| | (2.66)* | (0.79) |
| County mean firm mark-up | 0.168 | −0.012 |
| | (1.79) | (0.19) |
| Log of emp./sq. mile | 0.020 | 0.019 |
| | (4.55)** | (5.35)** |
| Constant | 0.008 | −0.054 |
| | (0.27) | (3.51)** |
| Observations | 58 | 67 |
| $R$-squared | 0.47 | 0.38 |

Unit is a county. Dependent variable is matching correlation, corr($\theta$, $\psi$). Absolute value of $t$-statistics in parentheses.
* Significant at 5%.
** Significant at 1%.

One possibility we need to discount is that certain types of industries might locate in dense areas to take advantage of knowledge spillovers.[6] If such industries also have a tendency to hire high quality workers, the previous results might only be picking up differences in industry mix. We examine this possibility by repeating the exercises reported in Tables 7 and 8 by industry. The results are reported in Tables 9 and 10. The evidence in Table 9 shows that the correlation between worker and firm quality increases with density in every case by two, with the only

[6] We are grateful to an anonymous referee for pointing this out.

Table 8
Matching regressions

| | California | Florida |
|---|---|---|
| (1) Firm effect | 0.103 | −0.131 |
| | (82.94)** | (34.72)** |
| (2) Log of employment/sq. mile | 0.030 | 0.032 |
| | (353.90)** | (132.34)** |
| Interaction between (1) and (2) | 0.020 | 0.045 |
| | (101.31)** | (69.59)** |
| Constant | −0.107 | −0.075 |
| | (196.51)** | (52.76)** |
| Observations | 17,049,560 | 6,233,180 |
| $R$-squared | 0.02 | 0.01 |

Unit is a worker. Dependent variable is estimated worker quality ($\theta$).
* Significant at 5%.
** Significant at 1%.

Table 9
Matching regressions by industry group

| | Construction | Manufacturing | Transportation and utilities | Wholesale trade | Retail trade | FIRE | Services |
|---|---|---|---|---|---|---|---|
| | | | CALIFORNIA | | | | |
| Mean worker effect in | −0.064 | 0.239 | −0.199 | 0.254 | 0.043 | 0.089 | −0.107 |
| industry and county | (0.42) | (1.97) | (1.59) | (1.96) | (0.21) | (0.77) | (1.06) |
| Mean firm effect in | −0.171 | 0.064 | 0.232 | 0.021 | −0.764 | 0.389 | 0.293 |
| industry and county | (2.65)* | (0.77) | (2.40)* | (0.16) | (7.08)** | (3.37)** | (4.43)** |
| Log of Emp. per sq. mile | 0.019 | 0.010 | 0.002 | 0.013 | 0.015 | −0.014 | 0.022 |
| in industry and county | (4.42)** | (1.74) | (0.29) | (1.75) | (1.68) | (2.16)* | (4.70)** |
| Constant | −0.022 | 0.037 | 0.036 | 0.023 | −0.237 | 0.087 | 0.058 |
| | (1.53) | (2.80)** | (2.99)** | (1.69) | (5.85)** | (5.35)** | (2.20)* |
| Observations | 58 | 57 | 57 | 57 | 58 | 57 | 58 |
| $R$-squared | 0.27 | 0.29 | 0.14 | 0.35 | 0.50 | 0.24 | 0.73 |
| | | | FLORIDA | | | | |
| Mean worker effect in | 0.314 | −0.105 | 0.403 | −0.051 | −0.011 | 0.302 | 0.064 |
| industry and county | (2.77)** | (1.22) | (2.81)** | (0.35) | (0.06) | (1.92) | (0.48) |
| Mean firm effect in | 0.044 | 0.062 | −0.046 | 0.076 | −0.067 | −0.284 | 0.096 |
| industry and county | (0.53) | (1.03) | (0.47) | (0.75) | (0.31) | (2.01)* | (1.27) |
| Log of Emp. per sq. mile | 0.000 | 0.016 | −0.013 | 0.030 | 0.004 | 0.025 | 0.016 |
| in industry and county | (0.05) | (2.46)* | (1.10) | (2.91)** | (0.60) | (2.46)* | (3.37)** |
| Constant | −0.049 | −0.026 | 0.014 | −0.017 | −0.056 | −0.005 | 0.001 |
| | (3.38)** | (2.37)* | (0.61) | (0.79) | (0.91) | (0.13) | (0.08) |
| Observations | 67 | 67 | 67 | 66 | 67 | 65 | 67 |
| $R$-squared | 0.13 | 0.19 | 0.11 | 0.20 | 0.01 | 0.16 | 0.20 |

Unit is county. Dependent variable is matching correlation, corr($\theta$, $\psi$). Absolute value of $t$-statistics in parentheses.
* Significant at 5%.
** Significant at 1%.

significant exception being the financial services industry in California. Similarly in Table 10, the interaction term is positive on an industry by industry basis for every industry but transportation in California.

