Invention in the City: Increasing Returns to Scale in Metropolitan Patenting

Luis M. A. Bettencourt
CCS-3 Computer and Computational Sciences
Los Alamos National Laboratory
MS B256
Los Alamos, NM 87545
lmbett@lanl.gov

José Lobo
Faculty of Economics
University of Modena and Reggio Emilia
Via Giglioli Valle, 9
Reggio Emilia, Italy
jose@santafe.edu

Deborah Strumsky *
Harvard Business School
E1-2A Gallatin Hall
Soldiers Field Road
Boston, MA 02163
dstrumsky@hbs.edu

December 2004

* : Corresponding author.

Acknowledgements: We thank Lee Fleming for making the patent data available to us, the Harvard Business School Division of Research for assistance in assembling the patent data set, and the Santa Fe Institute and the Information Society as a Complex System (ISCOM) Project for research support. We also thank Richard Florida and Kevin Stolarick for use of their data on the metropolitan distribution of supercreative professions. The work benefited from comments and suggestions by Geoffrey West, David Lane, Denise Pumain, Andrea Ginzburg and Erika Blumenfeld.

Invention in the City: Increasing Returns to Scale in Metropolitan Patenting

Abstract

We investigate the relationship between patenting activity and the size of metropolitan areas in the United States over the last two decades (1980-2001). We find a clear superlinear effect, whereby patents are granted disproportionately in larger urban centers, thus showing increasing returns in inventing activity with respect to population size. We characterize this relation quantitatively as a power law with an exponent larger than unity. This phenomenon is commensurate with the presence of larger numbers of inventors in larger metropolitan areas, which we find follows a quantitatively similar superlinear relationship to population, while the productivity of individual inventors stays essentially constant across metropolitan areas. Finally we show that R&D establishments and employment in other creative professions also follow superlinear relations to metropolitan population size, albeit possibly with different exponents.

1. Introduction

The crucial role that cities have played in the development of science and technology, and more broadly, in the generation of inventions and innovations—intellectual and material, cultural and political, institutional and organizational—is well documented by historiographical work (see for example Bairoch, 1988; Braudel, 1992; Hall, 1998; Jacobs, 1984; Landes, 1999; Mokyr, 2002; Mumford, 1968; Rosenberg and Birdzell, 1986; Spufford 2003). The role of cities as centers for the integration of human capital and as incubators of invention was rediscovered by the "new" economic growth theory, which posits that knowledge spillovers among individuals and firms are the necessary underpinnings of growth (Lucas, 1988; Romer, 1986,). As Glaeser (1996) points out, the idea that growth hinges on the flow and exchange of ideas naturally leads to a recognition of the social and economic role of urban centers in furthering intellectual cross-fertilization. Moreover the creation and reposition of knowledge in cities increases their attractive pull for educated, highly skilled, entrepreneurial and creative individuals who, by locating in urban centers, contribute in turn to the generation of further knowledge spillovers (Feldman and Florida, 1994; Florida, 2002, 2004; Glaeser, 1999). This spontaneous process, whereby knowledge produces growth and growth attracts knowledge, is the engine by which urban centers sustain their development through unfolding innovation.

It is a compelling question to ascertain which features of urban societies foment, or hinder, invention and innovation. Historical evidence notwithstanding, it is not easy to measure knowledge spillovers (a problem noted by Krugman (1991)). This difficulty hampers progress towards the quantitative understanding of the relationship between city characteristics and innovation. Some knowledge flows do nevertheless leave an evidentiary trail, for example in the form of patented inventions (Acs and Audretsch, 1989; Jaffe and Trajtenberg, 2002; Jaffe, Trajtenberg and Fogarty, 2000; Malerba and Orsenigo 1999,).

A clear finding from examining where patents originate in the United States is that patenting is largely a metropolitan phenomenon. Ullman (1958) found that inventive activity in the Unites States is linked to urban development and agglomeration. Pred (1961) examined U.S.

patent data for the mid-nineteenth century and found that patenting activity was significantly greater in the principal cities than the national average. Higgs (1971) found that the number of patents issued in the United States during the period from 1870 to 1920 was positively correlated to urbanization. More recently Jaffe, Trajtenberg and Henderson (1993) examined the pattern of citations by new to previously issued patents. They found that new patents are 5 to 10 times more likely to cite previous ones originating in the same metropolitan area. Acs, Anselin and Varga (2002) also find that patenting in the United States is overwhelmingly concentrated in metropolitan counties. Further examination of data from England (Beggs and Cameron 1988), Sweden (Sjöholm 1996) and the semiconductor industry (Almeida and Kogut 1997) confirms the urban character of patenting for other countries and in specific industries.

Based on this evidence it seems plausible, *a priori*, to assume a close and positive relationship between city (population) size and inventive activity, measured by way of number of new patents granted to metropolitan inventors[1]. Indeed, a higher concentration of individuals and firms in large cities can be expected to sustain a larger repertoire of intellectual capabilities thereby facilitating the creation and recombination of ideas. This environment in turn attracts creative individuals and firms to locate in cities thus sustaining a "virtuous" cycle of invention and innovation.[2] Surprisingly, the relationship between urban scale and patenting has remained largely unexamined. Of the few studies in this direction, Feldman and Audretsch (1999) find that the level of urban employment is positively correlated with invention in U.S. metropolitan areas, while Carlino, Chatterjee and Hunt (2001) find that the rate of metropolitan patenting is positively correlated to employment density. In the present discussion we directly investigate the relationship between patenting activity and population size of metropolitan area in the United States.

We recognize that cities differ not only in size but also in the characteristics of their populations. Boston, for example, has a large population of academics, researchers and technical workers, while proportionally New York City has less, and Los Angeles less still. Some large cities like Detroit have presently very low indices of high-tech development, although arguably Detroit is a very creative city in popular music and some of the arts. Although such considerations are important for a detailed understanding of each metropolitan area, in the present discussion we ask if there are *average* characteristics of cities that make them centers of inventive activity. The expectation is that if large cities are more innovative, then there should be an *average* trend for measures of invention to increase with city size. The quantification of this relationship should also clarify if larger population agglomerations give rise to increases of invention that are simply proportional to population, or if instead there are increasing returns to scale.

Before proceeding we make an explicit clarification. As emphasized by Romer (1990), what matters for growth is not simply an economy with a large number of people but rather how their capabilities are integrated by the environment they create and live in. This argument may be taken to point to specific characteristics of a city, beyond its population size. In line with this

[1] We will use the term "inventor," in a rather restrictive way, to refer to those individuals who have been granted a patent for their invention.

[2] This expectation is a variant of the familiar argument that increases in urban scale generate greater positive externalities (Jacobs, 1969; Marshall, 1890).

argument we reassert here that the size of a large city is not important by itself. Size matters instead because it is the most obvious manifestation of that city's success at attracting and maintaining human capital and engaging it in a myriad of competitive and interdependent activities. It is in this sense that we expect city size to correlate positively with measures of invention and innovation, of which patenting is all but one.

