The Geography of Research and Development Activity in the U.S.

Kristy Buzard

University of California-San Diego

Gerald Carlino

Federal Reserve Bank of Philadelphia

August 2009

Abstract

This study details the location patterns of R&D labs in the U.S., but it differs from past studies in a number of ways. First, rather than looking at the geographic concentration of manufacturing firms (e.g., Ellison and Glaeser, 1997; Rosenthal and Strange, 2001; and Duranton and Overman, 2005), we consider the spatial concentration of private R&D activity. Second, rather than focusing on the concentration of *employment* in a given industry, we look at the clustering of individual R&D *labs* by industry. Third, following Duranton and Overman (2005), we look for geographic clusters of labs that represent statistically significant departures from spatial randomness using simulation techniques. We find that R&D activity for most industries tends to be concentrated in the Northeast corridor, around the Great Lakes, in California's Bay Area, and in southern California. We argue that the high spatial concentration of R&D activity facilitates the exchange of ideas among firms and aids in the creation of new goods and new ways of producing existing goods. We run a regression of an Ellison and Glaeser (1997) style index measuring the spatial concentration of R&D labs on geographic proxies for knowledge spillovers and other characteristics and find evidence that localized knowledge spillovers are important for innovative activity.

JEL Codes: O31 and R11

Keywords: Agglomeration economies, Knowledge spillovers, Urban density, Innovation, R&D

This paper was prepared for the "Handbook of Economic Geography and Industry Studies" to be published by Edward Elgar Publishing.

The authors especially thank Jake Carr and Robert O'Loughlin for excellent research assistance. This paper has benefited from comments from Robert Hunt, Leonard Nakamura, and Tony Smith. We alone are responsible for any remaining errors. The views expressed here are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

Although metropolitan areas account for less than 20 percent of the total land area in the United States, they contain almost 80 percent of the nation's population and nearly 85 percent of its jobs. Put differently, the United States has, on average, 24 jobs per square mile, but metropolitan areas average about 124 jobs per square mile. According to Strange (2009), the population of six Canadian metropolitan areas (Toronto, Montreal Vancouver, Ottawa, Calgary, and Edmonton) account for almost one-half of the national population but less than 1 percent of Canada's land area. Very similar concentration patterns are evident in data from Europe (Combes and Overman, 2004) and Asia (Fujita, et al., 2004).

This high degree of spatial concentration of people and jobs leads to congestion costs, such as increased traffic and pollution, and higher housing costs. For example, according to the Texas Transport Institute's annual Mobility Report, U.S. drivers spent on average 4.2 billion hours in traffic delays in 2007; Los Angeles-area drivers sat the longest (70 hours per traveler), while motorists in the Wichita, Kansas, area spent only 6 hours per traveler in delays.[1] Congestion has become so severe in London that in February 2003 the city imposed a congestion fee, currently £8 a day, on all vehicles entering, leaving, driving, or parking on a public road inside the Charging Zone between 7:00 a.m. and 6 p.m., Monday through Friday. On January 3, 2006, Stockholm became the second European city to introduce a congestion charge. Similar fees are now in effect in Singapore, Oslo, and Rome. In the United States, New York City considered a similar plan and the city of Chicago is currently considering congestion fess for cars parked downtown. To offset these congestion costs, workers must receive higher wages, and higher wages increase firms' costs.

[1] 2007 Annual Urban Mobility Report, Texas Transport Institute: http://mobility.tamu.edu/ums/report/.

If congestion costs were the only effect of the spatial concentration of firms, firms could easily disperse to reduce these costs; yet they do not. This is because the negative effects of concentration make up only one side of the urban ledger. The positive effects of agglomeration economies — efficiency gains and cost savings that result from being close to suppliers, workers, customers, and even competitors — make up the other. Other things equal, firms will have little incentive to move if congestion costs are balanced by the benefits of agglomeration economies.

In this paper, we a use a new data set on the location of almost 3,500 research and development (R&D) labs in 1998 to establish some stylized facts regarding the spatial concentration of innovative activity in the United States. We use a variant of the Ellison and Glaeser (1997) index of agglomeration developed by Guimarães, Figueiredo, and Woodward (2007) to examine the spatial distribution of R&D labs. To address the issue of the significance of clusters, we follow the lead of Duranton and Overman (2005) in that we are able to identify R&D clusters that are significantly different from spatial randomness. Specifically, we use the geographic location of manufacturing employment as a benchmark against which to measure significant clustering. Hence, R&D clusters are identified as "significant" only when they contain more R&D labs than would be expected on the basis of manufacturing employment alone.

We show that while economic activity tends to be geographically concentrated, spatial concentration is even more pronounced among establishments doing basic R&D. We find, in particular, that R&D activity for most industries tends to be concentrated in the Northeast corridor, around the Great Lakes, in California's Bay Area, and in southern California.

We conjecture that more than most types of economic activity, R&D depends on a particular byproduct of agglomeration economies called knowledge spillovers — the continuing exchange of ideas among individuals and firms. A high geographic concentration of R&D labs creates an environment in which ideas move quickly from person to person and from lab to lab. Locations that are dense in R&D activity encourage knowledge spillovers, thus facilitating the exchange of ideas that underlies the creation of new goods and new ways of producing existing goods. We find several pieces of evidence that support this view. We run a regression of a variant of the Guimarães, Figueiredo, and Woodward (2007)—hereafter GFW—index measuring the spatial concentration of R&D labs on geographic proxies for knowledge spillovers and other location-specific characteristics and find evidence that localized knowledge spillovers are important for innovative activity. In particular, we show that a strong positive correlation exists between the geographic concentration of R&D labs and citation-weighted patent intensity (our proxy for knowledge spillovers).[2] All else equal, the index of agglomeration for R&D labs will be 15 percent greater in a county with twice the citation intensity of another county. In addition, we find evidence (although it's mixed) to support the existence of Jacobs externalities for R&D activity.

We also find evidence that human capital is highly correlated with the clustering of R&D labs. In fact, of the things we considered, by far the most powerful effect on spatial clustering of labs is generated by local human capital. Specifically, a 1 percent increase in the share of employment accounted for by professional and specialty

[2] We define citation-weighted patent intensity as citation-weighted patents that are normalized by the number of workers in professional and specialty occupations

occupations is associated with a 3.5 percent increase in the spatial concentration of R&D labs.

LITERATURE REVIEW

Krugman (1991) and Audretsch and Feldman (1996) developed a "locational Gini coefficient" to answer the question of which manufacturing industries cluster geographically. A locational Gini coefficient shows how similar (or dissimilar) the location pattern of employment in a particular industry is from the location pattern of overall employment. Let s_{ij} represent location i's share of employment in industry j, and x_i represent location i's share of aggregate employment or total population. The spatial Gini coefficient for in industry j is defined as:

$$G_j = \sum_i (x_i - s_{ij})^2 \tag{1}$$

$G = 0$ indicates that employment in a given manufacturing industry is no more or less geographically concentrated than overall manufacturing employment, and $G > 0$ implies that employment in the industry is over-concentrated. Audretsch and Feldman (1996) use the United States Small Business Administration's Innovation Data Base that consists of innovations compiled from the new product announcements sections in manufacturing trade journals. They found that innovation tends to be relatively more concentrated in industries where knowledge spillovers tend to be important.

