

**Center for Policy Research
Working Paper No. 56**

**GEOGRAPHY,
INDUSTRIAL ORGANIZATION,
AND AGGLOMERATION**

**Stuart S. Rosenthal
William C. Strange**

**Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu**

April 2003

\$5.00

Up-to-date information about CPR's research projects and other activities is available from our World Wide Web site at www-cpr.maxwell.syr.edu. All recent working papers and Policy Briefs can be read and/or printed from there as well.

CENTER FOR POLICY RESEARCH – Summer 2003

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director,
Budget and Administration

Douglas Holtz-Eakin
Professor of Economics
Associate Director, Center for Policy Research

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Dan Black..... Economics
Stacy Dickert-Conlin..... Economics
William Duncombe Public Administration
Gary Engelhardt Economics
Deborah Freund Public Administration
Madonna Harrington Meyer Sociology
Christine Himes..... Sociology
William C. Horrace Economics
Bernard Jump Public Administration
Duke Kao Economics
Eric Kingson Social Work
Thomas Kniesner Economics
Jeff Kubik Economics

Andrew London Sociology
Len Lopoo Public Administration
Jerry Miner Economics
John Moran Economics
Jan Ondrich Economics
John Palmer Public Administration
Lori Ploutz-Snyder.. Health and Physical Education
Jeff Racine Economics
Grant Reeher..... Political Science
Stuart Rosenthal..... Economics
Ross Rubinstein Public Administration
Michael Wasylenko..... Economics
Janet Wilmoth..... Sociology

GRADUATE ASSOCIATES

Anna Amirkhanyan..... Public Administration
Beth Ashby..... Economics
Dana Balter Public Administration
Gabby Chapman..... Economics
Yong Chen Economics
Christopher Cunningham Economics
Ana Dammert..... Economics
Tae Ho Eom..... Public Administration
Val Episcopo Sociology
Mike Eriksen Economics
Kate Farrar..... Public Administration
Garey Fuqua Public Administration
Jose Galdo..... Economics
Andrzej Grodner..... Economics
Glenda Gross Sociology
Anil Kumar Economics

Kristina Lambright Public Administration
Xiaoli Liang..... Economics
Liqun Liu Economics
Joseph Marchand..... Economics
Cristian Meghea Economics
Desmond Nation..... Public Administration
Emily Pas Economics
Seth Richards Public Administration
Cynthia Searcy Public Administration
Claudia Smith Economics
Sara Smits Sociology
Adriano Udani..... Public Administration
Wen Wang..... Public Administration
Bo Zhao..... Economics
Na Zhao..... Sociology

STAFF

Kelly Bogart Administrative Secretary
Martha Bonney..... Publications/Events Coordinator
Karen Cimilluca..... Librarian/Office Coordinator
Kim Desmond Administrative Secretary
Kati Foley Administrative Assistant, LIS

Kitty Nasto Administrative Secretary
Candi Patterson..... Computer Consultant
Mary Santy Administrative Secretary
Mindy Tanner Admin. Secretary/Receptionist

ABSTRACT

This paper makes two contributions to the empirical literature on agglomeration economies. First, the paper uses a unique and rich database in conjunction with mapping software to measure the geographic extent of agglomerative externalities. Previous papers have been forced to assume that agglomeration economies are club goods that operate at a metropolitan scale. Second, the paper tests for the existence of organizational agglomeration economies of the kind studied qualitatively by Saxenian (1994). This is a potentially important source of increasing returns that previous empirical work has not considered. Results indicate that localization economies attenuate rapidly and that industrial organization affects the benefits of agglomeration.

1. Introduction

The costs of cities can be seen in the skyscrapers, highways, and aqueducts that must be built to concentrate people in a small area. The benefits of cities – known as agglomeration economies – are less concrete but just as real. Marshall (1920) provides the first careful economic analysis of agglomeration economies, arguing that cities enhance productivity by allowing for labor market pooling, input sharing, and technological spillovers. An extensive empirical literature has considered agglomeration, including Sveikauskas (1975), Moomaw (1981), Henderson (1986), Nakamura (1985), Carlton (1983), Glaeser et al (1992), Henderson et al (1994), and Ciccone and Hall (1996) to name just a few.¹ These papers focus on whether the advantages of cities depend on city size or employment in a particular industry, whether agglomerative externalities are static or dynamic, and on the importance of urban diversity.

This paper addresses two important unanswered questions about agglomeration. First, what is the geographic scope of agglomerative externalities? In contrast to explicitly geographic theoretical work,² empirical work on agglomeration has been almost innocent of geography, and instead has implicitly modeled the city as a club. The economy is divided into geographic units, typically states, cities (more precisely, metropolitan statistical areas -- MSAs), or counties. Economic activity is then divided spatially according to the geographic partition, and the effects of the local economic environment on productivity are measured. This approach has the advantage of allowing the use of readily available aggregate data. However, it is somewhat unsatisfying since the benefits firms get from each other through labor market pooling, shared inputs, and technological spillovers are all likely to attenuate with distance. An important gap in our understanding of agglomeration economies, therefore, is that we do not know the geographic extent of agglomerative spillovers.

The second question that the paper addresses is how the organization of economic activity within a city affects the value of agglomeration. There is reason to believe that the productivity of a local economic environment does not depend just on the quantity of available inputs, but also on the way that such inputs are organized. In Saxenian's (1994) study of the computer industry, she points out that in the mid-1970s, both Boston (especially around Route 128) and the San Jose to Palo Alto corridor (Silicon Valley) were essentially equal in their positions as centers of electronics and high-technology. The next decade witnessed a movement offshore of semiconductor production, which hurt the Silicon Valley, and a shift away from minicomputers, which hurt Route 128. The Silicon Valley made a transition to software and other computer related industries that has been successful enough to make it one of the most productive economies on the planet. Route 128 did not make the transition as successfully.

There are two explanations for this divergence. One is that either location could have become dominant in software based on its local characteristics, but that the random hand of history selected the Silicon Valley as the industry core. The other explanation is that the locations did not have identical characteristics, and that the Silicon Valley offered a more productive environment. On the one hand, both locations had many of the characteristics that could be expected to attract high-technology employment including educated workforces and proximity to research universities. However, Saxenian argues that the key difference between the Silicon Valley and Route 128 is in their industrial systems. In her view (Saxenian, p. 7), a local industrial system has "three dimensions: local institutions and culture, industrial structure and corporate organization." Route 128 is presented by Saxenian as being relatively rigid and hierarchical, while the Silicon Valley is presented as being flexible and entrepreneurial. This certainly seems

to be the view of the industry. Saxenian quotes Jeffrey Kalb, an entrepreneurial refugee from Boston's Digital Electronics Corporation:

There's a fundamental difference in the nature of the industry between Route 128 and [the Silicon Valley]. Route 128 is organized into large companies that do their own thing...It's very difficult for a small company to survive in that environment...The Valley is very fast-moving and start-ups have to move fast. The whole culture of the Valley is one of change. We laugh about how often people change jobs. The joke is that you can change jobs and not change parking lots. There's a culture associated with that which says that moving is okay, that rapid change is the norm, that it's not considered negative on your resume...So you have this culture of rapid decisions, rapid changes, which is exactly the environment that you find yourself in as a startup.

Saxenian's analysis complements the work of Jacobs (1969) and Chinitz (1962), both of whom also suggest that urban efficiencies depend not just on numbers (i.e., city or industry size) but also on the nature of urban interactions. In the empirical literature, this issue has been considered obliquely in Glaeser et al (1992) and Henderson et al (1995) by including variables such as the number of employees per firm and the degree of urban specialization. However, these variables do not really capture the degree to which a location is blessed with a creative, entrepreneurial environment rather than an inflexible, hierarchical one. A further gap in the literature, therefore, concerns the impact of industrial organization on the value of agglomeration.

This paper addresses the geographic and organizational nature of agglomeration by examining the birth of new establishments and the employment levels that they choose. Specifically, we estimate the determinants of the number of births per square mile and their associated employment levels as functions of the economic environment when the location decisions were made. Because we focus on new establishments, we are able to treat the existing economic environment as exogenous. Because our data are available at the zipcode level, we are able to include metropolitan area fixed effects in our model. These fixed effects control for a wide range of metropolitan characteristics that might impact births. Such characteristics include

financial variables like local fiscal policies and wage rates, as well as natural advantages like climate, proximity to natural harbors, and mineral deposits. The fixed effects also control for the metropolitan area's "birth potential" (Carlton (1983)) arising from personnel let go by firms that fail, downsize, or relocate to another metropolitan area. Having controlled for all of these variables, a positive effect of existing employment on births or new-establishment employment is evidence of agglomeration economies.

We estimate our models using Dun and Bradstreet Marketplace data. This data set contains a wealth of information on over twelve million establishments in the United States. The version of the data available for our use includes, among other things, establishment location at the zipcode level, employment, sales, corporate structure, age of the establishment and more. Drawing on corporate status and establishment size variables enables us to evaluate the influence of industrial organization on the benefits of agglomeration. Drawing on the geographic detail in the data in conjunction with mapping software enables us to evaluate the geographic scope of the benefits of agglomeration.

The paper's most important finding is that agglomeration economies attenuate with distance. The initial attenuation is rapid, with the effect of own-industry employment in the first mile up to 10 to 1000 times larger than the effect two to five miles away. Beyond five miles attenuation is much less pronounced. This pattern is consistent with both theoretical models of the internal structure of cities and stylized facts: moving away from a city center, land and house rents, building heights, and population density all decline rapidly at first and slowly thereafter. These findings suggest that agglomeration should ideally be studied at a much more refined geographic level than has been the norm.

The paper also establishes that industrial organization affects the benefits of agglomeration. The marginal effect of an employee at a small establishment is greater than that generated by an employee at a large establishment. This result is broadly consistent with arguments by Saxenian that a more competitive and entrepreneurial environment enhances growth. It is also consistent with Audretsch et al (2000), who consider the productivity of small firms. In contrast, we obtain mixed results on corporate structure. Specifically, we do not find consistent evidence as to whether an employee at a subsidiary has a different effect on nearby firms than does an employee at a nonsubsidiary establishment.