In order to address questions about the nature and magnitude of the scaling relationship between metropolitan invention and population we use data for patents granted in the United States between 1980 and 2001, spatially aggregated into Metropolitan Statistical Areas (MSAs)[3]. Our use of patent counts as a measure of metropolitan inventive activity is by itself not novel. What constitutes a contribution is our use of patent data to quantify metropolitan populations of inventors and to make visible the location-specific *networks* of collaboration among inventors. Economic sociologists argue that economic interactions cannot be fully understood without attention to the social relationships in which these interactions are embedded (*c.f.*, Granovetter, 1985; Polanyi, 1957; Swedberg, 2003; Uzzi, 1996; White, 2002; Zuckerman, 2003). One can similarly argue that the process of invention cannot be well understood without paying attention to the social interactions among inventors (Arora and Gambardella, 1994; Orsenigo, Pammolli and Riccaboni, 2001; Powell, Doput and Smith-Doerr, 1996; Walker, Kogut and Shan, 1997). Inventors do not operate in isolation; the creation of new ideas is a process that often involves the integration and recombination of existing knowledge originating from different individuals, locations, institutions and organizations. Social networks play an important role in the diffusion of information and knowledge since they provide the formal connections and informal linkages through which ideas flow among individuals.[4] These knowledge spillovers often occur without the mediation of market mechanisms, transcend the institutional and workplace settings in which individuals operate, and cut across organizational boundaries.[5] Through the mapping of patent co-authorship relations we investigate below whether the structural features of metropolitan networks of inventors can help explain the quantitative scaling relationship between metropolitan patenting and population.[6]

The manuscript is organized as follows. The next section describes the U.S. patent data, how it was used to identify metropolitan inventors and networks of inventors, and the details of how it was spatially aggregated and matched with metropolitan population data. Section 3 presents our econometric estimations for the dependence of patents on metropolitan size. Section

[3] We are aware of the criticism that patents are not necessarily good indicators of generic innovative activity since not all new inventions are patented, and many economically important types of innovations, such as organizational forms and computer software, cannot even be patented (Griliches, 1979, 1990; Pakes and Griliches, 1980). While these caveats make us cautious about the use of patent data and prudent in the interpretation of our results, we nevertheless see patents as the "footprints" of some (by no means all) inventive activity.

[4] Social networks have been highlighted as an important facet of regional innovation (see, *e.g.*, Breschi and Lissoni, 2001; Piore and Sabel, 1984; Owen-Smith and Powell, 2004) and are believed to be the vital conduits for transferring knowledge and ideas between firms in a region. Much of Silicon Valley's success, for example, has been attributed to its informal networks of friendship and collaboration among scientists, engineers, and entrepreneurs in the area (Saxenian 1994).

[5] Scientists, specially those working in fields where commercial exploitation is common or expected, do, however, also exchange information on a market basis (see Zucker, Darby and Armstrong, 1998).

[6] For examples of using patent co-authorship data as evidence of knowledge spillovers and to construct social networks, see Balconi, Breschi, and Lissoni (2004), Fleming, Juda and King (2004), Murray (2002) and Newman, (2000).

4 tests whether features of the co-authorship networks help explain the observed scaling between patenting and population. Section 5 considers the relationship between the number of metropolitan R&D establishments, as proxies for employment opportunities for inventors, and metropolitan size. Section 6 investigates how employment in other creative activities scales with city size and how that employment correlates to numbers of inventors. Section 7 concludes with a discussion of our findings, their clear and potential consequences, and maps out directions for further research.

2. Metropolitan Population, Patens and Inventor Networks

Source data was extracted from the U.S. Patent Office (USPTO) records on all granted U.S. patents from 1980 to 2001 (U.S. Patent Office 2003). Every patent includes all inventors' last names (with varying degrees of first and middle names or initials), each inventor's home town, detailed information about the patent's technology in class and subclass references (over 100,000 subclasses exist), and the discrimination of the owner, or assignee, of the patent (generally a firm, and less often a university, if not owned by the inventor). Patent filings do not, however, provide consistent listings of inventor names or unique identifiers for the authors. Since the USPTO indexes source data by patent number and not by inventor, a variety of conditional matching algorithms were used to identify inventors, each inventor's patents and other inventors with whom the focal inventor has co-authored at least one patent.[7] The final database includes 2,058,823 unique individual inventors and their patent co-authors, and a total of 2,862,967 patents.

By identifying individual inventors, matching inventors with patents, assigning a location to each inventor—specifically a Metropolitan Statistical Area (MSA)[8]—and linking inventors who have co-authored a patent, it is possible to construct *patent co-authorship networks* for 331 MSAs in the continental United States.[9] Every inventor's hometown was matched to a zip code, which was then assigned to an MSA using the *ZIPList5* dataset[10]. County level population data was extracted from the Bureau of Economic Analysis' "Regional Economic Accounts Tables" (which are available online at www.bea.doc.gov). Counties were assigned to MSAs according to the MSA definitions used to create the metropolitan inventor networks. (Table 1 presents the summary statistics for the variables used in our analysis.)

[Table 1 about here.]

As is by now well known, the variable measuring metropolitan patents is characterized by a negative binomial distribution (a Poisson distribution with over dispersion), as is the variable

---

[7] The matching procedures, discussed in detail in Fleming, King and Juda (2004), refine the previous approach of Newman (2000).
[8] An MSA includes a core city and surrounding counties, which together form a local labor market area.
[9] Patent co-authorship refers to the situation where a patent is either applied for by more than one individual or lists more than one individual as a designated inventor. We will use the terms "co-authorship," "co- patenting," and "co-inventing" interchangeably.
[10] Ziplist5 is a commercially available dataset containing every active ZIP code currently defined by the U.S. Postal Service for the entire USA. Every zip code is assigned to an MSA if the zip code lies within a metropolitan county. The MSA county definitions used by ZIPList5 are consistent with the Census Bureau's 2000 MSA definitions. The ZipList5 database is a commercially available database produced by CD Light (www.zipinf.cm)

measuring metropolitan inventors. As a consequence these two variables exhibit great variation across metropolitan areas. This variation, combined with the skewness characteristic of the negative binomial distribution, entails that both patents and inventors are predominantly concentrated in a few metropolitan areas.

We note that a metropolitan network of inventors includes isolated "nodes" (inventors who are the sole authors of patents), small clusters of inventors connected to each other through shared co-authorship, and larger-sized components grouping many metropolitan inventors together. Individual clusters and components are often linked through key individuals (with high degree of "betweenness") who have connections to multiple inventive communities. The analyses presented here relied upon all patents with at least one inventor within a metropolitan area. Thus, if inventors from inside and outside an MSA co-authored the same patent, the patent would appear in each inventor's metropolitan area. The networks are constructed anew for each year on the basis of the new patents granted that year.