Importantly, Ellison and Glaeser (1997)—hereafter EG—have identified a potential problem with the locational Gini coefficient. They argue that if an industry consists of a small number of establishments, the locational Gini coefficient may indicate localization of the industry under consideration, even if there is no agglomeration force behind the industry's location.[3] EG have developed an alternative concentration measure that controls for an industry's organization:

$$\gamma = \frac{G_j - (1 - \Sigma_i x_i^2) H_j}{(1 - \sum_i x_i^2)(1 - H_j)} \tag{2}$$

where H_j represents the employment Herfindahl index for industry *j*:

$$H_j = \sum_k z_{jk}^2 \,,$$

where $(z_{jk} : k = 1, \cdots, m_j)$ denotes the distribution of employment across m_j plants in industry *j*. Using Proposition 1 in EG, the numerator of (2) can be expressed as:

$$G_j - (1 - \Sigma_i x_i^2)[\gamma + (1 - \gamma) H_j] \tag{3}$$

Let $\gamma \equiv \gamma^{na} + \gamma^{s} - \gamma^{na}\gamma^{s}$ where γ^{na} reflects the benefits of a location's natural advantages (ports, resource endowments, etc.) and γ^{s} index reflects spillovers (e.g., knowledge spillovers).

Consider the special case where plants are randomly distributed across locations. That is, they are neither influenced by natural advantage or spillover effects ($\gamma^{na} = \gamma^{s} = 0$). If we treat this spatial randomness case as the null hypothesis, then EG show that $(1 - \Sigma_i x_i^2) H_j$ is the expected value, $E(G_j)$, of G_j under this hypothesis. In addition, they show that $E(G_j) > (1 - \Sigma_i x_i^2) H_j$ otherwise. If G_j is taken as an estimate of $E(G_j)$ then the EG index provides a test of the null hypothesis. That is, EG compare the degree of spatial concentration of employment in an industry with what would arise if all

[3]EG use employment in the U.S. vacuum cleaner industry to illustrate the concern. Seventy-five percent of employment in the industry is *industrially* concentrated in just four plants. Thus, using equation (1) to calculate a measure of spatial concentration of this industry suggests strong *spatial* concentration. But the strong spatial concentration may reflect nothing more than the fact that employment is industrially concentrated in a small number of plants.

plants in the industry were randomly distributed across locations. EG, and more recently Rosenthal and Strange (2001), find evidence of the geographic concentration of employment in many U.S. manufacturing industries.

Duranton and Overman (2005)—hereafter DO— use plant-level data for manufacturing activity in the UK and show that the geographic concentration of manufacturing jobs is not simply an American phenomenon. Their data identify the postal codes for each manufacturing plant in the UK, allowing them to geo-code the data. This is important, since DO are not bound by a fixed geographical classification (such as states, MSAs, or counties) but base their approach on the actual distance between firms. Additionally, rather than using a specific index to measure geographic concentration, such as the EG index, DO take a non-parametric approach (i.e., kernel density methods). Essentially, DO construct frequency distributions of the pair-wise distances between plants in a given industry. When the mass of the distribution is concentrated on the left of the distribution, this represents a spatial concentration of plants in the industry. Alternatively, if the mass of the distribution is concentrated on the right of the distribution, this represents a more dispersed spatial pattern. Importantly, DO consider whether the number of plants at a given distance is *significantly* different from the number found if their locations were randomly chosen. This represents an important improvement over the rule-of-thumb approach used by EG to determine differing levels of the spatial concentration of an industry.

A study by Arzaghi and Henderson (2005) looks at the location pattern of firms in the advertising industry in Manhattan. They report that Manhattan accounts for 20 percent of total national employment in the ad industry, 24 percent of all advertising agency receipts, and 31 percent of media billings. They show that for an ad agency, knowledge spillovers and the benefits of networking with other nearby agencies are large but the benefits dissipate very quickly with distance from other ad agencies and are gone after roughly one-half of a mile.

Holmes and Stevens (2004) take a broader approach. They use employment data for all U.S. industries, not just manufacturing, and not for just a single industry, such as advertising. Among the 15 most concentrated industries, they find that six are in mining and seven are in manufacturing; only two industries fall outside mining and manufacturing (casino hotels and motion picture and video distribution).

Several other studies find that knowledge spillovers dissipate rapidly with distance. See, for example, Arzaghi and Henderson (2005), Audretsch and Feldman (1996), and Keller (2002). Agrawal, Kapur, and McHale (2008) find that every 1000-mile increase in the distance between inventors reduces the probability of knowledge flow as measured by patent citations by about 2 percent.

Our work differs from past studies in three ways. First, rather than looking at the geographic concentration of firms engaged in the production of goods (such as manufacturing) and services (such as advertising), we consider the spatial concentration of private R&D activity.[4] Second, rather than focusing on the concentration of employment in a given industry, we look at the clustering of individual R&D labs.[5]

[4] A number of other studies look at innovative output across cities, such as the study by Audretsch and Feldman (1996). What is unique about our article is that we look at the spatial clustering of *private* R&D activity.

[5] The study by Guimarães, Figueiredo, and Woodward (2007) is the only other study we are aware of that looks at spatial clustering at the establishment level. Specifically, they look at the geographic concentration of over 45,000 plants in 1999 for concelhos (counties) in Portugal.

Third, following DO, we look for geographic clusters of labs that represent statistically significant departures from spatial randomness using simulation techniques.

DATA AND MODEL

We used 1998 data on R&D labs from the *Directory of American Research and Technology* to electronically code the addresses and other information about R&D labs, data that were not previously available in a machine-readable format. Since the directory lists the complete address for each establishment, we were able to assign a geographic identifier (using geocoding techniques) to 3,446 R&D labs in the U.S. in 1998.[6] The data on manufacturing employment at the zip code level are found in the Department of Commerce's Zip Code Business Patterns 1998 data. We assigned these data to each point by assuming that manufacturing employment is uniformly distributed across points within a zip code.

Clustering of R&D Labs. Figure 1 shows a map of the spatial distribution of R&D labs that reveals a striking clustering of this activity. The dots shown on the map tend to represent concentrations of R&D labs. A prominent feature of the map is the high concentration of R&D activity in the Northeast corridor, stretching from northern Virginia to Massachusetts. There are other concentrations, such as the cluster around the Great Lakes and the concentration of labs in California's Bay Area and in southern California. But some states that account for a relatively large share of the nation's jobs account for a much smaller share of the nation's R&D labs. For example, Texas ranks second among states in terms of employment, but it ranks eighth in terms of labs. Similarly, Florida ranks fourth in employment but 13th in terms of labs.

[6] Our data on individual labs were limited to the top 1000 public companies in terms of expenditures on R&D. The 1000 labs in our data set cover over 95 percent of all R&D performed by public companies.

As already noted, recent studies have shown that economic activity, especially manufacturing, also tends to be geographically clustered. However, it appears that R&D labs are more highly concentrated than establishments in general or establishments in manufacturing. There are more than 3100 counties in the U.S., and all of them are engaged in some type of economic activity. All but 33 counties are engaged in some form of manufacturing activity. In contrast, only 519 of these counties have at least one R&D lab, and far fewer counties have a notable concentration of labs.

Another way to quantify relative concentrations is to compute each county's share of total R&D labs and rank counties by descending order of this share. Moving down this ranking, we compute a cumulative total for the share of R&D labs. We also construct a similar ranking for establishments in general and for manufacturing establishments in particular. The top 50 counties ranked by number of R&D labs account for 58 percent of all R&D labs, while the top 50 counties ranked by number of manufacturing establishments account for only 36 percent of all manufacturing establishments and only 32 percent of all establishments. It appears that R&D labs are more highly concentrated than economic activity in general and overall manufacturing activity in particular. This is important because it means the concentration of R&D labs doesn't simply reflect the concentration of manufacturing activity. Since R&D is more concentrated than manufacturing activity, this suggests that some factors, such as knowledge spillovers, may be a more centralizing force for R&D than they are for manufacturing activity.

