

Agglomeration and Innovation

Gerald Carlino and William R. Kerr

August 2014

Abstract: This paper reviews academic research on the connections between agglomeration and innovation. We first describe the conceptual distinctions between invention and innovation. We then discuss how these factors are frequently measured in the data and note some resulting empirical regularities. Innovative activity tends to be more concentrated than industrial activity, and we discuss important findings from the literature about why this is so. We highlight the traits of cities (e.g., size, industrial diversity) that theoretical and empirical work link to innovation, and we discuss factors that help sustain these features (e.g., the localization of entrepreneurial finance).

JEL: J2, J6, L1, L2, L6, O3, R1, R3

Keywords: Agglomeration, clusters, innovation, invention, entrepreneurship

Author institutions and contact details: Carlino: Federal Reserve Bank of Philadelphia, jerry.carlino@phil.frb.org (corresponding author); Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu.

Acknowledgments: This paper is a draft chapter for the *Handbook of Regional and Urban Economics*, Vols. 5A and 5B. We thank Gilles Duranton, Vernon Henderson, and Will Strange for their editorial guidance and Ufuk Akcigit and seminar participants for their comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

1. Introduction

This chapter summarizes the recent literature on agglomeration and innovation and how agglomeration and innovation relate to economic performance and growth. It is well established that population and economic activity are spatially concentrated. In addition, innovation is more spatially concentrated than manufacturing employment (Audretsch and Feldman, 1996), and there is evidence that research and development (R&D) activity is more concentrated than employment (Carlino et al., 2012). Why is innovative activity clustered? What is the best way to measure this concentration? What is the economic impact of this concentration? These and related questions are covered in this chapter.

The second section of this chapter begins with a discussion of invention versus innovation and how these forces are measured. The third section quickly reviews patterns of innovation and agglomeration. The fourth section describes formal theories linking agglomeration and innovation. Particular attention is paid to the role of knowledge spillovers in local areas, and extra attention is devoted to describing recent work to quantify these relationships. This section also makes connections to the models developed in endogenous growth theory. The fifth section discusses research on additional factors that work to sustain agglomeration clusters, link global clusters together, promote large versus small company innovation, and deal with similar phenomena. The last section concludes. In the final section and at several points throughout the chapter, we highlight important areas for future research.

Several key themes weave their way through this review:

- *Innovation comes in many shapes and sizes, except in economic studies*—The next section provides quick definitions of invention versus innovation versus commercialization, describes underlying varieties such as exploration versus exploitation research or product versus process work, and so on. Most nuances are then forgotten as the substantial majority of theoretical and empirical studies consider single-dimension outcomes such as patents. Hopefully, future research will develop a richer accounting of the variations of innovation and how these variations relate to the traits of clusters. We do not pretend that this will be easy, due in part to data constraints. More important, an ad hoc collection of facts will not be very useful, so instead we need a structure of innovation outcomes and how they relate to local traits.
- *Agglomeration theories apply especially for innovation clusters but often lack empirical confirmation*—This chapter details the various mechanisms and models that have been proposed for the formation of agglomerations of economic activity. We trace how these models apply to innovation specifically, often conjecturing that rationales such as input sharing and knowledge spillovers apply particularly well in this setting. Yet, our review of the empirical literature often finds limited verification of these theories with respect to innovative industries. This is an important area for future research and one that is facilitated by the many data sets recently developed. A subtheme of the chapter is how the existing evidence is spread across many

adjacent fields (e.g., entrepreneurial finance). This porous boundary reflects the complex nature of innovation and requires researchers to cover a larger domain in their work.

- *Moving to the frontier in terms of the urban economics toolkit*—Empirical measurement in urban economic studies has made substantial strides forward in the past two decades. Two prominent examples are the use of continuous distance metrics for measuring economic activity and the development of micro-record data sets for studying economic and social interactions at detailed levels. Research related to innovation clusters is moving toward this empirical frontier, but the gap is not yet fully closed.
- *Globalization of innovation*—The last part of this chapter describes some key research opportunities, one of which we emphasize here. We have assembled evidence about innovative clusters within countries, and separate lines of economic research consider flows of activity across countries. Increasingly, research in this area is merging these fields by considering the connections of clusters across countries (e.g., the particular flows of activity from Bangalore, India, or Seoul, South Korea, to Silicon Valley). With the ongoing globalization of innovation, theoretical and empirical refinements of this work are very important.

Our review focuses equally on highlighting classic studies and identifying recent contributions. Previous reviews such as Audretsch and Feldman (2004) and Feldman and Kogler (2010) contain more extensive documentation of the older literature. Our review also seeks to highlight how research on innovation in economic geography is now connecting with adjacent fields. This has been a very active research area over the past decade and will continue to be so for some time to come.

2. What is innovation?

This section introduces and distinguishes the concepts of innovation and invention. We describe the typical data sets used to study the phenomena and discuss some of the relevant advantages and limitations that researchers should be aware of.

2.1 Distinguishing invention from innovation

It has become popular for researchers and policymakers to talk about the role innovation plays in economic growth. But what is innovation, and how does it differ from invention? According to the Organisation for Economic Co-operation and Development's *Oslo Manual* (2005):

An innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations.

Invention is a term often used interchangeably with innovation. But invention is a separate item in the dictionary. According to the *New Oxford Dictionary of English*, invention can be defined as:

Creating something new that has never existed before.

Schumpeter (1939, pp. 84–86) made a clear distinction between invention and innovation:

The making of the invention and the carrying out of the corresponding innovations are, economically and sociologically, two entirely different things. They may, and often have been, performed by the same person; but this is merely a chance coincidence which does not affect the validity of the distinction. Personal aptitude—primarily intellectual in the case of the inventor, primarily volitional in the case of the businessman who turns the invention into an innovation—and the methods by which the one and the other work, belong to different spheres.

For Schumpeter, invention is the creation of a new product, service, or process, whereas innovation is the commercialization or introduction of that product or service into the market. Many inventions are patented, but most patents never reach the point of commercialization. Those that do often require a long gestation period. On the other hand, innovations are closely linked to commercialization and often do not require a corresponding invention or patents. Innovation takes place when something new or different is brought to fruition. Thus, conceptually and partly by definition, innovation has an immediate effect on welfare through the introduction of new products and improvements in existing products, often yielding a direct effect on regional and national growth.

Despite these nuances, most researchers do not distinguish between invention and innovation. Intuitively, invention seems to be an antecedent of innovation, but the conceptual basis for separating them is not so absolute, as both are necessary and complementary factors in the advancement of technology (Scherer, 1986). This perspective suggests that the concept of innovation has two parts: the generation of ideas and the conversion of these ideas into useful commercial applications. This is the sense in which most current researchers use the term “innovation.” For example, the National Science Foundation (NSF) defines innovation as “a series of steps that begins with imagination, and results in the creation of something of value for society.”¹

Despite this blurred boundary, regions within countries represent an important instance in which the distinction between invention and innovation has economic and policy relevance. A region’s growth depends upon the successful commercialization of new products and processes more so than on invention. The idea for a new product may have occurred in one city, but the commercialization of that product may

¹ The ambiguous definitions are also due in part to empirical considerations. Patent data are a useful, if somewhat limited, source of inventive data. High-quality data for innovation are unfortunately scarce, and thus most researchers utilize patent data even if their conceptual focus is innovation.

take place in an entirely different location. Thus, the benefits from local public subsidies for basic research may not stimulate growth in targeted communities, except for creating a few jobs for scientists and engineers.

While these distinctions between invention and innovation are important, our review generally focuses on the more inclusive definition of innovation. Likewise, this chapter will consider evidence related to knowledge transfer and spillovers more broadly. While these forces clearly extend beyond innovation per se, findings from studies of innovation are often used to describe knowledge flows, and vice versa.

2.2 Types of innovation

Innovation can take many forms. At one level, the results from innovative efforts can be physical or abstract, such as computer programs or new business methods versus a higher-definition TV. As we show in Section 2.3, our ability to measure these advances differs across fields and over time. At a second and arguably more important level, innovation is a very broad term that includes multiple types of efforts. The types of advances sought by startup companies can be quite different from those pursued by large universities or corporate R&D labs. The theoretical and empirical literature tends to lump everything together, as our review later conveys, but it is important to outline some of the distinctions made.

The oldest distinction is between incremental and radical innovation. Incremental innovation, sometimes called Usherian innovation, refers to variations on a theme, involving a learning process through cumulative experience and experimentation (Usher, 1929). One example of Usherian innovation is the introduction of the electronic ignitions that replaced the mechanically timed ignition used to ignite a fuel-air mixture in the internal combustion engine. Alternatively, radical innovation, sometimes called Schumpeterian or disruptive innovation, is trend-breaking and renders existing products or methods obsolete (Schumpeter, 1934, 1939; Christensen, 1997). For example, the transistor transformed the way people listened to music because the transistor led to the miniaturization of the radio, thereby making it portable.

A second distinction focuses on product versus process innovation (e.g., Cohen and Klepper, 1996b). Product innovation refers to a new or improved product, whereas process innovation represents an improvement in a firm's production technology. Innovation also encompasses the introduction of existing products to new locations. A third distinction, which relates to product-process separation, focuses on whether innovations are internal versus external to the firm, also known as exploration versus exploitation (e.g., March, 1991; Akcigit and Kerr, 2010).

Similar to the distinction between innovation and invention, these differences will mostly fade into the background for the rest of this review. This is not desirable but instead reflects the very limited degree to which these distinctions have been made in an economic geography context. We describe the available studies where applicable, but much more work needs to be done in this vein, as agglomeration will interface in quite different ways with various forms of innovation. For example, Duranton and Puga (2001) find evidence that French entrepreneurs tend to develop prototypes for new products in diversified cities and that these businesses relocate to cheaper specialized locations for mass production once the products are sufficiently mature. This process would suggest that the innovation in diversified cities would be systematically different from that in specialized locations.

2.3 Measuring innovation

Most studies measure technological innovation in one of three ways: (1) by the inputs used in the innovation process, such as R&D expenditures or venture capital (VC) investment; (2) by intermediate outputs of the innovation effort, such as the number of patents; or (3) by some final measure of innovative work, such as the count of new product announcements. Each measure has strengths and weaknesses that are outlined below. Studies on innovation tend to use a single indicator of innovation and argue that the chosen measure has fewer shortcomings than other measures.

2.3.1 Investments in the innovation process

Given the scarce data on innovative output, early studies frequently used R&D employment and expenditures as a measure of innovative activity. These factors certainly reflect innovative effort, and Feldman (1994, pp. 30–31) provides references to these studies. This approach has lost some favor over time, however, since input-based measures do not capture the efficiency of the innovation process. Moreover, newer measures such as patent data are easier to access and more granular, providing greater empirical traction, including the qualitative importance of the innovations produced. R&D data are particularly difficult to collect at the local level except through confidential surveys such as the NSF R&D Survey.²

While R&D expenditures are less frequently used, a surge of work today considers VC investment by area as a metric for how innovative different regions are. VC firms fund innovative startup companies in return for equity, with Gompers and Lerner (2006) providing a full introduction. Conceptually, VC-based

² One exception to highlight is that many growth frameworks model firms as choosing their innovation intensity in terms of R&D efforts per sales. In this setting, converting innovation metrics such as patents into R&D-equivalent terms can enable the calibration of models (e.g., Acemoglu et al., 2014).

metrics again reflect an input into the innovative process, rather than an output or end goal. One advantage of VC-based measures is that they are now available at the micro-level through sources such as VentureXpert. Similar to the patent data described next, this affords researchers extensive flexibility in metric design. The central challenge for metrics based upon VC activity is that these investments are concentrated in specific technological areas (e.g., computers, software, and biotechnology) and are classified by types of firms (e.g., younger startups), thus making them quite incomplete for describing innovation broadly.

2.3.2 Patents and citations

Most regional studies use patent data to measure innovation, and we accordingly devote more attention to the strengths and limitations of the patent data. Essentially, a patent constitutes a legal right to exclude others from making, using, or selling the patented invention or process for some period of time. Patents provide the economic incentives for inventors to pursue new opportunities by granting a temporary monopoly to successful efforts. To qualify for patent protection, an invention must be deemed useful (utility), new (novelty), and a nonobvious extension of the prior art (nonobviousness).

A patent publication contains a description of the invention, the names of the inventors, the institution (e.g., firm, university) to which the patent is assigned, references (citations) to related inventions or techniques, claims that define the rights of the new invention, and much more. The United States Patent and Trademark Office (USPTO) granted more than 250,000 patents per year in the past few years. Hall et al. (2001) provide more details about these data, and Griliches (1990) surveys the use of patent data for measures of technological change.

Advantages: An advantage of using patents over R&D expenditures is that they are a direct outcome of the inventive process. Moreover, patent data are available in full (for what is patented) and not subject to confidentiality restrictions or suppressed information. Patent data are released at the micro-level, and thus researchers can use the granularity to describe inventive activity in great detail (e.g., isolating Microsoft's patents from a specific city in a specific technological area). Due to recent efforts to digitize historical records, patent data are now available over long periods. These advantages and easy access make patent data the most used empirical approach.

Disadvantages: One concern with using patents as a measure of innovation is that patents reflect the first stage of innovation—that is, the invention. To what extent is the location where an invention takes place coincident with the successful commercialization of the invention? Feldman (1994) finds a correlation of 0.8 between patents and the locations where new products are introduced to the market. While this

correlation is positive and moderately high, Feldman and Kogler (2010) nonetheless stress that “studies that draw inferences about innovation by focusing on invention should be interpreted with caution.”

Another concern about using patents as an innovation indicator is that the value of patents is very highly skewed. Most patents are not worth very much, while a few are very valuable (e.g., Harhoff et al., 1999). If a patent has value, we would expect it to be renewed before the patent expires. Serrano (2010) calculates that 78 percent of U.S. patents granted during 1983 to 2001 were not renewed, indicating that most patents are of low value. In the words of Griliches (1979) and Pakes and Griliches (1980, p. 378), “patents are a flawed measure (of innovative output) particularly since not all new innovations are patented and since patents differ greatly in their economic impact.” In addition, patent owners often obtain patents on close substitutes for their patent in an attempt to block follow-on patenting by rival inventors. Rather than adding value, these “patent fences” are strategic and thus skew innovation measures.

Fortunately, researchers can adjust for patent quality in their innovation metrics by weighting patents by the number of citations they receive, similar to what economists do with journal articles. Most researchers exclude self-citations from these counts (i.e., a Microsoft patent that cites another Microsoft patent). An important recent practice is to make these measurements relative to each technology category and application year to adjust for differences in propensities to cite across fields and the truncation biases in citations for the most recent cohorts. As we describe much later, citations are also used to model knowledge spillovers (Jaffe et al., 1993; Carlino et al., 2012; Murata et al., 2014).

There are naturally worries about what is not patented. Based on a 1994 Carnegie Mellon University survey of almost 1,500 industrial R&D labs in the United States, Cohen et al. (2000) report that firms typically use a number of mechanisms to protect their returns to invention. In most manufacturing industries, patents are used less frequently than other approaches such as trade secrets and lead-time advantages. Cohen et al. (2000) find that the two most important reasons given by firms for not applying for a patent are the amount of information disclosed in a patent application and the ease of legally inventing around a patent. They also report that the cost of patent litigation deters smaller firms from pursuing patents.³

Another issue is that there are large differences in the propensity to patent across industries. For example, Cohen et al. (2000) report that firms in the chemicals, drugs, mineral products, and medical equipment

³ These selection issues are also critically important across countries. It is frequently noted that local firms are hesitant to file patents in nations with weak intellectual property protections (e.g., China) because of disclosure concerns.

industries applied for patents for more than two-thirds of their innovations. In contrast, firms in the food, textiles, glass, and steel and other metals industries applied for patents on fewer than 15 percent of their product innovations. Quantitative work considering raw differences across industries in patent rates needs to proceed with caution, while growth in innovation rates for industries is more secure.

A final consideration relates to the longitudinal adjustment of what forms of technologies are patentable. For example, one of the strongest trends in the USPTO data is the introduction and growth of software patents. Founded in 1975, Microsoft had just five patents by 1990 and more than \$1 billion in revenue; by 2009, the company held 10,000 patents and over \$58 billion in revenue. More broadly, the rate of patenting for a given set of inputs slowed in some fields over the past three decades, while it accelerated in others. As cities and regions differ in their technical and industrial compositions, a worry exists that the measurement of changes in innovation at the city level will be biased or suffer substantial measurement error due to these underlying forces (e.g., Seattle's rate of innovation appears to skyrocket, but this is mainly due to changes in the legal practice of patenting software). Lerner and Seru (2014) further discuss these challenges and the importance of designing area-based studies to control for the underlying composition of inventive activity.

2.3.3 Literature-based indicators of innovation

A third approach is broadly referred to as literature-based indicators, since the metrics are based on product announcements in trade, engineering, and technical publications. Acs and Audretsch (1988), Audretsch and Feldman (1996), Feldman and Audretsch (1999), and Acs et al. (2002), among others, use new product announcement data available from the U.S. Small Business Administration (SBA) as an indicator of innovative activity for state and metropolitan areas in the United States. An advantage that new product announcement data have relative to R&D expenditures and patents as measures of innovation is that they document the commercialization of a new product.⁴

Although the new product announcement data are interesting, they are subject to several important qualifications. First, the data are available only for 1982 and do not allow any follow-up analysis. According to Acs and Audretsch (1988), the data largely consist of product innovations to the exclusion of other important types of innovation, such as process, services, and managerial innovations. Another concern is that the new product announcements tend to be issued by the marketing departments of companies; therefore, these announcements are not subject to the same scrutiny as patents are to

⁴ Capello and Lenzi (2014) develop a measure of innovation based on Eurostat's Community Innovation Survey of firms introducing new products and processes. Many R&D surveys are now including questions of this variety.

determining the originality and contribution of the product. Also, the announcement data may suffer from selection bias, as journal editors of trade publications may select innovations that they think are influential.

In sum, the various indicators of innovative activity have strengths and weaknesses. Few studies that look at the spatial scope of innovation have used multiple indicators to see if the results are robust to indicator choice. Is there a strong correlation among the various indicators? Are the findings using patents similar to those based on R&D expenditures or to findings based on new product announcement data? One possibility for future research is to develop a composite indicator based on two or three of the individual indicators (Hagedoorn and Cloudt, 2003). It is also hoped that the ongoing explosion in data collection will afford new measures for future studies.

3. Patterns of agglomeration and innovation

There is an abundance of evidence that innovative activity is spatially concentrated. We quickly review some basic evidence, not attempting to catalog all of the measurements that have been made. Three points are generally made: (1) innovation is more concentrated at a point in time than general economic activity, (2) it appears more localized than other economic forces linked to agglomeration, and (3) there are some spatial movements of clusters over time, although the new clusters remain spatially concentrated (Detroit circa 1900 has remarkable parallels to Silicon Valley today).

3.1 Higher degree of spatial concentration at a point in time

Each of the innovation measures discussed in the prior section display substantial spatial concentration. For R&D activity, Buzard and Carlino (2013) show that the spatial concentration of establishments undertaking R&D efforts is more pronounced than it is for firms generally. Figure 1 is taken from this study. Buzard and Carlino (2013) find, in particular, that R&D activity for most industries tends to be concentrated in the Northeast corridor, around the Great Lakes, in California's Bay Area, and in Southern California. This concentration is not limited to the United States. For example, Carrincazeaux et al. (2001) find that six regions in France account for 75 percent of all corporate R&D workers, compared with 45 percent of the production workers.

VC investments and patents are similarly concentrated, as Figure 2 from Chatterji et al. (2014) shows. During the 1990s, three-quarters of the U.S. population resided in metropolitan areas. In contrast, 92 percent of the patents were granted to residents of metropolitan areas, and virtually all VC investments were made in major cities. Bairoch (1988) shows that patent originations tend to be concentrated in large

cities. Fornahl and Brenner (2009) find that patents tend to be concentrated in 11 of the 97 German regions they considered.