Table 10
Matching regressions by major industry group

| | Construction | Manufacturing | Transportation and utilities | Wholesale trade | Retail trade | FIRE | Services |
|---|---|---|---|---|---|---|---|
| | | | CALIFORNIA | | | | |
| (1) Firm effect | 0.007 | −0.330 | 0.194 | 0.270 | 0.031 | −0.105 | −0.048 |
| | (0.89) | (52.24)** | (40.93)** | (27.61)** | (10.63)** | (20.19)** | (20.40)** |
| (2) Log of emp./sq. | 0.002 | −0.008 | 0.023 | 0.015 | 0.016 | 0.039 | 0.031 |
| mile | (5.08)** | (24.36)** | (79.67)** | (32.35)** | (83.19)** | (117.03)** | (208.84)** |
| Interaction between | 0.013 | 0.117 | −0.018 | −0.002 | 0.008 | 0.055 | 0.048 |
| (1) and (2) | (10.06)** | (119.04)** | (25.19)** | (1.25) | (18.96)** | (71.05)** | (138.72)** |
| Constant | 0.062 | 0.011 | −0.050 | 0.006 | 0.012 | −0.078 | −0.097 |
| | (21.93)** | (5.23)** | (26.22)** | (1.88) | (9.47)** | (34.42)** | (98.14)** |
| Observations | 1,220,146 | 3,114,370 | 2,772,863 | 1,440,304 | 11,700,000 | 2,629,290 | 18,200,000 |
| $R$-squared | 0.00 | 0.04 | 0.00 | 0.01 | 0.00 | 0.03 | 0.02 |
| | | | FLORIDA | | | | |
| (1) Firm effect | −0.074 | −0.255 | −0.267 | −0.074 | −0.167 | 0.003 | −0.131 |
| | (3.70)** | (12.42)** | (13.07)** | (3.35)** | (23.06)** | (0.22) | (18.37)** |
| (2) Log of emp./sq. | −0.001 | −0.009 | 0.056 | 0.031 | 0.032 | 0.062 | 0.028 |
| mile | (1.15) | (7.96)** | (41.77)** | (21.55)** | (70.54)** | (61.61)** | (57.86)** |
| Interaction between | 0.006 | 0.072 | 0.068 | 0.034 | 0.021 | 0.039 | 0.052 |
| (1) and (2) | (1.61) | (20.55)** | (20.07)** | (9.06)** | (16.40)** | (14.36)** | (43.57)** |
| Constant | 0.065 | 0.031 | −0.226 | −0.034 | −0.069 | −0.152 | −0.031 |
| | (9.38)** | (4.91)** | (28.01)** | (4.03)** | (26.14)** | (25.74)** | (10.89)** |
| Observations | 492,981 | 574,529 | 555,623 | 510,714 | 5,375,844 | 917,524 | 3,865,822 |
| $R$-squared | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 |

Unit is a worker. Dependent variable is estimated worker quality ($\theta$). Absolute value of $t$-statistics in parentheses.
* Significant at 5%.
** Significant at 1%.

### *4.3. How much does density matter?*

The results in the previous sections have established the preconditions for assortative matching to contribute to the urban productivity premium. In this section we calibrate the effects across areas with different densities by examining differences across firms in two counties with different employment density levels—low density and high density. We use Eq. (3), combining the estimated production function and the expected worker quality from the matching outcome. We compute a given firm's expected productivity from Eq. (3) for different values of $\Psi$ and $\delta$. The results are in Table 11. The first panel of the table reviews the facts for our two area definitions. It shows that in California the mean log productivity of firms is about 0.09 higher in high density areas than low density areas; firm mark-ups are about 0.21 higher; and mean worker quality is about 0.12 higher. In Florida the mean log productivity of firms is 0.18 higher in dense areas; firm mark-ups 0.11 higher; and mean worker quality is about 0.09 higher. In the second panel of table, we see that the fitted difference for an average firm (firm mark-up of zero) is around 0.07 in California and 0.05 in Florida in favor of urban areas. As compared to the overall urban premium in the first panel it is clear that this approach predicts a substantial part of the urban productivity premium in both states. The importance varies depending on which part of the firm fixed effect distribution is fitted to Eq. (3)—ranging from 0.09 for firms one standard deviation above the mean to 0.07 for firms one standard deviation below the mean firm fixed effect in California and

Table 11a
Calibrating productivity differences—California

| | Employment density | | Difference |
|---|---|---|---|
| | Low density | High density | |
| Number of counties | 29 | 22 | |
| Mean density | 3.07 | 7.55 | 4.45 |
| Mean actual productivity | 4.43 | 4.52 | 0.09 |
| Mean $\Psi$ | −0.24 | −0.03 | 0.21 |
| Mean $\theta$ | −0.05 | 0.07 | 0.12 |
| | Difference in fitted marginal productivity | | |
| | | At $\Psi = 0$ (mean): | 0.07 |
| | | At $\Psi = 0.4$ (mean + 1 Standard deviation): | 0.09 |
| | | At $\Psi = 0.4$ (mean − 1 Standard deviation): | 0.05 |

*Notes.* Employment density is log employment per sq. mile. We define a low density county as one where this measure falls below 3.58 (the median across counties in California); high density as one where it falls above 6.84 (the 90th percentile across counties in California).