The paper is organized as follows. Section 2 presents a simple empirical model of births and of new establishment employment. Section 3 discusses our data and presents summary measures. Section 4 presents the results, and Section 5 concludes.

2. Model

2.1 Estimating births and new-establishment employment

If agglomeration economies are present, then births will occur near concentrations of existing employment, all else equal. If agglomeration economies are absent, then births will tend to disperse. Consequently, our approach to estimating agglomeration economies is to focus on births and their agglomeration is taken as evidence of agglomeration economies.

Normalizing the price of output to one, an establishment generates profit equal to $\pi(y) = a(y)f(x) - c(x)$, where $a(y)$ shifts the production function $f(x)$, y is a vector of local characteristics, the components of which will be clarified below, and x is a vector of factor inputs that cost $c(x)$. Input quantities will be chosen to maximize profits by satisfying the usual first order conditions. Employment (n), for example, is chosen such that $a(y)\partial f(x)/\partial n - \partial c(x)/\partial n = 0$.

An establishment will be born when it is possible to earn nonnegative profits with all inputs chosen at their profit-maximizing levels. Establishments are heterogeneous in their potential profitability. We express such heterogeneity by rewriting the profit function as $\pi(y, \varepsilon) = \max_x a(y)f(x)(1 + \varepsilon) - c(x)$. We suppose that ε is independent and identically distributed across establishments according to the cumulative distribution function $\Phi(\varepsilon)$. For any y , there is a critical level $\varepsilon^*(y)$ such that $\pi(y, \varepsilon^*(y)) = 0$ and $\pi(y, \varepsilon) > (<) 0$ as $\varepsilon > (<) \varepsilon^*(y)$. In this case, the probability that an establishment is created is $\Phi(\varepsilon^*(y))$.

We assume that new establishments are opened at locations chosen from among all of the zipcodes in the United States, $j = 1, \dots, J$. Moreover, to control for differences in zipcode size, both births and new-establishment employment are deflated by zipcode area and are interpreted as arrivals per square mile hereafter.³ Location and employment decisions are made at time $t-1$ taking the existing economic environment as given and are born one period later at time t . We suppose that the local characteristics of each zipcode, y_j , are partitioned into two parts, $y_{z,j}$ and $y_{m,j}$. The elements of y_z vary by zipcode while the elements of y_m vary by metropolitan area. Aggregating over establishments gives the per square mile number of births (B) and total new-establishment employment (N) in zipcode j , which we express as linear functions of y_z and y_m :

$$B_{j,t} = b_z y_{z,j,t-1} + b_m y_{m,j,t-1} + \varepsilon_{b,t}, \quad (1)$$

$$N_{j,t} = n_z y_{z,j,t-1} + n_m y_{m,j,t-1} + \varepsilon_{n,t}, \quad (2)$$

where ε_b and ε_n are error terms.

Any local characteristic that increases productivity will result in both more births and in more employment by the new establishments. Thus, key elements of $y_{z,j}$ include the spatial distribution of employment oriented around zipcode j ($j = 1, \dots, J$). For example, the level of employment within and outside the establishment's industry within one mile, two miles, etc. of

the zipcode centroid. These variables define the level of agglomeration associated with a given zipcode and can be measured with our data.

In contrast, some of the most important elements of y_m are difficult to measure given the wide range of city-specific variables that affect productivity. Note, however, that $b_{m,y_{m,j}}$ and $n_{m,y_{m,j}}$ from (1) and (2) are city-specific effects. Accordingly, rewriting equations (1) and (2) we obtain

$$B_{j,t} = b_{zz}y_{z,j,t-1} + \gamma_{m,b} + \varepsilon_{b,t}, \quad (3)$$

$$N_{j,t} = n_{zz}y_{z,j,t-1} + \gamma_{m,n} + \varepsilon_{n,t}, \quad (4)$$

where $\gamma_{m,b}$ and $\gamma_{m,n}$ (equal to $b_{m,y_{m,j}}$ and $n_{m,y_{m,j}}$) control for all attributes common to a metropolitan area that affect productivity. Such attributes include metropolitan area fiscal policies, quality of the workforce, and wage rates for different classes of labor, as well as natural advantages like climate, harbors, and proximity to important natural resources.

A more subtle metropolitan area attribute that is also captured by the fixed effects is what Carlton (1983) refers to as the “birth potential” of an area. When firms fail or relocate away from an area, they may let go workers for whom the costs of moving to an alternate city are high. For example, workers who have developed family or professional ties to their present metropolitan area may choose to establish new firms in their current city even though higher profits could be obtained elsewhere. The number of such displaced workers is likely to be greater in cities with large existing concentrations of firms. For that reason, births may occur in cities with large existing concentrations of employment for reasons unrelated to the benefits of agglomeration. Nevertheless, even in this case, it seems quite plausible that such individuals will still select locations *within* the city to maximize profit. As long as that condition is met, the fixed effects in

(3) and (4) control for metropolitan-level birth potential and the coefficients on y_z in (3) and (4) reflect the benefits of agglomeration.

2.2 Other approaches to measuring the benefits of agglomeration

Our approach to measuring the benefits of agglomeration is to look at the decisions of new establishments. There are, of course, other approaches. Specifically, the benefits of agglomeration have been measured using value added as a measure of economic productivity (e.g., Ciccone and Hall (1996)) and the growth of total employment in an industry (e.g., Glaeser et al (1992) and Henderson et al (1995)).

Studying value added requires data on the market value of both output and input quantities. Although value of output and labor quantities are feasible to obtain, capital stock measures are generally quite difficult to come by at the micro level, making the value added approach difficult to implement. Instead, studying growth of total employment has been much more common, but it presents different challenges. Data on total employment are often readily available and the analysis lends itself to linear regressions. However, existing employers are constrained by prior choices, most importantly the level and kind of capital of previously installed. Those fixed factors affect how the employer values the marginal worker, and consequently how it changes its employment level in response to a change in its environment. In principle, this difficulty can be overcome by looking at changes in total employment over a sufficiently long time frame so that there are no fixed factors and all establishments are effectively new. Even then, however, one still has to address a difficult endogeneity problem: not only is the growth of total employment in a given area sensitive to the composition of employment in the area (an agglomeration effect), but the reverse may hold as well.

Implementing this approach, therefore, ideally requires a long panel and effective instruments to control for endogenous variables.

Focusing on the birth of new establishments and their employment avoids the problems most often associated with the two approaches above. Data on capital inputs is not required, new establishments are unconstrained by previous decisions, and they make their location and employment decisions taking the existing economic environment as exogenously given. The principal drawback of focusing on births and new-establishment employment is that many locations do not receive any births in a given period which can lead to technical challenges on the econometric side. These challenges will be clarified later in the paper.

As will become apparent, our data are especially well suited to studying births and new-establishment employment. For that reason, we focus on births per square mile and new-establishment employment per square mile as our measures of the benefits of agglomeration. In effect, we ask: in which location and at what scale will new establishments choose to open?

3. Data and variables

3.1 The Database

Data for the analysis were drawn from the Dun & Bradstreet Marketplace database which provides a wealth of information on over 12 million establishments. Details of the data are provided in Appendix A. Data from the fourth quarter of 1997 are used to construct two alternative dependent variables, new establishments and their employment, where new establishments are those that are listed in the data as being one year or less in age as of 1997:4. Each of these variables was deflated by zipcode area (in square miles) and represents arrivals per

square mile.⁴ Data from the fourth quarter of 1996 are used to measure the “existing” level of employment upon which new establishments are assumed to have based their location decisions.

3.2 The Variables

For each industry we calculate the existing level of employment both within and outside the industry in question. To measure the geographic extent of agglomerative externalities we create a set of concentric ring variables for both types of employment. These variables are calculated as follows. First, employment in a given zipcode is treated as being uniformly distributed throughout the zipcode. Then, using mapping software, circles of radius r_i , $i = 1, \dots, 15$, are drawn around the geographic centroid of each zipcode in the United States. The level of own-industry employment contained within a given circle is then calculated by constructing a proportional (weighted) summation of the own-industry employment for those portions of the zipcodes intersected by the circle. For example, if a circle includes all of zipcode 1 and 10 percent of the area of zipcode 2, then employment in the circle is set equal to the employment in zipcode 1 plus 10 percent of the employment in zipcode 2.⁵ The same procedure is used to calculate the level of other-industry employment within each circle. Differencing employment levels for adjacent circles (by employment type) yields estimates of the levels of own- and other-industry employment within a given concentric ring. Thus, the 2-mile ring (r_2) reflects employment between the 1 and 2-mile circles, and so on out to 15 miles.⁶

The other variables in our model are calculated in obvious ways. The number of establishments per worker are calculated as in Glaeser et al (1992) to proxy for local competitiveness. This variable is calculated separately for own-industry employment and employment outside of the own industry. The diversity of economic activity is incorporated using

a Herfindahl index of employment by 2-digit SIC industries as in Henderson (1995).⁷ Both the competitiveness variables and the Herfindahl index are calculated only at the zipcode level as opposed to the concentric ring approach used for the agglomeration variables.

Finally, any other arguments of the cost function that vary regionally such as wage rates, the quality of the local labor force, or access to raw materials are also pertinent. As discussed in Section 2, these variables are controlled for using metropolitan area fixed effects. In addition, we allow for different fixed effects for non-metropolitan locations (rural areas) for each of the 50 states. In total, this yields 373 fixed effects in the model.⁸

3.3 The Industries

Three criteria were used in selecting industries to study. First, we selected industries whose output is consumed nationally or internationally. Second, we selected industries with substantial numbers of new establishments and consequently substantial new-establishment employment. Third, we selected industries that are important enough to have been the focus of other studies. Specifically, we estimate the determinants of new-establishment employment and births for six industries: Software (SIC 7371, 7372, 7373, and 7375)⁹, Food Processing (SIC 20), Apparel (SIC 23), Printing and Publishing (SIC 27), Fabricated Metals (SIC 34), and Industrial and Commercial Machinery (SIC 35). All of these industries meet the first two criteria. In addition, software has been studied by Saxenian (1994), while the two-digit manufacturing industries were considered by Nakamura (1985) and Henderson (1986).