Which R&D Labs Cluster? To answer the question of which R&D labs cluster, we will follow the lead of GFW, who have generalized the EG index to include the case where the data are in the form of establishments (labs, in our case) rather than employment shares, as in the EG index. The GFW locational Gini, or the GFW index, is constructed as follows: let: n_{ij} = the number of labs in county *i*, in industry *j*, n_j = the total number of labs in industry *j* across all *i* counties in the U.S., x_i = is county *i*'s share of aggregate manufacturing employment. Let the spatial Gini based on counts be defined as:

$$G_{j,c} = \sum_i \left[\frac{n_{ij}}{n_i} - x_i \right]^2$$

The adjusted spatial Gini $(\hat{G}_j)$ for industry *j* is given by:

$$\hat{G}_j = \frac{n_j G_{j,c} - (1 - \sum_i x_i^2)}{\left(n_j - 1\right)(1 - \sum_i x_i^2)}$$

$\hat{G}_j$ is similar to that of the EG index—equation (2)—except that in the above index, H_j is replaced by $1/n_j$ and the spatial Gini for manufacturing employment, G_j in the EG index, is replaced by its counterpart expressed in terms of counts of labs instead of employment. The expression $\left(n_j - 1\right)(1 - \sum_i x_i^2)$ is included so that the index has the property that $\hat{G}_j = 0$ when labs are no more or no less concentrated than manufacturing employment. For our county data, $(1 - \sum_i x_i^2)$ takes on values close to one, 0.99593.[7]

We use the adjusted GFW index as our measure of concentration for R&D by industry. Our sample consists of 376 four-digit Standard Industrial Classification industries at the county level. We find an adjusted GFW index of 0.0457 for R&D in the average industry at the *county* level. In studying the agglomeration patterns in the

[7] See GFW for details on the construction of the adjusted GFW index used in this article as well as a discussion of the EG index.

manufacturing industries, Rosenthal and Strange (2001) report an average adjusted Gini coefficient (using the EG index) of 0.0193 for manufacturing in 2000 at the county level.[8] Thus, our R&D labs appear to be more spatially concentrated, on average, than is manufacturing activity. Our measure of concentration, the adjusted GFW index, has a maximum value of about one for R&D in five industries.[9] However, there are only two R&D labs in each of these industries, so it's not surprising to find a large value for the adjusted Gini index if the two firms are located in proximity to one another.[10] Table 1 shows the values of the adjusted GFW index for industries with 20 or more labs. The table shows that R&D tends to be most concentrated in the oil and gas field machinery industry, the computer storage devices industry, and the electronic computer industry.

Our findings indicate that 256, or 68 percent, of all R&D labs have an adjusted Gini index greater than zero, suggesting R&D labs are appreciably more concentrated than manufacturing employment. Earlier we reported that the top 50 counties ranked by number of R&D labs account for 58 percent of all R&D labs, while the top 50 counties ranked by number of manufacturing establishments account for only 36 percent of all manufacturing establishments. Thus, the concentration of labs is broadly similar when looking at the top 50 counties or the adjusted Gini index. Only 1 percent of the labs are associated with an index that is negative, indicating dispersion.

While an adjusted Gini index could have a value greater than zero, an important question is, "Does this represent a significant departure from the average concentration of

---

[8] Ellison, Glaeser, and Kerr (2007) report an average adjusted Gini coefficient of 0.03 for manufacturing in 1997 at the *metropolitan area* level.

[9] They are R&D activity in hog production; the production of brooms and brushes; the production of fiber cans, tubes, drums; the bottled and canned soft drinks and carbonated waters industry; and the rolling mill machinery and equipment industry.

[10] There is a negative relationship between the size of the adjusted GFW index and the number of labs in an industry. However, this relationship is not strong: a correlation coefficient of -0.09 that is only marginally significant (at the 10 percent level).

manufacturing employment?" We performed a simulation procedure to determine what value of the adjusted GFW indexes constitutes a significant departure from the concentration of manufacturing employment.[11] We find R&D labs in 129 of the 376 industries considered (34.3 percent) are significantly more concentrated than is manufacturing employment.

Figure 2 shows the distribution of the adjusted GFW index for R&D activity. Each bar shows the number of industries associated with a particular value for the adjusted GFW index. Following EG, we consider R&D in an industry to be highly concentrated if the adjusted Gini index is at least 0.05 and moderately concentrated if the index is at least 0.02, but less than 0.05. R&D in an industry is considered to be dispersed if the index is less than 0.01. A prominent feature of Figure 2 is the large number of industries falling into the range we have classified as not very concentrated (an adjusted GFW index less than 0.02). The tallest bars tend to surround an adjusted GFW index around zero. In fact, 69 percent of the industries have an adjusted GFW index below 0.02.[12] Sixty-six of the 376 industries considered have an adjusted GFW index

[11] To develop measures of statistical significance for the adjusted GFW indexes, we divide our labs into six non-overlapping groups based on the number of labs in a given industry. The first group consists of industries with between 2 and 9 labs. The second group consists of industries with 10 to 30 labs, while the third group consists of industries with between 31 and 50 labs. The fourth group consists of industries with between 51 and 100 labs, while the fifth group consists of industries with 101 to 200 labs. The final group consists of industries composed of over 200 labs. For each group, we performed a simulation procedure to produce a probability distribution for the adjusted GFW index. In the simulation we randomly allocated labs to counties while maintaining the counties' shares of national manufacturing employment. Therefore, if a given county has a relatively high share of the nation's manufacturing jobs, the county is more likely to be randomly assigned more R&D labs, too. For each group the simulation produces a value for the adjusted GFW index. For each group, we performed 1000 simulations and formed a probability distribution for the adjusted GFW indexes. From the distribution we can calculate critical values (one that's positive and one that's negative) that allow us to say that we are 95 percent certain that any value that exceeds (falls below) the critical value indicates that labs in that grouping are significantly more concentrated (significantly more dispersed) than is the distribution of manufacturing employment.

[12] Similar to our finding that the largest percentage of R&D labs is generally not more concentrated than manufacturing employment, Ellison and Glaeser's finding shows that the largest number of manufacturing industries could also be classified as not very concentrated.

that is at least 0.05. That's only 18 percent of all industries. In addition, these 66 highly concentrated industries account for only 6 percent of all R&D labs. If we include industries that are moderately concentrated — that is, those with an adjusted GFW index that is at least 0.02 but less than 0.05 — we can add another 49 names to the list of industries that tend to be more concentrated relative to manufacturing employment. Still, these 115 concentrated industries (66 highly concentrated industries plus 49 moderately concentrated industries) account for only 31 percent of all industries and for only 29 percent of all R&D labs.

Until now, we have looked at the concentration of R&D labs relative to the concentration of manufacturing employment. We would also like to know whether labs in a particular industry (such as pharmaceuticals) are more or less concentrated than overall R&D labs. To get this information, we recalculated the adjusted GFW index to reflect the geographic concentration of labs in individual industries relative to the overall concentration of R&D labs (as opposed to the overall concentration of manufacturing employment). The distribution of the adjusted GFW index when the benchmark is overall R&D labs (Figure 3) is remarkably similar to the distribution when the benchmark is overall manufacturing employment (Figure 2). However, there tends to be slightly more concentration of labs in individual industries compared with the location of labs in general. Thirty-six percent of the industries are at least moderately concentrated, compared with 31 percent when the benchmark is manufacturing employment. Still, the tallest bars in Figure 3 tend to surround an adjusted GFW index around zero, suggesting that for the majority of industries, labs at the industry level tend not to be more spatially concentrated than labs overall.

Maps of R&D activity for individual industries (for example, software, Figure 4; pharmaceuticals, Figure 5; and chemicals, Figure 6) confirm the findings of the locational Gini coefficient in that the location pattern of R&D activity for the majority of industries is broadly similar to the location pattern of overall R&D activity. That is, R&D activity for most industries tends to be concentrated in the Northeast corridor, around the Great Lakes, in California's Bay Area, and in southern California.

As indicated, there are a number of exceptions to the general pattern of geographic concentration just described. One exception is R&D activity in the oil and gas field machinery industry, which tends to be concentrated in Texas, especially in the Houston area, and accounts for about 60 percent of the labs doing R&D in this industry (Figure 7). Another exception is the location of R&D activity in the motor vehicle and car body industry, which tends to be concentrated in Michigan, especially in the Detroit area, and which accounts for just under 40 percent of the labs doing R&D in this industry (Figure 8). This industry is composed of establishments primarily engaged in manufacturing motor vehicle parts and accessories.