In terms of final outputs, Acs et al. (1994) find that the introduction of new products is more spatially concentrated than patents. Feldman and Audretsch (1999) find that less than 4 percent of the product innovations took place outside of metropolitan areas and that half of the new product innovations in 1982 occurred in just four metropolitan areas (Boston, New York City, San Francisco, and Los Angeles). Feldman (1994) attributes this strong concentration of the commercialization of innovation to the need for specialized business services (e.g., firms specializing in market research and product testing, specialized patent lawyers, and the availability of financing) and similar infrastructure.

Studies have gone beyond these raw statistics to compare the concentration of innovation with the underlying spatial concentration of economic activity. Krugman (1991) and Audretsch and Feldman (1996) use a “locational Gini coefficient” to study the spatial clustering of manufacturing industries. A locational Gini coefficient shows how similar (or dissimilar) the location pattern of employment in a given industry is from the location pattern of aggregate employment. Let s_{ij} represent location i 's share of employment in industry j , and let x_i represent location i 's share of aggregate employment. The spatial Gini coefficient for industry j is defined as: $G_j = \sum_i (x_i - s_{ij})^2$. $G_j = 0$ indicates that employment in industry j is no more or less spatially concentrated than aggregate employment, and $G_j > 0$ implies that employment in industry j is overconcentrated. Audretsch and Feldman (1996) apply this approach at the state level using SBA data.

One problem with the locational Gini coefficient is that it may spuriously indicate the localization of an industry resulting from the lumpiness of plant employment. Ellison and Glaeser (1997) improve on the locational Gini coefficient by offering an alternative index (the EG index) that controls for the industrial organization of an industry. Conceptually, the EG index compares the degree of spatial concentration of manufacturing employment in industry j with what would arise if all plants in the industry were randomly distributed across locations. Using this approach, Ellison and Glaeser (1997) and Rosenthal and Strange (2001) find evidence of the geographic concentration of employment in many U.S. manufacturing industries. Although the EG index is widely adopted in the literature, few studies use it to determine the agglomerative forces for innovative activity; one exception is Buzard and Carlino (2013).

The locational Gini coefficient and the EG index suffer from important aggregation issues. The first is known as the modifiable area unit problem (MAUP).⁵ These metrics depend upon the boundaries used to demarcate regions, and conclusions may differ if counties versus states, for example, are used as boundaries. The MAUP grows in severity as the level of aggregation increases. A related issue is referred to as “border effects”—each region is considered an exclusive zone, and the closeness of regions is not factored in. Thus, while Philadelphia County and Montgomery County border each other and have activity spilling across them, they are treated as being as distant from each other as they are from Los Angeles County. These partitions often lead to underestimations of concentration. Third, the earlier metrics provide index values but do not assess the statistical significance of their results. Without further statistical analyses, it is unclear whether the concentrations reported are significantly different from ones that might result even if the locations of economic activity resulted from random draws. Recently, Cassey and Smith (2014) have developed a procedure to simulate confidence intervals for statistical tests of EG indices.

In response to these issues, Duranton and Overman (2005) develop a continuous-distance metric that does not rely on fixed spatial boundaries. This approach provides researchers with a range of scales in continuous space, rather than at a single predefined scale, on which to measure effects. In addition, these techniques utilize Monte Carlo methods to determine whether the number of plants at a given distance is significantly different from the number found if their locations were randomly chosen. The power of these techniques comes with two significant costs. First, researchers must access data on individual plant locations, which can be challenging. Second, the calculation of these metrics takes substantially more time and computational power.

A number of more recent researchers have used continuous methods to measure the spatial concentration of innovative activity (e.g., Carlino et al., 2012; Inoue, et al., 2013; Kerr and Kominers, 2014; and Murata et al., 2014), which we recommend. For example, Carlino et al. (2012) use point patent methods (Ripley’s *k*-function analysis) to analyze locational patterns of R&D labs over a range of spatial scales (e.g., within a half mile, one mile, five miles, etc.). This approach allows them to consider the spatial extent of the agglomeration of R&D labs and to measure how rapidly the clustering of labs attenuates with distance. It is important to note that they look for geographic clusters of labs that represent statistically significant

⁵ See Briant et al. (2010) and Menon (2012) for a discussion of the MAUP and its implications for spatial concentration measures. Duranton and Overman (2005) suggest five properties for a good index of concentration. The index should (1) be comparable across industries, (2) control for overall concentration of the industry, (3) control for industrial concentration, (4) be unbiased with respect to scale and aggregation, and (5) test for the significance of the results.

departures from spatial randomness using simulation techniques. In sum, continuous approaches appear to be the preferred method for measuring the geographic concentration of innovative activity going forward, although we later discuss how political boundaries are still important when using these techniques.

3.2 Innovation is more localized than other economic forces linked to agglomeration

Marshall (1890) first elaborated on how many different forces give rise to spatial concentration. As depicted in the next section, innovation is related to multiple Marshallian channels, but it also has the closest connection with knowledge spillovers. Comparative studies often find that knowledge spillovers operate at the smallest spatial scales of the types of agglomeration forces. Rosenthal and Strange (2001) regress the spatial concentration of an industry (based on an EG index) on characteristics designed to measure the individual forces of agglomeration. The analysis is carried out separately at the zip code, county, and state levels. They find knowledge spillovers positively affect agglomeration only at the zip code level, with other forms of agglomeration such as labor pooling operating at broader levels. Ellison et al. (2010) exploit patterns of industry collocation to measure the relative importance of the various mechanisms using discrete and continuous distance measures. They relate coagglomeration levels to the extent to which industry pairs share goods, workers, and knowledge. They find evidence for all three mechanisms, and knowledge spillovers again are the most localized. Using a new metric for analysis of location patterns within cities, Billings and Johnson (2014) also emphasize collocation for knowledge sharing. Capello and Lenzi (2014) find that the growth benefits from new knowledge are spatially concentrated when considering 262 regions in the EU.

Complementing these regional calculations, several studies estimate production functions of spillovers. This approach again emphasizes the exceptional concentration of knowledge flows and innovative activity. Figure 3 illustrates results presented in a comparative study of software and manufacturing industries by Rosenthal and Strange (2003). To allow easy interpretation of their findings, the vertical axis in the exhibit is set such that the spillover benefits in the software industry within one mile are equal to a value of one; all other spillover effects are measured relative to this value. The horizontal axis shows the distance between any two establishments in the industry.

For all industries, the localization effects of being near similar businesses decay rapidly with distance within cities—the positive localization effect from being within one mile of another company in one’s own industry is at least 10 times greater than the positive effect realized when locating two to five miles away from said company. After five miles, the benefits continue to decay with distance, albeit at a less dramatic rate. By 10 miles, there are no more within-city localization benefits. This statement does not preclude general spillover effects that all firms within the city enjoy from higher rates of industrial

concentration; it only means that a firm's specific site within the city no longer governs access to the benefits. It is important to note that Rosenthal and Strange (2003) identify dramatic differentials across industries. The decay functions of two representative manufacturing industries, fabricated metal and machinery, have a similar shape to the decay function of the software industry. The most dramatic difference, however, is that the magnitude of spillover effects in the two manufacturing examples tend to be only 20 percent or so of the software effect at any distance.

While the declines over the first mile are dramatic in Rosenthal and Strange (2003), they can be even sharper in a very highly networked industry. Arzaghi and Henderson (2008) study location choices of ad agencies in Manhattan. They show that for an ad agency, knowledge spillovers and the benefits of networking with nearby agencies are extensive, but the benefits dissipate quickly with distance from other ad agencies and are gone after roughly a half-mile. Figure 4 plots the decay function in localization effects using this study's results, with the localization effect for ad agencies for 0–250 meters being set equal to a value of one. The decline in Figure 4 is much more dramatic, with the localization effect losing 80 percent of its value by the time two firms are 500 meters apart. Rather than the 10 miles of the previous example, the cluster benefits for ad agencies run out at 750 meters! This study provides one of the sharpest depictions of tightly concentrated knowledge flows developed to date, and we believe it represents an important precedent for future research related to innovation more directly. We discuss these features, and especially studies building upon patents, in more detail as we introduce direct theories in the next section.

While these two approaches—regional-based evidence and production function regressions—end up with similar conclusions about the rapid attenuation of knowledge flows, they embody very different spatial scales and are not immediately reconcilable with each other. For example, can city- or county-level comparisons of software versus machinery industrial concentration yield much insight if the main clustering effect in both industries is only over one mile?

Kerr and Kominers (2014) consider this question using variation across patent clusters. A central element of their model and empirics is illustrated in Figure 5, which depicts technology flows in Silicon Valley. Downtown San Francisco and Oakland, CA, are to the north and off of the map. The triangle in the bottom right corner of the map is the core of Silicon Valley. This core contains three-quarters of the industrial patents filed from the San Francisco Bay area and 18 of the top 25 zip codes in terms of patenting. Plotted on the map are three of the four largest zip codes for patenting in the San Francisco area that are outside of the core. Each focal zip code is marked with a box, and the other points of the shape are the three zip codes that firms in the focal zip code cite most often in their work. Zone 1, which covers

Menlo Park, extends deepest into the core. Zone 2, for Redwood City, CA, shifts up and encompasses Menlo Park and Palo Alto but has less of the core. Zone 3, which covers South San Francisco, further shifts out and brushes the core.

These technology zones are characterized by small, overlapping regions. None of the technology sourcing zones traverse the whole core, much less the whole cluster, and only the closest zip code (Menlo Park) even reaches far enough into the core to include the area of Silicon Valley where the greatest number of patents are issued. While technology sourcing for individual firms is localized, the resulting cluster extends over a larger expanse of land. Kerr and Kominers (2014) use this insight to build a model of clusters comprising small, overlapping regions. This model reconciles the area-based approaches with the firm-level analyses by showing that agglomerative forces with longer spillover distances at the firm level yield macrostructures characterized by fewer, larger, and less dense clusters. Kerr and Kominers (2014) find these predictions empirically hold true when comparing the clusters of different technologies with each other. Extended out-of-sample, this logic also provides a foundation for why innovation clusters can be especially concentrated in small geographic regions.

3.3 Spatial movements of clusters over time

While most thoughts of innovation clusters today naturally begin with Silicon Valley, it is important to recall that innovation clusters do move over time. Lamoreaux et al. (2004) describe the role of Cleveland as a prominent and frontier hub for innovation related to the second industrial revolution, including early forms of angel financing and incubators. Klepper (2010) similarly describes how Detroit circa 1900 has remarkable parallels to Silicon Valley today. Saxenian (1994) also prominently discusses the migration of the semiconductor industry from Boston's Route 128 to Silicon Valley in the 1970s and 1980s. Kerr (2010a) describes the specific movement of technologies to new breakthrough centers following the Duranton (2007) model. With the development of better long-run patent data, we hope that further regularities can be defined that enhance this case study evidence.

4. Formal theories linking agglomeration and innovation

Often, innovative activity takes place inside industrial clusters, such as in the semiconductor industry in Silicon Valley, in the financial services and advertising industries in New York City, in the entertainment industry in Los Angeles, and in the production of country music in Nashville, TN, to name just a few such clusters. Industrial clusters are not only a source for innovation leading to productivity growth, but they also tend to stimulate the formation of new businesses, leading to ongoing growth in entrepreneurial activity (Porter, 1998).

How do we account for this geographic concentration of innovative activity? Much of the theoretical literature on urban agglomeration economies has focused on externalities in the production of goods and services rather than on invention itself. Nevertheless, the three formal mechanisms primarily explored in this literature—sharing, matching, and knowledge spillovers—are also relevant for innovative activity.⁶

This section discusses models and evidence on these traditional channels as they relate to innovation. The next section will discuss additional topics and adjacent fields that are more specific to innovation processes. We argue in this section that the traditional Marshallian externalities, even beyond knowledge spillovers, are especially important for the spatial concentration of innovative activity, over and above that evident for production. In addition, the “natural advantages” of an area, such as its unique culture and intuitions, explain the location of some innovative activity (Saxenian, 1994).

This chapter focuses mostly on evidence related to innovative clusters and local interactions across firms. These firms are usually presumed to be of small size relative to local aggregate activity, taking the local conditions around them as given. In Section 5, we return to this question and discuss research on whether innovation is concentrated in a single large firm, spread across many local firms, or something in between.

For this setting, an innovation production function is a convenient way to demonstrate the effects that these agglomeration forces have on innovative activity: $I_{ic} = g(A_c)F(RD_{ic}, K_{ic})$, where I indexes the innovative output of firm i in city c , and RD and K represent R&D and human capital usage, respectively, by firm i , and $g(A_c)$ represents agglomeration economies that are external to firm i but internal to the firm’s city.⁷ If firm i wanted to double its production of I , it could do so by doubling RD and K . It is important to note that it is unnecessary to double agglomeration economies in the city. If the agglomeration economies also doubled, the production of I would more than double.

One baseline empirical strategy is to model $g(A_c)$ as a function of urban size or density. This is roughly the approach taken by Carlino et al. (2007), who find that the rate of patenting per capita—or patent intensity—is about 20 percent higher in a metropolitan area with twice the employment density (jobs per

⁶ These themes are developed by Duranton and Puga (2004). Marshallian economies are often expressed as customer–supplier linkages, labor pooling, etc., and Duranton and Puga (2004) describe the theoretical advantages from formulating their common principles, such as sharing, matching, etc. Recent surveys of the empirical literature on agglomeration economies include Eberts and McMillen (1999) and Rosenthal and Strange (2004), and Combes and Gobillon (2014) provide a comprehensive update in this handbook.

⁷ See Audretsch and Feldman (2004) for further discussion of knowledge production functions in a regional setting.

square mile) of another metro area.⁸ While this elasticity is interesting, the approach does not tell which of the mechanisms are responsible for the increased patenting activity found in denser locations, or anything about the relative importance of the various mechanisms; this requires theoretical models of how each of the mechanisms operate. We consider each in turn using the Duranton and Puga (2004) taxonomy.

4.1 Sharing

The sharing of common inputs depends on the existence of economies of scale in the production of these inputs at a given location. Thick factor markets can arise when innovative activity clusters locally. These clusters allow each of their members to benefit as if they had greater scale (Porter, 1998). Thick local markets allow for the efficient sharing of a variety of specialized inputs and access to a pool of specialized and experienced workers and to a variety of specialized business services (e.g., patent attorneys, commercial labs for product testing, and trade organizations). Knowledge can flow rapidly through these linkages (Porter, 1990). The ability to outsource locally allows firms to lower production costs by maintaining leaner inventories, and Porter (1998) argues this effect is especially important for “advanced and specialized inputs involving embedded technology, information, and service content.” Industrial clusters allow members to experiment at lower costs, and, if they decide to innovate, firms can externally source what they require to quickly implement their innovations.

Helsley and Strange (2002) develop an interesting dynamic model of innovation in which a dense network of input suppliers facilitates innovation by lowering the costs needed to bring new ideas to realization. The model is dynamic, since previous innovations in an area produce inputs that make entrepreneurs more productive, leading to continuing growth in entrepreneurial activity. Gerlach et al. (2009) demonstrate that firms invest more in R&D and take greater risk in R&D choices in clusters compared with spatially isolated firms. They also demonstrate that similar firms choose different R&D projects when located in clusters, which creates a diversified portfolio of R&D investments.

The benefits that firms in a given industry gain from colocating are often enhanced using production technology based on modularity. Modularity is the method of making complex products or creating processes from smaller subsystems (modules) that are developed by a network of independent firms. Under this system, different suppliers are responsible for separate modules with the understanding that each supplier will follow “design rules” that ensure the modules fit and work together (Baldwin and

⁸ Carlino and Hunt (2009) find that the elasticity of patenting with respect to density remains about 0.2 regardless of whether they use unweighted patents or citation-weighted patents.

Clark, 1997).⁹ An important aspect of modularity is that innovative activity becomes decentralized across many firms instead of being concentrated in a single vertically integrated firm. The rate of technical progress can be enhanced by decentralization, since independent firms can focus more fully on innovation to their specific components compared with the divided attention each component receives when firms are vertically integrated. Sturgeon (2002) stresses that modular production allows colocated firms to use the benefits of spatial proximity to establish and manage global-scale production networks. Using evidence from case studies and interviews, Saxenian (1991) concludes that Silicon Valley–based computer system manufacturers became more efficient by spreading the costs and risks associated with the rapid changes in product designs and technologies by adopting modularity techniques.

In general, there is little empirical evidence on the importance of input sharing for firms engaged in innovative activity, even if the relationship is intuitive. Feldman (1994) shows that the presence of a local network of firms in related industries helps to refine new innovations and that the existence of specialized business services helps firms bring their ideas to fruition. Audretsch and Feldman (1996) find a greater tendency for innovative activity to collocate in industries in which skilled labor is an important input. We discuss the sharing of common labor pools in more detail later.

What is better established is the development and sharing of specialized business services. This has been especially true with the case of entrepreneurial finance (e.g., angels, VC). Traditional sources of financing, such as bank loans, may be unavailable to innovative startups due to their high risk, large financing requirements, and asymmetric information, especially in high-tech industries (Gompers and Lerner, 2001). As a result, VC organizations tend to invest locally in order to monitor their investments and to provide operating assistance to these firms (e.g., Gompers and Lerner, 2001; Horvath, 2001). Kolympiris et al. (2011) find VC investment within the biotech sector is particularly concentrated in areas of a 10-mile radius or less. These investors illustrate input sharing in many ways—their business models depend upon a certain scale and diversity of local opportunities, the value-added benefits they offer to

⁹ Modularity as a method of production has been around for some time. A classic example is automobile manufacturing in Detroit, in which car manufacturers purchased components such as car bodies, transmissions, brakes, etc. from largely local independent suppliers. More recently, computer system manufacturers have almost entirely relied on a network of local independent suppliers of modules in the production of final products. In related theoretical work, Hellmann and Perotti (2011) depict conditions under which innovative ideas are better developed in thick markets of innovative firms versus within larger organizations.

portfolio companies frequently stem from this local networking (e.g., introductions to key employees to hire), and so on.¹⁰

4.2 Matching

A second theory argues that thicker markets improve the quality of matches in local labor markets. Models that support this theory include Helsley and Strange (1990, 2002), Wheeler (2001), Berliant et al. (2006), Strange et al. (2006), and Papageorgiou (2013). For example, in Berliant et al. (2006), workers in large cities can be more selective in forming matches because the opportunity cost of waiting for a prospective partner is lower. Despite the fact that people are more selective, on average they form matches more quickly. As a result, the average output from matches is higher, and a higher share of the workforce is engaged in productive matches. Strange et al. (2006) find that firms with an (exogenous) uncertain demand for highly specialized workers benefit from locating in specialized industrial clusters, since these clusters provide a more diverse range of worker skills to meet unforeseen or challenging opportunities.¹¹

Also, specialized workers can readily find new positions without having to change locations in innovation clusters. For example, there is evidence of job-hopping in Silicon Valley (Fallick et al., 2006) and in software publishing (Freedman, 2008). This research shows that mobility increases with industrial concentration. Fallick et al. (2006) find much greater mobility of college-educated males employed in the computer industry in Silicon Valley compared with the interfirm mobility of similarly educated workers in the computer industry in other areas outside of California. Part of this turnover could be induced as firms and workers seek better matches. Moreover, Fallick et al. (2006) describe the important link between this mobility and the modularity described earlier.¹²

While labor market pooling may reduce firm costs, leading to greater profit, competition for workers and worker turnover can have the opposite effect (Matouschek and Robert-Nicoud, 2005; Combes and Duranton, 2006; Gerlach et al., 2009). Labor market pooling exposes firms to poaching of vital workers by rival firms and increases labor costs to retain employees. Matouschek and Robert-Nicoud (2005)

¹⁰ Samila and Sorenson (2011) provide evidence that VC investment has positive growth spillovers at the MSA level in the United States, but there is much more research that can be done on how VC activity connects with local economic development.

