Journal of Economic Geography **9** (2009) pp. 1–31
Advance Access Published on 29 October 2008
doi:10.1093/jeg/lbn042

Spatial diversity in invention: evidence from the early R&D labs

*Tom Nicholas**

Abstract

This article uses historical data on inventor and firm research and development (R&D) lab locations to examine the technological and geographic structure of corporate knowledge capital accumulation during a formative period in the organization of United States innovation. Despite the localization of inventive activity around the labs, one-quarter of inventors lived outside a 30 mile commuting radius of the nearest facility of the firm they assigned their patents to. A strong positive effect of distance from a lab on technological importance is identified, especially for inventors from large cities that were geographically separated from a firm's labs. A patent case–control method helps explain spatial sourcing by showing that the average quality of externally available inventions was high. Firms selected complementary, not substitute, inventions from non-lab urban locations, suggesting a link between the organization and the geography of innovation.

Keywords: R&D, invention, location, cities
JEL classifications: O31, O32, O14, R12
Date submitted: 8 December 2007 **Date accepted:** 30 September 2008

1. Introduction

One of the central findings in the literature on the geography of innovation is that location-based externalities significantly contribute to technological progress. Following Alfred Marshall's (1890) seminal insights, Jaffe et al. (1993) illustrate the presence of localized knowledge spillovers using late 20th century patent citations and Glaeser et al. (1992) find that the increasing returns which drive economic growth stem from knowledge spilling over between industries in diversified cities. The benefits of physical proximity at the country-level and within-country regions appear to be quite persistent over time even as modern transportation and communication advances have decreased geographic barriers to interaction (Keller, 2004; Thompson, 2006). Spatial agglomeration continues to be a target of economic policy for research and development (R&D) because closeness acts as a conduit through which tacit knowledge about innovation gets transferred (Griffith et al., 2007).

Yet, while geographic proximity facilitates the transfer of knowledge and is undoubtedly a key driver of innovation and economic growth, the literature has placed much less of an emphasis on the fact that firms frequently source their inventions more widely than any bounded geographic location. Corporations may have multiple locations, or engage in the external market for technology thereby increasing the

*Entrepreneurial Management Group, Harvard Business School, Boston MA, 02163, USA.
email <tnicholas@hbs.edu>

geographic scope of their inventive activity beyond a single area. Although research on present day corporate R&D has shown that firms are increasingly sourcing innovations at a distance (e.g. Huston and Sakkab, 2006), we have remarkably little empirical evidence on this spatial aspect of innovation at the firm-level, or its influence on the accumulation of corporate knowledge capital. Do firms acquire inventions across different bounded areas? If they do, what is the technological significance of these inventions and where are they selected from geographically?

This article attempts to answer these questions using a unique historical data set of almost 18,000 patented technologies assigned to 69 firms operating 94 R&D laboratories in 1920s America. The historical setting is especially useful because it provides a test of how inventive activity was spatially distributed when technological development was focused around laboratory locations, but where external markets for innovation were also important (Lamoreaux and Sokoloff, 2002). The concentration of invention in-house was one of the most significant structural changes to affect the organization of business (Mowery, 1990; Chandler, 1990) and this laid the foundations for United States industrial leadership (Nelson and Wright, 1992; Field, 2003). The functioning of traditional corporate laboratories has attracted a considerable degree of attention in the literature and they are frequently referenced in discussions on how firms organize for innovation today (MacGarvie and Furman, 2005; Lerner and Wulf, 2007; The Economist, 2007). The literature, however, is overwhelmingly 'organizational' in orientation with a focus on the integration of R&D within the boundaries of the firm, decision making or incentives (Rosenbloom and Spencer, 1996). A link between the organizational structure of R&D and the geography of innovation has been overlooked.

Since over 90% of the inventions covered in the current data set originated in the United States, the focus here is on the spatial pattern of domestic and not international invention. Notwithstanding modern multinational firms transfer technology across international borders, and information about innovation is beginning to diffuse faster internationally (Griffith et al., 2007), the bulk of technological development stemming from the transmission of ideas tends to be localized at the country level (Thompson, 2006). I use discrete address matches by country, state, county and Standard Metropolitan Area (SMA) to determine the spatial distribution of inventors with respect to the location of R&D labs of the firms they assigned their patents to.[1] Given the problem of aggregation bias when matching at the state and SMA level, due to differences in the units of measurement of geographic areas,[2] I also geocode the addresses of inventors and labs. Differences between inventor-lab location pairings within the United States are then calculated and analysed by mileage.

I use three patent-based metrics to explore the relationship between an invention and distance from a lab. First, I exploit a novel set of historical patent citations, taking

1 SMAs were first specified in 1950 by the Bureau of the Budget. The definition of these geograpgic areas has changed over time. The current term is 'core based statistical area' (CBSA), which became effective in 2000 and refers to both metropolitan and micropolitan statistical areas.

2 Duranton and Overman (2005) point out that traditional metrics based on geographic area matches can be biased. Thus in their words: 'analysing the localization of industries at the level of US states involves comparisons between Rhode Island and California, which is geographically more than 150 times as large'. (p. 1079).

advantage of the fact that many modern patents cite old patents as prior art. The idea is rooted in Trajtenberg's (1990) finding that citations can be used as a measure of a patent's importance or value. Alfred Marshall was aware of the type of technological dependence I measure when he commented that 'an invention that makes an epoch is very often a generation older than the epoch that makes it' (1890, 206). I show that historical citations are correlated with modern citations and can therefore be used to proxy for the quality of an invention. Second, I measure the generality of inventions based on the distribution of patent classes for citing patents to distinguish between inventions that influenced a wide range of subsequent technology fields from those that did not. Finally, following Jaffe (1986), I construct a technical proximity metric to test for differences in technology space between laboratory and non-laboratory inventions.

Using the location data, I show that despite the strong geographic concentration of inventive activity around the research labs, approximately one-quarter of inventors lived outside a 30 mile commuting distance from the nearest in-house facility of the firm they assigned their patents to. The 30-mile radius is established as a maximum distance for home-work separation based on contemporary commuting surveys. Although employment or non-employment of inventors cannot be observed for the whole sample, information on General Electric, the largest patent assignee in the data verifies the one-quarter share. Using personnel records to determine inventor-employment status shows that around 20% of patented inventions at the time were sourced from independent inventors operating outside of the firm's boundaries. The one-quarter share for the whole sample is the main descriptive statistic for the analysis and is almost double the share of inventors observed at this distance when tracing the leading R&D firms in the data set forwards in time to the late 20th century. At the within-country level, a quantitatively significant share of inventions were sourced from a distance by historical firms with R&D labs.

Adopting a baseline 30-mile radius around the labs, I test for differences in the characteristics of innovation for laboratory and non-laboratory inventions. Using the historical citations metric, I show that inventors who resided >30 miles away from a lab produced proportionately higher quality inventions than inventors working inside the labs. The expected number of historical citations to their patents increased by 19–21% relative to patents by inventors from laboratory locations. This effect, in turn, is associated with inventors who were outside the 30-mile radius, but close to a large city (more than 100,000 inhabitants) geographically remote from a firm's labs. Robustness checks show the result is not driven by a dominant non-lab location such as New York city. The finding is robust when shrinking the 30-mile radius to control for the location of the most productive laboratory inventors,[3] and for potential spillovers in the immediate vicinity of the R&D labs. The higher quality of inventions from distant cities also persists in the data relative to inventions from the hinterlands of the labs, therefore highlighting a difference in the quality of patents acquired from less local versus local lab regions.

Given this main result, an important empirical challenge is to distinguish between firms selecting only the high-quality distant inventions in the tail of the citation

3 Surveys reveal that residential proximity to the work place increased with the seniority of the technologist (Schnore, 1960).

distribution and the average quality of distant invention being high. For this purpose, I use a patent case–control method to match case patent—those sourced from outside the 30-mile lab radius—against control patents by individual inventors that were not sourced by R&D firms. The intuition follows Jaffe et al. (1993) use of a case-control method to identify localized spillovers from innovation. Whereas they look for closer discrete geographic matches between citing patent pairs relative to non-citing patent pairs within the same technology category, I look for differences in case–control patent historical citation counts. In accordance with Thompson and Fox Kean (2005), my matching method takes into account both patent class and sub-class to minimize the likelihood of confounds and additionally geographic area matching. I find that distant inventions acquired by the firms in the sample had insignificantly different historical citation counts compared to matching patents that were not sourced by the R&D firms. This result holds when establishing a patent quality threshold by excluding patents with zero citation counts. The results show that the average quality of sourced and non-sourced inventions was very high. A high-quality pool of external innovation helps to explain why inventions were acquired by R&D firms from geographically distant areas at this time.

Additional evidence makes a link between the organization and the geography of innovation. I show that firms were selecting inventions at a distance based on their technological characteristics relative to their overall technology profile. I find a negative effect of being outside the 30-mile radius, but close to a spatially separate large city, on the technical proximity of firm and inventor patents, with only patents sourced from New York being technologically more proximate. Furthermore, the patents from inventors close to all large cities were significantly more general in scope than those originating from a firm's in-house facilities according to the sectoral distribution of their citation counts. The technical proximity and generality metrics suggest that R&D firms as organizational entities were selectively choosing geographically distant innovations to augment complementary rather than substitute stocks of corporate knowledge capital.

Overall, the results show that while innovation was clustered at the lab-level, the high average quality of inventive activity outside the R&D facilities was an important driver of spatial diversity in the sourcing of invention. A link between the organizational and geographic structure of innovation helps to explain why sourcing technological innovations at a distance was economically important at this time. Firms with R&D labs strategically acquired inventions across bounded geographic locations. More broadly, the findings relate to the long run cycle associated with the corporate R&D function. As the 20th century progressed, technological development became increasingly internalized within the labs, but present day R&D has shifted in favour of dispersed innovation given the stocks of knowledge existing outside the boundaries of the firm (The Economist, 2007). From an historical perspective, the results presented here illustrate the prominence of externally available technological knowledge and illustrate that firms do benefit from spatial diversity in innovation.

The remainder of the article is organized as follows. Section 2 frames the research questions based on a brief historical background to the R&D labs. Sections 3 and 4 cover the data set and outline the method for measuring the geography of invention and the technological characteristics of patents. Section 5 discusses the empirical specifications, Section 6 presents the results and Section 7 concludes.