WHY DO R&D LABS CLUSTER?

A number of theories have been advanced to explain firms' (not just R&D labs') tendency to cluster spatially. Firms may attempt to minimize transport costs by locating close to a natural resource used as an input, or to their suppliers, or to their markets. Or firms may cluster to share inputs, such as specialized workers. Finally, firms may cluster to take advantage of knowledge that "spills over" when firms are located near one another. Among these, the sharing of inputs and especially of knowledge spillovers is likely to be most important for R&D labs when choosing a location.

Knowledge Spillovers. While theories of knowledge spillovers were originally developed to explain the concentration of industries in general, we think they are particularly important to an explanation of the clustering of R&D labs. More than most industries, R&D depends on new knowledge. Often, the latest knowledge about technological developments is valuable to firms but only for a short time. Thus, it behooves firms to set up shop as close as possible to the sources of information. The high spatial concentration of R&D activity facilitates the exchange of ideas among firms and aids in the creation of new goods and new ways of producing existing goods.

Two types of knowledge spillovers are thought to be important in understanding the location pattern of R&D labs: MAR spillovers and Jacobs spillovers. According to the MAR theory of spillovers, the concentration of establishments (labs in our case) in the same industry in a common area helps knowledge travel among labs and their workers and facilitates innovation and growth.[13] Using data for U.S. manufacturing, Rosenthal and Strange (2001) and Ellison, Glaeser, and Kerr (2007) consider the importance of input sharing, matching, and knowledge spillovers for manufacturing firms at various levels of geographic disaggregation (state, MSA, county, and zip code levels). These studies find evidence for all three mechanisms. Importantly, Rosenthal and Strange (2001) find that the effects of knowledge spillovers on the agglomeration of manufacturing firms tend to be quite localized, influencing agglomeration only at the zip code level.[14]

[13]Glaeser, Kallal, Scheinkman, and Shleifer (1992) coined the term MAR spillovers. MAR spillovers are so-called because Marshall (1870) developed a theory of knowledge spillovers that was later extended by Arrow (1962) and Romer (1986)—hence, MAR.

[14]Several other studies have found that knowledge spillovers dissipate rapidly with distance. See, for example, the articles by Arzaghi and Henderson (2005); Audretsch and Feldman (1996); Keller (2002); and Kolko (2007). The extent to which innovation in communications technologies are rendering face-to-face

Jacobs (1969) believed that knowledge spillovers are related to the diversity of industries (diversity of labs in our case) in an area, in contrast to MAR spillovers, which focus on firms in a common industry. Jacobs argued that an industrially diverse environment encourages innovation. Such environments include knowledge workers with varied backgrounds and interests, thereby facilitating the exchange of ideas among individuals with different perspectives. This exchange can lead to the development of new ideas, products, and processes. While other factors could be at work, the adjusted GFW indexes appear to offer support for Jacobs's diversity view, in that R&D labs for the vast majority of industries (about two-thirds) tend to exhibit a common overlapping pattern of concentration. Feldman and Audretsch (1999) used the U.S. Small Business Administration's Innovation Data Base and focused on innovative activity for particular industries within specific MSAs. They found less industry-specific innovation in MSAs that specialized in a given industry, a finding that also supports Jacobs's diversity thesis.

Other Forces for Concentration. While it's tempting to argue that the broadly similar geographic clustering of R&D labs in many different industries is suggestive of Jacobs externalities, this conjecture is simply based on visual inspection of a map (Figure 1). The finding that R&D labs tend to display a common overlapping pattern of concentration suggests that Jacobs spillovers may be more important for R&D labs than MAR spillovers. Jacobs spillovers are one possible way to account for the common overlapping pattern of concentration among R&D labs, but other forces might be at work. One such source is the natural advantages an area offers to firms that locate there. The

---

contacts obsolete is not so clear. Gaspar and Glaeser (1998) argue that improvements in telecommunications technology increase the demand for all interactions. So while technology may substitute for face-to face contacts, this effect is offset by the greater desire for all kinds of interactions, including face-to-face contacts.

natural advantages of an area, such as climate, soil, and mineral and ore deposits, could explain the location of some R&D labs. For example, oil deposits, an essential ingredient for testing equipment, may be largely responsible for the concentration of R&D labs in the oil and gas field machinery industry (the most highly concentrated industry according to our adjusted GFW indexes) in Texas, especially in the Houston area. But the draw of ore deposits seems to be industry specific and is therefore unlikely to account for the common overlapping pattern of concentration among R&D labs in many different industries. Of course, if R&D labs tend to be drawn to areas offering amenities such as pleasant weather, proximity to the ocean, and scenic views, this could explain the overlapping concentration in amenity-rich locations, such as the concentrations found in California. While local amenities might explain some of the concentrations of labs, the vast majority of R&D labs tend to be highly concentrated in the relatively low-amenity rust belt region of the country.

Another type of an area's natural advantage is its workers and its institutions, especially its universities. Universities are key players not only in creating new knowledge through the basic research produced by their faculties but also in supplying a pool of knowledge workers on which R&D depends. It is well known that Silicon Valley and the Route 128 corridor became important centers for R&D as a result of their proximities to Stanford and MIT. Saxenian (1994) describes how Stanford's support of local firms is an important reason for the Silicon Valley's success. Two of Stanford's star engineering professors, John Linvill and Fred Terman, not only drew some of the best and brightest students to Stanford but also trained their students (and encouraged them) to seek careers in the semiconductor industry.

There is also evidence that an area's human capital can be an important type of natural advantage. Carlino, Chatterjee, and Hunt (2007) looked at the effect of a metropolitan area's human capital (the share of the adult population with at least a college education) on an area's ability to innovate (measured by patents per capita). Of the explanatory factors they considered, by far the most powerful effect on local innovation is generated by local human capital. They find that a 10 percent increase in the share of the adult population with at least a college degree is associated with an 8.6 percent increase in patents per capita. Lin (2007) finds that applications and adaptations of new work are positively correlated with an area's human capital and industrial variety.

There is also general evidence that R&D at local universities is important for firms' innovative activity. Audretsch and Feldman (1996) and Anselin, Varga, and Acs (1997) found evidence of localized knowledge spillovers from university R&D to commercial innovation by private firms, even after controlling for the location of industrial R&D. However, Carlino, Chatterjee, and Hunt (2007) found that R&D at local universities has only modest effects on local innovative activity. They found that a 10 percent increase in R&D intensity of local universities is associated with less than a 1 percent increase in patent intensity.

Estimates of the Determinants of R&D Clustering. In this section, we consider the role of knowledge spillovers and access to skilled human capital on the spatial clustering of R&D labs. Recall that we have only one adjusted GFW index for each industry. These industry indexes can, however, be used to construct an *overall* adjusted GFW index for each county. We do this by weighting each industry's adjusted GFW index by the share of the county's R&D labs accounted for by that industry. Each of the industry-weighted adjusted GFW indexes for a given county are then summed to arrive at an overall adjusted GFW index for each county.

The overall adjusted GFW index for a county can be correlated with proxies for agglomeration economies. One such economy is the ability of firms in a common endeavor to share specialized inputs. Highly skilled labor is likely to be an important, if not the most important, input in the R&D process. We measure highly skilled labor using those jobs falling into the Census Bureau's classification of professional specialty occupations. This grouping includes engineers, scientists, social scientists, doctors, and other health professionals. But it also includes teachers, lawyers, artists, and athletes. This is a residency-based measure of employment in 1990. We expect a positive correlation between the GFW index and the professional specialty occupations variable.