What part of the invention process is captured by the co-authorship links between metropolitan inventors? What do these links imply in terms of information flow and the possible effects of such flows upon subsequent inventive productivity? Singh (2004) reports significant flow of information between patent co-authors, as measured by citations from future patents that are linked by direct—and even indirect—collaborative ties. The results hold even after econometrically controlling for the greater likelihood of a citation arising simply because it refers to work in similar technologies. Singh goes on to demonstrate that almost all of the geographical citation "spillovers" in the United States (*e.g.*, Jaffe *et. al.*, 1993) result from co-authorship networks. (Breschi and Lissoni (2004) find similar results for European inventors.) Thus agglomeration of connections among inventors can be expected to increase inventiveness if connectivity indeed enhances information flow and knowledge spillovers.

Collaborative patenting ties are potentially very effective vehicles for all types of information, especially information that is effectively transmitted only through direct interaction. We are of the view that information that flows along an observed collaborative tie in the years subsequent to its formation varies greatly, from none to a great deal, but that on average, the information flow between former patent co-authors is certainly positive and occasionally substantial. Furthermore, we believe that the distribution of these ties varies from exceptionally weak to extremely strong, such that they support a variety of different types of information flows. Since patent collaboration ties span such a wide spectrum of characteristics and strengths, we have not, in our statistical work, made any assumption regarding the content of information flows or the capability of a tie to transmit information.

Even if we see the links forged by inventors in the act of co-inventing as possible channels for knowledge spillovers, we also echo the cautionary remark made by Hussler (2004), who, using terminology from Hur and Watanabe (2001), views the spillovers evidenced by patents as "intentional spillovers." Inventors, aware that their patents are public documents that make their knowledge accessible, can be selective in what prior knowledge they chose to cite as relevant to their invention[11]. Here it is important to remember that patents are, in essence, ownership claims, sought by inventors not only for the sake of recording intellectual originality but also in the expectation of financial gain. Thus one would not expect patents to exhaustively record all of the individuals who positively influenced an inventor's thinking.

We will invoke four simple graph-theoretic descriptions of networks—connectivity, density, clustering and the size of the largest component—when exploring the impact of inventor network features on the scaling relationship between metropolitan patenting and population. *Connectivity* is simply a measure of how many connections, or ties, there are in the network; the higher this measure is, the greater the number of patenting collaborations present within a metropolitan area and, for fixed number of nodes, the greater the number of inventors who are linked to each other. The average *density* of connections among inventors in the network is defined as the actual number of ties, L, divided by the potential number of ties between N nodes (*i.e*., density = $L/(N(N-1))/2$). The naïve expectation is that more highly connected and denser networks will have greater flows of information among inventors thereby making the inventors in such networks more inventively productive.

In order to define average network *clustering* we follow Watts and Strogatz (1998) and first calculate individual clustering for each node as the number of actual "triples" for each inventor (*i.e.,* the number of different pairs of an inventor's collaborators that have worked with one another and are therefore linked). Inventors with one or zero ties receive a clustering score of zero. This single node clustering is then averaged over the whole set of inventors within an MSA. We further normalized this number to produce an averaged MSA clustering coefficient by dividing the average node clustering by the theoretical clustering of a graph of commensurate size (*i.e*., with the same number of inventors) and mean degree[12]. The *largest component* of a network is the largest set of inventors that can trace a direct or indirect collaborative connection to one another. The *size of the largest component* is calculated as the fraction of inventors that had a collaborative tie within the largest component in the MSA. The largest component is thus the largest inventive community within a metropolitan network of inventors.

Glancing at the summary statistics for network density and size of the largest component it is notable how non-dense metropolitan inventor networks are, and the small size of the largest component. These two observations, combined with a third one, the relatively high level of clustering, provide a first hint that overall network connectivity is not a significant determinant of patenting output. The high clustering levels insinuate a picture of inventors linked to ther inventors in small co-authorship groups.

3. Scaling of Metropolitan Invention with Metropolitan Size

Our initial question is the nature of the relationship between a metropolitan area's inventive output, proxied by the number of new patents granted to inventors residing in the corresponding MSA, and the size of its population. Specifically we want to know if there is a

[11] As discussed by Breschi and Lissoni (2004), Globerman, Kokko and Sjöholm (2000) and Jaffe, Trajtenberg and Henderson (1993) the large majority of patent citations are added by patent examiners rather than by the inventors authoring patents.

[12] The *degree* of a node in a graph is the number of edges linked to the node.

general *average* trend for the increase of invention with metropolitan population size. The use of the term *average* requires some explanation. We use it here to allow us to gloss over metropolitan specificities, which we neglect altogether. In this sense, by taking many cities with different characteristics, we expect such variables to be effectively averaged over. Thus we assume a very simple functional relation between metropolitan population (*N*) and newly granted metropolitan patents (*P*):

$$P_{i,t} = cN_{i,t}^{b}, \tag{1}$$

where c is a constant independent of P or N, i indexes the metropolitan area and t stands for time (in units of a year). The exponent β determines the (power-law) scaling relationship between patenting and population. Taking the natural logarithm of equation (1) and assuming the presence of i.i.d. Gaussian noise we have our basic estimation equation[13]:

$$\ln\left(P_{i,t}\right) = \ln c + b\,\ln\left(N_{i,t}\right) + e_{i,t}. \tag{2}$$

We estimated the exponent β using data for three single periods (1980, 1990 and 2000) and an Ordinary Least Squares (OLS) estimation procedure with a correction for heteroskedasticity; the statistically significant (at the 95% confidence level) coefficient values are shown in Figure 1 together with the scatter plots for the dependent and independent variables. We also availed ourselves of the richness of a dataset containing cross-sectional and time-series data by estimating parameters using a panel data fixed effects Feasible Generalized Least Squares (FGLS) framework (model 1 in Table 2). The estimated pooled coefficient for the effects of population on patenting output is 1.29 (with a 95% confidence level of $1.26 \le b \le 1.32$).

[Figure 1 about here.]

[Table 2 about here.]

Mathematically the scaling relationship between metropolitan population and metropolitan invention is "superlinear," or to use the language of economics, the relationship exhibits increasing returns to scale (*i.e.*, $\beta > 1$). Not surprisingly, a larger metropolitan population is associated with a greater output of new patents; what is surprising is the magnitude of the increasing returns to scale, corresponding an averaged $1.25 \le b \le 1.31$ (the 95% confidence interval). This means that on the average we can expect a city like Philadelphia, with about 1.5 million people to generate almost 18 times more patents than a city of the size of Eugene, OR or Springfield, MO, which are about 10 times smaller. This finding suggests two alternative explanations: either inventors are individually more productive the larger the city, or there are a disproportionate number of inventors the larger the metropolitan area. In the next section we investigate which of these effects is at the root of our scaling result.