¹¹ Interestingly, Strange et al. (2006) find that when firms are involved with creating new products or processes, they tend to locate in large cities but not in industrial clusters.

¹² Mobility could be unusually high in Silicon Valley because of the unenforceability of noncompete clauses under California law (Gibson, 1999). It is important to note that Fallick et al. (2006) find that employee turnover in other industries is no higher in California than in other locations, suggesting that noncompete clauses are not the primary reason for the job-hopping observed in California. We return to noncompete clauses in Section 5.

analyze the role that investments in human capital have on the location decisions of firms. They show that whether human capital investments act as a force for or against geographic concentration depends on who is making the investments. If firms undertake the investment in human capital, they are less likely to agglomerate in an attempt to protect their investments from rival firms. Firms are more likely to agglomerate when workers undertake the investment in human capital.

In the model of Rotemberg and Saloner (2000), workers are more likely to undertake investment in human capital when firms are agglomerated. In a small labor market, workers who invest in acquiring skills have weak bargaining power with local employers unless they are willing to undertake costly relocations. This setting can generate holdup problems by a monopsonistic employer attempting to extract the surplus from the workers' investments, thereby reducing the original incentive to invest in skill development. In thick labor markets, the presence of a large number of potential employers mitigates the holdup problem.

There is a lack of empirical evidence on these issues, and the endogeneity of location and investment choices make these models difficult to test. One mentioned route is to exploit exogenous cross-country differences in legislation requiring firms to provide worker training. For example, German firms are required to provide some general training for their workers. Evidence for the Matouschek and Robert-Nicoud (2005) hypothesis would be found if firms in Germany are more dispersed than they would be in countries in which firms are not required to provide such training, all other things being equal. Tests might also be feasible through investments in immigrant workers in settings in which the worker is tied to his employer for a period of time (e.g., the United States' H-1B visa program).

4.3 Knowledge spillovers

The third strand of the theory argues that the geographic concentration of people and jobs in cities facilitates the spread of tacit knowledge, which is not easily codified or transferred through written documents. While the exact mechanism is not well identified in theory, the underlying idea articulated in Marshall (1890) is that the geographic proximity to knowledge facilitates the transfer of such knowledge as well as unplanned or serendipitous interactions among individuals fostering the exchange of information among workers and firms.

Steve Jobs understood that people's physical proximity is a key ingredient in the innovative process when he helped to design the layout of Pixar Animation Studios. The original plan called for three buildings, with separate offices for animators, scientists, and executives. Jobs instead opted for a single building, with a vast atrium at its core. To ensure that animators, scientists, and executives frequently interacted

and exchanged ideas, Jobs moved the mailboxes, the cafeteria, and the meeting rooms to the center of the building. Pixar has been incredibly innovative, as demonstrated by its many important breakthroughs in the application of computer graphics to filmmaking. Pixar has also been quite productive, generating 14 feature films that have earned 27 Academy Awards.

Information sharing has become a hallmark of Silicon Valley firms. Facebook, Google, and Twitter, for example, have designed their work and recreational spaces to enhance the interaction and the sharing of ideas among workers. Liu (2010) quantifies how the physical layout of a biotech firm shaped its internal knowledge flows, and Olson and Olson (2003) describe the sharp attenuation of collaboration and distance. Information sharing is not just a Silicon Valley phenomenon, and the value of face-to-face contact for innovation has long been recognized. In the 1940s, Bell Labs' (AT&T's R&D lab) new building was specifically designed so that everyone would interact with one another (Gertner, 2012). In 17th-century England, people gathered in coffeehouses to share ideas, with different coffeehouses attracting specialized clientele. The London Stock Exchange began life in 1698 in a coffeehouse in which merchants met, while a second coffeehouse frequented by shippers and traders became recognized as the place to obtain marine insurance and gave rise to Lloyd's of London.

While theories of knowledge spillovers were originally developed to explain the concentration of industries in general, they are particularly important in explaining the clustering of innovative activity. R&D, more than most industries, depends on new knowledge. Often, the latest knowledge about technological developments is valuable to firms but only for a short time, and the reciprocal exchange of information among colocated firms engaged in innovation can reduce uncertainty (Feldman, 1994). Thus, it behooves innovative firms to locate near sources of information and each other.

This extended section reviews knowledge spillovers in three steps. We first outline its connection with growth theory and some of the economic geography models that have been influential in this context. These models identify a Mincerian wage relationship that has been frequently used in empirical work, and the second part highlights this body of work. We then review the large body of work using patents and other techniques.

4.3.1 Knowledge spillovers and endogenous growth

Solow (1957) shows that the accumulation of physical capital is insufficient for long-run growth of per capita output. In Solow's model, the long-run rate of growth of per capita output is determined largely by an exogenous rate of technical progress, providing no explanation for productivity improvements. Since

the rate of productivity growth is the most important determinant of long-run growth, treating such an important factor as given leaves many unanswered questions.

Beginning in the mid-1980s, economists developed theories of endogenous growth in which cities often play a prominent role. Lucas (1988) emphasizes the role that the stock of human capital in cities plays in economic growth. While the channels through which knowledge spillovers are transmitted are not well articulated, the basic idea is that skilled or educated workers are better able to receive new knowledge from others. The acquisition of knowledge is related to cities since the dense concentration of people and firms in cities creates an environment in which new ideas travel quickly. Thus, cities facilitate knowledge spillovers. The idea that the exchange of knowledge among individuals enhances productivity can be expressed by allowing production to depend on aggregate human capital in a city. Let Y_{ic} denote a homogeneous final good produced by worker i in city c according to:

$$Y_{ic} = H_c^\delta h_{ic},$$

where h_{ic} represents the human capital of worker i and H_c^δ is aggregate human capital in the city. Notice there are constant returns to scale in the human capital of worker i , and increasing returns are associated with aggregate human capital at the city level. The parameter δ reflects the importance of an external return in production. The aggregate measure of human capital can be expressed as:

$$H_c^\delta = \left\{ \frac{1}{n} \sum_{n=1}^N h_{ic}(n)^\rho \right\}^{\delta/\rho},$$

where $h_{ic}(n)$ is the human capital of the n -th worker and ρ is a parameter governing how the amount of the human capital of different workers is aggregated to form the city's stock of human capital. When $\rho = 1$, the average level of human capital in a city is what matters when creating the externality. When $\rho > 0$, the skills of the most talented workers create the externality. If the most talented workers sort into the largest cities, as Glaeser and Saiz (2004) find, externalities will not be equalized across cities. In most empirical specifications of Lucas's model, $\rho = 1$ is assumed.

Assuming the zero-profit condition holds, one obtains a relationship between the nominal wage of workers and the aggregate measure of human capital that can be expressed in a Mincerian wage regression:

$$\ln W_{ic} = \delta \ln H_c + \nu \ln h_{ic} + \varepsilon_{ic},$$

where the parameter of interest is δ .¹³ Later, we will review wage-based evidence in this regard. It is important to highlight that this specification leaves many open issues. For example, endogeneity bias is an important concern when estimating δ , since aggregate schooling may be determined simultaneously with wages. In addition, the specification does not depict whether the learning effect is within industries of cities or across them. Third, most empirical studies consider quantitative measures of human capital, such as years of schooling, but differences in the quality of human capital are not captured. A fourth issue with the previously mentioned formulation is that the spillovers are static.¹⁴ A fifth issue, which we articulate more fully when reviewing these studies, is that factors beyond knowledge spillovers can give rise to these Mincerian wage regressions, and so evidence derived from these approaches is consistent with knowledge spillovers but not complete proof of them.

Eaton and Eckstein (1997) embed the human capital accumulation structure of Lucas (1988) within a system of cities framework. Cities grow at a common rate in equilibrium, with the relative size for a given city being determined by the environment for learning that the city provides. In Eaton and Eckstein (1997), the agglomeration externality is static. Glaeser (1999) proposes a model with a dynamic externality in which the direct interactions among workers in cities are the basis for accumulation and diffusion of knowledge. In the model, young and unskilled workers learn through interactions with older and skilled workers. Cities provide greater opportunities for younger workers to meet and interact with older workers. Black and Henderson (1999) develop a model of human capital accumulation and population growth in cities. In the model, human capital externalities make workers more productive in larger cities. Workers devote some of their time to the accumulation of human capital, which, in turn, makes cities more attractive and leads to population growth. The growth in population, in turn, reinforces the externality in human capital. Black and Henderson's (1999) model comprises two sectors in which cities are specialized in a particular sector. Although cities differ along a variety of dimensions, all cities grow at the same rate in the long run. De la Roca and Puga (2012) use Spanish data on individual worker characteristics and their employment histories to investigate the urban wage premium and find that

¹³ Duranton and Puga (2014) provide a microfoundation for this model in which the externality in human capital arises through entrepreneurship. If entrepreneurs are overrepresented in the population of more-educated workers, then cities with relatively more-educated workers will be more entrepreneurial and grow faster. See also Jones (2005).

¹⁴ Lucas (1988) proposes a model with dynamic externalities. In Lucas's models, individuals decide how much time they will devote to current production and how much time they will spend acquiring skills. Growth is now driven by an externality in the accumulation of human capital in cities. In this setup, the human capital externality is the engine of growth and an agglomeration economy. Duranton and Puga (2014) provide further details.

workers receive an earnings premium in larger cities. They also find evidence of a dynamic externality in that workers tend to accumulate more valuable knowledge in larger cities, leading to faster earnings growth.

Most recently, Davis and Dingel (2013) develop a spatial equilibrium framework to show why skill premiums are higher in large cities. They develop a system of cities model in which the exchange of ideas is costly and leads to agglomeration economies. The model has tradable and nontradable goods, labor is the only factor of production, and idea exchanges are local. Agents allocate their time according to the expected gains from exchanging knowledge, where the gains are greater in places where prospective partners are more numerous and of higher ability. Agents with the greatest ability to receive new knowledge are most willing to pay for this knowledge. This framework leads to cities of various sizes, with larger cities exhibiting greater skill premiums and productivity and thus offering higher wages, but also higher housing costs.

In another branch of the literature, Romer (1986, 1990) formalized the relationship between knowledge and economic growth. Romer divided inputs into two categories: rivalrous inputs and nonrivalrous ideas. The notion is that once a new design is introduced by one firm, the design is subsequently available to all other firms. This nonrivalrous nature of knowledge is central for economic growth in Romer's model. Consider the constant returns to scale production function of a representative firm i at time t :

$$Y_{it} = K_{it}^{\alpha} (A_t L_{it})^{1-\alpha} ,$$

where Y indexes the output of firm i , A represents the aggregate amount of knowledge available to all firms at time t , and K and L are the inputs of capital and labor used by firm i . If firm i wanted to double its production of Y , it could do so by doubling K and L . Due to the nonrivalrous nature of knowledge, it is not necessary to double the stock of knowledge. If the stock of knowledge is also doubled, the production of Y would more than double. Therefore, incorporating the stock of knowledge into the production function leads to increasing returns to scale because of the nonrivalrous nature of knowledge.

One concern is that in a competitive market, once a new idea is discovered, the marginal cost to supplying the idea to other users is zero. But private firms undertake costly R&D to profit from this research. It is unlikely that R&D will be undertaken if a firm cannot profit from this research. Romer (1990) addresses this problem by treating knowledge as a quasi-public good in that, although new knowledge is nonrivalrous, the owner of a new idea can exclude others from its use through patents and trade secret protection, for example. The rents that the owners of patents receive compensate them for the cost of discovering new ideas.

The idea that patent and trade secret protection gives firms receiving such protection monopoly power suggests that models that depart from perfect competition are needed. This is accomplished in the innovation-based strand of the literature that formalizes Schumpeter's idea of growth through creative destruction (Aghion and Howitt, 1992; Grossman and Helpman, 1991; Aghion et al., 2014). In these models, innovation takes place along a "quality ladder" that is driven by temporary monopoly profits that innovating entrepreneurs require as compensation for the cost of innovating. Every product has its own current point on a quality ladder, and firms undertake R&D to improve the quality of a product. Once a product is introduced, however, the technology diffuses and rival entrepreneurs can produce the same quality product at a lower cost and hence offer a lower selling price. Entrepreneurs can recapture market advantages through innovation, which leads to climbing up the quality ladder.

Kelly and Hageman (1999) and Duranton (2007) extend the quality ladder approach to a regional setting. In Duranton (2007), cities are composed of industries, and production for an industry is concentrated at the site of the latest breakthrough invention. When a new discovery takes place for an industry in another city, the industry relocates to the new center. This quality ladder model provides a solid foundation for explaining the fast churn of industries across cities, the slower movement of cities up and down the size distribution (as industry-level shocks are partially netted out), and the stable long-term distribution of city sizes. Kerr (2010a) provides evidence of the Duranton (2007) model when looking at the locations of the breakthrough patents by technology field during the 1975–1985 period and the subsequent growth of technologies across locations. The link of these breakthrough locations to subsequent production is less established.

While cities have this long-standing theoretical connection to growth, an open question is the quantitative importance of the externalities associated with human capital in cities for national economic growth. Duranton and Puga (2014) develop a system of cities models in the tradition of Henderson (1974) and address this question. Using recent estimates found in the literature for the elasticity of agglomeration economies and the elasticity of congestion cost with respect to city size, they calculate that cities account for almost two-thirds of aggregate output growth. While this calibration exercise is tantalizing, it remains speculative and will hopefully be the groundwork for complete assessments.

One potential pitfall should be noted about the connections between endogenous growth theory and studies of innovation clusters. Researchers need to think hard about the role of mobility in their setting and the resulting equilibrium conditions (e.g., Glaeser, 2008). In many core growth models, population is fixed because the presumed application is to a country, and thus the impact of innovation is pushed toward the income per capita development. Cities can adjust their population, and the impact of

innovation may come through population growth instead. A full articulation requires thinking about mobility costs, housing prices, amenities, and similar factors alongside the standard issues such as regional convergence (e.g., Magrini, 2004). It is likewise important to consider how the diffusion of ideas across cities and nonpermanent relocations can influence the degree to which the cities in question are being governed by the logic and assumptions built into growth models.¹⁵

4.3.2 Empirical evidence on localized knowledge spillovers – wages

Since knowledge spillovers are invisible, empirical verification of them relays on indirect evidence. The two main approaches in regional studies are Mincerian wage regressions and analyses of patent citations. We start with the wage analyses that are motivated by the theoretical models of city and growth. From the outset, we caution that these studies relate local human capital stocks to wages, and thus while they are consistent with knowledge spillovers, they are not conclusive evidence for them. Combes and Gobillon (2014) discuss this issue extensively in their chapter in this handbook.

Many studies use metrics of educational attainment as a proxy for the human capital stock of cities. The conceptual idea is that a greater share of educated workers in a city make other workers in that city more productive. The share of the adult population with a college education differs dramatically across cities, with a three-fold range evident in 2010 between the high of 28 percent in Raleigh, NC, and the low of 9 percent in Visalia, CA. Among employed workers specifically, Moretti (2012) shows that there is an even larger five-fold differential. Most studies look for this increased productivity by examining worker wages, quantifying the additional earnings that similar workers (e.g., in terms of their age, education, occupation, industry, and experience) receive as the share of college graduates in their city increases.

Using 1980 census data, Rauch (1993) estimates that each additional year of average education for a worker in a city increases expected wages 3 percent to 5 percent. As already noted, endogeneity is an important concern, since higher wages might cause individuals to get more schooling instead of the other way around. This reverse causality would lead to an upward bias in ordinary least squares (OLS) estimates of the social return to schooling. Alternatively, OLS estimates of the social return to schooling can be biased downward if highly skilled workers are attracted to amenity-rich locations, since these workers accept lower wages in return for greater amenities. In addition, there could be unobserved factors

¹⁵ As an example, some studies apply growth theory frameworks to primary metropolitan statistical areas (PMSAs) in the United States. PMSAs separate San Jose, CA, home of Silicon Valley, from San Francisco. This separation is clearly inappropriate, as the tech cluster covers the whole region. In fact, many tech firms in Silicon Valley offer shuttle buses for employees who wish to live in downtown San Francisco.

that correlate with measures of city-level human capital that independently affect productivity and wages (e.g., good local governance), leading to biased estimates.

To address endogeneity bias, Acemoglu and Angrist (2000) use compulsory K–12 school attendance laws that existed in an individual’s state at the time of his birth and child labor laws as instruments for determining a state’s average level of schooling. They also use an individual’s quarter of birth as an instrument for individual-level schooling. Instrumental variable estimates suggest that a one-year increase in average schooling is associated with about a 1 percent to 2 percent increase in average wages compared with a similar OLS regression that yielded a 7 percent external return. One caveat to these findings is that they empirically identify the social return associated with the supply of workers with secondary schooling, as opposed to the supply of the most educated workers frequently emphasized in the theory on human capital spillovers. A second issue is that states are not the best geographic scale for estimating the social returns to education. These externalities are certainly stronger at the city or subcity level.

Moretti (2004a) estimates the social return associated with the share of an MSA’s population with a college degree during the period 1979–1994. Moretti (2004a) deals with endogeneity by using the presence of a land-grant college in the MSA as an instrument for college share.¹⁶ He finds that a 1 percent increase in the share of college graduates increases the wages of college graduates by 0.5 percent. The wages of high school graduates increase by 1.5 percent, and the wages of high school dropouts increase by almost 2 percent. One issue with this approach is that the spatial distribution of land-grant universities is highly uneven (Shapiro, 2006). Also, reverse causality remains a concern. Shapiro (2006) points out that it is hard to compare correlations between college attainment and land-grant schools today with correlations from the past because the Census Bureau did not begin asking about educational attainment before the 1940 census, “by which time the land-grant schools were already of significant size.”

Ciccone and Peri (2006) point out that the findings reported so far may overstate the social returns to human capital, as the estimates confound human capital externalities with factor demand effects. They show that an increased share of highly skilled workers in a city alters the city’s skill composition such that an increase in a city’s average wage can occur in the absence of human capital spillovers. Ciccone and Peri (2006) develop an approach that holds the labor force skill mix constant over time. They find no evidence of a return to a one-year increase in average schooling during the period from 1970 to 1990 once they account for imperfect substitutability between education groups. Combes and Gobillon (2014) raise

¹⁶ To account for unobserved worker characteristics, Moretti (2004a) uses longitudinal data from the National Longitudinal Survey of Youths (NLSY). Moretti reports that omitted worker characteristics are not a major source of bias and uses lagged age structures to account for unobserved MSA-specific demand shocks.

important concerns regarding this approach including, for example, that the evolution of wages in a local area may be influenced by changes in the local skill composition of workers.¹⁷

As pointed out by Duranton (2006) and Henderson (2007), the literature on the external returns to education may still be upwardly biased, since much of the work does not control for urban agglomeration economies. Glaeser and Saiz (2004) find that skilled workers sort into larger cities, indicating that the estimated effects of human capital on individual worker wages may be picking up urban scale effects. In fact, when Rauch (1993) includes metropolitan land area (an instrument for metropolitan population scale) in his regressions, the coefficient on the standard metropolitan statistical areas average education variable in the wage equation is still positive but only marginally significant (see Table 1, column 6 in Rauch). Ciccone and Peri's (2006) empirical analysis includes aggregate metropolitan employment, which can be interpreted as an agglomeration effect. Interestingly, as Henderson (2007) points out, although Ciccone and Peri (2006) find no evidence of a human capital externality, they do find evidence of an urban agglomeration effect, underscoring the need to distinguish these effects.