2. Historical background

R&D laboratories have a long history with important examples of corporate labs including those founded by Thomas Edison in Menlo Park, and Alexander Graham Bell in Boston, both in 1876. In-house facilities became more common with the complex new knowledge of the Second Industrial Revolution and they were central to the organizational transformation of American corporate enterprise during the early 20th century (Chandler, 1990; Peretto, 1998). Like industry as a whole, the labs were heavily clustered in the east coast manufacturing belt (Krugman, 1991) and growth was rapid during the time period of this study. Between 1921 and 1927, the number of scientists and engineers employed in industrial research laboratories more than doubled from 2775 to 6274 (Mowery and Rosenberg, 1998, 21–22).

The economic decision to organize R&D in-house revolved around the shift of innovation towards more complex capital intensive projects, problems of contracting in the market for technology due to asymmetric information and the desire to retain intellectual property rights on inventions (Mowery, 1990; Fisk, 1998). Furthermore, it was thought that the R&D labs would have advantages as repositories of knowledge, information and skills as inventors were hired to facilitate learning. The literature on the economics of science frames these advantages as externalities arising from location (Stephan, 1996). In the spirit of Marshall (1890), it is well-known that geographic proximity facilitates the flow of ideas thereby acting as a catalyst to the diffusion of innovation. Storper and Venables (2004) present a general model of face-to-face contact that can rationalize the positive benefits of closeness on innovation when the type of information being transmitted is non-codifiable and when interactions take place in high-density urban locations.

However, while Marshall described the positive externalities arising from environments where tacit knowledge can be communicated, the disadvantages of closeness have also been emphasized. Schumpeter argued that the R&D labs encouraged 'routinization' thereby stifling creative invention and the 'romance of earlier commercial adventure' (1950, 132). Writing during the heyday of the R&D labs, Gilfillan (1935) cautioned that localized knowledge undermined the creativity needed for making path-breaking inventions. Grosvenor (1929) provided empirical support showing that 83% of the most significant technological developments between 1889 and 1929 were made by independent inventors. Reflecting back on the early labs, Baumol (1993) suggests that they were much more likely to produce incremental rather than radical technological breakthroughs due to their focus on modifications or improvements to existing product lines.

One way that firms could simultaneously benefit from location-based externalities using labs and from externally generated ideas was to engage in the market for technology. Mowery (1995, 148) comments that in-house scientists and engineers, in addition to being active in commercial innovation within the firm, 'monitored technological developments outside of the firm' advising corporate managers on what technologies to acquire. That open market transactions constituted an important source of innovation for firms is highlighted by the fact that many corporations such as AT&T and Eastman Kodak maintained patent departments that engaged in the trade for invention (Reich, 1980). According to Kim (1995) regional specialization reached a high point in the United States during the early 20th century, which fits in with the idea that firms may have sourced innovations across multiple geographic locations in order to access different pools of technological knowledge.

Trade in invention during the early 20th century was aided by the institution of the US patent system, which allowed a wide spectrum of the population to protect their intellectual property rights (Khan and Sokoloff, 2004). By 1925, it was 10 times more expensive to carry a patent to full term in Britain than in the United States (Lerner, 2002), such that the relatively low cost of patenting in the United States acted as a spur to the democratization of invention (Khan, 2005). Furthermore, intermediaries such as patent lawyers and agents created a nexus of efficient contracting institutions connecting firms and inventors within and across geographic locations (Lamoreaux and Sokoloff, 2002). Transactions of this sort occurred to the extent where inventors outside of established firms made revolutionary contributions to technological progress.

3. The data

Despite important pools of knowledge existing outside of firms, we have no systematic quantitative evidence on the magnitude of external invention or the mechanisms by which distant knowledge capital was selected by firms. In the following two sections, I show how these gaps in our understanding of innovation at a distance and organized technology formation can be filled using data on corporate laboratories listed by the National Research Council (NRC) and both the locations of inventors from patent documents and the technological characteristics of their inventions.

3.1. R&D lab locations

Organized in 1916, the NRC surveyed and reported on research laboratories in the United States for various snapshot years between 1920 and 1985.[4] The NRC conducted two main surveys in the 1920s—in 1921 and 1927—which list approximately 500 and 1000 laboratories, respectively.[5] To assemble a data set of firms with corporate R&D facilities, I merged the NRC in-house labs against the 135 firms documented in *Moody's Manual of Industrials* as in Nicholas (2008). These data include every publicly traded United States corporation that was systematically covered by *Moody's* during the 1920s, capturing the most important in-house R&D firms such as General Electric, Eastman Kodak, AT&T and Du Pont as well as less research intensive concerns such as Otis Elevator and the Diamond Match Company. The matching method under samples small facilities in the NRC data, for which R&D was less central to the firms

4 In the early reports, the NRC's approach was to acquire information through direct correspondence and cross-check the accuracy of the data with the firm/laboratory once a draft list had been drawn up. Although some firms did not answer the NRC's questionnaire and others refused to be included in the publication, the NRC concluded that the survey provided a fairly complete picture of in-house R&D activities in America. The early surveys include information on the company affiliation and location of labs, the name of the research director, the number of staff and the main lines of research activity. The surveys reflect industrial laboratories. Governmental and educational facilities are excluded.

5 The first survey in 1920 contained 300 laboratories. I use the 1921 survey because it has a more comprehensive coverage.

Table 1. R&D lab sample descriptive statistics

	1921 Data	1927 Data
Number of firms	54	69
Number of research labs	57	94
Date research started	1909 (10.93)	1911 (11.76)
Research staff	103.74 (284.96)	107.02 (308.66)
Percent of total US research staff	48.6	36.0
Number of patents	16,506	17,620
Individual Inventors	4726	5186

Notes: Standard deviations in parentheses. If not stated otherwise, statistics are means.
Source: Research employment figures are taken from the 1921 and 1927 National Research Council surveys. The date that research started is from the 1946 survey.

functional activities. Approximately 52% of the labs in the 1921 survey and 42% in the 1927 survey list five or fewer research employees.

An initial match of the data revealed 41 of the 135 firms with in-house facilities in 1921 and 64 with in-house facilities in 1927. One reason for the difference between these years is the growth of in-house R&D during the 1920s and the inclusion of additional labs. Another reason is that the accuracy of the surveys increased over time and therefore labs that did exist are more likely to be documented as time goes by.[6] To address this problem, I used the 1946 NRC survey, which is both more complete and was the first to include the year that research activity started. Putting the data from the 1921, 1927 and 1946 surveys together increased the number of matches. Descriptive statistics are given in Table 1. These show that 54 of the 135 firms had research laboratories in 1921 and 69 firms had laboratories in 1927 with some firms operating more than one lab.[7] The firms included in the data set accounted for 49% and 36% of total research employment in 1921 and 1927, respectively. The average commencing year for a lab's operations was 1910.

3.2. Inventor locations

From the *Official Gazette* of the United States Patent and Trademark Office, I collected all the patents that were assigned to the firms with one or more research laboratories for the decade of the 1920s. I then restricted the sample to those patents that could have originated from a lab by eliminating any patent with a filing date prior to the establishment of an R&D facility at the firm. This resulted in 16,506 matching patents for the 1921 sample and 17,620 for the 1927 sample.

Despite some drawbacks, patents are a well-documented measure of innovation (Griliches, 1990) and are particularly suited to spatial analysis because the locations

6 For example, J.I. Case Threshing Machine established a laboratory in 1893 but it is missing from both the 1921 and 1927 surveys.

7 Some of the labs that existed in 1927 undoubtedly also existed in 1921, but since the 1946 survey does not include the year that research activity in a lab started for all firms, I only made corrections where I could be sure about chronology.

Table 2. Inventor locations for patents assigned to general electric, 1927–1929

Location of inventor	Patents (%)
Any GE Laboratory	23.4
Schenectady Lab	17.4
Other Lab	6.0
Non-Laboratory	76.6
GE division	56.5
GE division[1]	[41.6]
Independent inventor	20.1
Independent inventor[2]	[4.9]

Notes: Statistics reflect matches between patent holders and employment locations as given in editions of GE's *Organization Directory*. Figures in squared brackets are adjusted percentages of the ones immediately above: [1]inventors employed in a GE division collocated with an in-house lab. [2]Not including inventors as independent if they resided in the same or adjacent city to a lab.

of inventors are documented on the front page of the patent. Notwithstanding some inventors may have changed location while a patent application was processed, the patent residential address is a commonly used guide to the geographic location of inventors. United States patent law requires that a patent must be granted to an individual, who can subsequently keep or assign these rights in part or in whole to another individual or to a firm. All of the assignments considered here were made as of the patent grant date.[8]

Inventors who assigned their patents to firms fall into two categories: they were either employees and affiliates of the firm such as consulting engineers or they were independent inventors who had sold their intellectual property rights. Although patent records do not reveal occupational identity, a unique insight into the two categories of inventors can be gained from a rich set of personnel records that exist for General Electric, one of the most prolific patenting companies at the time. Table 2 provides descriptive statistics on inventor locations for 1208 General Electric patents granted to US-based inventors between 1927 and 1929 matched against editions of the company's *Organization Directory* from 1925 to 1929, which list all the main employees plus consulting and academic affiliates across the entire spectrum of the firm's functional activities.[9]

Table 2 summarizes the data by lab and non-lab location. Lab locations are disaggregated into shares of inventors employed at the main Schenectady lab or a peripheral lab. Non-lab locations are split between General Electric divisions and the

8 While patents could be assigned to another party after their grant date, this practice was less common by the early 20th century. Lamoreaux and Sokoloff (1996) consider post grant date assigments for their well-known list of B-inventors, but it is not possible to do this given the scale of the present study.

9 I use a lag of 2 years (between 1925 and 1927) for the matching because this was the average time lag of the patent application process. Thus, an employee listed in the *Organization Directory* in 1925 would have had a good chance of having their patent granted by 1927.