Table 11b
Calibrating productivity differences—Florida

| | Employment density | | Difference |
|---|---|---|---|
| | Low density | High density | |
| Number of counties | 33 | 6 | |
| Mean density | 2.58 | 6.55 | 4.45 |
| Mean actual productivity | 4.37 | 4.55 | 0.18 |
| Mean $\Psi$ | −0.17 | −0.06 | 0.11 |
| Mean $\theta$ | −0.05 | 0.04 | 0.09 |
| | Difference in fitted marginal productivity | | |
| | | At $\Psi = 0$ (mean): | 0.05 |
| | | At $\Psi = 0.4$ (mean + 1 Standard deviation): | 0.09 |
| | | At $\Psi = 0.4$ (mean − 1 Standard deviation): | 0.02 |

*Notes.* Employment density is log employment per sq. mile. We define a low density county as one where this measure falls below 3.50 (the median across counties in Florida); high density as one where it falls above 6.17 (the 90th percentile across counties in Florida).

from 0.09 to 0.02 in Florida. However, this exercise includes within the role of location the difference in mean worker quality between urban and rural areas. We pursue a different approach to isolate the pure direct role of matching.

We use the distributions of worker and firm quality and re-match them in different ways within each county to get at the contribution of matching. We then compute productivity for the counter-factual matched worker-firm pairs and compare with actual productivity. Throughout, the number of job slots in each county is held fixed to maintain area differences in density. This re-matching within counties is run in two contexts. First, we generate a random allocation of firms and workers across counties to simulate a situation without endogenous relocation of firms and workers in response to differential returns. Second, we use the actual allocation of firms and workers across counties, thus including the higher mean values in urban counties.

The re-matching within counties works as follows. We assign a firm mark-up to each job slot in the county and match a worker quality to that job according to three different matching regimes: random matching, perfect positive assortative matching, and actual matching. The lat-

Table 12
Decomposition of productivity effects across Californian Counties

| Mean of Log Productivity in: | Rural counties | Urban counties | Urban–Rural |
|---|---|---|---|
| Random sorting across counties | | | |
| Within counties: | | | |
| Random matching | 0.000 | 0.000 | 0.000 |
| Perfect positive assortative matching | 0.031 | 0.031 | 0.000 |
| Actual matching | −0.001 | 0.015 | 0.016 |
| Actual sorting across counties | | | |
| Within counties: | | | |
| Random matching | −0.142 | 0.147 | 0.289 |
| Actual matching | −0.109 | 0.194 | 0.303 |

*Notes*. The estimates are based on 100 boot-strapped samples. "Rural Counties" are defined as Counties with log of employment per square mile in the bottom 50th percentile. "Urban Counties" are defined as Counties with log of employment per square mile in the top 90th percentile.

ter is based on the actual correlation between worker and firm quality in each county. Finally, having created new worker-firm pairs, we use the estimated productivity function to calculate productivity for each pair. In each case, we repeat the probabilistic match of workers and firms one hundred times, and present the mean outcome.

The results are presented in Table 12. The first block refers to outcomes with the random allocation of workers and firms across counties, and the second block to the actual allocation across counties. Taking the top three rows, it is clear that with complementarity in the production function positive assortative matching produces higher productivity than the other matching regimes. Holding worker and firm fixed effects constant, the difference between positive and random matching is 0.031 log productivity points in urban counties. Thus patterns of worker-firm matching matter. Because of the random sorting in this part of the table, rural and urban areas have (in expectation) the same worker and firm quality, and so this difference between matching regimes is the same for both urban and rural areas. In row 3 we allow for the difference in actual matching patterns across rural/urban areas, reflecting the greater degree of positive assortative matching in cities. Productivity in rural areas is about the same as with random matching (0.000 random and −0.001 with actual matching). In urban areas, it is about half way between random and perfect positive assortative matching, (0.000 random, and 0.015 with actual matching). Thus, the direct productivity effect of differences in matching between urban and rural areas is 0.017, in this context with randomly assigned factors across counties.