Rosenthal and Strange (2008) use 2000 census data to estimate the impact of both agglomeration economies and very localized knowledge spillovers on individual worker wages. They recognize that knowledge spillovers attenuate rapidly within a few miles from the source of the externality and thus pay close attention to the rate at which the "wage-agglomeration" relationship decays with distance. Their approach involves differencing and the use of instrumental variables to address endogeneity and measurement error in their agglomeration and human capital variables. They find that the spatial concentration of employment (the measure of agglomeration economies) within a five-mile radius is positively related to wages. Second, they find that swapping 50,000 workers with less than a college degree within five miles with workers who have at least a college degree increases the wage of a typical worker by 6 percent to 15 percent. Finally, agglomeration and human capital effects attenuate sharply with distance. These are important findings in that they recognize that the human capital externalities are certainly stronger at the "neighborhood" level and that they will be substantially underestimated at even the metropolitan area level. As we will see, a growing number of studies report that knowledge spillovers attenuate rapidly with increasing distance from the source of that knowledge.

¹⁷ Combes and Gobillon (2014) provide a thorough discussion of the empirical issues associated with the imperfect substitutability between skill groups and approaches for dealing with the issue. Bacolod et al. (2010) point out that important aspects of skills can be missed under the usual approach of equating a worker's skills to education. In their framework, individual characteristics such as intelligence, personality, dexterity, and the like interact with a local environment to produce skills. They find education is an important part of the process in skill determination; however, a worker's innate abilities, along with urban agglomeration, contribute to the development of skills.

An important issue that is largely ignored in this literature strand is that theory suggests productivity differentials should be capitalized in land rents as well as in wages. The degree to which the externality is capitalized in wages or rent depends on the elasticities in local land and labor markets. Rauch (1993) and Conley et al. (2003) are the only studies that allow for the capitalization of the education externality into both wages and rents and find a positive effect on rents. The finding of a small or zero externality when simply looking at wages should not be interpreted as evidence that the externality is small or even zero; it is imperative to know the magnitude of human capital spillovers (the sum of its effect on wages and rents). Knowing the extent to which spillovers are capitalized into either wages or rents also has important welfare and policy implications. If the externality is largely capitalized in land values, then landowners will benefit, and a tax on land to finance education can be efficient. However, if some regions benefit more than others from increasing returns due to knowledge spillovers, a positive correlation could exist between public spending on higher education and the net exodus of skilled people (college graduates) in those regions with decreasing returns to higher education.¹⁸

Another issue is that static Mincerian regressions do not analyze growth in wages, but a primary advantage of cities (particularly bigger ones) is that they facilitate learning. These dynamic advantages lead to greater human capital formation and perhaps more experimentation and innovation. Glaeser and Maré (2001) were among the first to show that workers learn more quickly in dense metropolitan areas. The dynamic externality is such that the initial effect on wages is small upon workers arriving in a new city but accumulates over time. More recently, several studies confirm faster wage growth in larger cities (Baum-Snow and Pavan, 2013; De la Roca and Puga, 2012; Wang, 2014). De la Roca and Puga (2012) use longitudinal data for a sample of Spanish workers during the period 2004–2009 to investigate sources of urban wage premium. They find during the seven years of their sample that one-half of the urban wage

¹⁸ Recently, researchers have demonstrated how ignoring rent differentials across cities (a measure of cost of living differentials) can bias estimates of real income inequality. Several studies document that an increased share of college grads sorted into high-wage U.S. cities between 1980 and 2000 (Moretti, 2013; Lindley and Machin, 2014; Diamond, 2012). As college workers are increasingly located in expensive cities, Moretti (2013) calculates that at least 22 percent of the increase in spatial college wage premium between 1980 and 2000 can be explained by geographic differences in the cost of living. One possibility is that college graduates receive greater utility from city amenities than other workers, and the higher costs reflect access to these amenities. Alternatively, the relative demand for college graduates may have increased in high-cost locations as a result of an increase in the productivity of more skilled workers. If so, the benefits college grads receive from the spatial college wage premium is somewhat offset by higher living costs. Moretti (2013) finds that the geographical sorting by skill groups is mostly driven by changes in their relative demands and concludes that price-adjusted welfare inequality is lower than nominal income inequality. Diamond (2012) also studies changes in welfare inequality during the period from 1980 to 2000. In contrast to Moretti (2013), Diamond finds that despite higher rents, the well-being inequality gap between college grads and high school grads is actually 20 percent greater than the observed gap as a result of the increased utility college grads derive from the consumption of localized amenities. That is, the growth in welfare inequality between skill groups is greater than the growth in income inequality between these two groups.

premiums is static, received upon arriving in a city, while the other half accumulates over time associated with the dynamic benefits from learning. Wang (2014) finds college-educated workers who spend their early years in large cities tend to have faster wage growth.

In summary, this literature has progressed substantially in a very short period. From the perspective of this chapter, the most needed area of further inquiry is a sharper connection of wage levels and growth to knowledge spillovers in cities. To provide empirical traction and identification, such analyses will likely need to look at microareas within cities, exploit administrative data on wages, and combine these data with additional information about knowledge flows. These ingredients are fortunately becoming available, even to the point of imagining future studies of wages, learning, and knowledge flows within individual buildings and city blocks.

4.3.3 Empirical evidence on localized knowledge spillovers — patents

A second empirical route for studying knowledge flows looks at patents. Citations made in patent documents have received particular attention, as they directly relate to knowledge flows, and this literature has exploded in size since the digitization of patent publications in the late 1990s. Jaffe, Trajtenberg, and Henderson (1993, hereafter JTH) is a particularly famous paper. JTH study whether inventors are more likely to cite other inventors who are geographically close to them, making use of the detailed location information available on patents for inventors. If knowledge spillovers are localized within a given metropolitan area, then citations to patents within a given metropolitan area should come disproportionately from other inventors who are located within that metropolitan area.

JTH examine two samples of originating patents—950 patents from 1975 and 1,450 patents from 1980. They also consider institutional variation, separating university patents, those from the top 200 R&D performing firms, and those from other domestic corporate applicants. From the start, JTH are concerned that technologically related activity may be clustered geographically for reasons unrelated to knowledge spillovers. For example, the semiconductor industry could be concentrated in Silicon Valley because of the presence of VC firms, which might lead to a spurious inference about the role of knowledge spillovers. So, for each citation, JTH select a control citation that is technologically similar to the original citation and from the same time period.

This matched-sample approach tests whether citations are more localized relative to what would be expected based on the preexisting spatial distribution of technological activity. JTH calculate the geographic matching frequency between the citing patents and the originating patents and compare these with the control sample. JTH find a significant “home bias” in patent citations. Excluding self-citations,

citations are two to three times (for the corporate samples) to six times (for the university sample) more likely than control patents to come from the same metropolitan area. Since the observed citation-linked patent pairs are found to be in closer spatial proximity than are the control patent pairs, JTH take this as strong evidence that knowledge spillovers are geographically concentrated.¹⁹

This seminal study led to significant further inquiry and refinements. A first line of work seeks to clarify whether patent citations really measure knowledge flows. Jaffe et al. (2000) investigate this issue using a survey of inventors. They find evidence of communications among inventors but also note that citations are a “noisy signal of the presence of spillovers.” Citations can be listed on patents by inventors, or they can be added by patent examiners reviewing and approving the patent application. Since 2000, patent publications separate these types of citations. Thompson (2006) finds that citations added by inventors are much more spatially localized around the inventor than those citations added by examiners, which is particularly powerful evidence of the localized knowledge flows.

A second branch of inquiry involves the identification of the control group. Thompson and Fox-Kean (2005, hereafter TFK) revisit the JTH study and find the results are sensitive to the way the control patents are selected. JTH draw control patents using three-digit technological classifications. This approach works to the extent that the three-digit level represents appropriately aggregated technology fields; if in fact subfields are prominent and are themselves clustered for various reasons, then the control patents do not adequately account for underlying technology clustering. Continuing the earlier example, the three-digit semiconductors technology class may itself be composed of several variants that are clustered in different cities. If this is true, then it may be more appropriate to draw from the lower technology level. When TFK use six-digit technological classifications to select the control patents, they find substantially less evidence of knowledge spillovers at either the state or metropolitan area level.

There is tension, however, about the selection of control patents—it is important to make the control groups as accurate as possible, but one can narrow the groups from which the controls are drawn too much. Indeed, there is a sense in which every patent is deemed a unique contribution, and yet one cannot obviously define a peer group within each patent. This tension is especially salient for patent counterfactuals because the three-digit level of the system has 400–500 technologies, depending upon the period studied, while the six-digit level typically has more than 150,000 technologies. Thus, the

¹⁹ The selection of control citations versus other techniques to control for local economic activity may lead to understatement of agglomeration for innovation. In an extreme example, the need for spatial proximity for knowledge spillovers may be so high as to have a whole technological field cluster into a very narrow space. In this setting, looking for deviations in spatial concentration between actual and control citations from the same field would under-represent how important the localized knowledge flows are for the field.

narrowness of the control groups differs remarkably, and there are unfortunately no midpoints in the hierarchy. In a recent and provocative study, Murata et al. (2014) use sophisticated sensitivity techniques that allow for both levels of controls to illustrate the differences between JTH and TFK. Their work finds significant evidence for localization, suggesting TFK may have sliced too thinly.

A third branch of inquiry seeks to relax the fixed geographic boundaries used in the original studies, similar to the broader movement toward the continuous distance metrics described earlier. Since knowledge spillovers tend to be highly localized within metropolitan areas, states and metro areas may not be the appropriate geographies for studying them in many applications. Given the rapid distance decay in knowledge spillovers identified in these studies, researchers using labor market boundaries (such as MSAs) or administrative boundaries (such as states) run the risk of underestimating the importance of knowledge spillovers in the location of innovative activity. Murata et al. (2014) and Carlino et al. (2012) use distance-based approaches and find substantial evidence supporting the localization of patents and patent citations.

Going forward, we hope that researchers will continue to use these techniques. One methodological point is particularly worth emphasizing. Studies today often use the original JTH empirical strategy of drawing a single control citation for each real observation. With today's computing power, there is no reason to stop with a single counterfactual, but instead researchers should draw a distribution of counterfactuals similar to the Duranton and Overman (2005) technique. Murata et al. (2014) provide a comprehensive depiction. Drawing 100 counterfactuals improves upon a single draw by creating a more accurate baseline and, even more important, by providing the basis for determining if a deviation is statistically significant.²⁰

Other studies find evidence of patent localization without using citations. The Kerr and Kominers (2014) study depicted earlier uses Duranton and Overman (2005) techniques with the spatial location of patents themselves versus citations. This study then relates tightly packed clusters to those technologies displaying very short spillover ranges. Lin (2014) looks for evidence of knowledge spillovers in patent interferences, which are administrative proceedings to determine which applicant is entitled to the patent when multiple applications are submitted for the same invention. The basic idea is that inventors involved in an interference are likely to share certain knowledge, so patent interferences may offer evidence of knowledge spillovers among inventors. If localized knowledge spillovers are important, we should see

²⁰ Similarly, patents often list more than one inventor. While the location of the first inventor listed on the patent has typically been used in the past, there is little reason with today's tools to not factor in the locations of all inventors. This critique also applies to technologies, as most patents list multiple technology codes, but historically only the first one is considered.

that inventors in close geographic proximity should be disproportionately involved in interferences. Lin (2014) finds that patent interferences are more likely to be observed between inventors located close to one another as opposed to those located farther apart—evidence that common knowledge inputs among independent inventors are highly localized.

Carlino et al. (2012) describe how the geographic concentration of R&D labs can be used to determine more appropriate geographic boundaries in which knowledge spillovers are most likely to occur. They introduce a multiscale core cluster method to measure the localization of R&D labs across continuous space. Among others, they find a cluster of R&D labs centered in Cambridge, MA, and another in Silicon Valley. On the whole, corporate R&D labs are significantly more spatially concentrated than manufacturing employment is. Next, they assign patents and citations to these clusters and find evidence of a significant home bias in patent citations (excluding self-citations) in most clusters. Carlino et al. (2012) find that patent citations are over 12 times more likely to come from the San Jose, CA, cluster and more than six times more likely to come from the Cambridge cluster as from their respective control patents chosen to match the geographic concentration of technologically related activities. This finding provides not only evidence of localized knowledge spillovers in patent citations but also much stronger evidence than reported in prior studies.

As a whole, the patent-based evidence provides confirmation of the localized nature of knowledge flows. This evidence relates to the extreme attenuation of knowledge spillovers noted earlier in this study in Figures 3 and 4 (e.g., Rosenthal and Strange, 2001, 2003; Arzaghi and Henderson, 2008). Several other studies have found that knowledge spillovers dissipate rapidly with distance (e.g., Conley et al., 2003; Moretti, 2004b and 2004c; Audretsch and Feldman, 1996; and Adams and Jaffe, 1996).²¹ While there is much more detail that we need to learn, the general conclusion of this work on localized knowledge flows for innovation appears sound.

An important topic for future work is to evaluate how patent-based estimates of knowledge flow distances compare with other estimates. JTH suggest that since not all ideas are patented, their findings represent a lower bound on the incidence of information flows among inventors. It is also likely that patent-based studies underestimate the decay rates of knowledge flows. The transmission of tacit knowledge likely requires tighter spatial connections, for example, as evidenced by the many seminars and conferences that

²¹ Other nonspatial aspects of distance may play a role. Using input–output connections to measure economic distance among manufacturing establishments, Moretti (2004c) also finds evidence consistent with the view that human capital spillovers decay with economic distance.

accompany papers and journal articles. We see the very rapid Arzaghi and Henderson (2008) decay rates as evidence in this regard and hope for ongoing inquiry in this vein.

A second important topic is to link local knowledge stocks into patenting outcomes (and further into productivity and local growth). While evidence based on strong correlations exists in this regard, causal identification remains elusive. For example, Carlino et al. (2007) find that of the factors they considered, local human capital has the strongest correlation with patenting rates per capita in U.S. cities. A 1 percent increase in the adult population with a college degree is associated with an increase in the local patenting rate by about 1 percent. The focus of Carlino et al. (2007) is on the connection of local employment density and patenting, however, and thus the authors do not work to establish a causal relationship between human capital and the patenting rate. An extension of their paper would be to use the spatial distribution of land-grant universities as an instrument for local college share.

4.3.4 Empirical evidence on localized knowledge spillovers — other approaches

A number of approaches provide evidence on knowledge spillovers without using patents or wage regressions. Some studies consider how differences in education levels across cities translate into differences in firm productivity. The hypothesis is that firms situated in cities with high human capital levels will be able to produce more output using the same level of inputs compared with similar firms located in cities with low human capital levels. Moretti (2004c) looks at the growth in the productivity of manufacturing plants during the 1980s and finds that, on average, human capital spillovers account for a meager 0.1 percent increase in output per year, or about \$10,000 per year.

Lychagin et al. (2010) assess the relative contributions of localized knowledge spillovers on firm productivity using Compustat data for the period 1980–2000. Lacking data on the location of a firm’s R&D labs, Lychagin et al. (2010) use USPTO patent data to create a spatial distribution of the locations of a firm’s inventors, which is used as a proxy for the firm’s R&D. They find a positive correlation between this R&D proxy and productivity growth conditioning on technology and product-market spillovers. They use predetermined variables and Arellano and Bond (1990) techniques to argue a causal relationship.

A second group of studies consider the rise of the “skilled city.” Glaeser et al. (1995) find that postwar income per capita grew faster in cities with high initial levels of human capital. A one-year increase in median years of schooling in 1960 increased subsequent income growth by almost 3 percent. Glaeser and Saiz (2004) find that a one standard deviation increase in a metropolitan area’s initial college share results in a 2.5 percent increase in its size. Similarly, Shapiro (2006) finds that from 1940 to 1990, a 10 percent

increase in a metropolitan area's share of college-educated residents (e.g., from 20 percent to 22 percent) raised employment growth by approximately 2 percent.

Beyond the traditional agglomeration literature, several recent papers develop interesting identification approaches to investigate knowledge spillovers among scientists, often in unique historical settings. Moser (2011) uses an exogenous increase in patenting rates for chemicals to examine the effects this increase had on geographic concentration and the diffusion of ideas. She finds evidence of a spatial concentration among prize-winning inventors at four World's Fairs in the 19th century and early 20th century, although innovative activity subsequently became less localized after patenting rates increased. Ham and Weinberg (2014) find that scientists in chemistry, medicine, and physics who are proximate to Nobel laureates show a "small, but meaningful" increase in the probability of starting their own Nobel research agenda, suggestive of horizontal knowledge spillovers among individuals of comparable stature. Menon (2009) finds that prolific inventors have a positive and lagged effect on the patenting of inventors who are less active in the same city.

While these studies find benefits from proximity to prominent scientists, Azoulay et al. (2010) and Waldinger (2012) challenge the existence of localized knowledge spillovers originating from celebrated academics. Waldinger (2012) deals with endogeneity by using the firing of scientists by the Nazis as a source of exogenous variation in the peer group of scientists remaining in Germany. He finds that the publishing activity of the scientists in physics, chemistry, and mathematics in all German universities whose departments suffered losses during the period from 1925 to 1938 did not decline compared with that of other scientists. Azoulay et al. (2010) use the sudden and unexpected death of superstar collaborators as a source of exogenous variation. While they find that the death of a superstar scientist reduces coauthor productivity, the declines in output are similar regardless of the geographic distance among collaborators.

In sum, much of the extra evidence beyond wages and patents is also consistent with the existence of localized knowledge spillovers, although some studies find null results. A central limitation of the studies reviewed is that there is very little insight into how knowledge is transmitted among individuals living in close geographic proximity. Presumably this occurs through both professional and social networks, but this has not been confirmed. Henderson (2007), in particular, poses questions about how these networks form, how members are accepted, and what the influence of these phenomena on spatial outcomes is.

While our depiction has mostly presented spillovers operating outside of market-based interactions, geographic proximity may facilitate the exchange of knowledge through contractual and market-based channels. Knowledge can spread in a city as workers move across firms, and this is not a spillover to the

extent that these workers are compensated for the knowledge they bring to their new firms. Of course, even with competitive wages, there could remain uncompensated knowledge spillovers originating with these newly hired workers and benefiting the firm's existing workers. Employers use tactics ranging from employment law to stock options (e.g., Pakes and Nitzan, 1983) to limit this mobility and knowledge diffusion. Later, we depict some studies that suggest that worker mobility over firms may be an important part of the local knowledge effects discussed.

4.4 Marshallian equivalence

Duranton and Puga (2004) point out that the empirical identification of the various microfoundations (sharing, matching, and learning) can be problematic, since they all are related to the scale of activity at a given location (Marshallian equivalence). The Rosenthal and Strange (2001) and Ellison et al. (2010) studies mentioned an earlier attempt to comparatively assess the forces to overcome this issue. A calibration exercise, or quantitative approach, is another way to identify the outcomes of a particular mechanism. For example, Papageorgiou (2013) documents a number of observed features relating to occupational switching, moving patterns, and wages in large cities. Using these features, he designs a model in which larger cities offer workers better matches because they have more occupations. In the calibrated framework, Papageorgiou (2013) finds that better matching in large cities accounts for roughly 40 percent of the observed wage premium. Although the magnitude of the matching externality reported needs to be confirmed, quantitative methods are an underutilized tool in regional analysis.

4.5 City size and industrial composition

We next describe work on the link between innovation and the sizes of cities and their industrial compositions. This relates to the previously mentioned formal theories (e.g., thicker markets are more likely in larger cities), but there is also some distinction. The scale of a city in which an invention is found relates to the degree of spillovers in invention, but it is also influenced by factors such as the degree to which internal and external R&D of large firms is consolidated into a single location versus dispersed over plants. How the forces all empirically add up is intrinsically interesting too, especially as a guide for theory models.