[13] Transforming the patent count data by using the natural logarithmic function has the effect of changing the distribution into a normal one–as verified both by visual inspection of histograms and by performing the Wilks-Shapiro test (for individual periods and across all periods for all MSAs).

4. More inventors or network effects?

What lies at the origin of the increasing returns of scale of invention with metropolitan size? As briefly anticipated above, we advance two alternative hypotheses: the first is that the number of inventors in a city is roughly proportional to population (perhaps supported by a naive assertion that each individual has an equal probability to become an inventor), or that a larger metropolitan population leads to higher inventive productivity, *i.e.*, to more inventions per inventor. Increasing returns to scale in the latter case may be directly related the greater number of connections between inventors, which indeed generally grow superlinearly with the number of inventors: for N inventors connected all to all this grows quadratically with N, like $N(N-1)/2$, setting an upper bound on $b \leq 2$. In this way we may be able to explain a superlinear relation between invention and population even if the number of inventors remains a fixed fraction of the number of inhabitants across many different MSAs.

The second, alternative, hypothesis is that inventors are not more productive in cities but that their numbers increase superlinearly with metropolitan size. This hypothesis is commensurate with a theory of cities as attractors for the "creative class" (Florida 2002, 2004). It is also conceivable that both effects are at play. These two pure scenarios (but not their mixture) are easy to distinguish: either the data reveals that the relationship between the number of inventors and population is linear, and the number of patents per inventor vs. metropolitan size superlinear (hypothesis 1), or *vice versa* (hypothesis 2).

Regressing the number of metropolitan inventors on metropolitan population, we find that the relationship is superlinear, with a coefficient $b = 1.24$ and a 95% confidence interval of $1.22 \leq b \leq 1.28$ (see Figure 2 and model 2 in Table 2). The relationship between the number of new metropolitan patents and the number of metropolitan inventors (Figure 3 and model 3 in Table 2) is, however, remarkably linear with a coefficient β very close to unity ($0.97 \leq b \leq 0.99$). So, indeed more inventors result in more patents, but in a nearly one-to-one relationship, a result which vindicates hypothesis 2. The absence of agglomeration externalities for patenting productivity is already revealed by the summary statistics for the variable *patents per inventor* (simply defined as total metropolitan patents divided by total metropolitan inventors). As shown on table 1, the mean for metropolitan patents per inventor is close to or below one, which is possible because there can be several authors per patent, for the years 1980, 1990 and 2000, and with a small coefficient of variation indicating that this measure of productivity does not vary much across metropolitan areas. Regressing patents per inventor on population size reveals a very small coefficient, $b = 0.028$, thereby providing further evidence against the presence of strong agglomeration externalities on inventive activity (see model 4 in Table 2).

What about possible network effects? We examined the effect on patent output of the level of connectivity, the magnitude of clustering, connection density and size of the largest component computed for the patent co-authorship networks described above. We find that the relationship between the number of metropolitan patents and the level of connectivity of a metropolitan network of inventors is clearly positive but sublinear, with a pooled coefficient of $\beta = 0.82$ (model 5 in Table 2). Thus increasing connectivity, the measure of the extent to which inventors are linked to each other through co-inventing, does not result in proportionately greater patenting output, contradicting the main thrust of hypothesis 1. The effect of increasing the size of the network's largest component, while positive, is well below linear with a coefficient $\beta = 0.26$ (see model 6 in Table 2). The effect of network clustering on patenting is also sub-linear, with a coefficient $\beta = 0.72$ (model 7 in Table 2). The effects on patenting of network density are negligible (model 8 in Table 2). The combination of these results strongly suggest that in spite of positive correlations between some network properties and patenting, the quantitative features of co-authorship networks are not simply related to the scaling of inventiveness with metropolitan size.

5. A corollary: inventive establishments and metropolitan size

The single best explanation for the scaling of inventiveness with metropolitan size is that larger-sized MSAs have a disproportionate share of inventors. Most inventors, however, do not do their inventing in the privacy of their garages; inventors work within organizations and institutions, both private and public, profit and non-profit, which encourage and reward inventive activity. We may therefore expect larger-sized metropolitan areas to also have a disproportionate share of the type of establishments (the term used by the U.S. Census Bureau for a place of work) that employ inventive individuals.

Inventive activity takes place in many and varied establishments so it would be a daunting task to assemble the data required to comprehensively measure all of the metropolitan establishments were inventive activity could be expected to occur—and we have not taken on this challenge. What we have done instead is to use a measure of metropolitan establishments explicitly engaged in research and development work. The North American Industrial Classification System (NAICS), recently adopted by the U.S. Census Bureau, includes sector 5417, "Scientific Research and Development Services" (www.census.gov/epcd/naics.html) which is defined as:

> "This industry group comprises establishments engaged in conducting original investigation undertaken on a systematic basis to gain new knowledge (research) and/or the application of research findings or other scientific knowledge for the creation of new or significantly improved products or processes (experimental development). The industries within this industry group are defined on the basis of the domain of research; that is, on the scientific expertise of the establishment. This industry comprises establishments primarily engaged in conducting research and experimental development in the physical, engineering, and life sciences, such as agriculture, electronics, environmental, biology, botany, biotechnology, computers, chemistry, food, fisheries, forests, geology, health, mathematics, medicine, oceanography, pharmacy, physics, veterinary, and other allied subjects."

It seems reasonable to use data for this sector as providing a measure for metropolitan establishments engaged in patenting activity. Unfortunately, data on the number of metropolitan R&D establishments is currently available only for the years 1998-2001, and for only 287 MSAs. Nevertheless, the modest result made possible by using the available data is still illuminating.

We averaged the yearly data on metropolitan R&D establishments and metropolitan population (see Table 1 for the summary statistics for metropolitan R&D establishments), and then regressed the natural logarithm of R&D establishments on the natural logarithm of population (model 1 in Table 3)[14]. The relationship between the number of R&D establishments and metropolitan population is, as expected, superlinear, with a coefficient of $\beta = 1.19$ (and a 95% confidence interval of $1.11 \leq b \leq 1.27$)[15]. By regressing the natural logarithm of metropolitan inventors on metropolitan R&D establishments (model 2 in Table 3) we find an almost linear relationship with a coefficient ß = 0.91 (with a 95% confidence interval of $0.86 \leq b \leq 0.99$). These results reinforce our findings above that not only do larger cities have a disproportionate share of inventors, they also have a similarly disproportionate number of institutions that support inventive activities (inventors and inventive employment seem to be in balance). To be more quantitative we should not only compile more systematic data on R&D establishments but also take into account their size in terms of employment and numbers of inventors. Whether this infrastructure is the reason for inventors to locate in larger cities or the consequence of their presence is clearly an interesting question that we must defer to future work.