The industries are a mix of traditional industries with established products and innovative industries where new products are important. Innovation in the software industry is widely known given the explosion of computer technologies and use. Apparel and to a lesser degree

food processing involve fashion and are, therefore, almost by definition also innovative. Additionally, the six industries studied here are a mix of heavy and light industries, with machinery and metals both in the former category.

3.4 Summary Statistics

Table 1 provides selected summary statistics for the sample. Apart from the specific details of each industry, it is important to emphasize two points when viewing Table 1. First, there are 39,068 zipcodes and 373 identified metropolitan areas and rural zones. Because of the large number of zipcodes, many of our estimates are quite precise. Because of the large number of metropolitan/rural fixed effects, it is hoped that all regional attributes that affect productivity are controlled for.

Second, it is important to recognize that many of the observations are censored. Although there are births in many zipcodes for each of the six industries (the uncensored observations), there are zero births in the majority of the zipcodes in each case (the censored observations). The large number of zeros requires nonlinear estimation. Because we have a large sample, and because births and new-establishment employment per square mile vary widely, we estimate equations (3) and (4) by Tobit for each of the industries.¹⁰ This raises a technical issue because imprecise estimation of the fixed effects in nonlinear models typically leads to inconsistent estimates of the slope coefficients [e.g. Chamberlain (1980, 1985), Hsiao (1986)]. In addition, Tobit models are more sensitive to distributional assumptions than are linear regressions. We have two principal responses to this issue.

First, bias resulting from noisy estimates of fixed effects in nonlinear models goes to zero as the number of observations per fixed effect becomes arbitrarily large. Since our sample has

over 100 zipcodes per fixed effect, inconsistency arising from noisy estimates of the fixed effects is hoped to be small. Second, our results are very robust. Appendix B presents results from a Probit fixed effect model that examines whether individual zipcodes have positive or zero births (Table B-2). To facilitate review of the attenuation pattern, the partial derivatives based on the Probit coefficients are reported.¹¹ The results are qualitatively similar to those from the Tobit estimation. Appendix B also presents results for each industry based on a linear (ordinary least squares) fixed effect specification in which all zipcodes with zero births are omitted (Table B-3).¹² That approach suffers of course from a potential sample selection problem since most of the zipcodes are thrown out of the analysis. On the other hand, for linear models, noisy estimates of fixed effects do not bias estimates of the slope coefficients. It is notable, therefore, that the results continue to be robust. These findings suggest that the key qualitative results in this paper are robust to any issues related to econometric specification.¹³

4. Results

4.1 Initial Results

Tables 2a and 2b present estimates for the Tobit fixed effect models using respectively the number of new establishments per square mile and new-establishment employment per square mile as dependent variables. Likelihood ratio test statistics presented at the bottom of the table soundly reject the hypothesis that the fixed effects are jointly equal to zero for any industry.¹⁴ This confirms the importance of across-city variation in local attributes. On the other hand, although our model identifies agglomeration effects based on *within* city variation in the data, our results are broadly consistent with previous work that was based on *between* city variation in the data.¹⁵ Specifically, Tables 2a and 2b show that own-industry competition

encourages births and new-establishment employment in every industry but one. In contrast, other industry competition has a negative effect in both models for every industry. In addition, the tables show that a decrease in the diversity of employment – as measured by an increase in the Herfindahl index – decreases births and leads to less new-establishment employment. These results are consistent with Glaeser et al (1992).¹⁶

Observe also, that in nearly all cases for both births and new-establishment employment, localization effects (own-industry employment) are more important than urbanization effects (other-industry employment) at the margin. In particular, for any given concentric ring of employment (e.g. the mile 1 ring), the coefficient on the localization employment variable is typically at least one order of magnitude larger than the coefficient on the corresponding urbanization employment variable. In addition, most of the localization coefficients are highly significant while most of the urbanization coefficients are not significant. These results are consistent with Henderson's (1986) findings for Brazil and the U.S. and those of Nakamura (1985) for Japan.¹⁷

Given the comparability of our results to those of other studies, it is interesting to briefly characterize the degree to which systematic variation in the propensity for births can be attributed to the MSA fixed effects versus our within-city measures of agglomeration. Accordingly, Tables 3a and 3b present the distribution of the MSA fixed effect values across zipcodes as well as the distributions of $b_z y_z$ and $n_z y_z$ across zipcodes. These latter two terms measure the overall effect of the within-city agglomeration variables for the births and new-employment models, respectively (see Equations (3) and (4)), and are referred to here as the agglomeration effect. In nearly every case, the standard deviation of the agglomeration effect is much larger than for the city fixed effects, causing the range of values from the minimum to the maximum to be correspondingly

larger as well. That pattern indicates that our within city agglomeration variables drive a greater share of the systematic variation in birth propensities than do the city-wide fixed effects, although both are clearly important.

4.2 The Geography of Agglomeration Economies

We turn now to our most important results. In principle, one could use an arbitrarily large number of concentric rings when assessing how quickly agglomeration economies attenuate. In practice, however, it is necessary to aggregate the geographic detail in order to maintain a parsimonious specification. After some experimentation, the spatial distribution of employment was aggregated into four concentric rings: employment within 1 mile of the zipcode centroid, between 1 mile to 5 miles, between 5 miles to 10 miles, and between 10 miles to 15 miles. Condensing the geographic effects to these four variables greatly facilitated both estimation and presentation without changing the qualitative nature of the geographic patterns.¹⁸

Returning to Tables 2a and 2b, the key geographic result is that for most cases, localization economies attenuate rapidly in the first few miles but slowly thereafter. The exceptions to this generalization are Food Products (SIC 20) where only the new firm employment model exhibits consistent attenuation, and Apparel (SIC 23) where only the births model exhibits consistent attenuation. In addition, localization effects are largely insignificant (and negative) for Printing and Publishing (SIC 27).

To gain a sense of the magnitude of these estimates, consider first the software industry for which the localization effects are among the most pronounced. For that industry, adding one hundred software workers to the 1-mile ring would generate 0.04 births and 1.117 additional employees at new establishments. Adding one hundred additional employees to the 2-5 mile ring

would result in 0.005 births and 0.08 employees, while adding one hundred additional employees to the 6-10 mile ring would lead to 0.004 births and 0.08 workers.

The attenuation pattern implied by these estimates is highlighted at the bottom of the table where we present the change per mile (CPM) in the localization effects for each industry. This is measured by the difference between each of the adjacent concentric ring coefficients divided by the number of miles between the mid-points corresponding to the two rings. Averaging across all six industries, for births (Table 2a), the ratio of CPM values in the 0.5 to 3 mile range relative to the 3 to 7.5 mile range, $CPM_{(0.5 \text{ to } 3)}/CPM_{(3 \text{ to } 7.5)}$, is 48.23 while the ratio of $CPM_{(3 \text{ to } 7.5)}/CPM_{(7.5 \text{ to } 12.5)}$ is 0.62. For new establishment employment (Table 2b), the analogous values are 8.52 and 2.56, respectively.¹⁹ These values indicate the dominant pattern: localization economies attenuate rapidly for most of the industries studied here. Moreover, this pattern is consistent with theoretical models of urban areas. It is also consistent with well-known stylized facts on urban form. For example, building heights and population density both decline rapidly at first when moving away from the center of economic activity but decline slowly in more distant suburbs.

In contrast to the clear geographic pattern for the localization coefficients, the geographic pattern for the urbanization coefficients is obscure in most of the models. That difference serves to highlight an important distinction between urbanization and localization economies. Among industries that benefit from information spillovers and the ability to share both intermediate inputs and specialized labor, localization effects are expected to be positive and to diminish monotonically with distance. Our results largely support this. On the other hand, urbanization effects could be of any sign and are idiosyncratic to the individual industry. This is because urbanization effects reflect the tradeoff between the benefits of locating near densely developed

areas versus congestion costs. To the extent that industries differ in the net benefits they derive from proximity to employment centers, some industries will prefer more densely developed areas while others will prefer more outlying locations, *ceteris paribus*. As a result, the geographic pattern of urbanization effects is expected to differ across industries and can be quite varied. Our results support that argument as well.²⁰

4.3 Industrial Organization and Agglomeration

As noted earlier, Saxenian (1994) defines the local industrial system as having three dimensions: culture and institutions, corporate organization, and industrial structure. Absent hard data on culture and institutions, we focus on the latter two aspects of the industrial system. We will address two questions. First, does an industrial structure dominated by small establishments provide a more productive environment than one dominated by large establishments? Second, does a corporate organization based on parent-subsidiary links affect productivity to a different degree relative to one dominated by independent establishments?

If small establishments were more open and innovative as might be inferred from Saxenian's (1994) comparison of the Silicon Valley and Route 128, then an additional worker at a small establishment would enhance the productivity of neighboring establishments more than an additional worker at a medium or large establishment. To test that idea, we re-estimated the models in Tables 2a and 2b with the localization variables divided into three types: employment at small establishments (fewer than 25 employees), employment at medium establishments (25 to 99 employees), and employment at large establishments (100 or more employees). We focus here on own-industry effects since Saxenian's analysis is fundamentally about localization: how does the openness of a given industry impact its productivity? Since localization effects generally

attenuate rapidly, we aggregated own-industry employment from the zipcode centroid out to 5 miles and omitted more distant rings of own-industry employment from the regression. That simplification enables us to avoid a proliferation of localization variables and serves to highlight the effect of establishment size. All other regressors in Table 2 including the fixed effects were retained in the model.