Empirically, the rates of innovation generally increase across much of city size distribution, but this pattern does not hold true among the largest cities. Carlino et al. (2007) demonstrate this inverted-U shape to city size and patenting intensity, calculating the inflection point to be a city approximately the size of Austin, TX. Separate from aggregate city size, the density of activity in cities predicts innovation rates. Carlino et al. (2007) quantify that a doubling in the employment density of a city raises patents per capita

by about 20 percent. This pattern is also nonmonotonic, with patenting intensity maximized at a density level on par with that of Baltimore or Philadelphia.²²

Multiple pieces in the urban literature connect innovation to city sizes or densities from a theoretical perspective. The classic theories of Marshall (1890) and Jacobs (1969) link larger cities to the greater development of knowledge because of increased specialization and diversity, respectively. In these formulations, exogenous spillovers monotonically provide a greater foundation for knowledge development in larger cities. This monotonicity is true in the Glaeser (1999) and Helsley and Strange (2002) models discussed earlier, with, for example, congestion for access to a fixed factor ultimately being used to limit city size.

By contrast, Helsley and Strange (2004) depict how purposeful knowledge barter in cities can create a nonmonotonic relationship to city size and knowledge work. The nonmonotonic role of city size comes in their model because of the greater ease with which agents can evade punishment in large cities if they withhold knowledge that is only ex post verifiable in these endogenous exchanges. Berliant et al. (2006) model more-purposeful search and matching for knowledge exchange that is governed by the city's traits. Their work describes inefficiencies that can occur through endogenous migration choices to cities for knowledge exchange. Further, modeling that captures the nuances of innovation across the city size distribution is to be encouraged.

Holding the size of local areas constant, a large body of work discusses the role of city diversity for innovation and the identification of "new work." This concept dates to Jacobs (1969), who articulated how innovation often exploits the novel recombination of past ideas and tools. Cities that possess a more diverse industrial distribution may be a more fertile soil for these activities. Duranton and Puga (2001) formalize the nursery cities argument. Lin (2011) looks at which cities are the most creative, in that they generate "new work," measured by jobs that did not exist a decade earlier. Lin finds that 5 percent to 8 percent of U.S. workers are engaged in new work and that the percentage is higher in cities with a higher-than-average density of college graduates and a more diverse set of industries.

Beyond diversity of the industrial base, recent studies have also considered local firm size distributions and entrepreneurship levels. Agrawal et al. (2014) explore the optimal firm size distribution for innovation in cities. Their work emphasizes the need for both a large incumbent firm and many startups in

²² Albouy (2009) describes the uneven impact of federal taxation over cities based upon their prices and sizes, given that taxes are applied to nominal wages and not real wages. Albouy describes the impact this structure has on agglomeration returns. Agglomeration for innovation occurs most often in large and expensive cities, so Albouy's tax implications apply equally if not more for these areas.

the local area. The large incumbent firm is important for the production of potential spinoff firms, while the local cluster of startups provides the capacity for the spinoff to take root. This pattern relates to the themes about internal and external innovation that are discussed in the next section. In a related study, Glaeser et al. (2014) depicts how U.S. cities built up around historical mines developed industrial structures that were dominated by large incumbent firms that were not conducive to these types of dynamics. There is tremendous scope for more research in this regard.²³

Going forward, there is an unresolved tension about market sizes of cities and innovation. At one extreme, a necessary foundation for agglomeration is the efficient trade of goods over distances. Detroit can innovate in cars for a worldwide market, and Hollywood can do the same for movies, because of the trade that follows. This logic would suggest that city size does not influence innovation because of market size forces (i.e., larger markets create greater incentives to innovate), at least in the United States today. Yet, there are clear deviations from this frictionless starting point. Especially if property rights across places are not secure or are differentially enforced (e.g., developing economies, early U.S. history, sectors with rapid imitation), local market sizes may be what is relevant for inventors. Thomas Edison's aggressive litigation to support his own property rights was one of the forces that pushed the film industry to the West Coast. More broadly, greater attention needs to be paid to which market sizes are important for firms in their innovation choices and how this is changing over time.

4.6 "Natural advantages" for innovation

Traditional discussions of natural advantages focus on geographic features such as harbors and coal mines. For innovation, "natural advantages" can also form in terms of institutions, universities, and similar relatively fixed factors that are required for innovation. These factors are unevenly distributed over places and will continue to be so for some time horizon into the future, and clusters of innovation can form around them. We group these determinants separately from the more generic education and skills noted in the work previously mentioned.

A very visible factor in this regard is the location of research universities and academic R&D (as well as military facilities). An agglomeration of innovation can form around these points, as firms seek to benefit from spillovers out of the university (e.g., Jaffe, 1989; Anselin et al., 1997). Universities are key players not only in creating new knowledge through the basic research produced by their faculties but also in

²³ Broader studies of local entrepreneurship rates and clusters include Glaeser et al. (1992, 2010), Glaeser and Kerr (2009), and Delgado et al. (2010). It is important to recognize that the agglomeration of entrepreneurship for innovation and growth-oriented firms does not apply more broadly, where studies instead document a local bias of entrepreneurship (e.g., Michelacci and Silva, 2007).

supplying a pool of knowledge workers on which innovative activity depends. Silicon Valley and Boston became important centers for innovation in part because of their proximity to Stanford University and MIT (e.g., Saxenian, 1994; Lee and Nicholas, 2012).

There is general empirical evidence that R&D at local universities is important for firms' innovative activity. Audretsch and Feldman (1996) and Anselin et al. (1997) find localized knowledge spillovers from university R&D to commercial innovation by private firms, even after controlling for the location of industrial R&D. Andersson et al. (2009) find evidence that the expansion of the number of university-based researchers in a local labor market is positively associated with an increase in the number of patents granted in that area. Agrawal and Cockburn (2003) argue that local academic R&D is likely to be more productive, in terms of its contribution to additional patents, in the presence of a large research-intensive firm located nearby—the anchor tenant hypothesis. Taking this effect into account, they report a significant positive correlation between local patents and academic publications in the fields of medical imaging, neural networks, and signal processing. However, Carlino et al. (2007) find that R&D at local universities has only modest effects on local innovative activity. They found that a 10 percent increase in R&D intensity at local universities is associated with less than a 1 percent increase in patent intensity.

Recent efforts pursue causal identification of the role of universities through shocks in their activity levels. Kantor and Whalley (2014a) use interactions of lagged university endowments and stock market fluctuations to generate exogenous changes in local university activity. They find that universities have significant spillovers to their surrounding local areas, especially when they are engaged in research efforts and that the long-term composition of local activity may come to reflect the university's traits and spillovers. In a similar manner, Hausman (2012) quantifies how U.S. universities stimulate nearby economic activity using the interaction of a national shock to the spread of innovation from universities—the Bayh-Dole Act of 1980—with the initial academic strengths of universities and variations across universities in federal research funding. Hausman finds a connection to the long-run growth of the areas surrounding the universities, with particular complementarity evident for large establishments.

Related to these knowledge flows is the specific knowledge contained within scientists and inventors who are also unevenly distributed. Zucker et al. (1998) provide one of the earliest expressions of this phenomenon in the extreme concentration of early biotechnology activity around the locations of star scientists in the nascent field. On a more general level, studies such as Breschi and Lissoni (2009) and Singh and Agrawal (2011) point to substantial concentration of knowledge exchanges that are embedded in the workers moving across firms, most frequently at local distances. Breschi and Lissoni (2009) find that inventor mobility accounts for most of what would be perceived as knowledge flows in their sample.

The concentration of these workers has many antecedents similar to the standard labor pooling Marshallian economies (e.g., specialized skills, general investments in knowledge).

Agglomeration of innovation in the United States and some other countries also has another specific factor influencing its spatial location. The United States is very dependent upon immigrants for its innovation workforce, and these workers have general ties to certain U.S. cities (e.g., the Chinese population in San Francisco). As a consequence, adjustments in national immigration admissions differentially influence the rates of innovation in local areas of the country (e.g., Kerr and Lincoln, 2010; Hunt and Gauthier-Loiselle, 2010). Kerr (2010b) describes how the contributions of Chinese and Indian scientists and engineers to U.S. technology formation increased dramatically in the 1990s. At the same time, these ethnic inventors became more spatially concentrated across U.S. cities. The combination of these two factors reversed long-term declines in the overall agglomeration of U.S. inventors evident in the 1970s and 1980s. Similar effects are discussed in Europe, with a particular link to fostering local diversity (e.g., Nathan, 2014; Ozgen et al., 2011; Parrotta et al., 2014).

Looking across this work, fixed factors and “natural advantages” play a clear role in the formation of clusters. However, the extent of spatial concentration in innovation is too much to be explained only by these exogenous spatial differences in natural advantages (Buzard and Carlino, 2013; Murata et al., 2014). More research is particularly needed on the long-term dynamics of these local spillovers. The study of 19th-century agricultural experiment stations by Kantor and Whalley (2014b) is an important step in this regard.

5. Additional issues on innovation and agglomeration

In this section, we discuss some additional, interesting topics related to the agglomeration of innovation and entrepreneurship. These topics often reach into adjacent fields.

5.1 Sustaining mechanisms

There are several additional factors that are frequently observed in the operation of successful clusters for innovation. These factors are often linked to the persistence and growth of clusters, rather than the start of clusters, and thus we consider them separately from the previously mentioned origination theories. In most cases, these factors have the intuitive feel of the cluster feeding back upon itself.

The most natural expression of this sustaining role is localized spinoffs from innovative incumbent firms. Klepper (2010) documents this process in detail for the two very well-known examples of Silicon Valley and Detroit. Fairchild Semiconductor holds a central place in Silicon Valley lore for its pioneering work

on transistors and integrated circuit manufacturing. Fairchild was itself a spinoff from Shockley Semiconductor Laboratory, and various reports estimate that some 400 or more ventures subsequently spun out of Fairchild, including Intel and Advanced Micro Devices. Klepper (2010) describes how a similar process occurred in the development of Detroit. The Duranton (2007) model of cluster migration to where new innovations occur can be related to this process.

Several parallel strands to these case studies exist in recent work. Gompers et al. (2005) quantify that the spawning creation of VC-backed entrepreneurs is most aided by the presence of local publicly listed companies that were themselves VC-backed earlier. Chatterji (2009) finds that spawned firms perform better than new entrants in the medical devices industry, but this is more due to nontechnical transfers than technical knowledge.

Entrepreneurial finance is also thought to play a role in the agglomeration of innovation. We noted financial intermediaries earlier as a shared input, but many features have this feedback mechanism. Studies find spatial proximity to be important between young companies and their financiers, including banks (Petersen and Rajan, 1994), angel investors (Kerr et al., 2014b), and VC firms (Chen et al., 2010). Proximity can help with the screening of these ventures, monitoring of the portfolio, the transfer of value-added advice, and similar activities. Proximity also allows the formation of reputations and repeated interactions that are useful in situations with asymmetric information and moral hazard. As the markets can also be thin—e.g., the number of very qualified VC investors is limited—the formation of a critical mass of these investors can create a sustaining force for a cluster. Angel investing—e.g., the Google millionaires funding the next generation of Silicon Valley startups—is a particularly strong expression of this feedback loop.

Finally, differences may exist in the cultures of cities for entrepreneurship and innovation. Landier (2005) provides a formal model of an entrepreneurial culture, which is quite different across countries and cities, and Manso (2011) depicts the need for tolerance for failure in innovation and compensation practices. It is often noted that the legal or reputational costs of failure in many places are so high that entrepreneurs get only one shot, while in Silicon Valley an early failure merely makes an entrepreneur “experienced.” Qualitative discussions at the city level date back to at least Chinitz (1961), and Saxenian (1994) describes strong cultural differences between Boston and Silicon Valley in her arguments for why semiconductors grew at a more rapid pace in the latter. Florida (2005) also depicts large differences across cities in their organization and cultural embrace of innovation. Recent contributions include Hofstede (2001) and Falck et al. (2009).

5.2 Policy choices

Local and national policies and regulations play an important role in the agglomeration of innovation. At the local level, Singh and Marx (2013) show the continued importance of political boundaries even after using continuous distance techniques to model spillovers.

The most prominent example in the United States is state-level differences in noncompete and nondisclosure clauses, which limit the ability of workers to start a new venture that could compete with their past employer. Many states allow such employment contracts, with California being an important exception. Employers may use these clauses to protect proprietary knowledge from leaking to rival firms. While Fallick et al. (2006) find mixed evidence about the importance of these clauses, Marx et al. (2009) find sharper evidence of a decreased mobility of inventors specifically when the clauses are feasible. Bozkaya and Kerr (2014) more broadly show how rigid employment law can hinder the development of innovative sectors that rely on rapid labor turnover.

State-level bankruptcy laws also appear to matter (e.g., Berkowitz and White, 2004), although this impact appears to be determined through the competing forces of greater willingness to take on risk for innovation versus greater creditor caution in lending with stronger debtor rights. In a recent contribution, Cerqueiro et al. (2014) find that stronger debtor protection decreases the number of patents produced by small firms and makes them less explorative, especially in financially dependent sectors, consistent with a greater wariness on the part of creditors.

Beyond specific policies such as these two examples, general local regulations and investments play important roles. Much of our discussion connects agglomeration of innovation with local human capital levels, and education investments (both K–12 and tertiary) play an important role in this population base. Likewise, external infrastructure can play an important role for innovation and entrepreneurship. Zoning laws play a clear role at the neighborhood level and perhaps higher.

Many policymakers suggest that they want their cities to become “the next Silicon Valley.” Policymakers often view the success of areas such as Silicon Valley as a recipe for local economic development and growth. This chapter has repeatedly emphasized the substantial agglomeration of activity related to innovation. This agglomeration, of course, means that not every city, or even the majority of them, can become a central hub of innovation, such as Austin, TX. Moreover, Wilson (2009) finds that efforts by U.S. states to expand R&D tax credits to lure innovative efforts succeeded, but mainly by attracting R&D efforts away from neighboring states (“beggar thy neighbor” effects). For these reasons and others, innovation policy may require coordination across regions or national priorities.

More generally, the ecosystems that underlie clusters of innovation are difficult to create, and there is a very questionable record of targeted government interventions to do so (Lerner, 2009; Chatterji et al., 2014). As Duranton (2011) emphasizes, the coordination problems and market failures are difficult for policymakers to solve, and the benefits derived may be too small to justify the costs. Perhaps the best advice is to focus on getting the general conditions correct—e.g., functioning local infrastructure, efficient labor regulations, good schools, etc. These efforts to lower the costs of entry into entrepreneurship and innovative activities may be most powerful and make good sense regardless. Other chapters in this volume more systematically take up the roles of urban policy and placemaking efforts.

5.3 Connections across clusters

Recent research explores the special connections that exist across clusters of innovation. The backdrop for this work is the set of early economic models that focus on a worldwide technology frontier (e.g., Mankiw et al., 1992; Heckscher–Ohlin–Vanek trade models) or technology diffusion processes that are governed only by distance and time (e.g., S-curve diffusions in concentric rings). Empirical work shows, however, that global R&D spillovers and knowledge transfers are stronger when countries share a common language or ethnic bond, even if it is conditional on spatial distances (e.g., Keller, 2004; Jaffe and Trajtenberg, 1999). This line of research is being further pushed down to the subnational level to consider special relationships across global cities and their impacts on the heterogeneous transmission of knowledge.

This research has been most frequently connected to immigrant inventors and innovators in an advanced country and their potential interactions with their home countries, although other networks are quite likely too (e.g., alumni and professional networks). The particular focus on immigrants is due to 1) their importance for technology development in the United States as noted previously and 2) the desire by labor and development economists to understand whether this global migration is a net benefit or cost to the sending country (i.e., the brain drain versus brain gain debate). Despite originating in adjacent fields, this work increasingly speaks to connections across cities for innovation purposes.

A prominent voice in this area is Saxenian, who discusses in many settings the special relationships that exist between Silicon Valley and rising Asian tiger economies in countries such as Taiwan, Korea, and Bangalore, India. This is effectively city-to-city flows given the extreme spatial concentration of innovation in emerging economies. Saxenian et al. (2002) survey immigrant scientists and engineers in Silicon Valley to understand the linkages they maintain with their home countries. These surveys, while very unrepresentative for all immigrants even within Silicon Valley, offer some sense of the relative forms of these contributions. Eighty-two percent of the Chinese and Indian immigrant workers in the

science, technology, engineering, and mathematics disciplines report exchanging technical information with their respective nations, roughly 50 percent of the immigrants report aiding the development of contracts or business relationships between the United States and their home countries, and 18 percent report investing in overseas business partnerships. While the magnitudes of the figures should be treated with heavy caution, as the sample comes from immigrant groups in Silicon Valley who are seeking to maintain global ties (e.g., The Indus Entrepreneurs), the relative likelihoods of the various channels reported is interesting.

Moving beyond case studies and surveys, empirical work using patent citation data suggests that immigrant scientists and engineers in the United States do aid technology transfer to their home countries, albeit at a magnitude that is significantly smaller than what the showcase examples suggest. Kerr (2008) emphasizes that these ethnic transmission channels are particularly powerful in the first five to seven years after a new technology is developed. Agrawal et al. (2011) find that the Indian diaspora in the United States aids its home country the most with the development of big, highly cited inventions but that the diaspora is not as helpful in the production of average inventions as a larger domestic base of inventors.²⁴

While these studies have been important for establishing the connections of innovation clusters across countries, they have noticeable gaps. Perhaps the most prominent gap is a precise articulation of what form of migration matters for the connections. There are many different varieties: permanent migration to the new economy, return migration back to the home countries, regular travel across places, and so forth. Saxenian (2006, 2008) particularly emphasizes the latter with her colorful moniker Global Argonauts, but we have very little data that allow the systematic quantification of these features and their relative importance. Data development in this regard is most warranted, and Hovhannisyan and Keller (2010) provide a recent example with respect to business travel.

Second, these studies frequently have less to offer as to how the traits of places shape these connections. Nanda and Khanna (2012) provide a rare example of this form of study. They survey entrepreneurs in India who were members of the National Association of Software and Services Companies trade association. They find that entrepreneurs who had lived outside of India previously are more likely to activate overseas connections when living outside of the prominent software hubs. In this sense, their

²⁴ Related work explores these connections in the specific forms of enhanced trade flows, foreign direct investments by multinational companies, and outsourcing contract placement. These ethnic-based placements are particularly strong in industries and fields related to innovation. Kerr (2013) provides a review of this work and the particular role of diaspora connections across clusters.

work pointed to the social networks as substituting for domestic institutions and capabilities when they are very weak in India.

It is hoped that more research refines our understanding of how global agglomerations for innovation relate to each other and how the traits of cities in advanced countries and developing nations modulate these connections. Given the increasing globalization of innovation, these factors will increasingly shape tomorrow's economic geography of innovation.

5.4 Innovation in entrepreneurial clusters vs. large firms

This chapter has mostly focused on the development, organization, and interactions of clusters of small entrepreneurial firms. Yet, this intense local environment of startups is not the only way that innovation can be organized, and many examples of large firms and R&D laboratories abound. In this section, we review several pieces that consider this organizational choice. The studies that we cite are just the tip of the iceberg in regard to large company innovation.

What factors might favor a large, isolated R&D laboratory over a connected cluster of small firms? The discussion in Section 4 identifies several initial factors. The larger laboratory can be preferred when the gains from modularization and experimentation are small, when firm investment in worker human capital is central, when the firm is sufficiently large to bring typically shared resources in-house (e.g., general counsel lawyers), and so on. Moreover, to the extent that a branch of innovation requires large fixed costs, the bigger entity is more likely. It is important to note that the fixed costs relate to the innovation (e.g., basic R&D) and not to complementary assets (e.g., manufacturing and distribution capabilities). As the biotech-pharma world illustrates, a cluster of smaller startups can connect on the innovation side to large companies that hold complementary assets if an effective market for ideas exists (e.g., Gans et al., 2002). More broadly, researchers since Schumpeter note the longer investment horizons of large firms, their incentives to improve the markups on their products, the surplus resources they can devote to the innovative efforts, and so on. Very clearly, the likelihood of engaging in R&D or patenting increases with firm size (e.g., Cohen and Klepper, 1996a; Akcigit and Kerr, 2010).