[Table 3 about here.]

6. A speculation: inventors and supercreative professions

Finally we would like to know if our results for the superlinear scaling relationship between metropolitan inventors and population are a manifestation of larger phenomenon involving a broader categorization of "creative" activities. In order to investigate this we examined the statistical relationship between the number of people involved in a select few creative professions in metropolitan areas and the corresponding metropolitan population size, as well as the relationship between creative employment and number of inventors. To do this we adopt Richard Florida's definition of "supercreative" employment, which consists essentially of all scientific, artistic, educational and entertainment professionals (Florida 2002).[16]

We averaged the data on metropolitan supercreative employment for the three years 1999, 2000 and 2001 (see Table 1 for the summary statistics), and then regressed the natural logarithm of supercreative professionals against metropolitan population (model 1, Table 4). The result of this estimation indicates a superlinear relationship between supercreative professionals and metropolitan population size with $1.10 \leq b \leq 1.18$ at 95% confidence level. The relation

[14] The estimation was done using OLS with a correction for heteroskedasticity.

[15] To fully appreciate this result consider that the scaling relationship between all types of establishments and metropolitan population is linear with a pooled FGLS coefficient of 1.03 (estimated using data for the 1980 to 2001 period and 331 MSAs). Details of this result are available upon request.

[16] According to the definition put forward by Florida (2002, pp.327-329) "supercreative" professions are "Computer and Mathematical, Architecture and Engineering, Life, Physical and Social Science Occupations, Education, Training and Library, Arts, Design, Entertainment, Sports and Media Occupations." The occupational classifications were derived from the Standard Occupational Classification System (SOC) introduced by U.S. Bureau of Labor Statistics in 1998. The SOC classification data is constructed using the North American Industrial Classification System (NAICS). We believe it is reasonable to suppose that these professions include most inventors but we have not examined the question of relative numbers in detail.

between metropolitan inventors and supercreative professionals is in turn approximately linear with $1.01 \le b \le 1.16$ at 95% confidence level (see model 2, Table 4)[17].

[Table 4 about here.]

These results, taken over a much more limited data set than our inventors data, suggest that supercreative employment, like inventor numbers, scales superlinearly with metropolitan size, although with a smaller exponent. It also suggests that inventors may be a reasonable proxy for total number of professionals engaged in creative activities (albeit slightly over-represented relative to the more inclusive supercreative professions).

7. Conclusion

We started the present inquiry with the objective of quantifying the scaling relationship between metropolitan innovation and population, across cities with very different population sizes as well as many other distinct characteristics. Our robust statistical results indicate that larger metropolitan areas have *disproportionately* more inventors than smaller ones and generate more patents according to essentially the same relation. In fact larger cities are tangibly more inventive *per inhabitant* than smaller ones, thus producing increasing returns in invention to population scale. This property is quantified by exponents $\beta > 1$, characteristic of superlinear scaling in patents (and number inventors) with population size.

The totality of our results paints a picture of invention where agglomeration effects do not increase on average (or decrease, for that matter) the productivity of the individual inventor. Instead there are a range of informal interaction effects, not captured by co-authorship links but present in a large population, that lead to the concentration of inventive professionals in the largest metropolitan spaces. The fundamental question left open for future work is to explain these relationships quantitatively by integrating them into a predictive theory of endogenous (population and economic) growth. At the more qualitative level we have shown that there are good indications of a connection between the size of a city and its pull on intellectual capital. Larger cities are also disproportionately the seats of R&D institutions which, either as the cause or the effect, help support a range of innovative activities. Finally while we do not find a convincing link between co-authorship network effects and the relationship between patent output and metropolitan size, we have shown tentative indications that numbers of inventors and patents may be symptoms of a much larger, if more tenuous, phenomenon. Creative human capital resides *disproportionately* in larger cities.

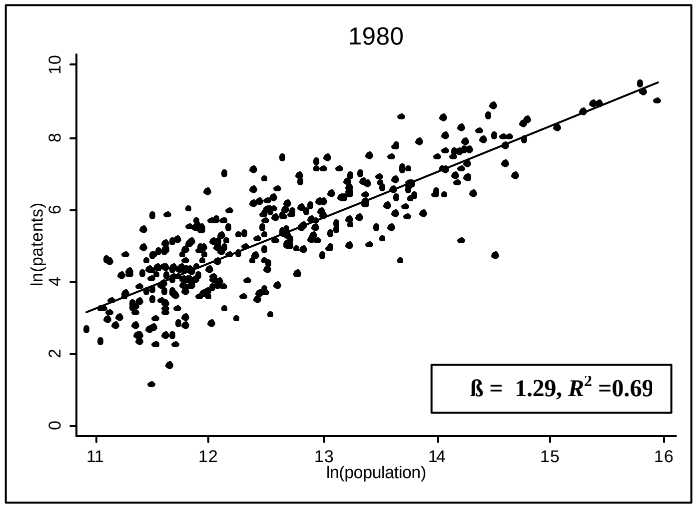
[17] Estimations performed for each individual year show substantial variation in the estimates for β, although its value remains in all cases above unity for the relationship between supercreative numbers and population, and closer and statistically consistent with unity for the relationship between inventors and suprecreatives.

Table 1. Summary statistics for metropolitan variables.