Table 4a reports results for the localization variables with all other coefficients suppressed to conserve space.²¹ A clear pattern emerges. For five of the six industries, employment at small establishments has a larger effect on births or new-establishment employment than does employment at medium establishments. Averaging across the six industries, the premium associated with employment at small versus medium sized establishments is roughly 90 percent for births and 60 percent for new establishment employment, respectively. In contrast, the localization effects of employment at medium sized establishments are similar to the effects of employment at large establishments.

These results have important implications for the study of localization. They suggest that efficiency arises not simply from the concentration of own-industry employment but also from the concentration of the right kind of own-industry employment. That finding is consistent with Saxenian's comparative systems analysis of the Silicon Valley and Route 128 and suggests that for software and other industries, small establishments make better neighbors.²²

In an analogous manner, because a subsidiary establishment is constrained by the rest of its corporation, it might be less flexible or innovative. Because a subsidiary may purchase its inputs or sell its outputs within the corporation, it may not be intimately involved with its neighbors. For both of these reasons, employment at a subsidiary could have a smaller effect on the productivity of nearby establishments than employment at a nonsubsidiary establishment.²³

On the other hand, access to a subsidiary might provide access to resources elsewhere in the subsidiary's parent corporation, including resources at other plants in other locations. In addition, subsidiaries may generate agglomeration economies through spin-offs. Jacobs (1969, p. 66), for example, notes that breakaways from Hughes Aircraft were important sources of entrepreneurship in the Los Angeles electronics industry after World War II.²⁴ For both of these reasons, employment at a subsidiary could have a larger agglomerative effect than employment at a nonsubsidiary establishment.

To evaluate the effect of subsidiary status on localization economies, the localization variables in Table 4a were replaced with two new localization variables: own-industry employment at subsidiaries of corporate parents and own-industry employment at non-subsidiaries. All other features of the models estimated in Table 4a were retained. Results for this new specification are presented in Table 4b where as before only the coefficients on the localization variables are presented to conserve space.

In contrast to our firm size results, no clear pattern emerges from the subsidiary/non-subsidiary model. There is evidence that non-subsidiary employees have a larger agglomeration effect for Fabricated Metals, but subsidiary employment is more important for Software. For the other industries the evidence is inconclusive, with a similar influence of subsidiary and non-subsidiary employment for many industries. These mixed findings may suggest that subsidiary status is too rough a measure to capture the influence of a hierarchical corporate structure. The findings may also suggest the presence of differing degrees of tradeoffs across industries with regard to the benefits and costs of subsidiary status.

5. Conclusion

This paper makes two important contributions to the empirical literature on agglomeration economies. First, we use a unique and rich database in conjunction with mapping software to measure the geographic extent and nature of agglomerative externalities. Previous papers have been forced to assume that agglomeration economies are club goods that operate at a metropolitan scale. Second, we test for the existence of organizational agglomeration economies of the kind studied qualitatively by Saxenian (1994). This is a potentially important source of increasing returns that previous empirical work has not considered.

Results from five of our six industries provide evidence that localization economies – agglomeration economies arising from spatial concentration within a given industry – attenuate rapidly over the first few miles and then attenuate much more slowly thereafter. While it is beyond the scope of this study to determine exactly which sources of agglomeration economies are responsible for this pattern, it is tempting to speculate. As discussed in the Introduction, three potential sources are information spillovers, labor market pooling, and shared inputs. Information spillovers that require frequent contact between workers may dissipate over a short distance as walking to a meeting place becomes difficult or as random encounters become rare. On the other hand, the benefits of labor market pooling and shared inputs might extend over a much greater distance since those benefits rely more on the ability of agents to conveniently drive from one location to another. Initial rapid attenuation of information spillovers followed by a more gradual attenuation of benefits from labor market pooling and shared inputs would produce an attenuation pattern consistent with that found in this paper. Systematic empirical support for that argument, however, is left for future research.

Our results also indicate that industrial structure and corporate organization affect the benefits that arise from clustering within a given industry. This finding is strongest with regard to the size distribution of establishments: own-industry employment at small establishments presents a much greater attraction to potential new arrivals than does a comparable level of own-industry employment at larger establishments. That pattern lends support to recent arguments that a more entrepreneurial industrial system promotes growth. In contrast, we obtain mixed results on the effects of the subsidiary versus non-subsidiary status of local establishments. On balance, therefore, the agglomerative effects of industrial structure and corporate organization is an issue that certainly warrants further study.

It is worth emphasizing that all of these results are robust to a variety of alternative specifications and estimation methods. Accordingly, future studies of agglomeration economies should be sensitive both to industrial organization and especially the micro geography of agglomeration.

References

- Audretsch, David B., George van Leeuwen, Bert Menkveld, and Roy Thurik, "Market Dynamics in the Netherlands: Competition Policy and the Response of Small Firms," 2000 (mimeo, Indiana University).
- Carlton, D. W., "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables," *Review of Economics and Statistics* 65 (1983), 440-449.
- Chamberlain, Gary, "Analysis of covariance with qualitative data," *Review of Economic Studies*, 47 (1980), 225-238.
- Chamberlain, Gary, "Panel Data" (Chapter 22, Vol 2) in Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, (New York: Elsevier Science Publishing Company, 1984).
- Chinitz, B.J., "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review* 51 (1961), 279-89.
- Ciccone, A. and R.E. Hall, "Productivity and the Density of Economic Activity," *American Economic Review* 86 (1996), 54-70.
- Ellison, G. and E. Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105 (1997), 889-927.
- Glaeser, E.L, H. D. Kallal, J. A. Scheinkman, and A. Shleifer, "Growth in Cities," *Journal of Political Economy* 100 (1992), 1126-1152.
- Helsley, R.W., "Knowledge and Production in the CBD," *Journal of Urban Economics* 28 (1990), 391-403.
- Henderson, J. V. (1986), "Efficiency of Resource Usage and City Size," *Journal of Urban Economics* 19 (1986), 47-70.
- Henderson, J.V. (1988), Urban Development: Theory, Fact, and Illusion, Oxford: Oxford University Press.
- Henderson, J. V., "Where Does an Industry Locate?" *Journal of Urban Economics* 35 (1994), 83-104.
- Henderson, J.V., A. Kuncoro and M. Turner, "Industrial Development in Cities." *Journal of Political Economy* 103 (1995), 1067-1085.
- Hoover. E. and R. Vernon (1959), Anatomy of a Metropolis, Cambridge: Harvard University Press.

Hsiao, Cheng (1986), Analysis of Panel Data, New York: Cambridge University Press.

Imai, H., "CBD Hypothesis and Economies of Agglomeration," *Journal of Economic Theory* 8 (1982), 275-299.

Jacobs, J. (1969), The Economy of Cities, New York: Vintage.

Jaffe, A. B. M. Trajtenberg, and R. Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* CVIII (1993), 577-598.

Krugman, P., "On The Number and Location of Cities," *European Economic Review* 37 (1993), 293-298.

Marshall, A. (1920), Principles of Economics, London: MacMillan.

Moomaw, R.L., "Productivity and City Size: A Critique of the Evidence," *Quarterly Journal of Economics* 96 (1981), 675-688.

Nakamura, R., "Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities," *Journal of Urban Economics* 17 (1985), 108-124.

Ogawa, H. and M. Fujita, "Equilibrium Land Use Patterns In A Nonmonocentric City," *Journal of Regional Science* 20 (1980), 455-475.

O'Hara, D.J., "Location Of Firms Within A Square Central Business District," *Journal of Political Economy* 85 (1977), 1189-1207.

Quigley, J. "Urban Diversity and Economic Growth," *Journal of Economic Perspectives* 12 (1998), 127-138.

Romer, P.M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94 (1986), 1002-1037.

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

Sveikauskas, L., "The Productivity Of Cities," *Quarterly Journal of Economics* 89 (1975), 393-413.

Vernon, R. (1962), Metropolis 1985, Cambridge, MA: Harvard University Press.

TABLE 1
SELECTED SUMMARY STATISTICS
All Values Calculated At The Zipcode Level Over 39,068 Zipcodes*

	Software: SIC 7371-7373, 7375				Food Products: SIC 20				Apparel: SIC 23			
	Uncensored Obs 6362				Uncensored Obs 1324				Uncensored Obs 1,639			
<i>Own Industry</i>	Mean	Std Dev	No. 0's**	Max	Mean	Std Dev	No. 0's**	Max	Mean	Std Dev	No. 0's**	Max
Births	0.38	1.26	32,706	39	0.04	0.23	37,744	10	0.06	0.63	38,429	107
New firm wrkrs	2.85	54.15	32,706	7,392	1.05	19.69	37,744	1,200	0.62	21.13	38,429	3,300
Firm/workers	0.10	0.19	26,198	1	0.06	0.15	26,355	1	0.07	0.19	28,212	1
Workers	30.40	310.09	26,198	39,764	44.40	240.56	26,355	14,261	23.39	192.16	28,212	25,555
Non-sub. wrkrs	26.55	290.32	26,245	39,727	38.32	211.47	26,520	14,225	21.45	180.51	28,285	24,175
Subsidiary wrkrs	3.85	65.21	37,966	4,066	6.08	78.04	37,850	5,120	1.94	34.88	38,605	2,700
Small firm wrkrs	8.76	32.71	26,400	1,033	3.90	11.06	28,114	376	3.54	47.09	29,583	8,101
Med. firm wrkrs	5.84	39.23	36,818	1,405	6.87	27.82	35,286	1,097	5.18	66.61	36,774	10,413
Large firm wrkrs	15.81	284.75	38,087	39,720	33.63	228.97	36,384	14,225	14.67	112.90	37,426	7,041
<i>Other Industry</i>												
Firm/workers	0.17	0.13	5	1	0.17	0.13	7	1	0.17	0.13	3	1
Workers	2,929.22	6,114.52	5	155,906	2,911.31	6,115.12	7	156,586	2,932.32	6,137.43	3	156,748
	Printing and Publishing: SIC 27				Fabricated Metal: SIC 34				Machinery: SIC 35			
	Uncensored Obs 5,006				Uncensored Obs 1,875				Uncensored Obs 3,280			
<i>Own Industry</i>	Mean	Std Dev	No. 0's**	Max	Mean	Std Dev	No. 0's**	Max	Mean	Std Dev	No. 0's**	Max
Births	0.20	0.64	34,062	13	0.06	0.27	37,193	6	0.12	0.48	35,788	16
New firm wrkrs	1.67	28.91	34,062	3,500	1.52	26.94	37,193	3,500	2.16	27.36	35,788	1,804
Firm/workers	0.11	0.19	20,805	1	0.07	0.17	24,606	1	0.09	0.19	20,869	1
Workers	49.50	264.70	20,805	14,981	48.50	223.77	24,606	12,799	76.15	422.86	20,869	26,401
Non-sub. wrkrs	41.84	212.72	20,871	12,697	41.61	192.91	24,755	8,962	64.83	386.22	20,990	26,401
Subsidiary wrkrs	7.66	105.31	37,242	8,447	6.89	87.33	37,172	12,424	11.32	106.45	36,471	8,125
Small firm wrkrs	13.46	40.08	21,233	2,519	7.38	22.95	25,989	862	12.44	38.14	21,762	1,293
Med. firm wrkrs	10.50	45.72	34,230	2,075	12.92	49.19	33,596	1,980	15.35	57.55	32,889	1,947
Large firm wrkrs	25.54	221.95	36,858	13,250	28.19	193.73	36,306	12,775	48.36	388.96	35,895	26,255
<i>Other Industry</i>												
Firm/workers	0.17	0.13	14	1	0.17	0.13	4	1	0.17	0.13	16	1
Workers	2,906.21	6,042.10	14	149,251	2,907.21	6,109.74	4	157,999	2,879.56	6,032.95	16	157,678