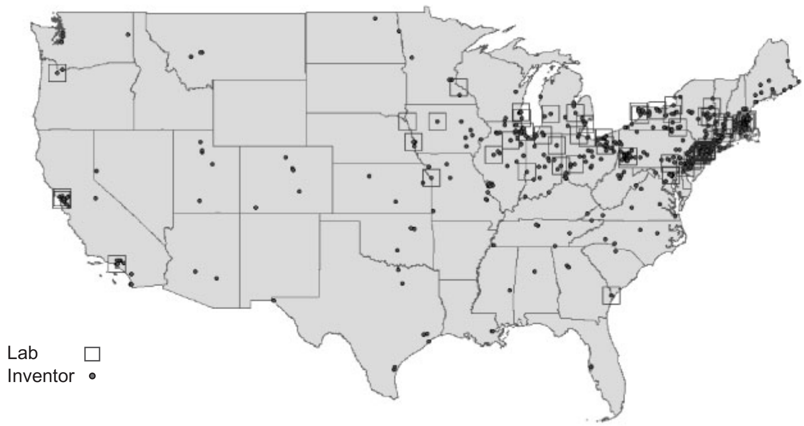

Figure 1. Geographic location of inventors and labs
Notes: Laboratory and inventor locations are geocoded addresses. Lab locations are for 1921 and 1927.
Source: NRC surveys and front-page patent addresses.

residential locations of independent inventors. Independent inventors are defined as patent holders assigning to General Electric who were not listed in the *Organization Directory* or complementary sources as employees or affiliates of the firm.

A surprising finding from Table 2 is that the vast majority of patented inventions were not granted to employees working in the in-house laboratories, although over half of the non-laboratory patents were granted to individuals working in a General Electric division collocated with an in-house lab (i.e. a division and a lab within the same plant). While the majority of innovation was spatially concentrated and inventive activity tended to increase with geographic proximity to an R&D facility, an important share of the firm's patentable assets came from outside the organizational boundaries of the labs. Independent inventors accounted for as much as 20% of patents granted to General Electric in the late 1920s. Figure 1 shows that General Electric was not unusual in the diverse geographic location of its inventions. It not only illustrates a marked concentration of both R&D labs and inventor activity in the prominent east coast and midwestern manufacturing belts, but also illustrates that inventors could be highly de-localized from a corporate research facility.[10]

3.3. A 30-mile radius

Given the spatial diversity identified in Figure 1, a key issue is to distinguish an area within which inventors may have been influenced by technological development taking

10 Some were independent inventors such as Walter T. Oxley who assigned his photographic printing machine patent (number 1,332,854) to Eastman Kodak, but lived in the small town of Fergus Falls, Otter Tail County, Minnesota 930 miles from the firm's Rochester laboratory. Others were employed in a non-lab division of the firm, such as Eugene E. Valk, an engineer in General Electric's Los Angeles office who invented an overload relay to safeguard electrical systems against short circuits or power surges (number, 1,347,767).

Table 3. Commuting distances for plant 'x' in upstate New York, 1921–1930

Distance between residence and workplace (miles)	Percent distribution of employees		
	1921	1925	1930
0–4.9	85	86.8	81.6
5–14.9	9.3	9.2	12.3
15–19.9	3.5	2.5	3.7
20 and over	2.2	1.5	2.4

Notes: Percentage distributions taken from Adams and Mackesey (1955, 15). The plant is anonymous because commuting distances were calculated from confidential management records.

place in the labs. For this purpose, I estimate the maximum distance to work. The basic assumption is that if inventors engage in search for new technologies they may be either directly influenced by the innovative activity of the lab through their employment status, or indirectly influenced through spillovers of technological knowledge.

I map a radius based on contemporary surveys on commuting.[11] An important study on commuting, which is summarized in Table 3 shows the percentage distribution of workers by distance travelled for an anonymous large manufacturing plant in Upstate New York. The data show a high degree of consistency for this particular plant during the 1920s. Over 85% of the workers resided <5 miles away from the plant with greater distances showing a large decline in the percentage of worker residencies. A small fraction of workers commuted to work from >20 miles away.

Further surveys reveal that employees with higher seniority tended to live closer to the work place (Schnore, 1960). This is verified by the locations of inventors working in General Electric's prestigious Schenectady lab. Using data in the firm's *Organization Directory* and patent residential addresses, I find that 97% of engineers and scientists resided within 5 miles of the lab with a maximum distance to work by one employee (Wesley F. Masey of Ballston Spa, New York) of 13.5 miles. Guided by both the commuting survey data and the specific information on Schenectady lab workers, I define a conservative area of localization as a radius of 30 miles around the in-house laboratories in the sample.

To measure distances between inventors (*INV*) and labs (*LAB*) in order to match them according to this radius, I geocoded the data and calculated the great circle distance (in miles) between the coordinate latitude (*lat*) and longitude (*lon*) pairs. Distance (D) of inventor i from lab j is given as $D_{ij} = R \cdot \arccos(X_{ij})$ where R is the radius of the earth (3963.17 miles) and X_{ij} is defined by Equation (1) with one radian being equal to 57.30 degrees. For multiple lab locations, I use $\min D_{ij}$ to define the distance between an inventor and a lab.

$$X_{ij} = \{\sin(INV_i^{lat}/57.30) \cdot \sin(LAB_j^{lat}/57.30)\} + \{\cos(INV_i^{lat}/57.30) \cdot \cos(LAB_j^{lat}/57.30) \cdot \cos(\mathrm{abs}[(INV_i^{\mathrm{lon}}/57.30) - (LAB_j^{\mathrm{lon}}/57.30)])\}, \quad 0 \leq X_{ij} \leq 1. \tag{1}$$

11 Prior to 1960 such data are not available from the U.S. Bureau of the Census.

Table 4. Inventor distance descriptive statistics

	1921 Data	1927 Data
Percent of inventors living outside of 30 miles of a lab and	23.9	25.6
within 30 miles of a big city	20.3	22.3
Miles between inventor and nearest lab	87.7 (284.0)	99.1 (301.8)
Percent of labs inside 30 miles of a big city	89.5	83.0
Miles between lab and nearest big city	10.9 (13.0)	13.6 (17.4)

Notes: Standard deviations in parentheses. Distances are geocoded great circle distances. Nearest big city is defined as those with 100,000 or more inhabitants as given in the 1921 and 1927 editions of the *Biennial Census of Manufactures*.

Additionally, given the pivotal role of cities in innovation, I used the same method to calculate distances between inventors, labs and large cities. I define large cities as those with 100,000 or more inhabitants as documented by the *Biennial Census of Manufactures* in its analysis of manufacturing establishment agglomerations. Table 4 reveals that around one-quarter of inventors resided >30 miles from an in-house lab of the firm to whom they assigned their patents. Outside of that radius, the vast majority of inventors were located close to a different large city to the location of any of the firm's labs.

How large is the proportion of inventors outside the 30 mile radius relative to the locations of inventions who assign their patents to modern R&D firms? Figure 2 provides benchmark estimates of distances to the nearest R&D facility for the top 10 firms in the data set that could be traced in the NBER data file covering patents in the late 20th century.[12] A striking pattern to emerge is the difference in the spatial distribution of inventors in the early relative to the late period. While a higher share of inventors were located in immediate proximity to the lab ($\leq$5 miles) in the 1927–1929 data, more than double the 1997–1999 share were located in excess of 30 miles from the nearest R&D facility. The distributions are statistically different under a non-parametric Kolmogorov–Smirnov test under the null of equality in the two populations ($D = 0.21$, $P = 0.000$).

Although the difference in the spatial distributions over time might be explained by the changing organizational structure of R&D,[13] it does highlight in a relative context just how much innovation was sourced from outside the boundaries of the labs during their early history. United States Patent and Trademark Office (USPTO) statistics show that independent inventors external to firms accounted for 53% of patented inventions

12 These firms are: Eastman Kodak, General Electric, General Motors, B.F. Goodrich, Goodyear Tire & Rubber, Ingersoll Rand and Westinghouse Electric and Manufacturing. The three companies excluded are AT&T and United Shoe Machinery (did not exist in approximately the same organizational form) and Westinghouse Air Brake (no patents 1997–1999). R&D lab locations are from the *Directory of American Research and Technology*.

13 From the late 1980s, firms moved towards divisional labs and away from centralized R&D (Rosenbloom and Spencer, 1996; Lerner and Wulf, 2007) which may lead to shorter distances between inventors and R&D facilities simply because the labs were more widely distributed.

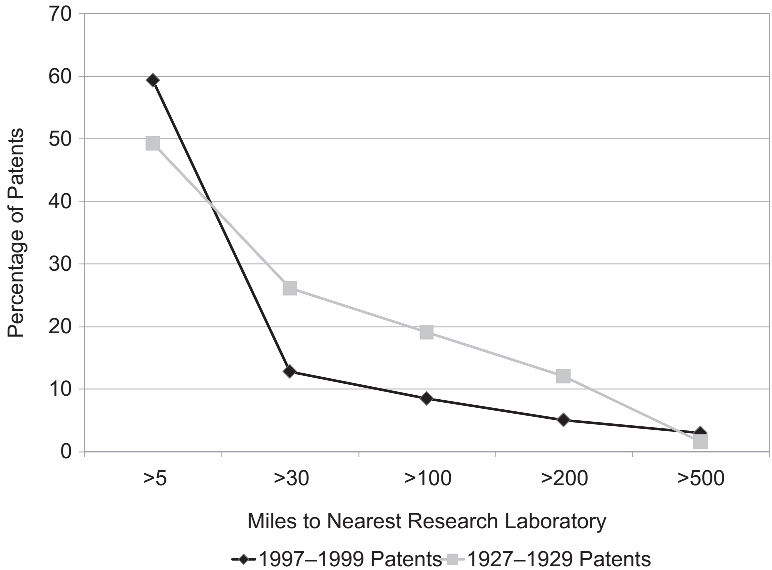

Figure 2. Spatial distribution of patentees for leading R&D firms 1927–1929 and 1997–1999
Notes: Data are from geocoded front-page patent addresses for all patents assigned to seven of the top ten R&D firms in the data set in the late 1920s, which are also listed as patentees in the 1990s in the NBER patent data file. These firms are: Eastman Kodak, General Electric, General Motors, B.F. Goodrich, Goodyear Tire & Rubber, Ingersoll Rand and Westinghouse Electric and Manufacturing. Number of patents is 3578 for 1927–1929 and 5086 for 1997–1999.

by 1930, more than three times the proportion at the end of the 20th century.[14] Interestingly, Huston and Sakkab (2006) show that 21st century trends in R&D are reverting back to a more diffuse spatial distribution pattern as exemplified by the early history of the labs. They show in the case of Protcer and Gamble in 2006 that 35% of innovations originate from outside the boundaries of the firm, a trend that is driven by high-quality external inventions and lower transactions costs associated with sourcing distant knowledge capital.