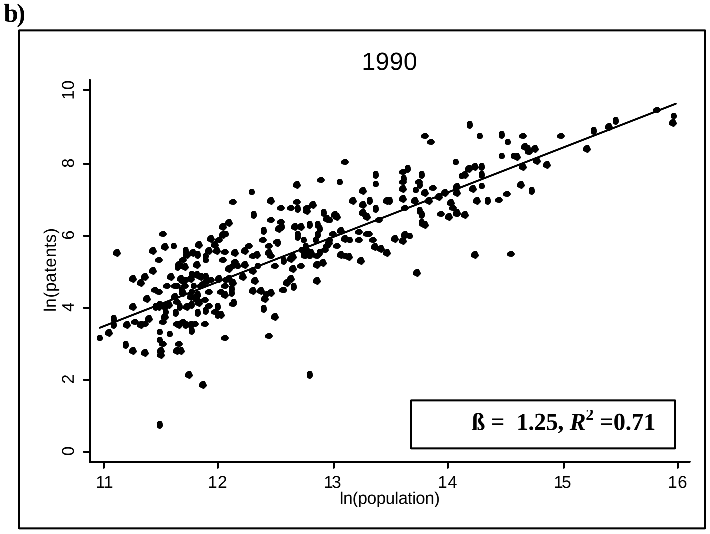
Year	Patents	Inventors	Population	Patents per Inventor	Connectivity	Density	Clustering	LC Size	R&D Establishments	Supercreative Professions
1980										
Mean	689.8	633.3	789104.8	1.03	733.7	0.01	0.43	0.09		
Std. Dev.	1472.1	1274.4	1334787	0.22	1276.2	0.02	0.21	0.11		
CoV	2.13	2.01	1.69	0.21	1.74	2.00	0.49	1.22		
Min	3	2	67707	0.57	1	0	0	0.01		
Max	12990	10220	7247840	2.49	7813	0.07	0.76	0.64		
1990										
Mean	858.3	955.1	776685.1	0.88	1262.8	0.02	0.56	0.11		
Std. Dev.	1690.9	1823.8	1273644.6	0.16	2370.3	0.06	0.23	0.15		
CoV	1.97	1.91	1.64	0.18	1.88	3.00	0.41	1.36		
Min	2	2	63851	0.47	3	0	0	0.08		
Max	12882	12594	7367683	1.55	15717	0.08	0.97	0.68		
2000									**1998 - 2001***	**1999 - 2001***
Mean	1724.2	2069.2	863789.2	0.79	4638.1	0.02	0.81	0.15	40.06	36985.47
Std. Dev.	3753.7	4217.7	1360917.9	0.18	8620.6	0.07	0.14	0.16	97.35	70483.01
CoV	2.18	2.04	1.58	0.23	1.86	3.50	0.17	1.07	2.43	1.91
Min	6	10	66230	0.42	39	0.01	0.43	0.15	1	296.67
Max	38170	37051	8130362	2.93	58014	0.09	1.11	0.80	1071.56	541033.33
No. of MSAs	331	331	331	331	331	331	331	331	287	331

***: yearly variable values averaged over this period.**

a)



b)



c)

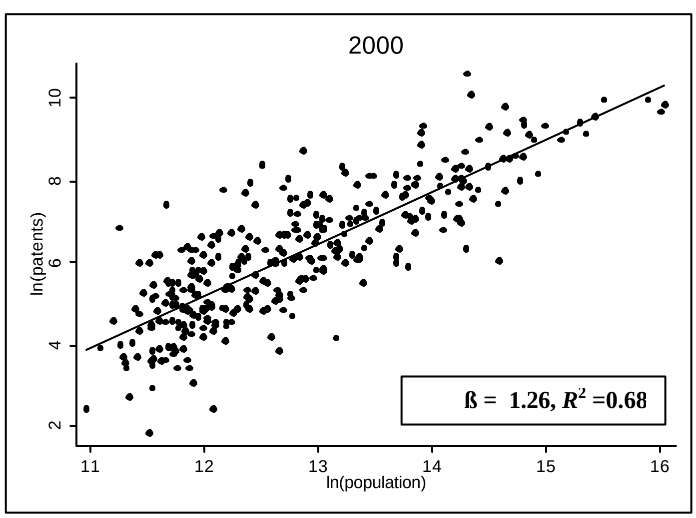
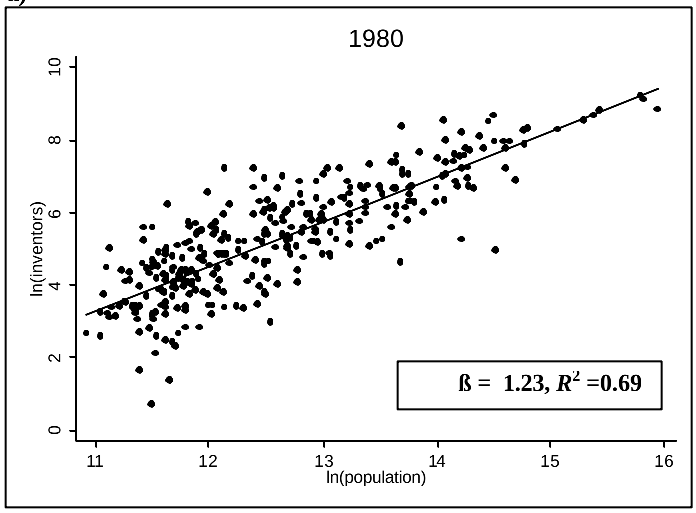
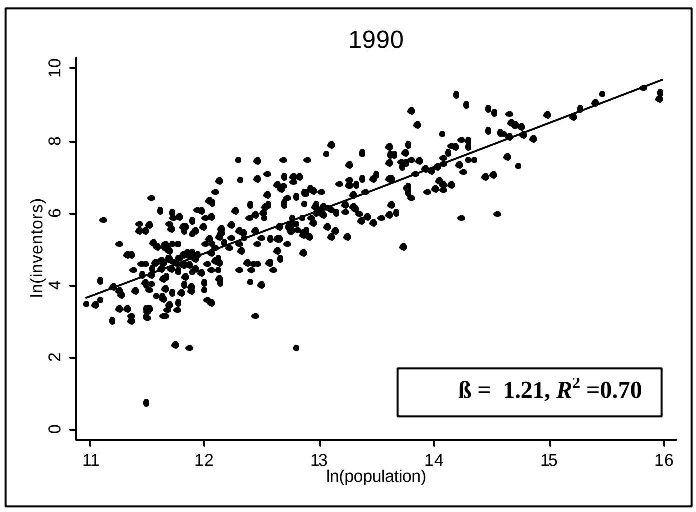


Figure 1. Metropolitan Patents and Population (331 MSAs).

a)



b)



c)

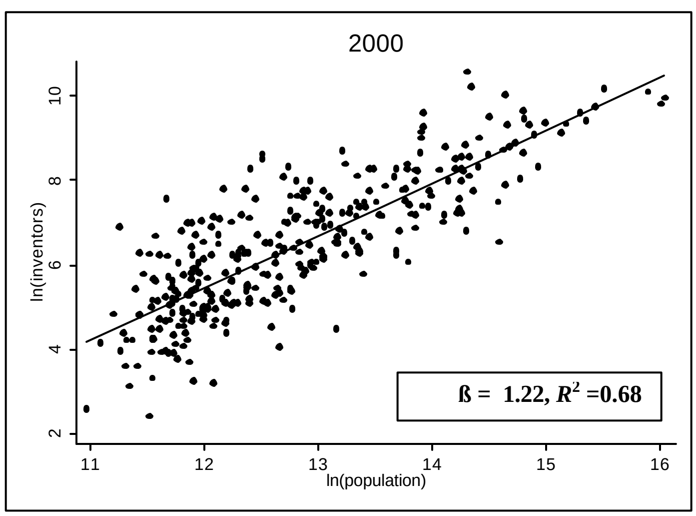


Figure 2. Metropolitan Inventors and Population (331 MSAs).