*Zipcode area (in square miles) and the 2-digit Herfindahl index are common to all industries. For zipcode area, the mean, standard deviation, and min and max are, respectively, 126.95, 486.09, 0.008, and 18,559. For the Herfindahl index the corresponding values are 0.23, 0.24, 0.029, and 1.

**No. 0's refers to the number of zipcodes for which the variable has a value of 0.

TABLE 2a
BIRTHS OF NEW ESTABLISHMENTS
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
Zipcode Area, Diversity, and Competition Effects						
<i>Zipcode Herfindahl Index</i>	-1.040E+01 (-36.369)	-3.634E+00 (-17.049)	-2.151E+01 (-17.551)	-6.036E+00 (-34.340)	-1.395E+00 (-20.190)	-2.265E+00 (-27.265)
<i>Zipcode firms per worker - other ind.</i>	-9.466E+00 (-22.272)	-3.349E+00 (-12.051)	-1.329E+01 (-8.838)	-5.149E+00 (-19.763)	-1.631E+00 (-16.241)	-2.160E+00 (-18.948)
<i>Zipcode firms per worker - own ind.</i>	1.473E+00 (13.151)	4.561E-01 (5.875)	2.936E+00 (7.969)	4.859E-01 (6.459)	5.230E-02 (1.924)	-1.694E-01 (-0.495)
Urbanization Effects: Other (Total - Own) Industry Employment In The ...						
0 to 1 Mile Ring	1.070E-06 (1.120)	1.060E-06 (2.195)	-2.000E-05 (-5.445)	7.000E-06 (7.529)	-1.050E-07 (-0.527)	5.010E-07 (2.235)
1 to 5 Mile Ring	1.600E-06 (6.308)	2.180E-07 (1.476)	6.040E-06 (5.113)	1.360E-06 (4.149)	-2.820E-08 (-0.494)	-2.260E-08 (-0.321)
5 to 10 Mile Ring	-2.570E-07 (-1.431)	1.350E-07 (1.243)	1.110E-06 (1.258)	5.460E-07 (2.304)	2.050E-08 (0.522)	7.960E-09 (0.174)
10 to 15 Mile Ring	3.270E-07 (2.094)	-4.560E-08 (-0.405)	1.780E-06 (-2.2040)	3.480E-07 (1.638)	3.610E-08 (0.939)	9.170E-09 (0.233)
Localization Effects: Own Industry Employment In The ...						
0 to 1 Mile Ring	3.843E-04 (7.446)	1.680E-05 (0.469)	8.054E-04 (11.033)	-5.950E-05 (-3.966)	6.150E-05 (4.145)	6.360E-05 (6.088)
1 to 5 Mile Ring	5.030E-05 (3.844)	5.690E-05 (4.830)	-8.770E-05 (-2.474)	-2.130E-06 (-0.340)	2.330E-05 (5.017)	1.540E-05 (4.818)
5 to 10 Mile Ring	3.380E-05 (3.669)	-1.160E-05 (-1.293)	-3.870E-06 (-0.139)	-8.170E-06 (-1.666)	2.150E-06 (0.679)	-1.620E-06 (-0.747)
10 to 15 Mile Ring	-1.470E-05 (-1.723)	5.700E-06 (0.643)	-6.100E-05 (-2.127)	-5.530E-06 (-1.190)	-1.170E-06 (-0.425)	6.520E-06 (3.483)
Average Change In Localization Effect Per Mile From ...*						
0.5 to 3 Miles	-1.34E-04	1.60E-05	-3.57E-04	2.29E-05	-1.53E-05	-1.93E-05
3 to 7.5 Miles	-3.67E-06	-1.52E-05	1.86E-05	-1.34E-06	-4.70E-06	-3.78E-06
7.5 to 12.5 Miles	-9.70E-06	3.46E-06	-1.14E-05	5.28E-07	-6.64E-07	1.63E-06
Summary Measures						
Std Error	2.59	1.02	6.50	1.49	0.41	0.58
Log-L	-20,178.23	-4,658.13	-8,273.64	-14,274.27	-4,570.31	-7,660.18
2(LogL - LogL _{NoFE})**	1,910.66	643.73	622.10	1,167.64	609.05	694.11
Uncensored	6,362	1,324	1,639	5,006	1,875	3,280
Total Obs	39,068	39,068	39,068	39,068	39,068	39,068
Fixed Effects	373	373	373	373	373	373

*Change per mile is computed by differencing the adjacent localization coefficients and dividing by the number of miles between the midpoints.

**The test statistic 2(LogL - LogL_{NoFE}) is distributed Chi-square with 372 degrees of freedom, where LogL_{NoFE} is the value of the log-likelihood function when the fixed effects are omitted but a constant is retained.

TABLE 2b
NEW-ESTABLISHMENT EMPLOYMENT
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
Zipcode Area, Diversity, and Competition Effects						
<i>Zipcode Herfindahl Index</i>	-2.208E+02 (-34.372)	-2.085E+02 (-16.419)	-9.332E+02 (-15.692)	-3.239E+02 (-31.236)	-1.189E+02 (-19.28)	-1.147E+02 (-26.216)
<i>Zipcode firms per worker - other ind.</i>	-2.204E+02 (-22.926)	-2.071E+02 (-12.37)	-6.910E+02 (-8.738)	-3.144E+02 (-20.267)	-1.435E+02 (-15.893)	-1.200E+02 (-19.577)
<i>Zipcode firms per worker - own ind.</i>	3.188E+01 (12.615)	2.275E+01 (4.700)	1.602E+02 (8.159)	2.718E+01 (6.094)	3.986E+00 (1.603)	-1.316E+00 (-0.716)
Urbanization Effects: Other (Total - Own) Industry Employment In The ...						
0 to 1 Mile Ring	-7.830E-05 (-3.606)	-7.760E-05 (-2.424)	-1.241E-04 (-0.723)	2.538E-04 (4.383)	-4.800E-05 (-2.344)	-1.500E-05 (-0.995)
1 to 5 Mile Ring	2.610E-05 (4.510)	1.670E-06 (0.177)	2.549E-04 (4.088)	3.630E-05 (1.807)	-1.400E-05 (-2.505)	-7.450E-06 (-1.777)
5 to 10 Mile Ring	-1.030E-05 (-2.52)	6.230E-06 0.930	4.810E-05 (1.016)	5.390E-05 (3.805)	-4.550E-06 (-1.203)	-1.790E-06 (-0.69)
10 to 15 Mile Ring	4.070E-06 (1.156)	2.120E-06 (0.3140)	8.220E-05 (-1.9030)	1.500E-05 (1.184)	4.510E-06 (1.264)	-2.690E-07 (-0.124)
Localization Effects: Own Industry Employment In The ...						
0 to 1 Mile Ring	1.171E-02 (10.385)	1.245E-02 6.272	4.034E-03 (1.061)	-1.346E-03 (-1.433)	6.613E-03 4.602	3.802E-03 (6.738)
1 to 5 Mile Ring	8.574E-04 (2.895)	2.945E-03 4.032	-5.261E-04 (-0.288)	-5.659E-04 (-1.458)	2.689E-03 6.368	6.897E-04 (3.973)
5 to 10 Mile Ring	8.545E-04 (4.110)	-5.531E-04 (-1.015)	-3.604E-04 (-0.242)	-9.037E-04 (-3.086)	4.750E-04 1.647	-9.560E-06 (-0.081)
10 to 15 Mile Ring	-1.215E-04 (-0.63)	6.180E-05 (0.115)	-2.819E-03 (-1.838)	-3.661E-04 (-1.32)	-3.765E-04 (1.480)	2.666E-04 (2.637)
Average Change In Localization Effect Per Mile From ...*						
0.5 to 3 Miles	-4.34E-03	-3.80E-03	-1.82E-03	3.12E-04	-1.57E-03	-1.24E-03
3 to 7.5 Miles	-6.44E-07	-7.77E-04	3.68E-05	-7.51E-05	-4.92E-04	-1.55E-04
7.5 to 12.5 Miles	-1.95E-04	1.23E-04	-4.92E-04	1.08E-04	-1.70E-04	5.52E-05
Summary Measures						
Std Error	58.17	60.76	346.65	87.44	36.26	30.78
Log-L	-39,825.85	-9,963.03	-14,809.64	-34,435.40	-12,710.56	-20,398.02
2(LogL - LogL _{NoFE})**	1,703.27	648.03	618.00	1,105.36	552.54	714.70
Uncensored	6,362	1,324	1,639	5,006	1,875	3,280
Total Obs	39,068	39,068	39,068	39,068	39,068	39,068
Fixed Effects	373	373	373	373	373	373

*Change per mile is computed by differencing the adjacent localization coefficients and dividing by the number of miles between the midpoints.