On the other hand, extensive work in the management literature also emphasizes limits for innovations at large companies. This has often been expressed in the exploration versus exploitation distinction highlighted earlier. Large companies are frequently accused of having overreaching bureaucracies that limit innovation. Even if the hierarchies of the company are appropriately sized, larger companies may naturally favor forms of innovation that enhance their current products, spending less time and resources on exploratory work. It is also frequently discussed how separate, independent firms can help resolve

incentive and compensation constraints; for example, while an entrepreneur's equity can become worth billions of dollars with a breakthrough discovery, large firms face political constraints on whether they could pay an in-house innovator more than the firm's CEO. Because they have resources and do not need to appeal to external markets, large companies can also struggle with the termination of poorly performing projects. This termination is essential for experimentation to be effective and the generation of sufficient returns to innovative efforts (e.g., Kerr et al., 2014a).

Several theory models consider explicitly the trade-offs between these alternative landscapes, beyond those noted earlier (e.g., modularity, worker training). Gromb and Scharfstein (2002) model the trade-off between the high-powered incentives of independent entrepreneurs and the better retention of knowledge about employees in large firms. If external markets cannot fully discern the reason why past projects failed, large firms can derive advantages from their ongoing interactions and monitoring of workers. The authors show that multiple equilibria can arise and even coexist. In one setting, entrepreneurship is weak, and so the market for failed entrepreneurs is quite thin, making internal labor markets and innovation within large firms particularly valuable. In a high entrepreneurship equilibrium, the robust labor market and high incentives encourage innovation outside of large firms.

Hellmann and Perotti (2011) consider a second trade-off. In their model, innovative ideas are initially incomplete and must be circulated to provide complementary pieces. Although circulating ideas is essential, innovators face the risks that their ideas will be stolen by those seeing them. Large markets of firms provide innovators with better prospects for finding the best matches compared with efforts to circulate ideas among a single, large company. On the other hand, the internal sharing of the idea within a large company limits the risk of expropriation compared with exposing ideas through markets. Similar to Gromb and Scharfstein (2002), this model can lead to the coexistence of multiple organizational forms related to innovation (internal ventures, spinouts, and startups), with the possibility in the Hellmann and Perotti (2011) model of these forms mutually reinforcing each other. Markets for idea generation work best when the cost of idea generation is low. Hellmann and Perotti (2011) further discuss the literature in this area.

Empirically, Agrawal et al. (2010) quantify how the insular nature of "company towns" affects the types of innovations undertaken locally. Large, isolated plants are more likely to take on lines of innovation that exploit the firm's past work and reflect less on the newer technologies being created. This work connects with the studies of firm size and innovation choices (e.g., Rosen, 1991; Akcigit and Kerr, 2010) and the Duranton and Puga (2001) model of nursery cities. Further, development of these cross-correlations of city types, firm sizes, and innovation choices will greatly enhance our understanding of these issues.

5.5 Multiunit firm locations and agglomeration clusters

A significant body of work in the management literature touches on the location choices of multiunit firms. In an important sense, this work connects to the industrial organization issues identified earlier. This research often uses a foreign direct investment lens, since the overseas operations of multinational firms are feasible to identify. Alcacer and Chung (2007, 2014) provide examples of work that specifically relate these placement decisions to local agglomeration benefits, and Keller and Yeaple (2013) describe knowledge flows in particular. These papers also outline the longer earlier literature. Choudhury (2014) and Choudhury and Khanna (2014) flip the lens around and demonstrate in India how the movement of resources and labor in multiunit and multinational firms affects the innovation of the local areas in which these firms operate.

As greater data on plant-level operations have become available, researchers have also considered the optimal placement of domestic facilities with respect to agglomeration. Alcacer and Delgado (2013) contrast intrafirm linkages that multiunit firms have with the localized external agglomeration economies available in their areas. They find that intrafirm linkages matter a lot and that empirical estimates of agglomeration economies might be overstated as a consequence of not sufficiently modeling them. Tecu (2012) considers how multiunit firms place their industrial innovation. She too finds that internal linkages among the plants of a firm matter a lot. Her estimates suggest that the elasticity of R&D productivity with respect to the firm's production workers is almost as large as the elasticity with respect to the host city's total patents. These internal–external studies remain relatively rare, yet the existing work suggests great promise for future work.

6. Conclusions

Marshall (1890) noted long ago that, in industrial clusters, “[t]he mysteries of the trade become no mysteries; but are as it were in the air, and children may learn them unconsciously” (p. 271). Much progress has been made since Marshall made this famous statement about understanding exactly how agglomeration, knowledge spillovers, and innovation fit together, and what their impact on local and national economies might be. Over the past two decades, research in this field has particularly surged. This has been unquestionably due in large part to the substantial data now available to better study these questions. It also surely has some connection to the growth of the knowledge economy, the fascination with Silicon Valley and related clusters, and similar popular themes. Particularly encouraging is the extensive set of connections being made from economic geography to adjacent fields such as labor economics, entrepreneurial finance, and business management.

Much, however, remains to be accomplished. With the handful of important exceptions reviewed previously, we still have not opened the black box of how clusters operate. Most of our empirical work has instead been comparisons over places. Better empirical guidance about the microinteractions within clusters with respect to innovation will allow us to differentiate among models and build stronger theoretical frameworks. Fortunately, the data are within reach to do so. Many employer–employee data sets are being linked to information on innovation (e.g., inventor identifiers), and these data sets often allow precise spatial coordinates of firms. This should prove a powerful lever for seeing further inside local areas, and it is likely that the line of work will increasingly draw upon network theory (e.g., Jackson, 2008; Dempwolf, 2012; and Kerr and Kominers, 2014).

We also need better insight into the long-term nature of agglomeration and innovation—the life cycles of innovative places. This is true within countries—innovation cores have shifted between Detroit, Boston, Silicon Valley, etc., and will continue to do so—but also true across countries. How does the rise of Bangalore impact Boston? Does the development of global innovation centers in rapidly emerging countries complement or substitute for those in advanced economies? The study of agglomeration and innovation is exciting because we are starting to make some progress at understanding the local and global nature of these phenomena.

Finally (although by no means closing off a comprehensive list), we need to develop better guidance about what local policymakers can do to foster agglomeration and innovation in their cities. This is a big and difficult question, and we are right to be cautious that we do not have all the answers. However, governments have spent billions of dollars on this, and many will continue to fund “be the next Silicon Valley” initiatives. This is true in advanced economies, in nations currently looking to transition from resource-dependent to a knowledge-based economy, in developing countries looking to leapfrog growth stages, and everywhere in between. Economists must continue to provide insight on these critical matters, and ideally our insights will get sharper faster.

REFERENCES

- Acemoglu, A., Akcigit, U., Bloom, N., Kerr, W. (2014). "Innovation, reallocation, and growth." NBER Working Paper 18993.
- Acemoglu, A., Angrist, J. (2000). "How large are human-capital externalities? Evidence from compulsory schooling laws." In Bernanke, B., Rogoff, K. (Eds.), *NBER Macroeconomic Annuals*, MIT Press, Cambridge, MA.
- Acs, Z., Anselin, L., Varga, A. (2002). "Patents and innovation counts as measures of regional production of new knowledge." *Research Policy*, 3, 1069-85.
- Acs, Z., Audretsch, D. (1988). "Innovation in large and small firms: an empirical analysis." *American Economic Review*, 78, 678-90.
- Acs, Z., Audretsch, D., Feldman M. (1994). "Resource and output trends in the United States since 1870." *American Economic Review*, 46, 5-23.
- Adams, J., Jaffe, A. (1996). "Bounding the effects of R&D: an investigation using matched establishment-firm data." *RAND Journal of Economics*, 27 (4), 700-21.
- Aghion, P., Akcigit, U., Howitt, P. (2014). "What do we learn from Schumpeterian growth theory?" NBER Working Paper 18824.
- Aghion, P., Howitt, P. (1992). "A model of growth through creative destruction." *Econometrica*, 60, 323-51.
- Agrawal, A., Cockburn, I. (2003). "The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems." *International Journal of Industrial Organization*, 21 (9), 1217-53.
- Agrawal, A., Cockburn, I., Galasso, A., Oettl, A. (2014). "Why are some regions more innovative than others? The role of firm size diversity." *Journal of Urban Economics*, 81 (1), 149-65.
- Agrawal, A., Cockburn, I., Rosell, C. (2010). "Not invented here? Innovation in company towns." *Journal of Urban Economics*, 67 (1), 78-89.
- Agrawal, A., Kapur, D., McHale, J., Oettl, A. (2011). "Brain drain or brain bank? The impact of skilled emigration on poor-country innovation." *Journal of Urban Economics*, 69, 43-55.
- Akcigit, U., Kerr, W. (2010). "Growth through heterogeneous innovations." NBER Working Paper 16443.
- Albouy, D. (2009). "The unequal geographic burden of federal taxation." *Journal of Political Economy*, 117 (4), 635-667.
- Alcacer, J., Chung, W. (2007). "Location strategies and knowledge spillovers." *Management Science*, 53 (5), 760-76.
- Alcacer, J., Chung, W. (2014). "Location strategies for agglomeration economies." *Strategic Management Journal*. Forthcoming.
- Alcacer, J., Delgado, M. (2013). "Spatial organization of firms and location choices through the value chain." HBS Working Paper 13-025.
- Andersson, R., Quigley, J., Wilhelmsson, M. (2009). "Higher education, localization and innovation: evidence from a natural experiment." *Journal of Urban Economics*, 66 (1), 2-15.
- Anselin, L., Varga, A., Zoltan, A. (1997). "Local geographic spillovers between university research and high technology innovations." *Journal of Urban Economics*, 42, 442-48.
- Arellano, M., Bond, S. (1990). "Some tests of specification for panel data: monte carlo evidence and an application to employment equations." *Review of Economic Studies*, 58 (2), 277-97.
- Arzaghi, M., Henderson, J.V. (2008). "Networking off Madison Avenue." *Review of Economic Studies*, 75, 1011-38.
- Audretsch, D., Feldman, M. (1996). "R&D spillovers and the geography of innovation and production." *American Economic Review*, 86, 630-640.
- Audretsch, D., Feldman, M. (2004). "Knowledge spillovers and the geography of innovation," In: Henderson, J.V., Thisse, J-F. (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4. North-Holland, Amsterdam, 2713-39.
- Azoulay, P., Zivin, J., Wang, J. (2010). "Superstar extinction." *Quarterly Journal of Economics*, 125 (2), 549-589.
- Bacolod, M., Blum, B., Strange, W., (2010). "Elements of skill: traits, intelligences, education, and agglomeration." *Journal of Regional Science*, 50 (1), 245-80.

- Bairoch, P. (1988). *Cities and Economic Development*. University of Chicago Press, Chicago.
- Baldwin, C., Clark, K. (1997). "Managing in an age of modularity." *Harvard Business Review* (September-October).
- Baum-Snow, N., Pavan, R. (2013). "Inequality and city size." *Review of Economics and Statistics*, 95 (5), 1535-1548.
- Berkowitz, J., White, M. (2004). "Bankruptcy and small firms' access to credit." *Rand Journal of Economics*, 35, 69-84.
- Berliant, M., Reed, R., Wang, P. (2006). "Knowledge exchange, matching, and agglomeration." *Journal of Urban Economics*, 60, 69-95.
- Billings, S., Johnson, E. (2014). "Agglomeration within an urban area." Mimeo (University of North Carolina Charlotte).
- Black, D., Henderson, J.V. (1999). "A theory of urban growth." *Journal of Political Economy*, 107, 252-84.
- Bozkaya, A., Kerr, W. (2014). "Labor regulations and European venture capital." *Journal of Economics & Management Strategy*. Forthcoming.
- Breschi, S., Lissoni, F. (2009). "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography*, 9 (4), 439-68.
- Briant, A., Combes, P.P., Lafourcade, M. (2010). "Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimates?" *Journal of Urban Economics*, 67, 287-302.
- Buzard, K., Carlino, G. (2013). "The geography of research and development activity in the U.S." In: Giarratani, F., Hewings, G., McCann, P. (Eds.), *Handbook of Economic Geography and Industry Studies*, Edward Elgar, London.
- Capello, R., Lenzi, C. (2014). "Spatial heterogeneity in knowledge, innovation, and economic growth nexus: conceptual reflections and empirical evidence." *Journal of Regional Science*, 54 (2), 186-214.
- Carlino, G., Carr, J., Hunt, R., Smith, T. (2012). "The agglomeration of R&D labs." Federal Reserve Bank of Philadelphia Working Paper 12-22.
- Carlino, G., Chatterjee, S., Hunt, R. (2007). "Urban density and the rate of invention." *Journal of Urban Economics*, 61 (3), 389-419.
- Carlino, G., Hunt, R. (2009). "What explains the quantity and quality of local inventive activity?" In Burtless, G., Pack, J.R. (Eds.), *Brookings-Wharton Papers on Urban Affairs*, Brookings Institution Press, Washington, D.C.
- Carrincazeaux, C., Lunga, Y., Rallet, A. (2001). "Proximity and localisation of corporate R&D activities." *Research Policy*, 30, 777-789.
- Cassey, A., Smith, B. (2014). "Simulating confidence for the Ellison-Glaeser index," *Journal of Urban Economics*, 81, 85-103.
- Cerqueiro, G., Hegde, D., Penas, M., Seamans, R., (2014). "Debtor rights, credit supply, and innovation." *Tilburg Law and Economics Center Discussion Paper No. 2014-011*.
- Chatterji, A. (2009). "Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry." *Strategic Management Journal*, 30 (2), 185-206.
- Chatterji, A., Glaeser, E., Kerr, W. (2014). "Clusters of entrepreneurship and innovation." In: Lerner, J., Stern, S. (Eds.), *Innovation Policy and the Economy*, Vol. 14, University of Chicago Press, Chicago, IL.
- Chen, H., Gompers, P., Kovner, A., Lerner, J. (2010). "Buy local? The geography of venture capital." *Journal of Urban Economics*, 90-110.
- Chinitz, B. (1961). "Contrasts in agglomeration: New York and Pittsburgh." *American Economic Review*, 51 (2), 279-89.
- Choudhury, P. (2014). "Return migration and geography of innovation in MNEs: a natural experiment of on-the-job learning of knowledge production by local workers reporting to return migrants." HBS Working Paper 14-078.
- Choudhury, P., Khanna, T. (2014). "The role of firms in fostering within country migration: evidence from a natural experiment in India." HBS Working Paper 14-080.
- Christensen, C. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston, MA.

- Ciccone, A., Peri, G. (2006). "Identifying human-capital externalities: theory with applications." *Review of Economic Studies*, 73, 381-412.
- Cohen, W., Klepper, S. (1996a). "A reprise of size and R&D." *Economic Journal*, 106 (437), 925-51.
- Cohen, W., Klepper, S. (1996b). "Firm size and the nature of innovation within industries: the case of process and product R&D." *Review of Economics and Statistics*, 232-43.
- Cohen, W., Nelson, R., Walsh, J. (2000). "Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not)." NBER Working Paper 7552.
- Combes, P., Duranton, G. (2006). "Labour pooling, labour poaching and spatial clustering," *Regional Science and Urban Economics*, 36 (1), 1-28.
- Combes, P., Gobillon, L. (2014). "The empirics of agglomeration economies." In: Henderson, J.V., Duranton, G., Strange, W. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 5. North Holland, Amsterdam.
- Conley, T., Flyer, F., Tsiang, G. (2003). "Spillovers from local market human capital and the spatial distribution of productivity in Malaysia." *Advances in Economic Analysis & Policy*, 3 (1), 1-45.
- Davis, D., Dingel, J. (2013). "A spatial knowledge economy." NBER Working Paper 18188.
- De la Roca, J., Puga, D. (2012). "Learning by working in big cities." CEPR Discussion Paper 9243.
- Delgado, M., Porter, M., Stern, S. (2010). "Clusters and entrepreneurship," *Journal of Economic Geography*, 10 (4), 495-518.
- Dempwolf, C. (2012). "A network model of regional innovation clusters and their influence on economic growth." University of Maryland Ph.D. Dissertation.
- Diamond, R. (2012). "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." Mimeo (Graduate School of Business, Stanford University).
- Directory of American Research and Technology. (1999). 23rd Ed. R.R. Bowker, New York.
- Duranton, G. (2006). "Human capital externalities in cities." In: Arnott, R., McMillen D. (Eds.), *A Companion to Urban Economics*, Blackwell Publishing, Ltd, Malden, MA.
- Duranton, G. (2007). "Urban evolutions: the fast, the slow, and the still." *American Economic Review*, 97, 197-221.
- Duranton, G. (2011). "California dreamin': the feeble case for cluster policies." *Review of Economic Analysis*, 3, 3-45.
- Duranton, G., Overman, H. (2005). "Testing for localization using micro-geographic data." *Review of Economic Studies*, 72, 1077-106.
- Duranton, G., Puga, D. (2001). "Nursery cities: urban diversity, process innovation, and the life cycle of products." *American Economic Review*, 91, 1454-77.
- Duranton, G., Puga, D. (2004). "Micro-foundations of urban agglomeration economies." In: Henderson, J.V., Thisse, J.-F. (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4, North-Holland, Amsterdam.
- Duranton, G., Puga, D. (2014). "The growth of cities." In: Aghion, P., Durlauf, S. (Eds.), *Handbook of Economic Growth*, Vol. 2, North-Holland, Amsterdam.
- Eaton, J., Eckstein, Z. (1997). "Cities and growth: theory and evidence from France and Japan." *Regional Science and Urban Economics*, 27, 443-74.
- Eberts, R., McMillen, D. (1999). "Agglomeration economies and urban public infrastructure." In: Cheshire, P., Mills, E. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 3, Elsevier Sciences B.V., New York.
- Ellison, G., Glaeser, E. (1997). "Geographic concentration in U.S. manufacturing industries: a dartboard approach." *Journal of Political Economy*, 105, 889-927.
- Ellison, G., Glaeser, E., Kerr, W. (2010). "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economic Review*, 100 (3), 1195-1213.
- Falck, O., Fritsch, M., Heblich, S. (2009). "Bohemians, human capital, and regional economic growth." *Jena Economic Research Papers* 2009-049.
- Fallick, B., Fleischman, C., Rebitzer, J. (2006). "Job-hopping in Silicon Valley: some evidence concerning the microfoundations of a high-technology cluster." *Review of Economics and Statistics*, 88 (3), 472-81.
- Feldman, M. (1994). *The Geography of Innovation*, Kluwer Academic, Boston.