As a further indicator of the spatial dimension of invention sourcing for the historical data, Table 5 reports geographic matching rates for inventors and laboratory locations. The percentages refer to the proportions of inventors that shared a location with a laboratory of a firm they assigned their patents to. Since SMAs were first defined in 1950, I extrapolate the US Census Bureau's 1950 metropolitan area definitions back to the 1920s.[15] The data show that >90% of the inventors assigning patents to the firms in the sample resided within the United States. Whereas one-third of inventors matched the city of the lab, between 30% and 35% of inventors who resided in an SMA were located *outside* of the lab's SMA. This is significant given that

14 *Historical Statistics of the United States*, Part 2, Series W 96–106. 'Patent Counts by class by Year: Independent Inventor Patents, January 1977 to December 2004', USPTO.
15 SMAs were only first defined in 1950.

Table 5. Geographic matching for inventor-laboratory locations

	1921 Data	1927 Data
Inventor matches country of lab (%)	93.95	93.44
	[16,506]	[17,620]
Inventor matches state of lab (%)	70.48	70.88
	[16,238]	[17,510]
Inventor matches SMA of lab (%)	65.11	69.58
	[5597]	[6249]
Inventor matches county of lab (%)	64.34	63.17
	[15,251]	[16,247]
Inventor matches city of lab (%)	32.89	33.50
	[16,238]	[17,510]

Notes: Figures in square brackets are the number of observations on which the matches are based. SMA boundaries are extrapolated backwards from the U.S. Bureau of the Census' 1950 definitions.

the SMA is an area where localized knowledge spillovers are more likely to be observed (Jaffe et al., 1993; Thompson, 2006). The data suggest that firms were drawing on inventors from multiple SMAs.[16]

4. Patent metrics

The descriptive statistics show a considerable degree of spatial diversity in the sourcing of invention. To test for differences in the characteristics of inventive activity conditional upon distance from a R&D facility, I use three patent-based metrics: historical patent citations, a Herfindhal–Hirschman measure of patent generality and the technological proximity of inventor and firm patents.

4.1. Historical citations

Citations to previous patents provide a useful indicator of technological importance because the quality of patents varies widely across inventions. Following Hall et al. (2005), the conventional method to identify the significance of an invention is to weight patents by counts of the citations they receive. Here, I exploit a novel set of patent citations—those made in modern patents granted that cite the patents of older generations of inventors. I use these citations (as summarized in Table 6) to identify historically significant technologies.

The motivation for the historical citations measure is that, while most citations to patents occur with a lag of approximately a decade of a patent grant date, citations also continue into the past. Although the patents in the current sample have long exceeded

16 This finding can be related to other researchers that have noted a more nunaced pattern of localization. For the biotech industry Audretsch and Stephan (1996) find that relationships between university-based scientists and companies can take place over large distances. They make a distinction between informal knowledge spillovers that depend on interpersonal interactions and formal knowledge that can often be acquired in the absence of geographic proximity. Agrawal et al. (2005) note in the context of knowledge flows that social relationships harnessed in a particular location can negate the influence of distance when inventors subsequently move.

Table 6. Patent descriptive statistics

	1921 Data	1927 Data
USPTO Data (including self-cites)		
Percent patents cited	21.8	21.7
Number of citations	7241	7676
Citations of cited patents	2.01 (2.05)	2.00 (2.04)
NBER data (excluding self-cites)		
Percent Patents cited	19.9	19.8
Number of citations	5785	6138
Citations of cited patents	1.76 (1.46)	1.76 (1.45)
Generality	0.11 (0.21)	0.11 (0.21)
Technical distance	0.03 (0.04)	0.03 (0.04)
Patent classes	1.58 (0.90)	1.57 (0.89)

Notes: Standard deviations in parentheses.

their protection term, there is no stipulation in US patent law that limits citations to older technological knowledge. That old patents should be cited is unsurprising given the literature on the historical origins of technological development. Innovation is rarely distinguished by isolated discoveries; rather it is an evolutionary process as each generation of inventors builds on the ideas of others (Nelson and Winter, 1982). Historical technological dependence is a consequence of cumulative innovation and intertemporal spillovers from R&D (Scotchmer, 1991; Mokyr, 2002). Technological progress frequently depends upon society's intellectual heritage.[17]

The use of long-lagged citations as a measure of patent quality is supported by the fact that they are quantitatively associated with modern citations which are commonly used to identify the technological significance of an invention. Exploiting the extremes of the time series on patents in the most recent NBER patent data shows that citations with short lags are a good predictor of citations with long lags, especially in the upper tail of the citation distribution.[18] For example, if a 1975 patent was cited between 1976 and 1978, the odds of it being cited between 2000 and 2002 are 1.44 times greater relative to patents that were not cited, or almost two times greater for 1975 patents in the upper decile of 1976–1978 citations.[19] Furthermore, the fact that long-lagged citations can be observed so widely in the data suggests that they do not reflect just a propensity to 'cite the classics'.

17 The scientific revolution of the 1870s and the 1880s laid the foundations for 20th century industrial development, shaping in particular the evolution of the modern chemical and pharmaceutical industries (Murmann, 2003). Modern information and communications technology developed out of electronic and computer-related science of the 1950s (Mowery and Rosenberg, 1998).

18 The most recent update allows citations to be traced for the period 1975–2002.

19 The odds ratios are from a logit model with 1976–1978 cites of 1975 patents predicting 2000–2002 cites of the same patents (n = 72,000). The odds ratios for all citations and upper decile citations—1.44 and 1.93—are highly statistically significant, with respective standard errors of 0.02 and 0.110. A 3-year citation window was chosen to give a reasonable estimation span over the time series.

As an additional check on the data, I exploit a change in the patenting process to examine the composition of historical citations counts. Based on a random sample of 100 citing patents granted after the USPTO's 2001 disclosure of who makes citations on patents,[20] I find that applicants as opposed to examiners are more likely to reference older patents. Alcácer and Gittelman (2006) report that on the average patent between January 2001 and August 2003, 63% of citations are added by examiners. In my random sample of 100 patents citing old patents from January 2001 to February 2006, 45% of citations were added by examiners. These data indicate that historical citations, relative to modern citations, are less likely to be an artifact of the examination system.

Table 6 provides descriptive data on citations for the patented inventions in the sample. The first three rows of statistics are calculated from the USPTO database from 1976 to 2006 and show that 22% of the patents in the sample are referenced as prior art in modern patents granted, despite a lag of at least 47 years between cited and citing patents. Because many of the firms in the sample survived and may have cited their own patents independently of the merits of the technology, I purge the data of any self-citations using patent assignee matches in the NBER Compustat matching file.[21] Self-citations account for 4% of the citations observed. Since the matching file is only available for 1975–1999 patents, this is the time period that the NBER citations in Table 6 refer to.

Putting together the geocoded data on R&D labs and inventors and the historical citation data, Figure 3a plots citations against miles from an inventor residence to the nearest in-house laboratory of the firm they assigned their patents to for all distances; whereas, Figure 3b plots citations for within the 30-mile boundary of a lab and Figure 3c for within 30 miles of a large city. Many highly cited patents were granted to inventors who were geographically close to the location of an in-house laboratory although clusters of historically significant technologies can also be identified for inventors who resided some distances away. The basic descriptive evidence suggests both clustering and spatial diversity in the geography of patents assigned to R&D firms.

4.2. Generality & technical proximity

To complement the patent citations data, I use two further metrics to examine the types of inventions that inventors were patenting both within and outside of the vicinity of the R&D labs. Using citation data on 3-digit USPTO classes, I calculate a Herfindahl–Hirschman index to measure patent generality (Jaffe and Trajtenberg, 2002). This measures the range of later generations of inventors that benefit from an earlier technology, where C is citations in each of j citing 3-digit patent classes, C_i is total citations and the sum of the shares over all citing classes (N_i) is subtracted from 1. Thus,

$$G_i = 1 - \sum_{j=1}^{Ni} (C_{ij}/C_i)^2, \qquad 0 \leq G_i \leq 1. \tag{2}$$

20 Starting in 2001, the USPTO distinguishes between applicant and examiner added citations.

21 Specifically I match all of the company names in my 1920s data with assignee names in the NBER datafile of patents from 1976 to 1999. I then exclude any surviving firm citations to its own 1920s patents.

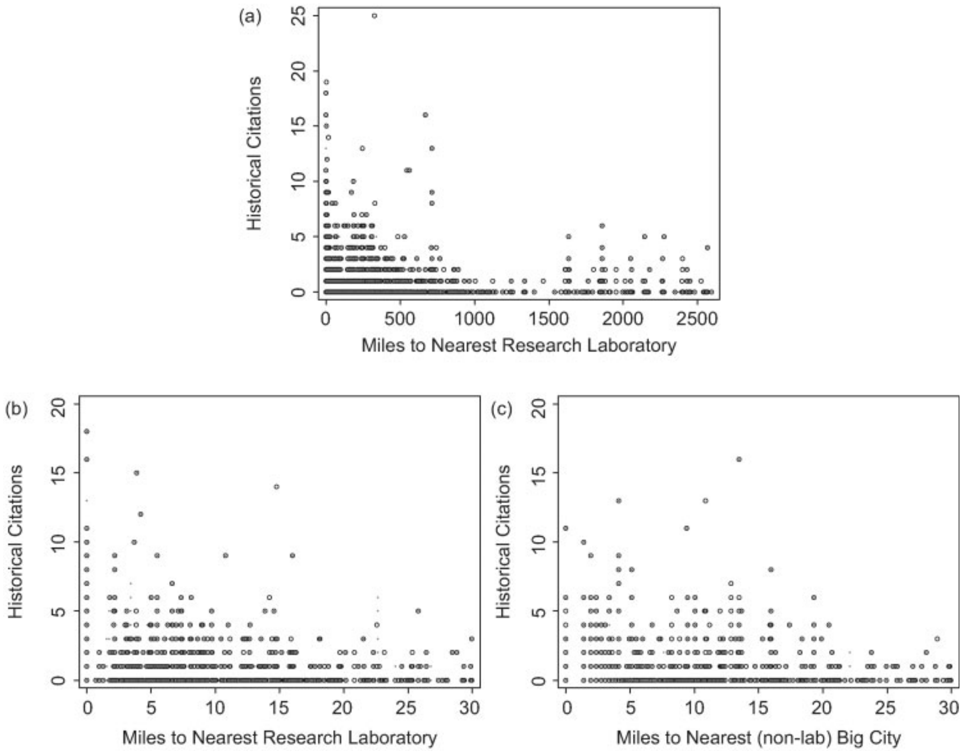

Figure 3. Historical citations and distance of patentee from a lab or big city
Notes: Hollow circles are 1921 data and dots are 1927 data.