Two proxies are used for knowledge spillovers available to labs in the same county. Citation-weighted patents normalized by the number of workers in professional specialty occupations are one variable used to proxy for knowledge spillovers. As Rosenthal and Strange (2001) point out, one concern about using patents as an indicator of knowledge spillovers is that the value of patents is very highly skewed. Most are not worth very much, while some have values that are higher by several orders of magnitude (see, for example, Harhoff et al., 1999). Fortunately, there are ways to introduce an adjustment for quality into these counts, just as is done for journal articles—by counting the number of citations a patent receives in subsequent patents. A number of empirical studies document a strong positive correlation between these "forward" citations and the economic value these patents contribute to the firms that own them. For example, Hall, et al. (2005) show that a one-citation increase in the average number of patents in a publicly held firm's portfolio increases its market value by 3 percent. In addition, these citations present a concrete illustration of knowledge spillovers.[15] Thus, we use citation-weighted patents per worker in the professional specialty occupations in 1990 at the county level as one of our proxy variables for knowledge spillovers. In addition, we use universities' R&D in the years 1987-89 as a proxy variable for information spillovers from universities to private R&D labs. R&D performed by academic institutions is normalized by total full-time enrollment at colleges and universities in the MA in those years.

A Herfindahl index of industrial diversity is used to address the question as to whether knowledge spillovers are largely Jacobs or largely MAR. By construction, a county is said to be more highly specialized or less diversified as the value of the diversity index increases.[16] Recall that as the value of the adjusted GFW index increases, the extent of the spatial concentration of labs in the industry also increases. A positive correlation between the overall county adjusted GFW index and the specialization index means that as the county becomes more specialized industrially its labs are also becoming more geographically concentrated. This is evidence favoring MAR spillovers. On the other hand, if the geographic concentration of labs tends to increase as the specialization index decreases—indicating a more industrially diverse (or less specialized) area—this negative correlation provides evidence in favor of Jacobs spillovers.

[15] In a survey of 1,300 inventors, Jaffe, Trajtenberg, and Fogarty (2000) found that approximately one-half of the patent citations refer to some sort of knowledge spillovers, of which 28 percent correspond to a very substantial spillover. Jaffe et al. (1993) provide evidence that these spillovers are at least initially localized.

[16] The Herfindahl index is calculated by squaring and summing the share of establishments accounted for by each industry in a given county. The squaring of industry shares means that the larger industries contribute more than proportionately to the overall value of the index. Thus, as the index increases in value for a given county, this implies that the county is more highly specialized or less diversified industrially.

The share of employment in the manufacturing industry is included in all regressions to control for the fact that the vast majority of patents are issued to firms in the manufacturing sector. Finally, we used eight census region dummy variables to broadly control for an area's natural advantages, such as climate (New England is the excluded region in all regressions).

Estimates of the Determinants of Concentration of R&D Labs. The forces underlying the agglomeration of R&D labs are estimated using the following regression framework:

$$GFW_i = \beta X_i + \varepsilon_i \qquad (4)$$

where GFW_i is the industry-weighed adjusted GFW index based on R&D labs in 1998 for county *i*; X_i is a vector of county characteristics; and ε_i is the usual error term. To mitigate any bias induced by endogeneity or reverse causation, all the independent variables are lagged—none reflect economic activity after 1990. Our results are presented using robust standard errors (White corrected) to control for any heteroskedasticity.

Findings. The findings of the regression results are summarized in Table 2. Column (1) of the table summarizes the results for a robust OLS regression for all counties (approximately 3100 counties). An important finding is that the citation-weighted patents intensity variable is positive and highly significant, indicating the importance of knowledge spillovers for R&D labs. The relationship between the county-adjusted GFW index and the citation-weighted patents intensity variable is economically significant, displaying an elasticity (evaluated at the mean) of 0.15—a 10 percent increase in citation intensity results in almost a 1.5 percent increase in the weighted adjusted GFW index.

The coefficient on the university R&D variable is positive, as expected, but not statistically significant, suggesting little information spillovers from university R&D to private lab R&D. This finding is consistent with Carlino, Chatterjee, and Hunt (2007), who found that R&D at local universities has only modest effects on local innovative activity. Interestingly, we found a negative and highly significant correlation between the overall county adjusted GFW index and the specialization index, evidence favoring Jacobs spillovers. The evidence from the regressions favors Jacobs spillovers and supports the intuition from our maps showing broadly similar geographic clustering of R&D labs in many different industries, also suggestive of Jacobs externalities.

By far the most powerful effect is generated by human capital (the share of the employment accounted for by professional specialty occupations). This variable is positive, highly significant, and economically significant too, with an elasticity of 3.5. Finally, the coefficients on the regional dummy variables are not statistically significant; suggesting that an area's natural advantages do not influence the location decisions of private R&D labs.

Given that the vast majority of labs tend to be located in metropolitan counties, the regression summarized in Column (2) of Table 2 adds a dummy variable that is unity for metropolitan counties and zero otherwise. The coefficient on the metropolitan county dummy variable is positive and highly significant. With the exception of the Herfindahl index, the results after adding the metropolitan dummy variable are broadly similar to the regression without this dummy variable. Interestingly, the coefficient on the Herfindahl index is negative but no longer statistically significant. One possibility is that metropolitan counties are more industrially diverse than non-metropolitan counties and that this greater diversity is being captured by the metropolitan dummy variable. To investigate this possibility, Column (3) of Table 2 summarizes the results when only the 752 metropolitan counties are used in the regression. The coefficient on the Herfindahl index is now negative and highly significant.

Spatial dependence. There is a very high degree of spatial inequality in the distribution of R&D labs. As Figure 1 shows, labs tend to be highly concentrated in the metropolitan areas of the Northeast corridor, around the Research Triangle in North Carolina, and in California's Silicon Valley. Even though the coefficients on our regional dummy variables are typically insignificant, this clustering of innovative activity suggests there could be strong spatial dependence at a more localized level and, if so, it should be controlled for in our empirical analysis.

The conjecture, then, is that a cluster of labs in one county may be highly correlated with a cluster in nearby counties; this is especially likely for counties in the same MSA. The consequences of spatial autocorrelation are the same as those associated with serial correlation and heteroskedasticity: When the error terms across counties in our sample are correlated, OLS estimation is unbiased but inefficient. However, if the spatial correlation is due to the direct influence of neighboring counties, OLS estimation is biased and inefficient (Anselin, 1988).

The literature suggests two approaches to dealing with spatial dependence. In the first approach, spatial dependence is modeled as a spatial autoregressive process in the error term:

$$\varepsilon = \lambda W \varepsilon + \mu$$
$$\mu \sim N(0, \sigma^2)$$

where λ is the spatial autoregressive parameter and μ is the uncorrelated error term. W is a spatial weighting matrix where nonzero off-diagonal elements represent the strength of the potential interaction between the s^{th} and t^{th} counties. We use the inverse of the square of the geographic distance between counties to fill in the off-diagonal elements of W. (The spatial weights are row standardized.) The null hypothesis of no spatial error dependence is $H_0 : \lambda = 0$.

The second approach models the spatial dependence in R&D labs via a spatially "lagged" dependent variable:

$$GFW = \rho WGFW + X\beta + \varepsilon$$

where *GFW* is an *Nx1* vector and *N* is the number of locations in our study; ρ is the autoregressive parameter (a scalar); W is the *NxN* spatial weight matrix described above; *X* is an *NxK* matrix of other explanatory variables from before; and ε is the *Nx1* random error term. The null hypothesis of no spatial lag is $H_0 : \rho = 0$.

Following Anselin and Hudak (1992), we perform three tests for spatial autocorrelated errors: Moran's I test, the Lagrange multiplier (LM) test, and a robust Lagrange multiplier test (robust LM). We also perform two tests for the spatial lag model (LM test and a robust LM test). The Moran's I test is normally distributed, while the LM tests are distributed χ^2 with *k* and one degree of freedom, respectively.