- Feldman, M., Audretsch, D. (1999). "Innovation in cities: science-based diversity, specialization, and localized competition." *European Economic Review*, 43, 409-29.
- Feldman, M., Kogler, D. (2010). "Stylized facts in the geography of innovation." In: Hall, B., Rosenberg, N. (Eds.), *Handbook of the Economics of Innovation*, Vol. 1. Elsevier, Oxford, 381-410.
- Florida, R. (2005). *Cities and the Creative Class*, New York: Routledge.
- Fornahl, D., Brenner, T. (2009). "Geographic concentration of innovative activity in Germany." *Structural Change and Economic Dynamics*, 20, 163-82.
- Freedman, M. (2008). "Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry." *Journal of Urban Economics*, 64 (3), 590-600.
- Gans, J., Hsu, D., Stern, S. (2002). "When does start-up innovation spur the gale of creative destruction?" *RAND Journal of Economics*, 33 (4), 571-86.
- Gerlach, H., Ronde, T., Stahl, K. (2009). "Labor pooling in R&D intensive industries." *Journal of Urban Economics*, 65 (1), 99-111.
- Gertner, J. (2012). *The Idea Factory*, The Penguin Press, New York.
- Gibson, R. (1999). "The legal infrastructure of high technology industrial districts: Silicon Valley, Route 128, and covenants not to compete." *New York University Law Review*, 74, 575-629.
- Glaeser, E. (1999). "Learning in cities." *Journal of Urban Economics*, 46 (2), 254-77.
- Glaeser, E. (2008). *Cities, Agglomeration and Spatial Equilibrium*, Oxford University Press, Oxford.
- Glaeser, E.L., Maré, D. (2001). "Cities and skills." *Journal of Labor Economics*, 19 (2), 316-42.
- Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A. (1992). "Growth in cities." *Journal of Political Economy*, 100 (6), 1126-52.
- Glaeser, E., Kerr, W. (2009). "Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?" *Journal of Economics and Management Strategy*, 18 (3), 623-63.
- Glaeser, E., Kerr, S., Kerr, W. (2014). "Entrepreneurship and urban growth: an empirical assessment with historical mines." *Review of Economics and Statistics*. Forthcoming.
- Glaeser, E., Kerr, W., Ponzetto, G. (2010). "Clusters of entrepreneurship." *Journal of Urban Economics*, 67 (1), 150-68.
- Glaeser, E., Saiz A. (2004). "The rise of the skilled city." *Brookings-Wharton Papers on Urban Affairs*, 5, 47-94.
- Glaeser, E., Scheinkman, J., Shleifer, S. (1995). "Economic growth in a cross-section of cities." *Journal of Monetary Economics*, 36, 117-43.
- Gompers, P., Lerner, J. (2001). "The venture capital revolution." *Journal of Economic Perspectives*, 15 (2), 169-92.
- Gompers, P., Lerner, J. (2006). *The Venture Capital Cycle*, 2nd Ed., MIT Press, Boston, MA.
- Gompers, P., Lerner, J., Scharfstein, D. (2005). "Entrepreneurial spawning: public corporations and the genesis of new ventures, 1986 to 1999." *Journal of Finance*, 60 (2), 577-614.
- Griliches, Z. (1979). "Issues in assessing the contribution of research and development to productivity growth." *Bell Journal of Economics*, 10, 92-116.
- Griliches, Z. (1990). "Patent statistics as economic indicators: a survey." *Journal of Economic Literature*, 28, 1661-707.
- Gromb, D., Scharfstein, D. (2002). "Entrepreneurship in equilibrium." NBER Working Paper 9001.
- Grossman, G., Helpman, E. (1991). "Quality ladders in the theory of growth." *Review of Economic Studies*, 58 (1), 43-61.
- Hagedoorn, J., Cloudt, M. (2003). "Measuring innovative performance: is there an advantage in using multiple indicators?" *Research Policy*, 32, 1365-79.
- Hall, B., Jaffe, A., Trajtenberg, M. (2001). "The NBER patent citation data file: lessons, insights and methodological tools." NBER Working Paper 8498.
- Ham, J., Weinberg, B. (2014). "Geography and innovation: evidence from Nobel laureate." Mimeo (Economics Department, The Ohio State University).
- Harhoff, D., Narin, F., Scherer, F., Vopel, K. (1999). "Citation frequency and the value of patented inventions." *Review of Economics and Statistics*, 81, 511-15.

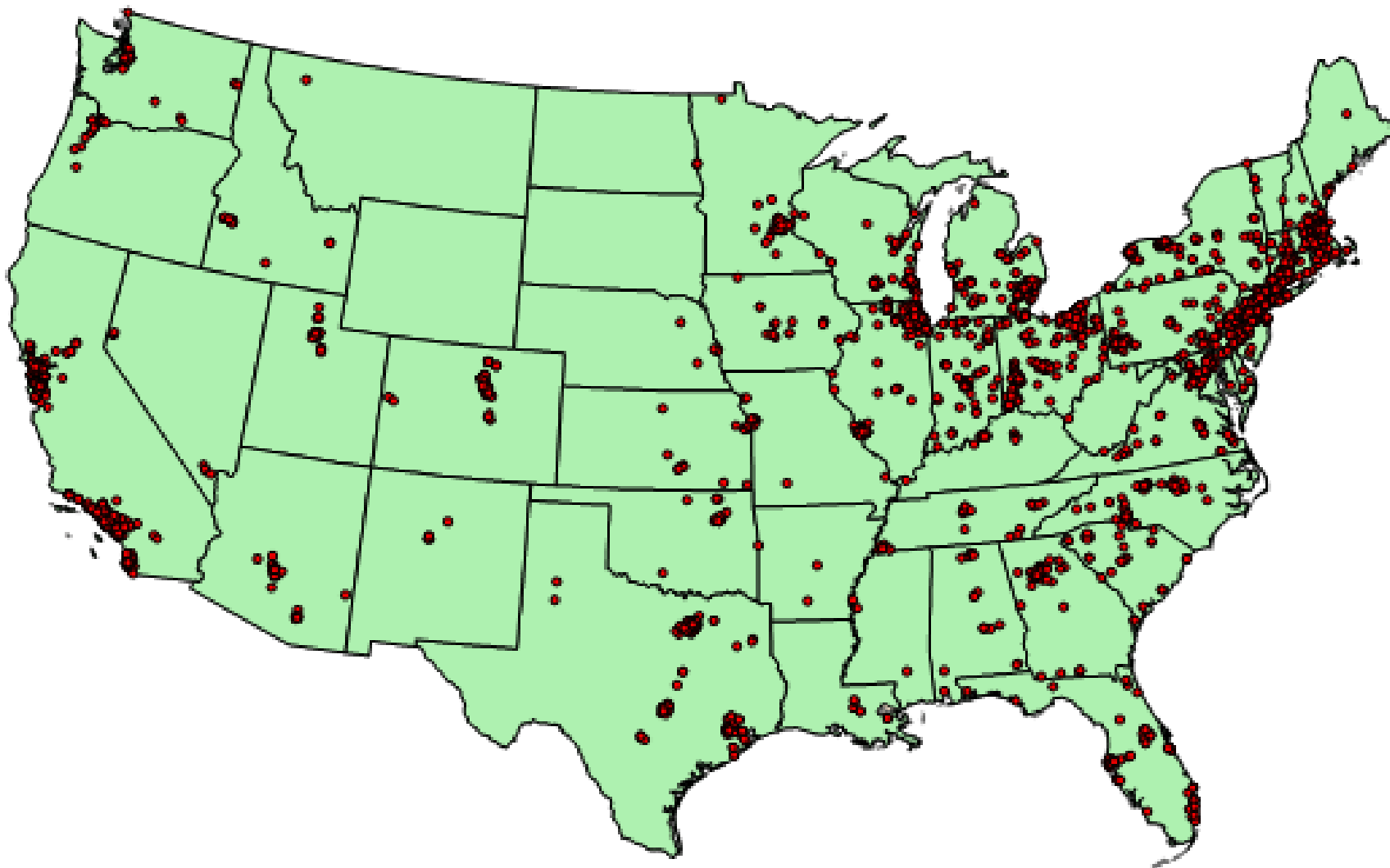
- Hausman, N. (2012). "University innovation, local economic growth, and entrepreneurship." US Census Bureau Center for Economic Studies Paper No. CES-WP- 12-10.
- Hellmann, T., Perotti, E. (2011). "The circulation of ideas in firms and markets." *Management Science*, 57 (10), 1813-26.
- Helsley, R., Strange W. (1990). "Matching and agglomeration economies in a system of cities." *Regional Science and Urban Economics*, 20 (2), 189-212.
- Helsley, R., Strange, W. (2002). "Innovation and input sharing." *Journal of Urban Economics*, 51, 25-45.
- Helsley, R., Strange, W. (2004). "Knowledge barter in cities." *Journal of Urban Economics*, 56 (2), 327-45.
- Henderson, J.V. (1974). "The size and types of cities." *American Economic Review*, 61, 640-56.
- Henderson, J.V. (2007). "Understanding knowledge spillovers." *Regional Science and Urban Economics*, 37 (4), 497-508.
- Hofstede, G. (2001). *Culture and Organizations*, Harper Collins, London.
- Horvath, M. (2001). "Imitating Silicon Valley: regional comparisons of innovation activity based on venture capital flows." In: Bresnahan, T, Gambardella, A. (Eds.), *Building High-Tech Clusters: Silicon Valley and Beyond*, Cambridge University Press, Cambridge.
- Hovhannisyanyan, N., Keller, W. (2010). "International business travel: an engine of innovation?" Center for Economic and Policy Research Discussion Paper No. DP7829.
- Hunt, J., Gauthier-Loiselle, M. (2010). "How much does immigration boost innovation?" *American Economic Journal: Macroeconomics*, 2 (2), 31-56.
- Inoue, H., Nakajima, K., Saito, Y.U. (2013). "Localization of collaborations in knowledge creation," *The Research Institute of Economy, Trade and Industry Discussion Papers Series No. 13-E-070*.
- Jackson, M. (2008). *Social and Economic Networks*, Princeton University Press, Princeton.
- Jacobs, J. (1969). *The Economy of Cities*, Vintage Books, New York.
- Jaffe, A., (1989). "Real effects of academic research." *American Economic Review*, 79, 957-70.
- Jaffe, A., Trajtenberg, M. (1999). "International knowledge flows: evidence from patent citations." *Economics of Innovation and New Technology*, 8, 105-36.
- Jaffe, A., Trajtenberg, M., Fogarty, M. (2000). "Knowledge spillovers and patent citations: evidence from a survey of inventors." *American Economic Review*, 90 (2), 215-8.
- Jaffe, A., Trajtenberg, M., Henderson, R. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." *Quarterly Journal of Economics*, 108, 577-98.
- Jones, C. (2005). "Growth and ideas." In: Aghion, P., Durlauf, S. (Eds.), *Handbook of Economic Growth*, Vol. 1B. North-Holland, Amsterdam.
- Kantor, S., Whalley, A. (2014a). "Knowledge spillovers from research universities: evidence from endowment value shocks." *Review of Economics and Statistics*. Forthcoming.
- Kantor, S., Whalley, A. (2014b). "Research proximity and productivity: long-term evidence from agriculture." *Review of Economics and Statistics*. Forthcoming.
- Keller, W. (2004). "International technology diffusion." *Journal of Economic Literature*, 42 (3), 752-82.
- Keller, W., Yeaple, S. (2013). "The gravity of knowledge." *American Economic Review*, 103 (4), 1414-44.
- Kelly, M., Hageman, A. (1999). "Marshallian externalities in innovation." *Journal of Economic Growth*, 4 (1), 39-54.
- Kerr, W. (2008). "Ethnic scientific communities and international technology diffusion." *Review of Economics and Statistics*, 90 (3), 518-37.
- Kerr, W. (2010a). "Breakthrough inventions and migrating clusters of innovation." *Journal of Urban Economics*, 67 (1), 46-60.
- Kerr, W. (2010b). "The agglomeration of U.S. ethnic inventors." In: Glaeser, E. (Ed.), *Agglomeration Economics*, University of Chicago Press, Chicago, IL, 237-76.
- Kerr, W. (2013). "U.S. high-skilled immigration, innovation, and entrepreneurship: empirical approaches and evidence." NBER Working Paper 19377.

- Kerr, W., Kominers, S. (2014). "Agglomerative forces and cluster shapes." *Review of Economics and Statistics*. Forthcoming.
- Kerr, W., Lerner, J., Schoar, A. (2014a). "The consequences of entrepreneurial finance: evidence from angel financings." *Review of Financial Studies*, 27 (1), 20-55.
- Kerr, W., Lincoln, W. (2010). "The supply side of innovation: H-1B visa reforms and U.S. ethnic invention." *Journal of Labor Economics*, 28 (3), 473-508.
- Kerr, W., Nanda, R., Rhodes-Kropf, M. (2014b). "Entrepreneurship as experimentation." *Journal of Economic Perspectives*. Forthcoming.
- Klepper, S. (2010). "The origin and growth of industry clusters: the making of Silicon Valley and Detroit." *Journal of Urban Economics*, 67, 15-32.
- Kolympiris, C., Kalaitzandonakes, N., Miller, D. (2011). "Spatial collocation and venture capital in the US biotechnology industry." *Research Policy*, 40 (9), 1188-99.
- Krugman, P. (1991). *Geography and Trade*, MIT Press, Cambridge, MA.
- Lamoreaux, N., Levenstein, M., Sokoloff, K. (2004). "Financing invention during the second industrial revolution: Cleveland, Ohio, 1870-1920." NBER Working Paper 10923.
- Landier, A. (2005). "Entrepreneurship and the stigma of failure." Mimeo (Stern School of Business, New York University).
- Lee, J., Nicholas, T. (2012). "The origins and development of Silicon Valley." Harvard Business School Case 813-098.
- Lerner, J. (2009). *Boulevard of Broken Dreams*, Princeton University Press, Princeton, NJ.
- Lerner, J., Seru, A. (2014). "The use and misuse of patent data." Paper Presented at the Allied Social Science Association Meeting, January 3-5, 2014, Philadelphia, PA.
- Lin, J. (2011). "Technological adaptation, cities, and new work." *Review of Economics and Statistics*, 93 (2), 554-74.
- Lin, J. (2014). "The paper trail of knowledge transfers." Federal Reserve Bank of Philadelphia Business Review. Second Quarter.
- Lindley, J., Machin, S. (2014). "Spatial changes in labour market inequality." *Journal of Urban Economics*, 85, 121-38.
- Liu, C. (2010). "A spatial ecology of structure holes: scientists and communication at a biotechnology firm." *Academy of Management, Proceedings*, 1, 1-6.
- Lucas, R.E. (1988). "On the mechanics of economic development." *Journal of Monetary Economics*, 22, 3-42.
- Lychagin, S., Pinkse, J., Slade, M., Van Reenen, J. (2010). "Spillovers in space: does geography matter?" NBER Working Paper No. 16188.
- Magrini, S. (2004). "Regional (di)convergence." In: Henderson, J.V., Thisse J-F. (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4, North-Holland, Amsterdam.
- Mankiw, G., Romer, D., Weil, D. (1992). "A contribution to the empirics of economic growth." *Quarterly Journal of Economics*, 107 (2), 407-37.
- Manso, G. (2011). "Motivating innovation." *Journal of Finance*, 66, 1823-69.
- March, J. (1991). "Exploration and exploitation in organizational learning." *Organizational Science*, 2 (1), 71-87.
- Marshall, A. (1890). *Principles of Economics*. Macmillan, London.
- Marx, M., Strumsky, D., Fleming, L. (2009). "Mobility, skills, and the Michigan non-compete experiment." *Management Science*, 55 (6), 875-89.
- Matouschek, N., Robert-Nicoud, F. (2005). "The role of human capital investments in the location decision of firms." *Regional Science and Urban Economics*, 35 (5), 570-83.
- Menon, C. (2009). "Star and comets: an exploration of the patent universe." London School of Economics SERC Discussion Paper 37.
- Menon, C. (2012). "The bright side of MAUP: Defining new measures of industrial agglomeration." *Papers in Regional Science*, 91 (1).

- Michelacci, C., Silva, O. (2007). "Why so many local entrepreneurs?" *Review of Economics and Statistics*, 89 (4), 615-33.
- Moretti, E. (2004a). "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *Journal of Econometrics*, 121, 175-212.
- Moretti, E. (2004b). "Human capital externalities in cities." In: Henderson, J.V., Thisse, J-F. (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4. North-Holland, Amsterdam.
- Moretti, E. (2004c). "Workers' education, spillovers and productivity: evidence from plant-level production functions." *American Economic Review*, 94, 656-90.
- Moretti, E. (2012). *The New Geography of Jobs*. Houghton Mifflin Harcourt, New York.
- Moretti, E. (2013). "Real wage inequality." *American Economic Journal: Applied Economics*, 5 (1), 65-103.
- Moser, P. (2011). "Did patents weaken the localization of innovations? Evidence from World's Fairs." *Journal of Economic History*, 71 (2), 363-81.
- Murata, Y., Nakajima, R., Okamoto, R., Tamura, R. (2014). "Localized knowledge spillovers and patent citations: a distance-based approach." *Review of Economics and Statistics*. Forthcoming.
- Nanda, R., Khanna, T. (2012). "Diasporas and domestic entrepreneurs: evidence from the Indian software industry." *Journal of Economics and Management Strategy*, 19 (4), 991-1012.
- Nathan, M. (2014). "Ethnic inventors, diversity and innovation in the UK: evidence from patents microdata." *Journal of Economic Geography*. Forthcoming.
- OECD (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, 3rd Ed. Organization of Economically Developed Countries, Paris.
- Olson, G., Olson, J. (2003). "Mitigating the effects of distance on collaborative intellectual work." *Economics of Innovation and New Technology*, 12 (1), 27-42.
- Ozgen, C., Nijkamp, P., Poot, J. (2011). "Immigration and innovation in European regions." IZA Working Paper 5676.
- Pakes, A., Griliches, Z. (1980). "Patents and R&D at the firm level: a first report." *Economics Letters*, 4, 377-81.
- Pakes, A., Nitzan, S. (1983). "Optimum contracts for research personnel, research employment, and the establishment of "rival" enterprises." *Journal of Labor Economics*, 1 (4), 345-65.
- Papageorgiou, T. (2013). "Worker sorting and agglomeration economies." Penn State University Working Paper.
- Parrotta, P., Pozzoli, D., Pytlikova, M. (2014). "The nexus between labor diversity and firm's innovation." *Journal of Population Economics*. Forthcoming.
- Petersen, M., Rajan, R. (1994). "The benefits of lending relationships: evidence from small business data." *Journal of Finance*, 49 (1), 3-37.
- Porter, M. (1990). *The Competitive Advantage of Nations*. The Free Press, New York.
- Porter, M. (1998). "Clusters and the new economics of competition." *Harvard Business Review*, November-December, 77-90.
- Rauch, J. (1993). "Productivity gains from geographic concentration in cities," *Journal of Urban Economics*, 34, 380-400.
- Romer, P. (1986). "Increasing returns and long-run growth." *Journal of Political Economy*, 94 (5), 1002-37.
- Romer, P. (1990). "Endogenous technical change." *Journal of Political Economy*, 98 (5), S71-S102.
- Rosen, R. (1991). "Research and development with asymmetric firm sizes." *RAND Journal of Economics*, 22 (3), 411-29.
- Rosenthal, S., Strange, W. (2001). "The determinants of agglomeration." *Journal of Urban Economics*, 50, 191-229.
- Rosenthal, S., Strange, W. (2003). "Geography, industrial organization, and agglomeration." *Review of Economics and Statistics*, 85 (2), 377-93.
- Rosenthal, S., Strange, W. (2004). "Evidence on the nature and sources of agglomeration economies." In: Henderson, J.V., Thisse, J.F. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4. North-Holland, Amsterdam, 2119-71.
- Rosenthal, S., Strange, W. (2008). "The attenuation of human capital spillovers." *Journal of Urban Economics*, 64 (2), 373-89.