A value of $G = 1$ implies that the benefits of a technology were spread widely over subsequent patented inventions; whereas, a value of $G = 0$ implies the benefits were narrowly confined to inventions in a single patent class.

For the second measure, I follow Jaffe's (1986) method to calculate the technical proximity between an inventor and the firm to whom they assigned their patent using the following formula:

$$P_{ij} = \frac{S_i S_j^{'}}{(S_i S_i^{'})^{\frac{1}{2}} (S_j S_j^{'})^{\frac{1}{2}}}, \qquad 0 \le \mathrm{P_{ij}} \le 1, \tag{3}$$

where $S_i = (S_{i1}, S_{i2}, ..., S_{in})$ is a vector containing the profile of inventor *i*'s patent with respect each of the USPTO's *n* 3-digit patent classes and S_j is a vector containing the profile for the firm. For inventor *i* one of the elements of the vector is coded 1 for the main 3-digit class while all other elements of the vector are set to zero. For firm *j* the elements are counts of patents in each of the 3-digit classes. P_{ij} is the uncentred correlation, between the vectors S_i and S_j with a high (low) value implying more (less) technological alignment between an inventor and a firm. I use the measure to test for differences in the technical proximity between lab and non-lab originating inventions.

Table 7. Correlation matrix

	Citations	Generality	Technical proximity
Generality	0.44***		
	(0.000)		
Technical proximity	0.06***	−0.04**	
	(0.000)	(0.025)	
Patent classes	0.18***	0.23***	−0.09***
	(0.000)	(0.000)	(0.000)

Notes: Standard errors in parentheses. Significance at the *10%, **5% and ***1% levels.

5. Estimation setup

Empirically, I use the setup below, which for simplicity abstracts from the different units of measurement (counts and continuous) used to construct the outcome variables, citations, generality and technical proximity. Using 1920s patents and 1921 and 1927 R&D lab locations, I adopt a pooled cross-sectional approach with technology dummies γ_i and year dummies μ_t.[22] Although I cannot estimate the model efficiently with inventor fixed effects, because almost a third of inventors were granted only one patent, I use a cluster adjustment to estimate the standard errors as a check against serial correlation within inventor units.[23] The main specification is,

$$M_{it} = \pi D30_{it}^{lab} + \kappa D30_{it}^{city} + \beta' Z_{it} + \gamma_i + \mu_t + \varepsilon_{it}, \tag{4}$$

where M is a patent metric (citations, generality or technical proximity—Table 7 provides a correlation matrix) and **Z** is a vector of two control variables. First, I use the number of USPTO 3-digit patent classes an invention was allocated into at the examination stage in an effort to filter out spurious historical patent citations. Examiners, applicants and their attorneys use the patent classification system to search for prior art, and therefore patents with higher class counts (as many as 10 for patents in the current sample) might have a higher probability of being cited regardless of their quality.[24] As a second control, I use technical proximity [as in Equation (3)] of inventor patents to the firm's stock of patents in regressions with historical citations and generality as dependent variables. Technologically less proximate patents may be more, or less, likely to be cited simply because they are in a different technological area.

The main variables of interest codify the geography of inventor locations. $D30^{lab}$ is a dummy variable for inventor locations outside (coded 1) or inside (coded 0) a radius of 30 miles from the nearest in-house laboratory of the firm a patent was assigned to.

22 Technology dummies are calibrated off the USPTO classification system following Hall et al. (2002). I define four technology categories: chemicals, communications, electricity, mechanical and other.

23 Over half of the inventors in the sample were granted three patents each or less. The most prolific patentee is Clyde Farmer of Westinghouse Air Brake who was granted 178 patents during the 1920s.

24 On the other hand, this variable might over-correct for spurious citations if the number of patent class allocations proxies for the scope of the invention's coverage. The theoretical and empirical literature shows that broad patents are especially valuable to inventors (Klemperer, 1990; Lerner, 1994).

$D30^{city}$ is a dummy coded 1 for inventors who were both *outside* a 30-miles radius from the nearest lab, but *inside* a 30-mile radius of a big city. I use these radii as boundaries to separate laboratory and non-laboratory inventors. The key parameters, π and κ measure the conditional expected value of M given the inventor's distance to the nearest in-house R&D facility, or large city geographically independent from the firm's labs.

An important issue with respect to Equation (4) is to assign a causal interpretation to π and κ. In particular, if firms only select the best inventions outside the boundaries of the labs π and κ will be biased upwards relative to if distant inventions were randomly selected from the feasible set of inventions that could be sourced. To address this issue, I use the following patent case–control specification to identify the distance effects,

$$M_{it} = \lambda CASE_{it} + \varepsilon_{it}, \tag{5}$$

where M_{it} is historical citations, $CASE$ is a dummy variable coded 1 for case patents and 0 for control patents. For each case patent in the sample—those from inventors outside the 30-mile radius of the labs—I use a matching control patent, where the match is determined by the characteristics of the case patent. I match patents by the geographic location of inventors, and by both main patent class and subclass to avoid any confounds arising from differential citation propensities at the USPTO technology classification levels (Thompson and Fox Kean, 2005). Each control patent is unassigned to a firm as of its grant date so the comparison is between distant inventions assigned to R&D firms (case patents) and distant inventions by individual inventors that were not sourced by firms (control patents).

If patents that were sourced by firms from inventors outside the 30-mile radius of the labs had higher citations than otherwise equivalent patents by individual inventors that were not sourced, $\lambda > 0$, whereas if the average quality of distant invention is high, $\lambda = 0$. The approach can also be extended to account for differences in the propensity to patent inventions of a given quality. Firms may patent more lower quality inventions than individual inventors because the marginal cost to them is lower. Constraining M_{it} to be a positive non-zero citation provides a way of testing for citation count differences between sourced and non-sourced inventions that are technologically important enough to be cited at least once.

6. Results

Beginning with the historical citations metric, Table 8 contains five columns of results for each pooled cross-sections of 1920s patents matched against 1921 and 1927 in-house laboratory locations. Citations are a non-negative integer so the coefficients are from count data regressions with a negative binomial specification to account for overdispersion in the data. The first column restricts the regression to the subset of observations where the inventor could be identified as residing in an SMA. All other coefficients are from regressions of citations on variables measuring distances in miles between inventors and R&D labs (column 2) and distances between inventors, labs and high-density urban locations (columns 3–5). The objective is to test for differences in the quality of innovation given distances from in-house R&D lab locations.

The main result to emerge is that inventors outside the vicinity of a lab patented more highly cited inventions relative to those who were proximate to the labs. Beginning with the results for SMAs, column one is a dummy variable coded 1 for inventors outside

Table 8. Historical citation regressions

	Dependent variable is historical patent citations (1921 Data)					Dependent variable is historical patent citations (1927 Data)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Inventor outside SMA of lab	0.2222***					0.1764**				
	[0.0838]					[0.0765]				
Inventor >30 miles from Lab		0.1712***	0.0453	0.0149	0.0187		0.1944***	0.0185	−0.0086	−0.0057
		[0.0554]	[0.0991]	[0.0921]	[0.0922]		[0.0521]	[0.0966]	[0.0901]	[0.0902]
Inventor >30 miles from Lab and ≤30 miles from a big city			0.1578					0.2148**		
			[0.1018]					[0.1008]		
Inventor >30 miles from Lab and ≤20 miles from a big city				0.2038***	0.1850*				0.2559***	0.2460**
				[0.0952]	[0.0957]				[0.0952]	[0.0957]
Inventor >30 miles from Lab and within New York city					0.1568					0.0913
					[0.1202]					[0.1159]
Number of patent classes	0.3386***	0.3610***	0.3607***	0.3604***	0.3606***	0.3591***	0.3729***	0.3722***	0.3722***	0.3722***
	[0.0318]	[0.0222]	[0.0222]	[0.0221]	[0.0220]	[0.0309]	[0.0217]	[0.0216]	[0.0216]	[0.0215]
Technical proximity	1.2952	−0.3321	−0.2963	−0.2867	−0.3245	1.1772	0.0121	0.064	0.0783	0.0472
	[0.8756]	[0.6170]	[0.6172]	[0.6163]	[0.6169]	[0.7318]	[0.5057]	[0.5050]	[0.5043]	[0.5074]
Technology and year dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5594	15,247	15,247	15,247	15,247	6241	16,344	16,344	16,344	16,344
Wald	234.8	610.6	613.1	614.4	607.9	257.6	624.9	629.0	630.31	628.17

Notes: Standard errors in square brackets are heteroskedasticity consistent and clustered by inventor. Significance is at the *10%, **5% and ***1% levels. Regressions are negative binomial where the change in the dummy distance variables equates to an [exp(coefficient)-1]* 100 percentage change in expected citations counts. The comparison group is patents by inventors ≤30 miles from a lab (columns 2–5) or patents by inventors residing within the lab's SMA (column 1).

of the SMA of the lab and 0 for inventors located within the lab's SMA. The citation premium when the dummy comes on is estimated to be between 19% (*exp* [0.1764]−1) and 25% (*exp* [0.2222]−1)% according to the 1927 and 1921 coefficients, respectively. This result is not driven entirely by higher quality innovations sourced from closer SMAs. Replicating the regression in column 1, but excluding within state non-lab SMAs gives a less precisely estimated, but still quantitatively large, increase in estimated citations of between 12% and 23%.[25] That is, the main effect of a citation premium on inventions sourced further away from the labs is only slightly smaller when comparing lab SMAs with distant SMAs.