**The test statistic 2(LogL - LogL_{NoFE}) is distributed Chi-square with 372 degrees of freedom, where LogL_{NoFE} is the value of the log-likelihood function when the fixed effects are omitted but a constant is retained.

TABLE 3
CITY FIXED EFFECTS VERSUS AGGLOMERATION EFFECTS*

		Software SIC 7371-73 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
BIRTHS OF NEW ESTABLISHMENTS**							
Mean	Fixed Effect	-1.17	-1.22	-9.42	-0.79	-0.38	-0.33
	Agglom Effect	-3.73	-1.39	-6.73	-2.08	-0.58	-0.87
Std Dev	Fixed Effect	1.18	0.33	2.61	0.51	0.14	0.16
	Agglom Effect	3.09	1.10	6.08	1.76	0.43	0.66
Minimum	Fixed Effect	-4.47	-3.10	-20.26	-2.17	-1.13	-0.78
	Agglom Effect	-19.95	-7.23	-35.67	-11.16	-3.05	-4.46
Maximum	Fixed Effect	3.13	0.40	-2.70	0.83	0.04	0.29
	Agglom Effect	10.18	1.68	31.57	4.56	0.46	1.09
NEW-ESTABLISHMENT EMPLOYMENT**							
Mean	Fixed Effect	-26.69	-71.86	-529.64	-49.05	-34.59	-17.33
	Agglom Effect	-83.40	-83.53	-304.87	-119.34	-51.19	-46.67
Std Dev	Fixed Effect	25.76	19.74	140.35	29.29	12.12	8.27
	Agglom Effect	66.65	64.20	272.68	96.08	37.29	33.95
Minimum	Fixed Effect	-100.68	-180.80	-1116.64	-118.95	-98.65	-41.07
	Agglom Effect	-444.69	-430.64	-1641.85	-636.91	-276.63	-244.69
Maximum	Fixed Effect	45.84	23.46	-166.97	62.15	6.05	27.33
	Agglom Effect	288.45	127.85	604.45	160.18	40.04	59.26

*Agglomeration effects are calculated as the linear combination of the slope coefficients from Table 2 and their corresponding variables in the model. Fixed effects are the model's MSA and rural fixed effects as described in the text. Values in the table are the distribution of the fixed effects and agglomeration effects across the zipcodes in the sample

**Dependent variable is measured per square mile.

TABLE 4a*
LOCALIZATION EFFECTS CONTROLLING FOR ESTABLISHMENT SIZE
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
BIRTHS OF NEW ESTABLISHMENTS						
Employment at:						
Small firms (1-24 workers)	1.76E-03 (10.733)	1.29E-03 (4.155)	-5.74E-04 (-1.303)	4.15E-04 (4.365)	1.66E-04 (2.754)	2.80E-04 (5.544)
Medium firms (25-99 workers)	-6.58E-04 (-4.453)	9.58E-06 (0.083)	-1.92E-04 (-0.601)	-3.41E-04 (-4.164)	4.30E-05 (1.409)	-2.49E-05 (-0.674)
Large Firms (100+ workers)	4.84E-05 (3.502)	1.96E-05 (1.396)	6.09E-04 (3.589)	-5.33E-07 (-0.049)	3.31E-06 (0.504)	4.40E-06 (1.267)
NEW-ESTABLISHMENT EMPLOYMENT						
Employment at:						
Small firms (1-24 workers)	2.48E-02 (6.614)	3.76E-02 (1.957)	-6.85E-02 (-2.551)	1.81E-02 (3.163)	1.65E-02 (2.901)	1.23E-02 (4.463)
Medium firms (25-99 workers)	6.88E-04 (0.205)	1.52E-02 (2.162)	4.17E-02 (2.539)	-1.48E-02 (-3.017)	7.74E-04 (0.275)	6.53E-04 (0.326)
Large Firms (100+ workers)	7.97E-04 (2.519)	1.88E-03 (2.212)	-3.52E-03 (-0.405)	-1.68E-04 (-0.254)	2.35E-03 (4.161)	1.67E-04 (0.881)

*The estimates above are the coefficients on the localization variable for the indicated employment type having summed the localization concentric ring variables out to 5 miles and omitted the remaining localization concentric rings (miles 6 to 15). All other variables listed in Table 2 were included in the model: coefficients for those variables are not reported to conserve space.

TABLE 4b*
LOCALIZATION EFFECTS CONTROLLING FOR SUBSIDIARY STATUS
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
BIRTHS OF NEW ESTABLISHMENTS						
Employment at:						
Non-Subsidiaries	7.95E-05 (6.223)	5.07E-05 (3.237)	5.75E-05 (1.628)	6.36E-06 (0.063)	3.17E-05 (7.230)	2.01E-05 (6.116)
Subsidiaries	2.53E-04 (3.521)	4.73E-05 (1.080)	-8.84E-04 (-1.737)	-4.02E-05 (-2.072)	7.91E-06 (0.694)	2.02E-05 (1.665)
NEW-ESTABLISHMENT EMPLOYMENT						
Employment at:						
Non-Subsidiaries	1.36E-03 (4.653)	2.58E-03 (2.732)	1.39E-03 (0.760)	1.58E-04 (0.259)	3.53E-03 (8.888)	6.40E-04 (3.559)
Subsidiaries	1.22E-02 (7.585)	8.82E-03 (3.439)	-1.93E-02 (-0.737)	-1.85E-03 (-1.572)	2.24E-03 (2.275)	3.37E-03 (5.310)

*The estimates above are the coefficients on the localization variable for the indicated employment type having summed the localization concentric ring variables out to 5 miles and omitted the remaining localization concentric rings (miles 6 to 15). All other variables listed in Table 2 were included in the model: coefficients for those variables are not reported to conserve space.

Appendix A: Data Description

Our principal data source is Dun and Bradstreet's Marketplace file. The data include information on over ten million establishments in the United States reporting, among other things, location at the zipcode level (much smaller than a county), corporate status (subsidiary or non-subsidiary), age of establishment, employment, and sales. The “complete” D&B database includes establishment-specific information on over twelve million establishments in the United States and is based on public and commercially available sources and D&B phone surveys. That data set, however, was prohibitively expensive. Instead, we obtained a more limited but still enormously rich version of the database in which all of the firm specific data was aggregated up to the zipcode level.²⁵

In phone conversations with analysts at D&B, we were advised that firms requesting not to be in the database are omitted from the data file. Partly for that reason, the D&B database, while immense, does not contain the entire universe of establishments in the United States. Nevertheless, the D&B analysts felt that the omissions from the data set are sufficiently random that the D&B database is representative of the spatial distribution of establishments in the United States.²⁶ Moreover, measurement error associated with the distribution of employment across industries within a given geographic zone is likely to be small if one aggregates up by even a modest amount, as to the zipcode level.²⁷ Accordingly, we assume that the data set provides an accurate measure of the spatial distribution of establishments at the zipcode level.²⁸

Appendix B: Supplemental Tables

TABLE B-1
NEW-ESTABLISHMENT EMPLOYMENT - CMSAs ONLY
 (Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
Zipcode Area, Diversity, and Competition Effects						
<i>Zipcode Herfindahl Index</i>	-2.812E+02 (-21.207)	-3.809E+02 (-9.899)	-1.202E+03 (-8.871)	-4.479E+02 (-18.888)	-1.809E+02 (-10.587)	-1.789E+02 (-14.914)
<i>Zipcode firms per worker - other ind.</i>	-2.449E+02 (-10.819)	-3.927E+02 (-7.739)	-9.032E+02 (-4.021)	-3.614E+02 (-9.493)	-2.457E+02 (-8.699)	-1.723E+02 (-10.339)
<i>Zipcode firms per worker - own ind.</i>	1.676E+01 (2.631)	2.982E+01 (2.357)	1.632E+02 (3.518)	1.603E+01 (1.468)	2.141E+00 (0.305)	-9.243E+00 (-1.906)
Urbanization Effects: Other (Total - Own) Industry Employment In The ...						
0 to 1 Mile Ring	-9.040E-05 (-2.804)	-1.821E-04 (-3.187)	-3.927E-04 (-1.407)	2.850E-04 (3.311)	-1.045E-04 (-2.558)	-1.800E-05 (-0.884)
1 to 5 Mile Ring	2.830E-05 (3.367)	7.470E-06 (0.488)	3.689E-04 (3.644)	2.690E-05 (0.874)	-2.040E-05 (-2.148)	-7.690E-06 (-1.380)
5 to 10 Mile Ring	-1.540E-05 (-2.626)	8.130E-06 (0.763)	1.860E-05 (0.243)	3.290E-05 (1.580)	-3.360E-06 (-0.533)	-2.250E-06 (-0.658)
10 to 15 Mile Ring	2.420E-06 (0.485)	4.000E-06 (0.369)	8.480E-05 (1.239)	1.440E-05 (0.794)	5.750E-06 (0.976)	-1.440E-06 (-0.501)
Localization Effects: Own Industry Employment In The ...						
0 to 1 Mile Ring	1.202E-02 (7.098)	2.478E-02 (5.716)	8.405E-03 (1.410)	-1.949E-03 (-1.423)	1.376E-02 (3.734)	3.943E-03 (4.519)
1 to 5 Mile Ring	8.293E-04 (1.930)	3.121E-03 (2.428)	-1.250E-03 (-0.425)	-1.355E-04 (-0.233)	3.908E-03 (4.879)	8.832E-04 (3.528)
5 to 10 Mile Ring	8.888E-04 (2.959)	-4.400E-04 (-0.482)	1.297E-03 (0.547)	-4.392E-04 (-1.039)	2.120E-04 (0.401)	-9.880E-05 (-0.597)
10 to 15 Mile Ring	-4.150E-05 (-0.151)	-3.834E-04 (-0.429)	-2.916E-03 (-1.224)	-2.718E-04 (-0.699)	-6.340E-04 (-1.448)	2.254E-04 (1.621)
Average Change In Localization Effect Per Mile From ...*						
0.5 to 3 Miles	-4.48E-03	-8.66E-03	-3.86E-03	7.25E-04	-3.94E-03	-1.22E-03
3 to 7.5 Miles	1.32E-05	-7.91E-04	5.66E-04	-6.75E-05	-8.21E-04	-2.18E-04
7.5 to 12.5 Miles	-1.86E-04	1.13E-05	-8.43E-04	3.35E-05	-1.69E-04	6.48E-05
Summary Measures						
Std Error	80.391	92.860	516.597	115.202	57.218	39.077
Log-L	-18,736.47	-4,195.01	-6,865.94	-15,521.66	-5,110.12	-7,710.47
Uncensored	2,959	550	754	2,265	735	1,254
Total Obs	7,431	7,431	7,431	7,431	7,431	7,431
Fixed Effects	73	73	73	73	73	73

*Change per mile is computed by differencing the adjacent localization coefficients and dividing by the number of miles between the midpoints.