We estimate the OLS specifications previously reported in column 2 of Table 2 using these various tests for spatial dependence. The null hypothesis of zero spatial lag cannot be rejected in any specification. The results for spatial error are somewhat more ambiguous. The null hypothesis is clearly rejected according to the Moran's I test and the LM test, but not according to the robust LM test. Still, spatial lag dependence is likely to be an issue for our specifications. Given this, we re-estimate the OLS specification reported in column 2 of Table 2, incorporating a correction for either spatial error or spatial lag. Columns 2 and 3 of Table 4 present the results for the regressions that correct for spatial dependence. As columns 2 and 3 show, the results after correcting for spatial error are virtually identical with the results after correcting for a spatial lag. Importantly, the evidence for knowledge spillovers is strengthened. First, as before, the correlation between citation-weighted patents per worker in the professional specialty occupations and the GFW index remains positive and highly significant. In addition, the coefficient on the university R&D variable is now positive and significant, suggesting that R&D at local universities is important in the location decisions of private R&D labs. However, the elasticity is small—0.04. The results are still consistent with the view that private R&D labs may be drawn to areas that have a relative abundance of highly skilled workers that the industry requires. Unfortunately, after correcting for spatial dependence, the coefficient on the Herfindahl index, while remaining negative, is no longer significant. Thus, our results regarding the importance of Jacobs spillovers for R&D labs are mixed.

Interestingly, as Table 3 shows, the null hypothesis of zero spatial error ($\lambda = 0$) as well as the null hypothesis of zero spatial lag ($\rho = 0$) cannot be rejected when the sample is restricted to the 752 MSA counties. The lack of spatial dependence in the MSA sample suggests that the spatial dependence found in the sample using all 3108 counties is largely due to spatial dependence among labs in counties within the same metropolitan area.

In sum, we find evidence for agglomeration economies in the clustering of R&D labs at the county level. The strongest evidence for agglomeration economies is found for the citation-weighted patents per worker in the professional and specialty occupations variable, a proxy for localized knowledge spillovers. The coefficient on this variable is positive and highly significant and is robust to sample size (a sample of all counties and a sample of only MSA counties) and robust to corrections for spatial dependence. The evidence for knowledge spillovers from academic R&D to private lab R&D is mixed. The coefficient on the university R&D variable is positive but not significant in either the full sample or the MSA sample. Interestingly, the coefficient on the academic R&D variable is positive and significant after correcting for spatial dependence. Still, the estimated magnitude of the effect of academic R&D on the clustering of private R&D labs is small. We find evidence that human capital in an area may be a draw for private R&D labs. An elasticity of 3.5 is found for the spatial concentration of labs with respect to the share of employment accounted for by professional specialty occupations. Finally, the evidence for Jacobs-style spillovers is mixed. We found evidence favorable to Jacobs in both the full sample and in the MSA sample, but this evidence is not robust to corrections for spatial dependence in the full sample.

CONCLUSION

In this paper, we show that R&D activity for most industries tends to be concentrated in the Northeast corridor, around the Great Lakes, in California's Bay Area, and in southern California. We hypothesize that the relatively high spatial concentration of R&D activity is due to the fact that R&D depends on knowledge spillovers more than most types of business activities. We find evidence that supports this view. We run a regression of a variant of the EG index measuring the spatial concentration of private R&D labs at the county level on geographic proxies for knowledge spillovers as well as other characteristics and find evidence that localized knowledge spillovers are important for the spatial clustering of innovative activity. In particular, we show that a strong positive correlation exists between the geographic concentration of R&D labs and citation-weighted patents per worker in professional and specialty occupations (a proxy for knowledge spillovers). All else equal, the index of agglomeration for R&D labs will be 15 percent greater in a county with twice the citations intensity of another county. In addition, we find evidence that supports the idea that Jacobs externalities are important for R&D activity, although the evidence is mixed. By far the most powerful effect is generated by human capital (the share of the employment accounted for by professional specialty occupations). This variable is positive and highly significant and it's economically significant too—an elasticity of 3.5.

Policymakers view the success of areas such as Silicon Valley in California, the Route 128 corridor in Boston, and North Carolina's Research Triangle as a miraculous recipe for local economic development and growth. But are these examples exceptions rather than the rule? The answer appears to be no. We show that equally remarkable concentrations may be found in many other types of R&D activity, such as the concentration of R&D in the pharmaceutical industry in Northern New Jersey and Southeastern Pennsylvania. In this article, we show that many types of R&D establishments are highly concentrated geographically. However, studies by Saxenian (1994) and Duranton (2008) provide a cautionary note for policymakers who view the success of areas such as Silicon Valley as a recipe for local economic development and growth. While investing in science centers to attract R&D activity is fairly common in the U.S., Saxenian's study suggests that creating the right corporate culture to make the centers successful is more challenging. Duranton (2008) points out that designing cluster policy is extremely tricky, and even if policymakers get it right, often the benefits of clustering may be too small to justify the cost of bringing them about. We suggest that policymakers, instead of targeting industries, consider strategies that help to establish a good business environment and the efficient provision of locally provided public goods and services.

REFERENCES

Agrawal, Ajay, Devesh Kapur, and John McHale. "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data," *Journal of Urban Economics*, Vol. 64 (2008), pp. 258-69.

Anselin, Luc, Attila Varga, and Zoltan. Acs. "Local Geographic Spillovers between University and High Technology Innovations," *Journal of Urban Economics*, 42 (1997), pp. 442-48.

Anselin, Luc, S. Hudak. "Spatial Econometrics in Practice: A Review of Software Options," *Regional Science and Urban Economics* 72 (1992), 509-36.

Anselin, Luc. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Boston, 1988.

Arrow, Kenneth J. "The Economics of Learning by Doing," *Review of Economic Studies*, 29 (1962), pp. 155-73.

Arzaghi, Mohammad, and J. Vernon. Henderson. "Networking Off Madison Avenue," unpublished manuscript (2005).

Audretsch, David B., and Maryann P. Feldman. "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86 (1996), pp. 630-40.

Carlino, Gerald A., Chatterjee and Robert M. Hunt. "Urban Density and the Rate of Invention," *Journal of Urban Economics*, 61 (2007), pp. 389-419.

Combes, Pierre-Philippe, and Henry G. Overman. "The Spatial Distribution of Economic Activities in the European Union," in J.V. Henderson and J.-F Thisse, eds., *Handbook of Regional and Urban Economics, Vol. IV: Cities and Geography*. Amsterdam: Elsevier, 2004.

Directory of American Research and Technology, 23rd ed. New York: R.R. Bowker, 1999.

Duranton, Gilles. "California Dreamin': The Feeble Case for Cluster Policies," Working Paper, University of Toronto (2008).

Duranton, Gilles, and Henry G. Overman. "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 72 (2005), pp. 1077-1106.

Ellison, Glenn, and Edward. L. Glaeser. “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 105 (1997), pp. 889-927.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns,” Discussion Paper 2133, Harvard Institute of Economic Research (April 2007).

Feldman, Maryann P., and David B. Audretsch. “Innovation in Cities: Science-Based Diversity, Specialization, and Localized Competition,” *European Economic Review*, 43 (1999), pp. 409-429.

Fujita, Masahisa, Tomoya Mori, J. Vernon Henderson, and Yoshitsugu Kanemoto. “Spatial Distribution of Economic Activities in Japan and China,” in J.V. Henderson and J.-F Thisse, eds., *Handbook of Regional and Urban Economics, Vol. IV: Cities and Geography*. Amsterdam: Elsevier, 2004.

Gaspar, Jess, and Edward Glaeser. “Information Technology and the Future of Cities,” *Journal of Urban Economics*, 43 (1998), pp. 136-56.

Glaeser, Edward, Hedi Kallal, Jose Scheinkman, and Andrei Shleifer. “Growth in Cities,” *Journal of Political Economy,* 100 (1992), pp. 1126-53.

Guimarães, Paulo, Octávio Figueiredo, and Douglas Woodward. “Measuring the Localization of Economic Activity: A Parametric Approach,” *Journal of Regional Science*, 47 (2007), pp. 753-44.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, “Market Value and Patent Citations,” *RAND Journal of Economics*, 36 (2005), pp. 16-38.