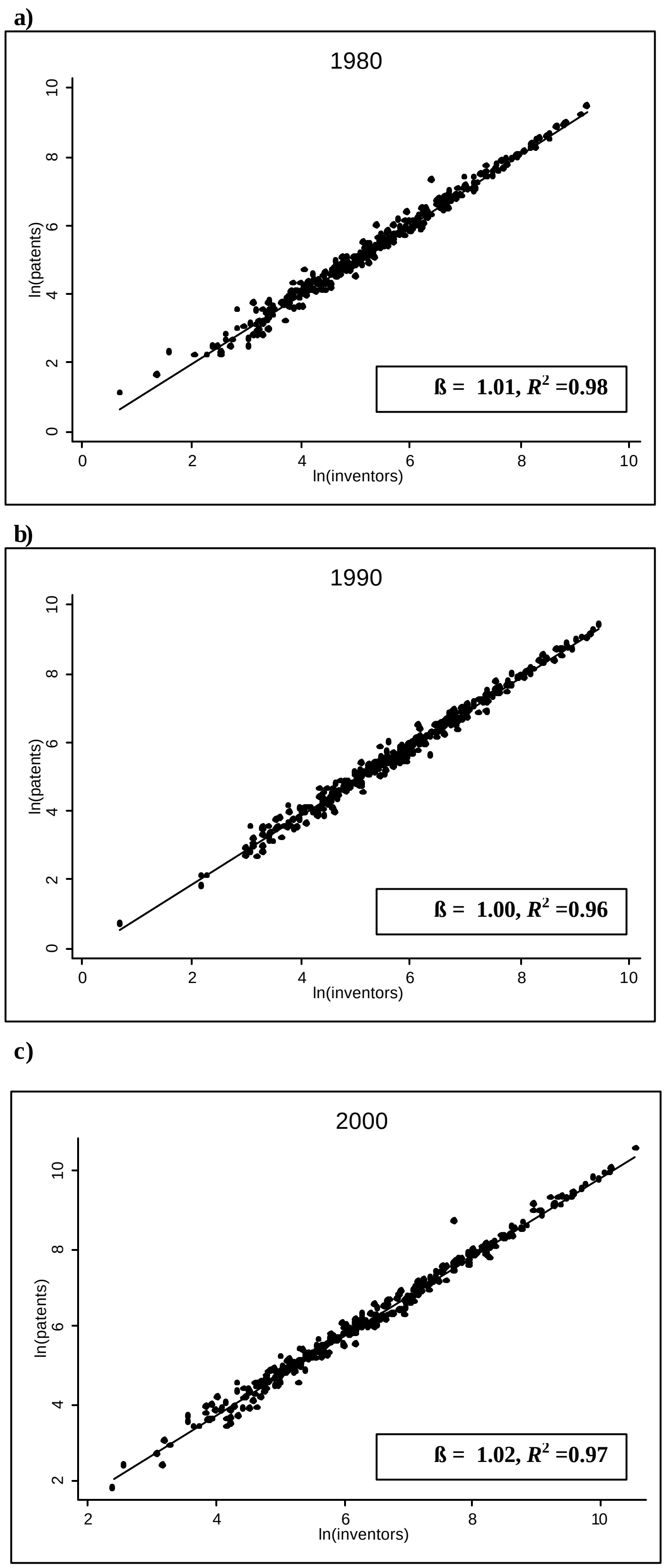


Figure 3. Metropolitan Inventors and Patents (331 MSAs).

Table 2. Estimation results using panel data and fixed effects FGLS by Metropolitan Statistical Area, 1980 - 2001.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	**Patents**	**Inventors**	**Patents**	**Patents per Inventor**	**Patents**	**Patents**	**Patents**
constant	**-10.82**	**-10.41**	**0.04**	**-0.47**	**5.59**	**6.02**	**5.68**
Population	**1.291**	**1.238**		**0.028**			
	(0.015)	**(0.014)**		**(0.003)**			
Inventors			**0.981**				
			(0.002)				
Connectivity					**0.823**		
					(0.001)		
Size of LC						**0.256**	
						(0.009)	
Clustering							**0.716**
							(0.013)
No. cross-sections	**331**	**331**	**331**	**331**	**331**	**331**	**331**
No. observations	**6951**	**6951**	**6951**	**6951**	**6951**	**6951**	**6951**
Log likelihood	**5637**	**5649**	**10011**	**10044**	**9080**	**5358**	**10464**

All of the variables are in natural logarithmic form. Estimations done assuming heteroskedastic error structure across cross-sectional units (MSAs) and *AR*(1) serial autocorrelation within cross-sectional units (with an *AR*(1) coefficient specific to each cross-section).

Standard errors in parentheses. All of the coefficients reported in the table are significant at the 99% confidence level.

Table 3. Estimation results for metropolitan R&D establishments (data averaged for 1998 – 2001).

	(1)	(2)
Dependent Variable:	**R&D Establishments**	**Inventors**
constant	**-12.81**	**4.31**
Population	**1.19**	
	(0.041)	
R&D Establishments		**0.91**
		(0.035)
No. MSAs	**287**	**287**
Adjusted R^2	**0.71**	**0.69**

All of the variables are in natural logarithmic form. Estimations done using OLS and a correction for heteroskedasticity.

Standard errors in parentheses. All of the coefficients reported in the table are significant at the 99% confidence level.

Table 4. Estimation results for metropolitan "supercreative" employment (data averaged for 1999 – 2001).

	(1)	(2)
Dependent Variable:	**Supercreatives**	**Inventors**
constant	**-4.97**	**-4.08**
Population	**1.14** **(0.021)**	
Supercreaives		**1.08** **(0.037)**
No. MSAs	**331**	**331**
Adjusted R^2	**0.89**	**0.75**

All of the variables are in natural logarithmic form. Estimations done using OLS and a correction for heteroskedasticity.

Standard errors in parentheses. All of the coefficients reported in the table are significant at the 99% confidence level.

References

Acs, Z. and Audretsch, D. (1989) "Patents as a Measure of Innovative Activity," *Kylos*, 42, 171-180.

Acs, Z.J., Anselin, L. and Varga, A. (2002) "Patents and Innovation Counts as Measures of Regional Production of New Knowledge," *Research Policy*, 31, 1069-1085.

Almeida, P. and Kogut, B. (1997) "The Exploration of Technological Diversity and the Geographic Localization of Innovation," *Small Business Economics*, 9, 21-31.

Arora, A., and Gambardella, A. (1994) "The changing technology of technical change: general and abstract knowledge and the division of innovative labor." *Research Policy*, 23, 523-532.

Bairoch, P. (1988) *Cities and Economic Development: From the Dawn of History to the Present*. Chicago: University of Chicago Press.

Balconi, M., Breschi, S., and Lissoni, F. (2004) "Networks of inventors and the role of academia: an exploration of Italian patent data," *Research Policy*, 33, 127–145.

Beggs, I. and Cameron, G. (1988) "High Technology Location and the Urban Areas of Great Britain," *Urban Studies*, 25, 361-379.