TABLE B-2
PROBIT MODEL PARTIAL DERIVATIVES WITH MSA FIXED EFFECTS*
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
Zipcode Area, Diversity, and Competition Effects						
<i>Zipcode Herfindahl Index</i>	-3.686E-01 (-41.99)	-7.781E-01 (-18.81)	-9.888E-02 (-19.69)	-3.171E-01 (-37.71)	-1.081E-01 (-22.71)	-2.145E-01 (-31.00)
<i>Zipcode firms per worker - other ind.</i>	-3.353E-01 (-26.86)	-7.022E-02 (-13.31)	-6.290E-01 (-10.24)	-2.878E-01 (-23.63)	-1.218E-01 (-18.05)	-2.119E-01 (-22.51)
<i>Zipcode firms per worker - own ind.</i>	5.645E-02 (17.430)	8.720E-03 (5.620)	1.389E-02 (9.330)	2.633E-02 (7.570)	3.816E-03 (2.020)	-2.615E-03 (-0.92)
Urbanization Effects: Other (Total - Own) Industry Employment In The ...						
0 to 1 Mile Ring	-2.230E-07 (-5.41)	-3.200E-08 (-2.73)	-8.120E-08 (-4.59)	-1.500E-07 (-2.66)	-7.160E-08 (-3.91)	-1.470E-07 (-4.23)
1 to 5 Mile Ring	1.190E-08 (1.280)	4.300E-09 (1.350)	3.180E-08 (6.270)	6.190E-08 (3.570)	-1.420E-09 (-0.32)	-3.310E-09 (-0.48)
5 to 10 Mile Ring	-1.240E-08 (-2.02)	1.580E-09 (0.071)	2.950E-09 (0.790)	5.040E-08 (4.160)	-3.250E-09 (-1.09)	-4.270E-09 (-1.02)
10 to 15 Mile Ring	1.070E-08 (2.080)	-1.860E-10 (-0.08)	8.510E-09 (2.540)	3.200E-08 (2.990)	2.390E-09 (0.860)	-2.100E-09 (-0.60)
Localization Effects: Own Industry Employment In The ...						
0 to 1 Mile Ring	6.220E-06 (2.660)	2.360E-06 (3.110)	1.280E-06 (3.460)	1.130E-07 (0.120)	3.780E-06 (3.040)	1.440E-06 (1.250)
1 to 5 Mile Ring	2.210E-06 (4.400)	7.180E-07 (2.880)	-3.560E-07 (-2.35)	-6.950E-07 (-2.04)	1.280E-06 (3.690)	1.020E-06 (3.490)
5 to 10 Mile Ring	1.760E-06 (5.240)	-1.010E-07 (-0.56)	3.660E-08 (0.310)	-8.640E-07 (-3.44)	2.800E-07 (1.210)	-1.070E-07 (-0.55)
10 to 15 Mile Ring	2.310E-08 (0.080)	5.610E-08 (0.320)	-2.920E-07 (-2.45)	-6.460E-07 (-2.75)	-6.500E-08 (-0.32)	5.990E-07 (3.590)
Average Change In Localization Effect Per Mile From ...**						
0.5 to 3 Miles	-1.60E-06	-6.57E-07	-6.54E-07	-3.23E-07	-1.00E-06	-1.68E-07
3 to 7.5 Miles	-1.00E-07	-1.82E-07	8.72E-08	-3.76E-08	-2.22E-07	-2.50E-07
7.5 to 12.5 Miles	-3.47E-07	3.14E-08	-6.57E-08	4.36E-08	-6.90E-08	1.41E-07
Summary Measures						
Log Likelihood	-11,102.69	-4,587.62	-5,221.26	-10,525.06	-5,934.96	-8,670.30
Pseudo R-square	0.3586	0.1921	0.2169	0.2930	0.2028	0.2266
Obs	38,781	36,162	36,063	38,568	37,510	38,454
Fixed Effects	357	274	278	350	318	341

*Estimates were obtained using the Dprobit option in Stata.

**Change per mile is computed by differencing the adjacent localization coefficients and dividing by the number of miles between the midpoints.

TABLE B-3
NEW-ESTABLISHMENT EMPLOYMENT - OLS MSA FIXED EFFECTS EXCLUDING ZEROS
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
Zipcode Area, Diversity, and Competition Effects						
<i>Zipcode Herfindahl Index</i>	-2.663E+00 (-0.281)	1.404E+01 (0.539)	7.564E+02 (5.954)	-6.344E+00 (-0.397)	1.466E+01 (1.266)	2.129E+00 (0.292)
<i>Zipcode firms per worker - other ind.</i>	-2.088E+01 (-1.304)	-6.250E+01 (-1.539)	-2.984E+01 (-0.166)	-3.237E+01 (-1.127)	-1.610E+01 (-0.819)	-1.978E+01 (-1.630)
<i>Zipcode firms per worker - own ind.</i>	-7.815E+00 (-1.885)	-9.504E+00 (-0.811)	-6.489E+00 (-0.157)	-1.180E+01 (-1.373)	-2.029E+00 (-0.355)	-2.287E+00 (-0.589)
Urbanization Effects: Other (Total - Own) Industry Employment In The ...						
0 to 1 Mile Ring	3.440E-04 (10.874)	2.030E-05 (0.329)	3.884E-03 (10.342)	9.605E-04 (10.737)	1.255E-04 (3.098)	2.014E-04 (6.715)
1 to 5 Mile Ring	1.310E-05 (1.880)	-2.300E-05 (-1.480)	-3.299E-04 (-3.263)	-4.580E-05 (-1.662)	-3.620E-05 (-3.670)	-1.800E-05 (-2.717)
5 to 10 Mile Ring	-8.240E-06 (-1.721)	7.050E-06 (0.610)	2.020E-05 (0.249)	2.830E-05 (1.519)	-1.060E-06 (-0.156)	2.720E-06 (0.681)
10 to 15 Mile Ring	-4.220E-07 (-0.101)	9.030E-06 (0.742)	-5.160E-05 (-0.674)	-1.800E-05 (-1.086)	2.300E-06 (0.350)	1.670E-06 (0.496)
Localization Effects: Own Industry Employment In The ...						
0 to 1 Mile Ring	1.071E-02 (7.990)	2.519E-02 (6.063)	-4.503E-02 (-7.224)	-3.432E-03 (-2.399)	8.278E-03 (2.867)	1.026E-02 (9.473)
1 to 5 Mile Ring	2.059E-04 (0.613)	2.486E-03 (1.943)	1.044E-02 (3.512)	1.928E-04 (0.363)	3.584E-03 (4.888)	2.909E-04 (1.144)
5 to 10 Mile Ring	2.679E-04 (1.157)	-9.127E-04 (-0.917)	-4.615E-04 (-0.186)	-4.216E-04 (-1.098)	3.628E-04 (0.748)	6.460E-05 (0.387)
10 to 15 Mile Ring	-1.556E-04 (-0.710)	-4.344E-04 (-0.441)	1.557E-03 (0.578)	2.435E-04 (0.668)	-6.957E-04 (-1.661)	-2.090E-05 (-0.147)
Average Change In Localization Effect Per Mile From ...*						
0.5 to 3 Miles	-4.20E-03	-9.08E-03	2.22E-02	1.45E-03	-1.88E-03	-3.99E-03
3 to 7.5 Miles	1.38E-05	-7.55E-04	-2.42E-03	-1.37E-04	-7.16E-04	-5.03E-05
7.5 to 12.5 Miles	-8.47E-05	9.57E-05	4.04E-04	1.33E-04	-2.12E-04	-1.71E-05
Summary Measures						
Std Error	56.072	58.873	346.029	85.806	35.484	29.326
R-sq within	0.1126	0.0831	0.1255	0.0922	0.0632	0.0911
R-sq between	0.2461	0.1128	0.0586	0.2139	0.1491	0.0630
R-sq overall	0.1355	0.0919	0.1359	0.1013	0.0815	0.0959
Obs	6,362	1,324	1,639	5,006	1,875	3,280
Fixed Effects	357	274	278	350	318	341

*Change per mile is computed by differencing the adjacent localization coefficients and dividing by the number of miles between the midpoints.