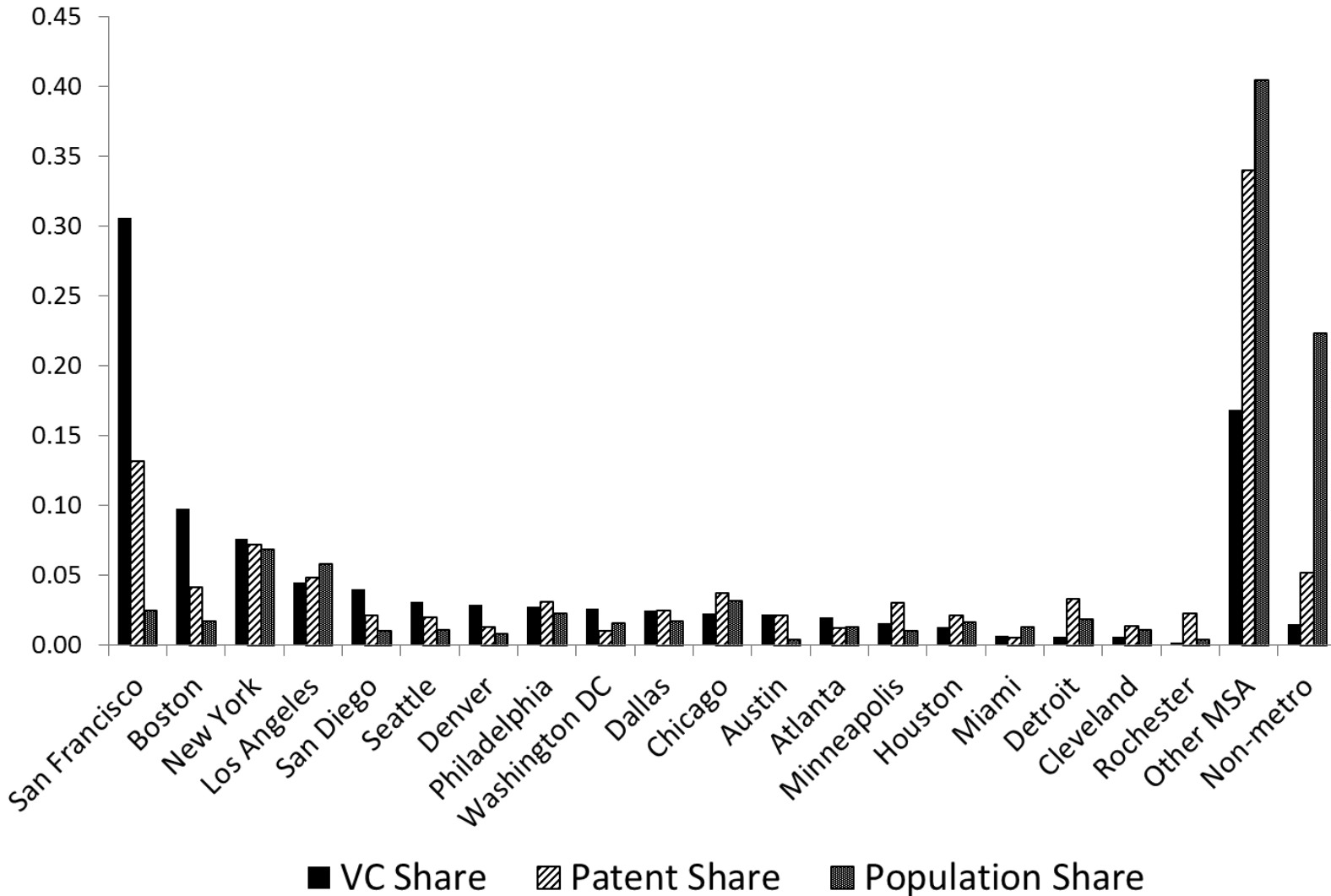
- Rotemberg, J., Saloner, G. (2000). "Competition and human capital accumulation: a theory of interregional specialization and trade." *Regional Science and Urban Economics*, 30, 373-404.
- Samila, S., Sorenson, O. (2011). "Venture capital, entrepreneurship and economic growth." *Review of Economics and Statistics*, 93 (1), 338-49.
- Saxenian, A. (1991). "The origins and dynamics of production networks in Silicon Valley." *Research Policy*, 20 (1), 423-37.
- Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard University Press, Cambridge, MA.
- Saxenian, A. (2006). *The New Argonauts*, Harvard University Press, Cambridge, MA.
- Saxenian, A. (2008). "Transnational communities and the evolution of global production networks: the cases of Taiwan, China and India." *Industry and Innovation*.
- Saxenian, A., Motoyama, Y., Quan, X. (2002). *Local and Global Networks of Immigrant Professionals in Silicon Valley*, Public Policy Institute of California, San Francisco, CA.
- Scherer, F.M. (1986). *Innovation and Growth*. MIT Press, Cambridge, MA.
- Schumpeter, J. (1934). *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.
- Schumpeter, J. (1939). *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, Vol. 1. Yale University Press, New Haven.
- Serrano, C. (2010). "The dynamics of the transfer and renewal of patents." *Rand Journal of Economics*, 41 (1), 686-708.
- Shapiro, J. (2006). "Smart cities: quality of life, productivity, and the growth effects of human capital." *Review of Economics and Statistics*, 88 (2), 324-35.
- Singh, J., Agrawal, A. (2011). "Recruiting for ideas: how firms exploit the prior inventions of new hires." *Management Science*, 57 (1), 129-50.
- Singh, J., Marx, M. (2013). "Geographic constraints on knowledge spillovers: political borders vs. spatial proximity." *Management Science*, 59 (9), 2056-78.
- Solow, R. (1957). "Technical change and the aggregate production function." *Review of Economics and Statistics*, 39, 312-20.
- Strange, W., Hejazi, W., Tang, J. (2006). "The uncertain city: competitive instability, skills, innovation, and the strategy of agglomeration," *Journal of Urban Economics*, 59 (3), 331-51.
- Sturgeon, T. (2002). "Modular production networks: a new American model of industrial organization." *Industrial and Corporate Change*, 11 (3), 451-96.
- Tecu, I. (2012). "The location of industrial innovation: does manufacturing matter?" PhD. Thesis, Brown University.
- Thompson, P., (2006). "Patent citations and the geography of knowledge spillovers: evidence from inventor- and examiner-added citations." *Review of Economics and Statistics*, 88 (2), 383-8.
- Thompson, P., Fox-Kean, M. (2005). "Patent citations and the geography of knowledge spillovers: a reassessment." *American Economic Review*, 95 (1), 450-60.
- Usher, A. (1929). *A History of Mechanical Inventions*. Harvard University Press, Cambridge, MA.
- Waldinger, F. (2012). "Peer effects in science: evidence from the dismissal of scientists in Nazi Germany." *Review of Economic Studies*, 79, 838-61.
- Wang, Z. (2014). "Location choice at labor force entry and new estimates of selection, growth, and level effects from U.S. census data." Mimeo (Fudan University).
- Wheeler, C. (2001). "Search, sorting, and urban agglomeration." *Journal of Labor Economics*, 19, 879-99.
- Wilson, D. (2009). "Beggars thy neighbor? The in-state, out-of-state, and aggregate effects of R&D tax credits." *Review of Economics and Statistics*, 91 (2), 431-36.
- Zucker, L., Darby, M., Brewer, M. (1998). "Intellectual human capital and the birth of U.S. biotechnology enterprises." *American Economic Review*, 88 (1), 290-306.

Figure 1: The Location of R&D Labs in 1998



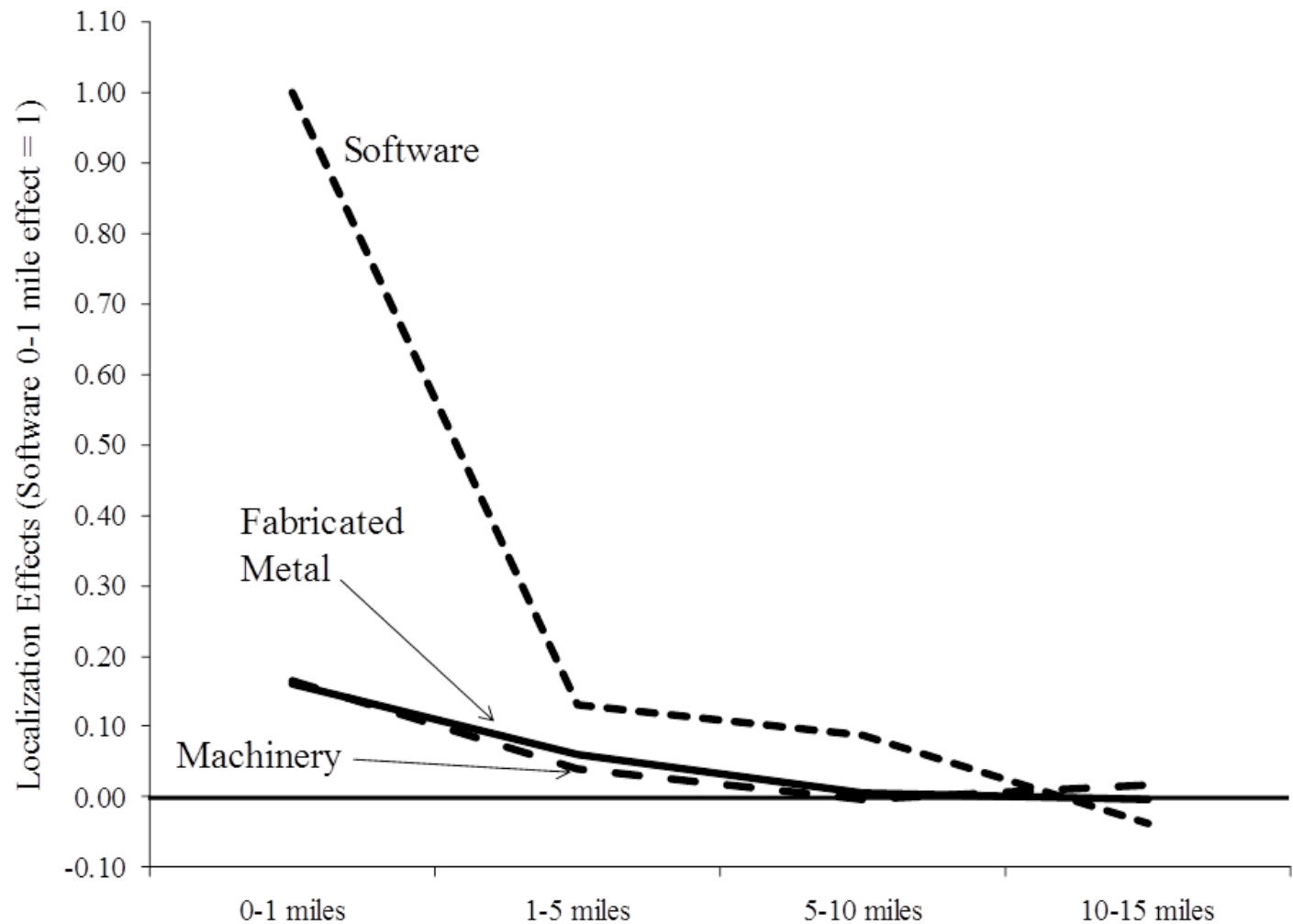
Notes: The figure is taken from Carlino et al. (2012). Each dot on the map represents the location of a single R&D lab. In areas with a dense cluster of labs, the dots tend to sit on top of one another, representing a spatial cluster of labs. Source data are from the Directory of American Research and Technology.

Figure 2: Spatial Concentration of U.S. Patenting and Venture Capital Shares over the 1990—2005 period



Notes: The figure is taken from Chatterji et al. (2104). Source data are from VentureXpert, USPTO patent data, and county-level population statistics. VC calculations use the share of deals over the 1990—2005 period. Patent calculations use the share of granted patents applied for from each city during 1990—2005. The population share is from 1999.

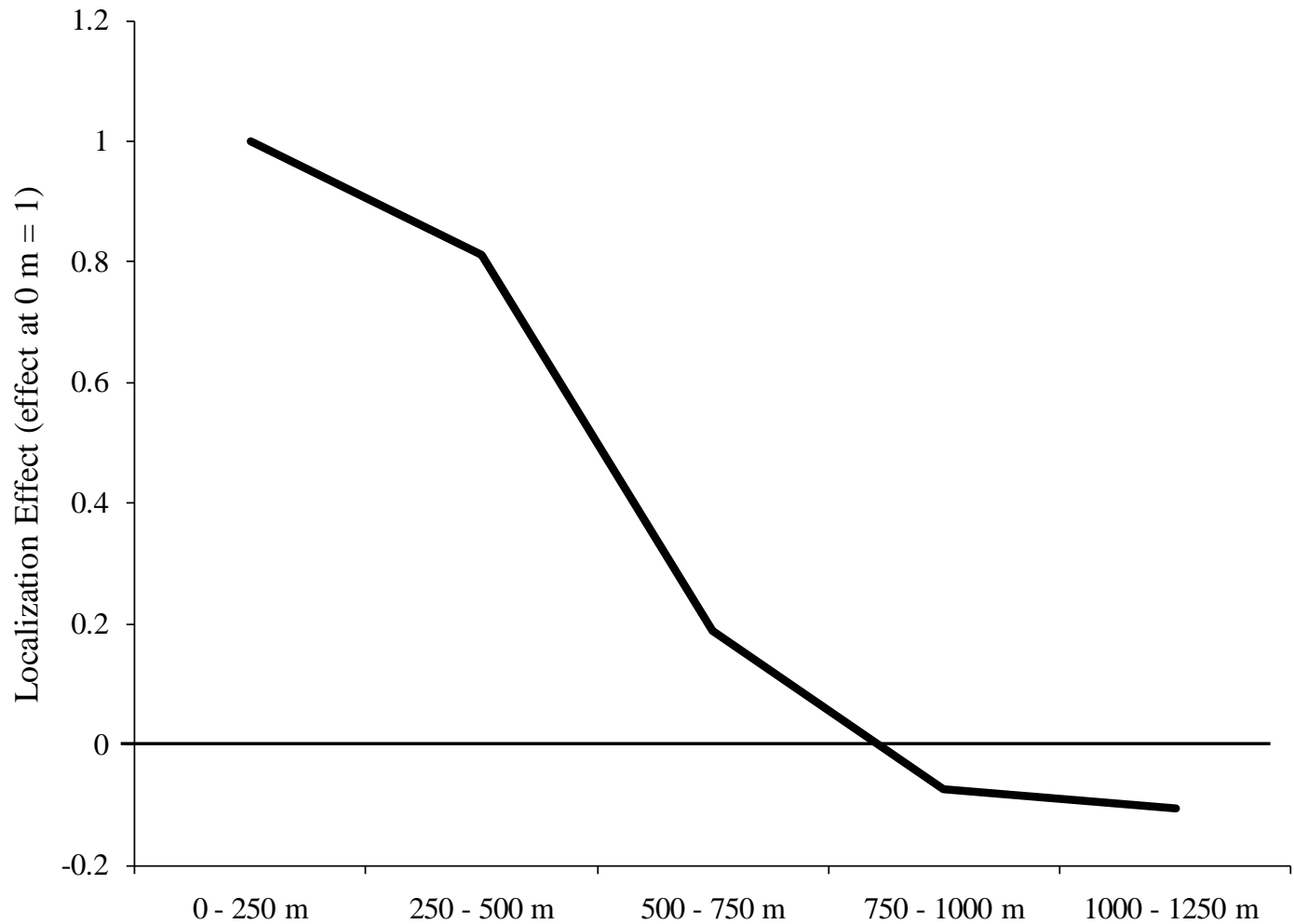
Figure 3: Localization Effects by Distance Between Establishments
Comparative analysis across industries relative to software 0—1 miles



Notes: The figure is derived from localization results presented in Rosenthal and Strange (2003). The horizontal axis measures distance between two establishments in the industry. The vertical axis measures localization effects, with the scale set such that the localization effect for the software industry in the 0—1 mile distance is equal to one. Clustering is more important for the software industry than for the fabricated metal or machinery industries, particularly at the 0—1 mile distance. In all industries, the benefit of proximity to firms in one’s own industry drops off sharply when located 1—5 miles apart rather than 0—1 miles apart. Once firms are 10—15 miles apart, there are no more differential benefits due to site selection within the city, although generalized city-level benefits may still exist.

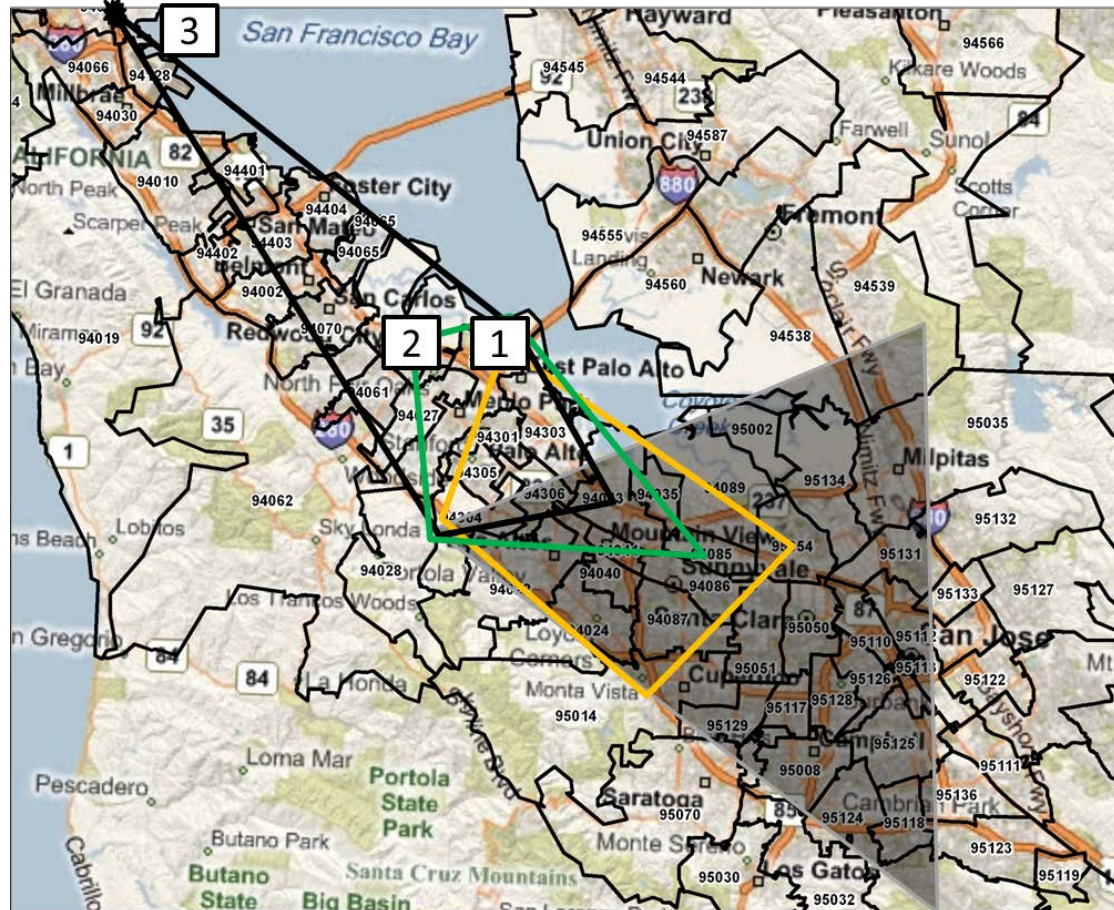
Figure 4: Localization Effects Among Ad Agencies in Manhattan

Localization relative to 0—250 meters



Notes: The figure is derived from localization results presented in Arzaghi and Henderson (2008). The horizontal axis measures the distance between two establishments in the industry. The vertical axis measures localization effects, with the scale set such that the localization effect at 0—250 meters is equal to one. Compared with Figure 3, the localization effects in the ad agencies attenuate extremely fast, reaching zero by 750 meters. Ad agencies and the people who work in them place a very high importance on networking, collaboration and information sharing, and knowledge spillovers, all of which are much more effective over very short distances.

Figure 5: Technology Sourcing from Silicon Valley
Top patenting zip codes outside of the core and their sourcing zones



Notes: The figure is taken from Kerr and Kominers (2014). The exhibit shows the shapes of technology spillover zones in the area surrounding Silicon Valley. The core of Silicon Valley is depicted by the shaded triangle. The Silicon Valley core contains 76 percent of the patenting for the San Francisco region. This map describes the technology sourcing for three of the four largest zip codes for patenting *not* included in the core itself. Technology sourcing zones are determined through patent citations. The boxes indicate the focal zip codes, and the shape of each technology sourcing zone is determined by the three zip codes that firms in the focal zip code cite most in their work. The yellow zone (1) for Menlo Park extends deepest into the core. The green zone (2) for Redwood City shifts up and encompasses Palo Alto but less of the core. The black zone (3) for South San Francisco further shifts out and brushes the core. These technology zones are characterized by small, overlapping regions. None of the technology sourcing zones traverse the whole core, and only the technology zone with the closest zip code (Menlo Park) reaches far enough into the core to include the area of the core where the greatest number of patents occur. Transportation routes and geographic features influence the shapes and lengths of these sourcing zones.