The second block of Table 12 relates to the actual distribution of worker and firm qualities between urban and rural areas. We see that the sorting of higher quality workers and firms into urban areas is important for productivity differences and dominates the direct effect of matching. The impact of matching can be seen by comparing the random matching row and the actual matching row. Productivity increases 0.047 log points (from 0.147 to 0.194) in urban areas once we apply the actual matching patterns in the data, and 0.033 (from −0.142 to −0.109) in rural areas. The difference in these differences, the direct contribution of matching to the urban productivity premium given the actual sorting of workers and firms, is 0.016, essentially the same as the 0.017 figure based on random allocation. Overall, these results show that pure differences in matching patterns are quantitatively important for productivity. Unsurprisingly, they also show that the endogenous re-allocation of high quality workers and firms to urban areas is more important.

## 5. Conclusions

In this paper we address the puzzle of the urban productivity premium. While it is clear that it is substantial, the literature is unclear what it derives from. We take one of the main contenders and test it using a new micro data set. Our results suggest that assortative matching in dense urban labor markets plus complementarities in production play an important role in generating high productivity in cities. Using a unique data set, we show that the degree of matching of firm and worker quality does vary with labor market density, and we establish that there is evidence of complementarity in production. Putting these together, we show that this process contributes to the urban premium.

The paper also illustrates the insights of the search and matching approach to labor markets, and the power that the new emerging data sets offer in addressing long-standing questions. There are other related issues that we can tackle: for example, segregation and networks in cities, earnings and local labor markets, residential and commuting patterns. Complementarity in production plus assortative matching also imply greater wage inequality in denser labor markets. We leave all these topics to future work.

## Acknowledgments

This document reports the results of research and analysis undertaken by the US Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications, and is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The data used are a part of the US Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging (R01-AG18854-01), and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the US Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The US Census Bureau is preparing to support external researchers' use of these data; please contact US Census Bureau, LEHD Program, FB 2138-3, 4700 Silver Hill Rd., Suitland, MD 20233, USA. Thanks are due to Harry Holzer, and seminar participants at the University of Bristol for helpful comments.

## References

[1] J. Abowd, P. Lengermann, P. McKinney, K. McKinney, The measurement of human capital in the US economy, Technical working paper, LEHD, US Bureau of the Census, Washington, DC, 2003, March.

[2] J. Abowd, J. Haltiwanger, R. Jarmin, J. Lane, P. Lengermann, K. McCue, K. McKinney, K. Sandusky, Relationship between human capital, productivity and market value: Building up from microeconomic evidence, in: C. Corrado, J. Haltiwanger, D. Sichel (Eds.), Measuring Capital in the New Economy, Univ. of Chicago Press, Chicago, 2006, pp. 153–198.

[3] J. Abowd, J. Haltiwanger, J. Lane, Integrated longitudinal employer–employee data for the United States, American Economic Review 94 (2004) 224–229.

[4] F. Andersson, H. Holzer, J. Lane, Moving Up Or Moving On Workers, Firms and Advancement in the Low-Wage Labor Market, Russell Sage Foundation, New York, 2005.

[5] J.R. Baumgartner, Physicians' services and the division of labor across local markets, Journal of Political Economy 96 (1988) 948–982.

[6] G.S. Becker, A theory of marriage: Part I, Journal of Political Economy 81 (1973) 813–846.

[7] K. Burdett, M.G. Coles, Long-term partnership formation: Marriage and employment, Economic Journal 109 (1999) F307–F334.
[8] P. Combes, G. Duranton, L. Gobillon, Spatial wage disparities: Sorting matters! Discussion paper 4240, CEPR, February, 2004.
[9] A. Delacroix, Heterogeneous matching with transferable utility: Two labor market applications, International Economic Review 44 (2003) 313–342.
[10] G. Ellison, E. Glaeser, The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review 89 (1999) 311–316.
[11] J.C. Haltiwanger, J.I. Lane, J.R. Spletzer, Wages, productivity and the dynamic interaction of businesses and workers, Labour Economics, 2006, in press. Available online 14 November 2005.
[12] M. Kremer, E. Maskin, Wage inequality and segregation by skill. Quarterly Journal of Economics, in press.
[13] E. Moretti, Human capital externalities, in: J.V. Henderson, J.F. Thisse (Eds.), Handbook of Regional and Urban Economics, North-Holland, Amsterdam, 2004, pp. 2243–2291.
[14] S. Rosenthal, W. Strange, Evidence on the nature and sources of agglomeration economies, in: J.V. Henderson, J.F. Thisse (Eds.), Handbook of Regional and Urban Economics, North-Holland, Amsterdam, 2004, pp. 2119–2172.
[15] R. Shimer, L. Smith, Assortative matching and search, Econometrica 68 (2000) 371–398.
[16] W.C. Wheaton, M.J. Lewis, Urban wages and labor market agglomeration, Journal of Urban Economics 51 (2002) 542–562.