Since the SMA results are reported for only a subset of observations where an inventor could be matched with a metropolitan area, the remaining estimates are for distances calculated in mileage, which leads to more than a 2.5-fold increase in the number of observations. Setting the boundary around the lab at 30 miles, in accordance with the discussion on commuting distances in Section 3.3, the parameters in column 2 are close in size and significance to the baseline SMA result suggesting that the 30-mile radius is largely capturing inventors within the metropolitan area of the labs. The parameter estimates reveal that being outside a 30-mile radius of a lab is associated with a 19–21% increase in citation counts. Given that approximately one-quarter of patents were sourced from outside the boundaries of the labs (Table 4), the estimates are economically significant. They imply that corporate stocks of knowledge capital were composed of a quantitatively important share of high-quality innovations sourced at a distance.

If inventors outside the 30-mile boundary produced technologically important inventions, where did they originate from geographically? Note that the coefficient in column 2 is estimated using a single spatial cut-off around the labs. All areas—rural or urban—outside of the 30-mile boundary of a firm's R&D facilities are treated in an equal manner. Columns 3–5 provide fuller specifications by adding spatial controls to identify inventors located in high-density urban environments. Thus, column 3 of Table 8 adds an additional dummy variable for inventors located within 30 miles of a big city with 100,000 or more inhabitants. While none of the coefficients on the spatial dummy variables is statistically significant in the 1921 data, for the 1927 specifications the parameter on the big city dummy is more precisely estimated indicating a positive relationship between citations and patents sourced from inventors in the area around large cities. Moreover, column 4 provides robust estimates across specifications by re-running the specification in column 3 with a tighter 20-mile boundary around the large cities. The 1921 and 1927 coefficients imply inventions sourced from within this area had 23–29% higher citations than inventions originating from the firm's in-house R&D facilities.

As a further check on the data, column 5 of Table 8 adds a dummy variable for inventors located within New York to determine if the big city result from column 4 is caused by a dominant urban location. New York accounts for 8–9% of inventors who assigned patents to 14 different companies in the 1921 data and 20 different companies in the 1927 data. Since New York is itself a big city, the coefficient in column 5 should be interpreted as an offset to the main big city effect. It can be seen that the offset

25 The estimated coefficients are, with standard errors in parentheses, 0.208 (0.086) for the 1921 data and 0.117 (0.079) for the 1927 data.

Table 9. Historical citation regressions with different comparison groups

	Dependent variable is historical patent Citations (1921 Data)				Dependent variable is historical patent Citations (1927 Data)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Inventor >30 miles from Lab	−0.0997	−0.0894	0.1658	−0.1202	−0.1317	−0.1211**	0.1926	−0.115
	[0.0970]	[0.0591]	[0.1442]	[0.0802]	[0.0942]	[0.0572]	[0.1322]	[0.0745]
Inventor >30 miles from Lab and ≤20 miles from a big city	0.2120**	0.1495***	0.2132***	0.1406**	0.2612***	0.1741***	0.2688***	0.2688***
	[0.0947]	[0.0617]	[0.0934]	[0.0613]	[0.0943]	[0.0590]	[0.0931]	[0.0590]
Number of patent classes	0.3470***	0.0605***	0.4023***	0.0810***	0.3547***	0.0607***	0.4004***	0.0781***
	[0.0263]	[0.0180]	[0.0404]	[0.0256]	[0.0251]	[0.0172]	[0.0411]	[0.0251]
Technical proximity	0.1266	0.8552*	−0.0952	0.9322	−0.3049	0.5247	−0.1127	0.6923
	[0.7105]	[0.4599]	[0.8933]	[0.6243]	[0.6044]	[0.3918]	[0.8763]	[0.5809]
Citations	All	>0	All	>0	All	>0	All	>0
Comparison group	Inventors <1 mile from Lab		Inventors >10 and ≤30 miles from Lab		Inventors <1 mile from Lab		Inventors >10 and ≤30 miles from Lab	
Technology and year dummies	YES	YES	YES	YES	YES	YES	YES	YES
Observations	9015	2048	4737	1000	10,076	2315	5400	1166
Wald	280.9	60.0	224.2	45.6	327.7	67.3	231.7	46.7

Notes: See notes to Table 8. Comparison group is determined by excluding inventors within a given range of the labs. That is, in columns 1 to 2 it is inventors co-located with an R&D lab (i.e. <1 mile away), with inventors 1 to 30 miles from a lab excluded from the regression. In columns 3 to 4, the comparison group is inventors >10—but within 30—miles of the lab with those within 10 miles excluded from the regression.

coefficient is insignificantly different from zero at the customary levels, while the main effect is close to that estimated in column 4. These results indicate that the positive correlation between citations and large cities is not driven by an outlying urban location. In summary, the results from columns 3–5 of Table 8 show that the positive effect of distance on cited inventions from column 2 can be attributed to an average, not an outlying, big city effect.

6.1. Varying spatial boundaries

Table 9 provides additional checks on the specification in column 4 of Table 8 by varying the spatial boundaries around the labs. If spillovers are decreasing in distance from a focal point of innovation, as shown by Rosenthal and Strange (2003), the highest quality innovations may be observed in very close proximity to corporate R&D facilities rather than at the extremes of the 30-mile radius.[26] To address this issue column 1 of Table 9 for the 1921 and 1927 data compares distant inventions to a comparison group of patents of inventors who were co-located with the R&D labs. Inventors located 1–30 miles from the labs are excluded from the regression to estimate cleaner effects of differences between lab-based and distant sourced inventions.

A second concern is that the difference in citation counts between local and distant inventions may reflect differences between the hinterlands of the labs and other city regions rather than differences between patents from laboratory and non-laboratory urban locations. Column 3 of Table 9 excludes inventors within a 10-mile radius of the labs so the resulting coefficient compares citations to patents from distant urban locations to patents from the lab hinterlands (i.e. >10 and <= 30 miles away).[27] Finally, with economies of scale in patenting even marginal inventions from the labs could be patented or lower transactions costs may increase the likelihood of lower quality inventions being sourced from the hinterlands. Columns 2 and 4 address these potential sources of bias by filtering out the lower quality non-cited patents from both sets of estimates. The respective specifications are run on patents with greater than zero citation counts.

A clear finding to emerge from the coefficients in Table 9 is that the results are robust to these different specifications. The coefficients in columns 1 and 3 for inventors within 20 miles of a big city are very close in magnitude to the coefficients in column 4 of Table 8. They imply a citation premium of around 24–31% across the 1921 and 1927 data to distant urban-based inventions. This finding holds when comparing sourced inventions at a distance to both inventions from the exact location of the labs (column 1) and their hinterlands (column 3). The estimates in columns 2 and 4, which condition

26 A negative binomial regression of citations on miles from a lab shows that the expected number of citations to a patent falls by a statistically significant 1.8–2.3% for each additional mile out to the 30-mile boundary (see Figure 3b for a plot of the citations). Although the effect of mileage is not significantly different from zero in the same regression for distances up to the 30-mile boundary of a non-lab city (Figure 3c), columns 3 and 4 of Table 8 clearly show sensitivities to different boundary specifications. Tightening the big city radius has a positive effect on predicted citation counts. The most highly cited patents originated from within 20 miles of dense urban locations.

27 Based on the discussion of laboratory worker locations in Section 3.3, I use a 10–30 mile ring around the labs to capture the hinterland because this is more likely to exclude inventors who worked at the labs and is less likely to contaminate the estimates with spillover effects from the labs.

on greater than zero citation counts, reveal smaller differences in average citations (between 15% and 19%), but these still reflect an economically and statistically significant citation premium for distant sourced inventions. Taken together, these results strengthen the main finding of Table 8 of a positive effect of geographic distance on cited inventions.

6.2. Case-control regressions

The results, so far, have established that firms sourced significant technological innovations at a distance, and that urban areas were especially important locations in the acquisition of corporate patentable assets. The case–control results in Table 10 attempt to econometrically identify the effect of distance on innovation by determining whether the results are being caused by firms acquiring only the best inventions at a distance, or by the average quality of distant invention being high.

Recall that the exercise matches case patents—those outside a 30-mile radius of the R&D facility—with control patents that were not sourced by the firms but were otherwise equivalent to the case patents based on their technological characteristics and the geographic location of the first named patentee.[28] Table 10 reports regression results where case patents are matched one-to-one with control patents.[29] In all the specifications, case and control patents are matched by their USPTO patent class and subclass and then in columns 2 and 3 the matching criteria also includes the geographic location of both the case and control patentee. Column 2 matches case patents against control patents if both were patented by an individual residing within the United States. Column 3 matches case and control patents conditional upon patentees being located within 20 miles of a large city, an area of particularly highly cited inventions relative to inventions originating from in-house laboratory locations (Tables 8 and Table 9).

A key result to emerge from Table 10 is the statistical insignificance on the case dummy. Although the reported coefficients are positive, their economic magnitude is small and all are insignificantly different from zero at the standard confidence intervals. This result holds at all levels of patent matching and shows that although the inventions sourced by firms at a distance had high citation counts relative to in-house lab inventions, they were not of a higher quality than the average distant invention. Note especially the third column when case and control patent citations are matched by proximity to a large city. The coefficient indicates no significant difference in quality between patents that were sourced and not sourced from distant urban areas.

The case dummy is also statistically insignificant when the regressions are run on only patents that received one or more citations as a control for potentially 'excess zero' citations in the data if firms have a lower quality threshold for patenting than individual inventors. The case dummy in these specifications (not reported) across the 1921 and 1927 data range from 0.0119 to 0.1005 and with standard errors of 0.0461

28 Of the 3585 patents in the 1921 and 1927 data set that identify inventors as being located >30-miles away from the nearest in-house lab of the firm they assigned their patents to, I was able to find 2745 matching patents. Case and control patents were initially matched if they shared the same USPTO patent class and subclass. Additional data on the matching patents were subsequently compiled giving the location of the first named inventor, latitude and longitude pairings, and historical patent citations.