Harhoff, Dietmar, Francis Narin, F. M. Scherer, and Katrin Vopel. “Citation Frequency and the Value of Patented Inventions,” *Review of Economics and Statistics*, 81 (1999), pp. 511–15.

Holmes, Thomas J., and John J. Stevens. “Spatial Distribution of Economic Activities in North America,” in J.V. Henderson and J.-F Thisse, eds., *Handbook of Regional and Urban Economics, Vol. IV: Cities and Geography*. Amsterdam: Elsevier, 2004.

Jacobs, Jane. *The Economy of Cities*. New York: Vintage Books, 1969.

Jaffe, Adam B., Manual Trajtenberg, and Michael S. Fogarty. “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors,” *American Economic Review*, 90 (2000), Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association, pp. 215-18.

Jaffe, Adam B., M. Trajtenberg, R. Henderson, Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* 108 (1993) pp. 577-98.

Keller, Wolfgang. "Geographic Localization of International Technology Diffusion," *American Economic Review*, 92 (2002), pp. 120-42.

Kolko, Jed. "Agglomeration and Co-Agglomeration of Services Industries," unpublished manuscript (April 2007).

Krugman, Paul. *Geography and Trade*. Cambridge: MIT Press, 1991.

Lin, Jeffery. "Innovation, Cities, and New Work," Federal Reserve Bank of Philadelphia Working Paper 07-25, October 2007.

Marshall, Alfred. *Principles of Economics*: London, Macmillan, 1890.

Romer, Paul M. "Increasing Returns and Long Run Growth," *Journal of Political Economy,* 94 (1986), pp. 1002-37.

Rosenthal, Stuart, and William C. Strange. "The Determinants of Agglomeration," *Journal of Urban Economics*, 50 (2001), pp. 191-229.

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

Strange, William C. "Viewpoint: Agglomeration Research in the Age of Disaggregation," *Canadian Journal of Economics*, 42 (2009), pp. 1-27.

Table 1: Concentration of R&D Labs for Selected Industries

INDUSTRY	NUMBER OF LABS	Adjusted GFW Index[a]
Highly Concentrated Industries[b]		
Oil & Gas Field Machinery	22	0.33
Tires and Tubes	14	0.12
Computer Storage Devices	34	0.08
Motor Vehicles & Car Bodies	26	0.06
Electronic Computers	57	0.06
Moderately Concentrated Industries[c]		
Semiconductors	278	0.03
Prepackaged Software	359	0.03
Motor Vehicles Parts	134	0.03
Optical Instruments and Lenses	36	0.03
Radio and TV Communication Equipment	185	0.02
Dispersed Industries[d]		
Search, Detection, Navigation	155	0.01
Paints, Varnishes, etc.	131	0.01
Refrigeration, Heating Equipment	36	0.00
Printed Circuit Boards	64	0.00
Electronic Connectors	66	0.00

[a] The adjusted GFW index for a given industry shows the sum of the squared differences of the share of employment in manufacturing from the share of labs in given industry, adjusted to account for the industrial organization of the industry under consideration.
[b] R&D in an industry is highly concentrated if the adjusted Gini coefficient is at least 0.05.
[c] R&D in an industry is moderately concentrated if the adjusted Gini coefficient is at least 0.02, but less than 0.05.
[d] R&D in an industry is dispersed if the adjusted Gini coefficient is less than 0.02.

Table 2: Determinants of Spatial Concentration of R&D Labs†

	Robust OLS		
	(1)	(2)	(3)
	All Counties	All Counties	MSA Counties
Herf. Index	-0.0012 (1.79)*	-0.0004 (0.79)	-0.0947 (2.74)***
Professional Share,1990	0.4086 (4.37)***	0.2633 (2.88)**	0.2543 (2.60)***
Citation-Weighted Patents, 1990	0.6294 (2.28)**	0.5030 (1.97)**	1.3145 (1.65)*
University R&D per Student, 1987-1989	0.0001 (1.39)	0.0001 (1.37)	0.0001 (1.40)
MSA Dummy		0.0129 (3.33)***	
No. of Obs.	3108	3108	752
R^2	0.0344	0.0401	0.0469

† All regressions include census region dummy variables (New England is the excluded region) and the share of county employment in manufacturing.
*,**,*** indicates significance at the 10 percent, 5 percent, and 1 percent levels.

Table 3: Spatial Dependence Tests[a] (P-values)

	All Counties		MSA Counties	
Test for:	Spatial Error	Spatial Lag	Spatial Error	Spatial Lag
Moran's I $\lambda = 0$	0.0144		0.6330	
LM - $\lambda = 0$	0.059		0.4908	
Robust LM- $\lambda = 0$	0.1301		0.8781	
LM - $\rho = 0$		0.025		0.5015
Robust LM- $\rho = 0$		0.054		0.9827

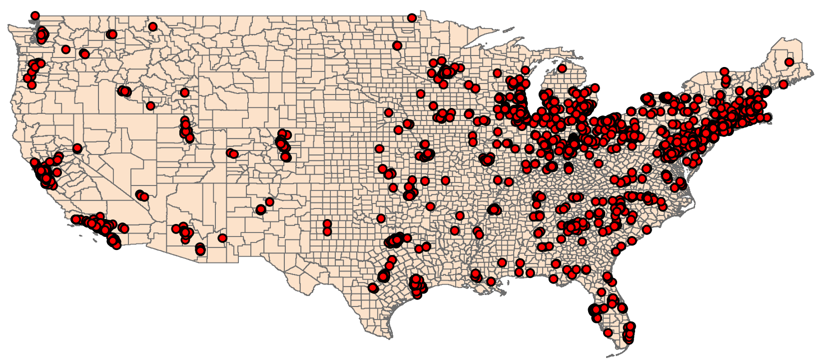
Notes: N = 3108. [a] Moran's I is based on standardized z-values that follow a normal distribution. The Lagrange multiplier (LM) tests are distributed as χ_1^2 with critical levels of 3.84 (p = 0.05).

Table 4: Correcting for Spatial Effects

	(1)	(2)
	All Counties	All Counties
	Spatial Error	Spatial Lag
Herf. Index	-0.0012 (0.473)	-0.0012 (0.468)
Professional Share,1990	0.4000 (5.51)***	0.3950 (5.50)***
Citation Weighted Patents, 1990	0.6028 (1.72)*	0.5960 (1.70)*
University R&D per student, 1987-1989	0.0001 (2.97)**	0.0001 (2.99)**
λ	0.1306 (2.19)**	
ρ		0.1456 (2.51)**
No. of Obs.	3108	3108
R^2	0.0358	0.0369

† All regressions include census region dummy variables (New England is the excluded region) and the share of county employment in manufacturing.
*,**,*** indicates significance at the 10 percent, 5 percent, and 1 percent levels .

Figure 1
Location of Total
R&D Labs



Each dot represents a spatial concentration of overall R&D labs.