TABLE B-4*
LOCALIZATION EFFECTS CONTROLLING FOR ESTABLISHMENT SIZE
BASED ON OWN-INDUSTRY ESTABLISHMENTS FIVE YEARS OR OLDER
(Numbers in Parentheses are t-ratios)

	Software SIC 7371-73, 75	Food Products SIC 20	Apparel SIC 23	Printing & Publishing SIC 27	Fabricated Metal SIC 34	Machinery SIC 35
BIRTHS OF NEW ESTABLISHMENTS						
Small (1-24 emp)	3.58E-02 (27.457)	1.04E-02 (8.748)	3.36E-02 (23.487)	1.08E-02 (24.763)	2.53E-03 (11.477)	2.78E-03 (15.212)
Medium (25-99 emp)	-7.94E-04 (-1.002)	2.26E-03 (4.437)	1.25E-02 (15.308)	1.48E-03 (4.296)	3.98E-04 (4.028)	3.40E-04 (2.869)
Large (100+ emp)	3.87E-05 (2.106)	2.93E-05 (1.582)	-4.84E-04 (-8.404)	2.83E-06 (0.305)	6.12E-06 (0.770)	1.02E-05 (2.469)
NEW-ESTABLISHMENT EMPLOYMENT						
Small (1-24 emp)	5.90E-01 (19.658)	6.51E-01 (9.298)	5.75E-01 (3.440)	4.07E-01 (14.510)	1.80E-01 (8.792)	9.84E-02 (9.913)
Medium (25-99 emp)	7.32E-02 (4.026)	1.39E-01 (4.555)	1.27E-01 (1.337)	2.32E-02 (1.047)	3.41E-02 (3.707)	2.79E-02 (4.331)
Large (100+ emp)	9.60E-04 (2.253)	3.44E-03 (3.084)	9.36E-05 (0.015)	-2.52E-04 (-0.411)	1.59E-03 (2.200)	4.12E-04 (1.766)

*The estimates above are the coefficients on the localization variable for the indicated employment type having summed the localization concentric ring variables out to 5 miles and omitted the remaining localization concentric rings (miles 6 to 15). All other variables listed in Table 2 were included in the model: coefficients for those variables are not reported to conserve space.

ENDNOTES

¹See Quigley (1998) for a more complete survey of this literature.

²See O'Hara (1977), Ogawa and Fujita (1980), Imai (1982), Helsley (1990), or Krugman (1993), for example.

³These assumptions are made in part because our data are observed at the zipcode level.

⁴We have also estimated the models without deflating by zipcode area and instead including zipcode area as a regressor on the right hand side of the regression. Although the key qualitative results of the analysis are robust to that specification (attenuation of localization effects and small establishment effects), the specification described in the text was preferred for two reasons. First, zipcode area is itself a function of the level of urbanization – densely developed areas have smaller zipcodes – but the level of urbanization is already directly included in the model. Second, zipcode area is not a determinant of agglomeration per se, but instead is a geographic unit of measure and, in that respect, can be used to measure births or new firm employment for a standardized area such as per square mile.

⁵Various MapInfo software products were used to geocode the data and create the concentric ring variables.

⁶We have also estimated our models based on concentric ring variables constructed under the assumption that all employment in a given zipcode is located at the geographic centroid of the zipcode. In that case, either all or none of the employment in a zipcode was allocated to a circle depending on whether the zipcode centroid was contained within the circle. Although the key qualitative results of the paper are robust to that specification, the proportional sum approach described above provides a more accurate measure of the agglomeration variables and was preferred.

⁷The Herfindahl index of specialization (the inverse of diversity) is defined as $\sum s_i^2$, where s_i is industry i 's share of total employment and $i = 1, 2, \dots, 90$ are the two-digit industries.

⁸We also estimated the models in Tables 2a and 2b omitting zipcodes that did not belong to Consolidated Metropolitan Statistical Areas (CMSAs). This focuses attention on the most densely developed regions of the country. Estimates from the CMSA-only model were qualitatively similar to those based on the entire United States and are presented in Appendix B, Table B-1 for the new establishment model (the birth model is not presented to conserve space).

⁹This definition restricts software to outputs that are sold nationally or internationally.

¹⁰Note, in contrast, that Table 1 displays the distribution of births and new establishment employment across zipcodes without deflating by zipcode area.

¹¹The estimates in Table B-2 were obtained using the Dprobit option in Stata. This routine omits data for which all zipcodes corresponding to a given fixed effect have zero births because the observed discrete 0-1 outcomes for such regions are perfectly predicted. In contrast, those areas are retained in the Stata Tobit routine because the focus of that model is to estimate the level of unobserved birth propensities. Fixed effects for these areas do not perfectly predict observed outcomes in the Tobit model and help

instead to evaluate birth propensities as will be illustrated in Tables 3a and 3b. This accounts for the different sample sizes in Table B-2 relative to the other tables in the paper.

¹²Only estimates from the new establishment employment model are presented to conserve space. Results from the birth model were qualitatively similar.

¹³Two other models were also considered but rejected as solutions to the Tobit-fixed effect problem. First, Chamberlain (1980) developed a conditional logit solution to this problem in which the fixed effects are integrated out of the model allowing one to obtain consistent slope coefficients with finite T (where T is the number of observations per fixed effect). Unfortunately, the conditional logit approach is not computationally feasible for large samples such as ours in which roughly 39,000 zipcodes are spread over 373 fixed effects. As an alternative, one could omit the fixed effects and in their place include those city-specific variables that are thought to affect firm location and employment decisions directly in the model. In a sense, this is what previous agglomeration studies have done when making intermetropolitan comparisons of employment growth. That approach, however, could suffer from omitted variable bias and was not preferred for that reason.

¹⁴The test statistics are given by twice the difference in the log-likelihood functions for the unrestricted model (LogL) less the restricted model that omits the fixed effects ($\text{LogL}_{\text{NoFE}}$). The values of the Chi-square test statistics in Tables 2a and 2b range from roughly 550 to 1,900 with 372 degrees of freedom. In comparison, the critical value at the .001 level is 149.

¹⁵In addition, when the fixed effects were omitted from the Tobit model the key qualitative results discussed below were largely unchanged.

¹⁶The diversity result is somewhat at variance with Henderson et al's (1995) results for innovative industries. Although they find a positive effect of diversity when it is interacted with a dummy representing historical concentration, when diversity is included without the interaction the coefficient on diversity is negative in their paper.

¹⁷In addition, in Tables 2a and 2b, note that printing and publishing exhibit the weakest localization economies. This is consistent with Ellison and Glaeser (1997), who found that printing and publishing exhibited little geographic localization in the U.S. It is interesting that when we control for firm size in the next section, even this industry exhibits significant localization economies.

¹⁸In earlier versions we included mile-by-mile geographic variables for each employment type out to fifteen miles. Results from that estimation were qualitatively similar to those in Tables 2a and 2b but were much more difficult to produce and present given the very large number of coefficients.

¹⁹These estimates were obtained by first averaging the CPM measures across industries and then dividing the averaged CPM values for adjacent rings.

²⁰For example, for Printing and Publishing, the urbanization coefficients are always positive and decline monotonically with distance. This suggests that Printing and Publishing benefits from locating in central areas, *ceteris paribus*. In contrast, the negative urbanization effects for Fabricated Metal in the 0 to 1 and 1 to 5 mile rings may suggest that that industry does best locating farther from densely developed areas, *ceteris paribus*.

²¹Coefficients on the other variables were similar to those in Tables 2a and 2b.

²²If young establishments tend to be small and if there are unmeasured zipcode attributes that attract new establishments, then our small establishment effect could reflect the influence of those unmeasured attributes. As a robustness check, we re-estimated the model in Table 4a using employment at establishments at least five years old for all of the localization (own-industry) variables in the model. Results from that regression are presented in Appendix B (Table B-4). The patterns discussed in the text are even stronger when establishment age is taken into account, with all six industries exhibiting small establishment effects.

²³See Saxenian (1994) for a discussion of the open industrial system of the Silicon Valley and its advantages relative to the closed industrial system of the Route 128 area.

²⁴Jacobs (p. 153) further notes that sometimes the new firm produced a product that was far removed from aircraft manufacture, such as sliding doors. In addition, Saxenian (1994, p. 52) also discusses the important role of spin-offs from Fairchild Semiconductor in the Silicon Valley.

²⁵The complete version of the data set for the core attributes of the individual establishments costs over \$600,000 for one quarter. Those same data aggregated to the zipcode level were available for less than \$1,000.

²⁶Additional details on the Dun and Bradstreet (D&B) MarketPlace file are provided at the Dun and Bradstreet web site, www.dnb.com. As described by Dun and Bradstreet, there are several important benefits to establishments from listing themselves in the D&B database and obtaining a D-U-N-S identification number. These benefits arise primarily because of the incredible size of the D&B data file. Because the D&B file is such an effective source of information on firms throughout the economy, businesses use the D&B file to do market analysis and search out potential trading partners. Individual establishments therefore have an incentive to list themselves with D&B in much the way establishments have an incentive to voluntarily list themselves in the yellow pages. In addition, DUNS identification numbers are rapidly becoming a standard identification device in the economy, and many companies including the Federal Government require that clients obtain a D-U-N-S number as a precondition for engaging in trade. As noted in the D&B website, "It [the D-U-N-S number] is now the standard for all United States Federal Government electronic commerce transactions to help streamline and reduce federal procurement costs."

²⁷Because some establishments are omitted from the data set, our regressors – which reflect various measures of the existing level of employment – could suffer from an errors in variables problem which would bias the estimated coefficients towards zero. Assuming, however, that the spatial distribution of the data set is representative of the United States, then aggregating up to the zipcode level likely averages away any errors in the data, at least as regards the relative magnitude of employment in one industry versus another.

²⁸Carlton (1983) is the only other study we are aware of that employs Dun and Bradstreet data. Carlton too concludes that the D&B data are reasonably representative.

**The authors gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada, the UBC Centre for Real Estate and Urban Land Economics, the Real Estate Foundation of British Columbia, and the Center for Policy Research at Syracuse University. We thank Richard Arnott, Jan Brueckner, Iain Cockburn, Ed Coulson, Ed Glaeser, Keith Head, Vernon Henderson, Daniel McMillen, Jay Wilson, and seminar participants at Tulane University and Washington University for helpful comments.