29 Where multiple matches existed, I randomly selected a matching patent.

Table 10. Case–control regression results

	Dependent variable is historical patent citations (1921 Data)			Dependent variable is historical patent citations (1927 Data)		
	(1)	(2)	(3)	(1)	(2)	(3)
Case dummy	0.0282 [0.0659]	0.0699 [0.0740]	0.0386 [0.1051]	0.0205 [0.0651]	0.0578 [0.0731]	−0.0073 [0.0893]
Case patents	Inventor >30 miles from a Lab	Inventor >30 miles from a Lab	Inventor >30 miles from a Lab and ≤20 miles from a big city	Inventor >30 miles from a Lab	Inventor >30 miles from a Lab	Inventor >30 miles from a Lab and ≤20 miles from a big city
Control patent matching criteria	Class, Subclass	Class, Subclass, Country	Class, Subclass, Country, Inventor ≤20 miles from a big city	Class, Subclass	Class, Subclass, Country	Class, Subclass, Country, Inventor ≤20 miles from a big city
Observations	5360	4428	2376	5490	4538	2936
Wald	0.18	0.89	0.14	0.180	0.63	0.01

Notes: Standard errors in square brackets are heteroskedasticity consistent. Significance is at the *10%, **5% and ***1% levels. Control patents are matched against case patents according to the criteria specified in the table.

and 0.0765, respectively. A finding of no difference between average citations to distant sourced and distant non-sourced inventions, even above the technology quality threshold for a patent to be cited, suggests that the average quality of geographically distant technological knowledge was high.

6.3. Generality and technical proximity regressions

If the average quality of external invention was high, how did firms select inventions at a distance? Cohen and Levinthal (1990) argue that R&D intensive firms place a high value on external knowledge, which they find easier to assimilate due to their absorptive capacity. Baumol (1993, 2002) asserts that technology choice revolves around a specialization in invention with routine efforts concentrated inside the boundaries of large R&D firms and non-routine developments outside. According to Becker and Murphy (1992) as the pool of knowledge expands in an economy and coordination costs decline, firms benefit from a division of labour in technology. These theories suggest a complementarity between lab and non-lab inventions.

In order to test for complementarities, Tables 11 and 12 report the results of technical proximity and generality regressions which identify the technological structure of distant inventions relative to the overall stock of technologies held by the firm. The specifications are aligned with the main specifications in Tables 8 and 9, so that the regressions include the main 20 mile big city dummy, an offset for New York city, and spatial boundary variations to pick up relative differences between distant inventions and those originating from the labs and their hinterlands.

Table 11 reports Ordinary Least Squares (OLS) coefficients with technical proximity as a dependent variable. Recall that technical proximity measures the uncentered correlation between 3-digit USPTO technology classes for inventor and firm patents, so the coefficients locate the patents of distant inventions in technology space relative to the overall patent portfolio of the firm. The first column of estimates for the 1921 and 1927 data indicates that the highly cited patents sourced at a distance from the labs, but close to big cities were also technologically less proximate to a firm's technology profile.[30] The 1927 estimate of -0.01 for inventors within 20 miles of a big city is significant at better than the 1% level and economically meaningful relative to the sample mean of 0.03.

An important caveat to this result is that the New York offset coefficient enters positively and significantly, which runs contrary to the average big city effect. One explanation is that New York was one of the most diverse agglomerations and therefore firms may have acquired substitute (i.e. technologically more proximate) inventions from this particular location. With the exception of New York, the main result of less technological proximity between distant patents close to big cities and firm patents is confirmed in columns 2 and 3 when the comparison group is either inventors co-located with the labs or inventions originating from the hinterlands of the labs. The coefficients are very close in size and significance across the different specifications and indicate that firms were generally selecting technologically less proximate inventions from distant urban locations.

30 These include inventions such as a method for laying underground wires patented by Howard H. Jewell of Los Angeles (number 1,679,427), who assigned his invention to AT&T a company with a central research facility in New York.

Table 11. Regression results: technical distance

	Dependent variable is technical proximity (1921 Data)			Dependent variable is technical proximity (1927 Data)		
	(1)	(2)	(3)	(1)	(2)	(3)
Inventor >30 miles from Lab	0.0145***	0.0119***	0.0180***	0.0133***	0.0102***	0.0153***
	[0.0041]	[0.0040]	[0.0045]	[0.0039]	[0.0039]	[0.0045]
Inventor >30 miles from Lab	−0.0071*	−0.0077*	−0.0076*	−0.0114***	−0.0114***	−0.0116***
and ≤20 miles from a Big City	[0.0042]	[0.0042]	[0.0041]	[0.0041]	[0.0040]	[0.0040]
Inventor >30 miles from Lab	0.0069***	0.0104**	0.0061	0.0129***	0.0098***	0.0097**
and within New York City	[0.0023]	[0.0032]	[0.0047]	[0.0032]	[0.0032]	[0.0046]
Number of Patent Classes	−0.0037***	−0.0043***	−0.0053***	−0.0042***	−0.0053***	−0.0051***
	[0.0004]	[0.0005]	[0.0008]	[0.0005]	[0.0006]	[0.0007]
Comparison group	Inventors ≤30 miles from Lab	Inventors <1 mile from Lab	Inventors >10 and ≤30 miles from Lab	Inventors ≤30 miles from Lab	Inventors <1 mile from Lab	Inventors >10 and ≤30 miles from Lab
Technology and Year Dummies	YES	YES	YES	YES	YES	YES
Observations	15,247	9015	4737	16,344	10,076	5400
R^2	0.07	0.07	0.08	0.06	0.07	0.07

Notes: Standard errors in square brackets are heteroskedasticity consistent and clustered by inventor. Significance is at the *10%, **5% and ***1% levels. Regressions are OLS where the change in the dummy distance variables measures the effect on the correlation between inventor and firm 3-digit USPTO class allocations based on Jaffe's (1986) measure of technical proximity. Comparison group in columns 2 and 3 is determined by excluding inventors within a given range of the labs. In column 2, the comparison group is inventors co-located with an R&D lab (i.e. <1 mile away), with inventors 1–30 miles from a lab excluded. In column 3, the comparison group is inventors >10—but within 30—miles of the lab with those within a 10 miles radius excluded from the regression.

Turning to Table 12, the coefficients are from OLS regressions with patent generality as a dependent variable. Generality measures the scope of the technologies that cite the sample patents. If a patent is subsequently cited by patents across a wide range of technology classes, generality will be high whereas it will be low if the citations are more concentrated.[31] According to the coefficient in column 1 for the 1921 and 1927 data, patents sourced from within 20 miles of a big city had higher generality relative to inventions from within the 30-mile boundary of the labs. Although the offset coefficient measuring generality for patents originating from New York city is negative, it is imprecisely estimated. The hypothesis that patents sourced from this location were different from other city regions cannot be rejected at the customary significance levels.

In the second column of results, the regressions estimate that distant inventions sourced from big cities are more general than those developed by inventors co-located with the in-house labs in both the 1921 and 1927 results. In the third column where the comparison group is the hinterlands around the labs, the 1927 coefficient falls below the threshold for statistical significance, but the same estimate for the 1921 data is close in size and is statistically significant at the 10% level. If firms engaged in search for innovation to complement existing stocks of knowledge, we would expect to observe technological differences between close and distance inventions. While the coefficient estimates are quite small in terms of economic magnitude, when taken with the historical citation and technical proximity results they add weight to the hypothesis that firms were sourcing different types of high-quality innovations at a distance.

7. Conclusion

The growth and development of the in-house research laboratories represents one of the most significant structural changes in the history of American corporate organization. The early labs are a benchmark case for how firms organize their innovation today. While the literature has extensively examined the organizational characteristics of these in-house R&D facilities, no study has considered the further link with the spatial distribution of innovation. This article has been concerned with examining the size and significance of lab versus non-lab inventions given both agglomeration at the level of individual research laboratories and the existence of important pools of technological knowledge at a distance from the labs.

While agglomeration in R&D is generally considered to yield positive externalities that drive innovation—and in this sample the majority of inventive activity was geographically clustered around laboratories—the first key finding is that a large share of innovations came from outside the organizational boundaries of the labs. Around one-quarter of patents held by firms originated from outside a 30-mile radius of any of the firm's in-house research facilities, which is approximately double the proportion observed for late 20th century R&D firms. Inventors outside of firms played a critical role in the process of technological development. Detailed information on one firm in the data set, General Electric, suggests that market-based transactions between firms

31 For an example of a general invention acquired by a firm, Joseph C. Theberath of Cleveland, Ohio, assigned his patent for smoothing the rims of pneumatic tires to prevent puncturing (number 1,518,283) to General Motors, which maintained a principal research facility 178 miles away in Detroit, Michigan.

Table 12. Regression results: generality

	Dependent variable is generality (1921 Data)			Dependent variable is generality (1927 Data)		
	(1)	(2)	(3)	(1)	(2)	(3)
Inventor >30 miles from Lab	−0.0127	−0.0074	−0.0294	−0.0183	−0.0158	−0.0215
	[0.0135]	[0.0143]	[0.0229]	[0.0131]	[0.0137]	[0.0211]
Inventor >30 miles from Lab and ≤20 miles from a big city	0.0321**	0.0321**	0.0296*	0.0272*	0.0265*	0.0217
	[0.0153]	[0.0154]	[0.0157]	[0.0144]	[0.0157]	[0.0148]
Inventor >30 miles from Lab and within New York city	−0.0152	−0.007	0.0065	−0.0158	−0.0034	−0.0034
	[0.0179]	[0.0219]	[0.0280]	[0.0167]	[0.0216]	[0.0265]
Number of patent classes	0.0307***	0.0284***	0.0309***	0.0304***	0.0287***	0.0259***
	[0.0039]	[0.0049]	[0.0084]	[0.0038]	[0.0046]	[0.0073]
Technical proximity	−0.0831	−0.0147	−0.0926	−0.043	−0.0245	−0.1576
	[0.0981]	[0.1196]	[0.1641]	[0.0755]	[0.0963]	[0.1477]
Comparison group	Inventors ≤30 miles from Lab	Inventors <1 mile from Lab	Inventors >10 and ≤30 miles from Lab	Inventors ≤30 miles from Lab	Inventors <1 mile from Lab	Inventors >10 and ≤30 miles from Lab
Technology and year dummies	YES	YES	YES	YES	YES	YES
Observations	3045	2048	1000	3265	2315	1166
R^2	0.03	0.03	0.03	0.03	0.03	0.03

Notes: Standard errors in square brackets are heteroskedasticity consistent and clustered by inventor. Significance is at the *10%, **5% and ***1% levels. Regressions are OLS where the change in the dummy distance variables measures the effect on the generality of patents. Comparison group in columns 2 and 3 is determined by excluding inventors within a given range of the labs. In column 2, the comparison group is inventors co-located with an R&D lab (i.e. <1 mile away), with inventors 1 to 30 miles from a lab excluded. In column 3, the comparison group is inventors >10—but within 30—miles of the lab with those within a 10 miles radius excluded from the regression.

and independent inventors were significant, accounting for as much as 20% of the firm's stock of knowledge assets.