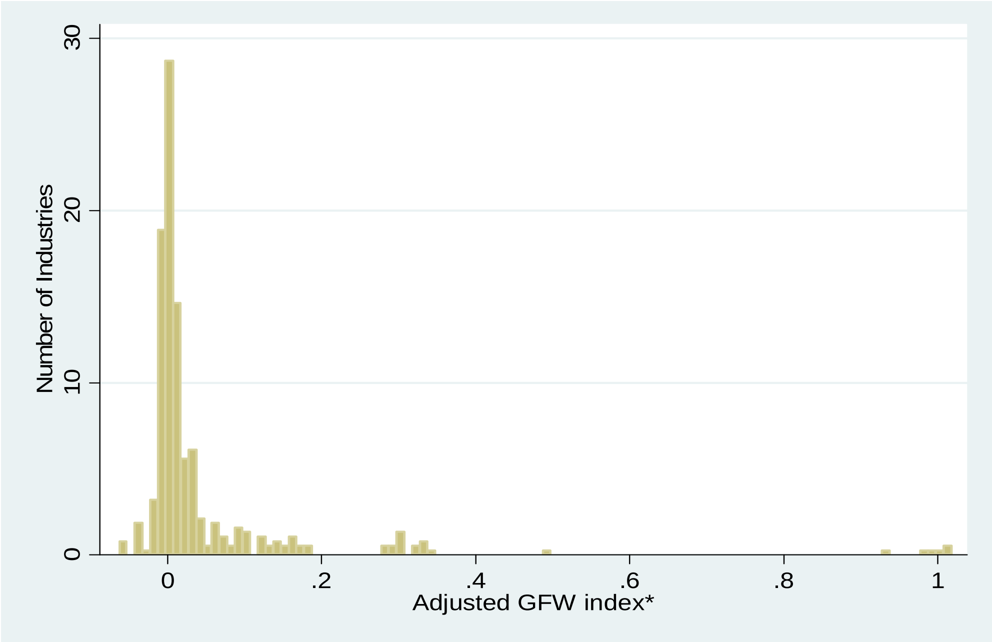


Figure 2. Histogram showing the frequency distribution of the adjusted GFW index that compares the concentration of R&D labs in a given industry with the concentration of manufacturing employment.

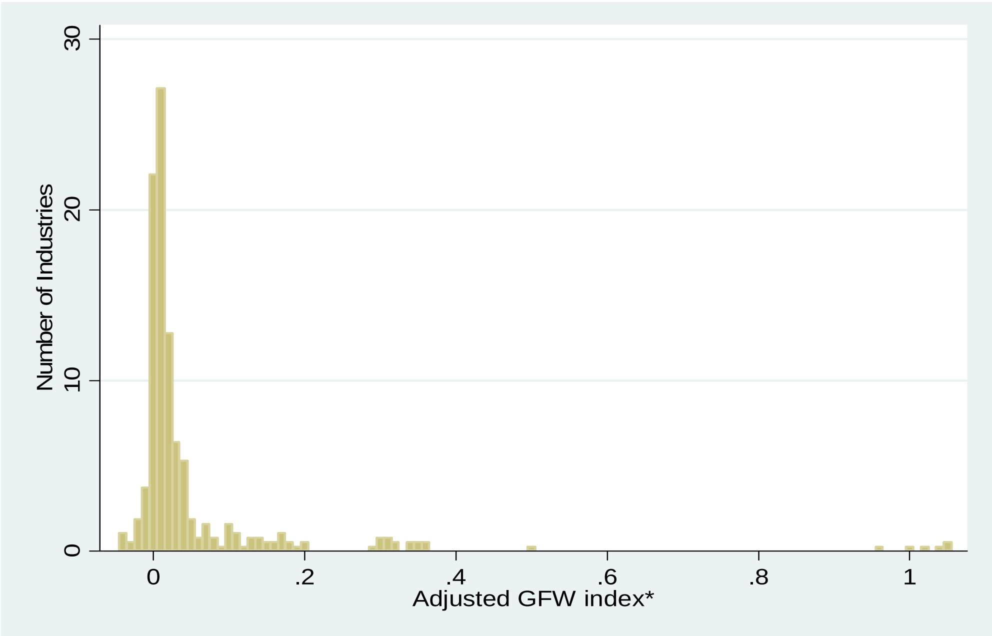


Figure 3. Histogram showing the frequency distribution of the adjusted GFW index that compares the concentration of R&D labs in a given industry with the concentration of overall labs.

The adjusted GFW index shows the sum of squared differences of the share of employment in manufacturing (Figure 2) or the share of overall labs (Figure 3) from the share of labs in a given industry, adjusted to account for industry structure.

Figure 4
Location of Software
R&D Labs

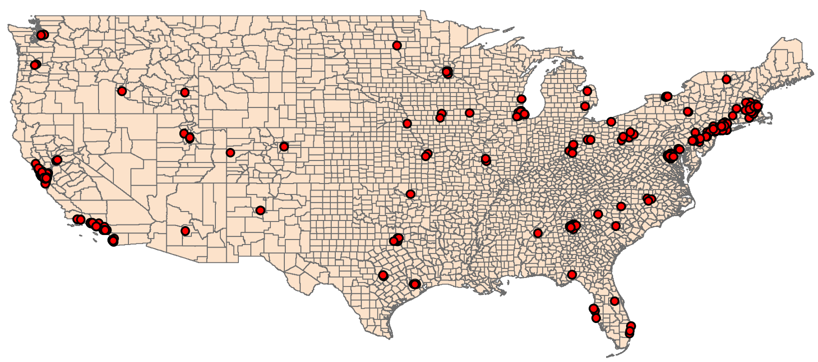


Figure 5
Location of Pharmaceutical
R&D Labs

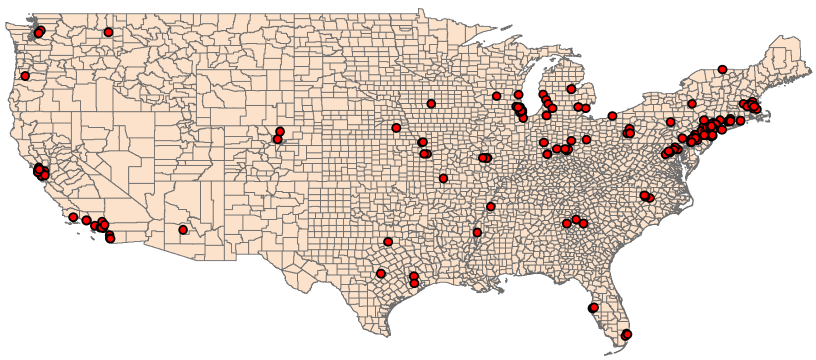


Figure 6
Location of Chemistry
R&D Labs

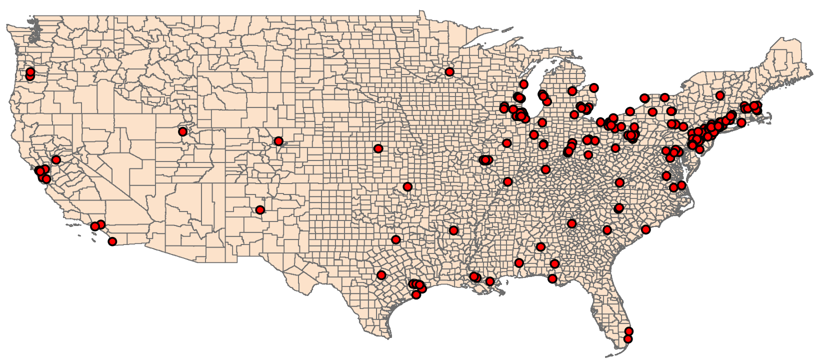


Figure 7
Location of Oil and Gas Field Machinery
R&D Labs

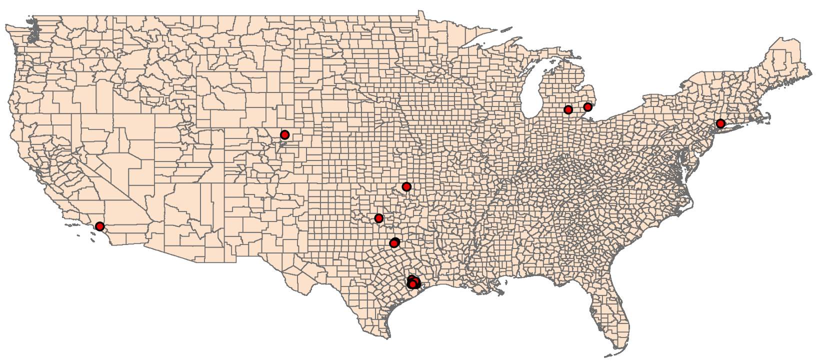


Figure 8
Location of Motor Vehicle and Car Body
R&D Labs