A second finding is that the highest quality distant inventions were sourced from dense urban areas. Distant innovations sourced by firms had proportionately higher historical citation counts than both inventions originating from the labs and from their hinterlands, a finding that holds under a variety of robustness checks. Although the regressions do not identify whether externalities from cities boosted innovation or whether creative inventors endogenously located in cities, the results are consistent with the literature's emphasis on urban environments as hubs of technological development (Henderson, 1988; Glaeser and Mare, 2001; Storper and Venables, 2004). The case–control method shows that the average quality of both sourced and non-sourced inventions from city regions was high. This finding, in turn, provides an explanation for why firms sourced a large component of their patent capital externally.

A third finding is that firms strategically chose the inventions that they sourced from a distance to be complementary to the overall technology profile of the firm. A plausible interpretation of the results is that while exploiting the benefits of knowledge spillovers in laboratory settings and localizing innovation around the labs, firms simultaneously engaged in search for technologically less proximate and more general inventions from outside the boundaries of their in-house R&D facilities. Since laboratory scientists and engineers were influential in the acquisition of external knowledge (Mowery, 1995), the labs played a critical role in determining optimal technology choices and in coordinating the geography of inventive activity.

The findings are based on a critical historical epoch in the organization of innovation and have important implications for present day R&D. According to the analysis presented here, links between the organization and the geography of innovation significantly determine both the quality and composition of knowledge capital accumulation. Current trends towards the wider geographic spread of research investment by firms, an expansion in the market for ideas, and lower transactions costs associated with communicating at a distance, suggest that associations between the organizational and geographic structure of innovation may become increasingly important aspects of corporate R&D.

Acknowledgement

I am extremely grateful to the editor and referees for their many comments and ideas which have been integrated into this paper and to numerous others who read early drafts. Chris Allen and Luke Archer helped with access to data and funding came, via Kash Rangan, from the Division of Research at the Harvard Business School.

References

Adams, L. and Mackesey, T. (1955) *Commuting Patterns of Industrial Workers*. Ithaca: Cornell University Housing Reseach Center.

Agrawal, A., Cockburn, I., McHale, J. (2005) Gone but not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships. *Journal of Economic Geography*, 6(5): 571–591.

Alcacer, J. and Gittelman, M. (2006) How do I know what you know? Patent examiners and the generation of patent citations. *Review of Economics and Statistics*, 88(2): 383–388.

Audretsch, D. and Stephan, P. (1996) Company-scientist locational links: the case of biotechnology. *American Economic Review*, 86(3): 641–652.

Baumol, W. (1993) *Entrepreneurship, Management and the Structure of Payoffs*. Cambridge: MIT Press.

Baumol, W. (2002) *The Free-Market Innovation Machine: Analyzing the Growth. Miracle of Capitalism*. Princeton: Princeton University Press.

Becker, G. S. and Murphy, K. M. (1992) The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, 107(4): 1137–1160.

Chandler, A. (1990) *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, MA: Harvard University Press.

Cohen, W. and Levinthal, D. (1990) Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1): 128–152.

Duranton, G. and Overman, H. (2005) Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4): 1077–1106.

Field, A. (2003) The most technologically progressive decade of the twentieth century. *American Economic Review*, 93(4): 1399–1413.

Fisk, C. (1998) Removing the fuel of interest from the fire of genius: law and the employee-inventor, 1830-1930. *University of Chicago Law Review*, 65(4): 1127–1198.

General Electric. (1925–1929) *Organization Directory*. Schenectady: General Electric.

Gilfillan, S. C. 1935. The Sociology of Invention. Cambridge: MIT Press.

Glaeser, E. and Mare, D. (2001) Cities and skills. *Journal of Labor Economics*, 19(2): 316–342.

Glaeser, E., Kallal, H. Scheinkman, J., Shleifer, A. (1992) Growth in cities. *Journal of Political Economy*, 100(6): 1126–1152.

Griffith, R., Lee, S., Van Reenen, J. (2007) Is distance dying at last? Falling home bias in fixed effects models of patent citations. NBER Working Paper Series no. 13338. Cambridge, MA, USA.

Griliches, Z. Patent statistics as economic indicators: a survey. *Journal of Economic Literature* 28(4): 1661–1707.

Grosvenor, W. N. (1929) The seeds of progress. *Chemical Markets*, 1(1): 1–30.

Hall, B., Jaffe, A., Trajtenberg, M. (2002) The NBER patent citations data file: lessons, insights, and methodological tools. In A. Jaffe and M. Trajtenberg (eds) *Patents, Citations and Innovations: A Window on the Knowledge Economy*, pp. 403–459. Cambridge: MIT Press.

Hall, B., Jaffe, A., Trajtenberg, M. (2005) Market value and patent citations. *Rand Journal of Economics*, 36(1): 16–38.

Henderson, J. V. (1988) *Urban Development: Theory Fact and Illusion*. Oxford: Oxford University Press.

Huston, L. and Sakkab, N. (2006) Connect and develop: inside procter and gamble's new model for innovation. *Harvard Business Review*, 84(3): 58–66.

Jaffe, A. (1986) Technological opportunity and spillovers of research and development: evidence from firms' patents profits and market value. *American Economic Review*, 76(5): 984–1001.

Jaffe, A. B. and Trajtenberg, M. (2002) *Patents, Citations and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.

Jaffe, A., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3): 577–598.

Keller, W. (2004) International Technology Diffusion. *Journal of Economic Literature*, 42(3): 752–782.

Khan, Z. (2005) *The Democratization of Invention: Patents and Copyrights in American Economic Development, 1790–1920*. New York: Cambridge University Press.

Khan, Z. and Sokoloff, K. (2004) Institutions and democratic invention in 19th-century America: evidence from great inventors, 1790–1930. *American Economic Review*, 94(2): 395–401.

Kim, S. (1995) Expansion of markets and the geographic distribution of economic activities: the trends in U.S. regional manufacturing structure, 1860–1987. *Quarterly Journal of Economics*, 110(4): 881–908.

Klemperer, P. (1990) How broad should the scope of patent protection be? *Rand Journal of Economics*, 21(1): 113–130.

Krugman, P. (1991) History and industry location: The case of the manufacturing belt. *American Economic Review*, 81(2): 81–83.

Lamoreaux, N. and Sokoloff, K. (1996) Long-term change in the organization of inventive activity. *Proceedings of the National Academy of Science*, 93(1): 12686–12692.

Lamoreaux, N. and Sokoloff, K. (2002) Intermediaries in the U.S. Market for Technology, 1870–1920. NBER Working Paper Series no. 9017. Cambridge, MA, USA.
Lerner, J. (1994) The importance of patent scope: an empirical analysis. *Rand Journal of Economics*, 25(2): 319–333.
Lerner, J. (2002) 150 years of patent protection. *American Economic Review* 92(2): 221–225.
Lerner, J. and Wulf, J. (2007) Innovation and incentives: evidence from corporate research and development. *Review of Economics and Statistics*, 89(4): 634–644.
MacGarvie, M. and Furman, J. (2005) Early academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry. NBER Working Paper Series no. 11470. Cambridge, MA, USA.
Marshall, A. (1890) *Principles of Economics*. London: Macmillan and Co.
Mokyr, J. (2002) *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton: Princeton University Press.
Mowery, D. (1990) Development of industrial research in U.S. manufacturing. *American Economic Review*, 80(2): 345–349.
Mowery, D. (1995) The boundaries of the U.S. firm in research and development. In N. Lamoreaux and D. Raff. (ed.) *Coordination and Information: Historical Perspectives on the Organisation of Enterprise*, Chicago: University of Chicago Press, 147–182.
Mowery, D. and Rosenberg, N. (1998) *Paths of Innovation: Technological Change in Twentieth Century America*. Cambridge, MA: Cambridge University Press.
Murmann, J. P. (2003) *Knowledge and Competitive Advantage: The Coevolution of Firms, Technology, and National Institutions in the Synthetic Dye Industry, 1850–1914*. New York: Cambridge University Press.
National Research Council. (1921–1946) *Industrial Research Laboratories of the United States*. Washington DC: National Research Council.
Nelson, R. and Wright, G. (1992) The rise and fall of American technological leadership: the postwar era in historical perspective. *Journal of Economic Literature*, 30(4): 1931–1964.
Nelson, R. and Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.
Nicholas, T. (2008) Does innovation cause stock market runups? Evidence from the great crash. *American Economic Review*, 98(4): 1370–96.
Perreto, P. (1998) Technological change, market rivalry, and the evolution of the capitalist engine of growth. *Journal of Economic Growth*, 3(1): 53–80.
Reich, L. C. (1980) Industrial research and the pursuit of corporate security: the early years of bell labs. *Business History Review*, 54(4): 504–529.
Rosnebloom, R. and Spencer, W. J. (1996) *Engines of Innovation: U.S. Industrial Research at the End of an Era*. Cambridge, MA: Harvard business School Press.
Rosenthal, S. and Strange, W. (2003) Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2): 377–393.
Schnore, L. (1960) Three sources of data on commuting: problems and possibilities. *Journal of the American Statistical Association*, 55(289): 8–22.
Schumpeter, J. (1950) *Capitalism Socialism and Democracy*. New York: Harper Torchbooks.
Scotchmer, S. (1991) Standing on the shoulders of giants: cumulative research and the patent law. *Journal of the American Statistical Association*, 5(1): 25–41.
Stephan, P. (1996) The economics of science. *Journal of Economic Literature*, 34(3): 1199–1235.
Storper, M. and Venables, T. (2004) Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4): 351–370.
The Economist. (2007) The rise and fall of corporate research and development: out of the dusty labs. *The Economist*, 1 March.
Thompson, P. (2006) Patent citations and the geography of knowledge spillovers: evidence from inventor and examiner added citations. *Review of Economics and Statistics*, 82(2): 383–389.
Thompson, P. and Fox Kean, M. (2005) Patent citations and the geography of knowledge spillovers: a reassessment. *American Economic Review*, 95(1): 450–460.
Trajtenberg, M. (1990) A penny for your quotes: Patent citations and the value of innovations. *Rand Journal of Economics*, 21(1): 172–187.
United States Patent and Trademark Office (2005) *Independent inventor patents*. Washington, DC: U.S. Government Printing Office.