Braudel, F. (1992) *Civilization and Capitalism 15th-18th Century, Vols. I, II, III.* Berkeley: University of California Press.

Breschi, S. and Lissoni, F. (2001) "Knowledge Spillovers and Local Innovation Systems: A Critical Survey," *Industrial and Corporate Change*, 10, 975-1005.

Breschi, S. and Lissoni, F. (2004) "Knowledge Networks from Patent Data: Methodological Issues and Research Targets." Working Paper 150, Centro di Ricerca sui Processi di Innovazione e Internazionalizzazione, Universita Commerciale "Luigi Bocconi."

Carlino, G., Chatterjee, S. and Hunt, R. (2001) "Knowledge Spillovers and the New Economy of Cities." Federal Reserve Bank of Philadelphia Working Paper No. 01-14.

Feldman, M.P. and Audretsch, D.B. (1999) "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition," *European Economic Review*, 2, 409-429.

Feldman, M. and Florida, R. (1994) "The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States," *Annals of the Association of American Geographers*, 84, 210-229.

Fleming, L., King, C. and A. Juda (2004) "Small Worlds and Regional Innovative Advantage," Harvard Business School Working Paper 04-008.

Florida, R. (2002) *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life.* New York: Perseus Books Group.

Florida, R. (2004) *Cities and the Creative Class*. New York: Routledge

Glaeser, E. (1996) “Economic Growth and Urban Density: A Review Essay,” in *Cities and the New Global Economy Conference Proceedings,* Vol.1, 227-246. Melbourne: Government Printing Office.

Glaeser, E. (1999) “Learning in Cities,” *Journal of Urban Economics*, 46, 254-277.

Globerman, S., Kokko, A. and Sjöholm, F. (2001) “International Technology Diffusion: Evidence from Swedish Patent Data,” *Kylos*, 53, 17-38.

Granovetter, M.S. (1985) “Economic Action and Social Structure: The Problem of Embeddedness,” *American Journal of Sociology,* 91, 481-510.

Griliches, Z. (1979) “Issues in Assessing the Contribution of R&D 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, 1661-1707.

Hall, P. (1998) *Cities in Civilization.* London: Trafalgar Square.

Higgs, R. (1971) “American Inventiveness, 1870-1920,” *Journal of Political Economy*, 79, 661-667.

Hur, K.I. and Watanabe, C. (2001) “Unintentional Technology Spillover Between Sectors: Kinetic Approach,” *Technovation*, 21, 227-235.

Hussler, C. (2004) “Culture and Knowledge Spillovers in Europe: New perspectives for Innovation and Convergence Policies?,” *Economics of Innovation and New Technology*, 13, 523-541.

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

Jacobs, J. (1984) *Cities and the Wealth of Nations.* New York: Vintage Books.

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

Jaffe, A.B., Trajtenberg M., and Fogarty M.S. (2000), “Knowledge spillovers and patents citations: evidence from a survey of inventors,” *American Economic Review* 90, 215-218.

Jaffe, A.B., Trajtenberg, M. and Henderson, R. (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577-598.

Krugman, P. (1991) *Geography and Trade*. Cambridge: The MIT Press.

Landes, D. (1999) *The Wealth and Poverty of Nations: Why Some Nations are Rich and Some Poor*. New York: W.W. Norton & Company.

Lucas, R.E. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3-42.

Malerba, F. and Orsenigo, L. (1999) "Technological entry, exit and survival: an empirical analysis of patent data," *Research Policy,* 28, 643-660.

Marshall, A. (1890) *Principles of Economics*. London: Macmillan.

Mokyr, J. (2002) *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton: Princeton University Press.

Mumford, L. (1968) *The City in History: It's Origins, It's Transformation, and It's Prospects.* New York: Harvest Books.

Murray, F. (2002) "Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering," *Research Policy*, 31, 1389-1403.

Newman, M. (2000) "The Structure of Scientific Collaboration Networks," *Proceedings of the National Academy of Sciences of the United States of America*, 98, 404-409.

Orsenigo, L., Pammolli, F. and Riccaboni, M. (2001) "Technological change and network dynamics: lessons from the pharmaceutical industry," *Research Policy,* 30, 485-508.

Owen-Smith, J. and Powell, W. (2004) "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community," *Organization Science,*15, 5-21.

Pakes, A. and Griliches, Z. (1980) "Patents and R&D at the Firm Level: A First Report," *Economics Letters*, 5, 377-381.

Piore, M. and Sabel, C. (1984) *The Second Industrial Divide.* New York: Basic Books.

Polanyi, K. (1957) "The economy as an instituted process," in Polanyi, K., Arensberg, C. and Pearson, H. (eds.) *Trade and Market in the Early Empires: Economies in History and Theory.* Chicago: Free Press.

Powell, W.W., Doput, K.W. and Smith-Doerr, L. (1996) "Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology," *Administrative Science Quarterly*, 41, 116-145.

Pred, A.R. (1966) *The Spatial Dynamics of U.S. Urban Industrial Growth, 1800-1914.* Cambridge: Harvard University Press.

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

Romer, P. M. (1990) "Endogenous technological change", *Journal of Political Economy*, 98, S71-S102.

Rosenberg, N. and Birdzell, Jr., L.E. (1986) *How the West Grew Rich*. New York: Basic Books.

Saxenian, A. (1994) *Regional Advantage*. Cambridge: Harvard University Press.

Singh, J. (2004) "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." Working Paper, INSEAD.

Sjöholm, F. (1996) "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity," *Weltwirtschaftliches Archiv*, 132, 97-115.

Spufford, P. (2003) *Power and Profit: the Merchant in Medieval Europe*. London: Thames & Hudson.

Swedberg, R. (2003) *Principles of Economic Sociology*. Princeton: Princeton University Press.

Ullman, E. (1958) "Regional Development and the Geography of Concentration," *Papers and Proceedings of the Regional Science Association*, 4, 179-207.

Uzzi, B. (1996) "The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect," *American Sociological Review*, 61, 674-698.

Varga, A. (1999) "Time-Space Patterns of U.S. Innovation: Stability or Change? A Detailed Analysis Based on Patent Data," in Fischer, M., Suarez-Villa, L. and Steiner, M. (eds.), *Innovation, Networks and Localities*. Berlin: Springer.

Walker, G., Kogut, B., and Shan, W. (1997) "Social capital, structural holes and the formation of an industry network," *Organization Science*, 8, 109-125.

Watts, D. and Strogatz, S. (1998) "Collective Dynamics of Small-World Networks," *Nature*, 393, 440-442.

White, H.C. (2002) *Markets from Networks: Socioeconomic Models of Production*. Princeton: Princeton University Press.

Zucker, L.G., Darby, M.R. and Armstrong, J. (1998) "Geographically localized knowledge: spillovers or markets?," *Economic Inquiry*, 36, 65-86.

Zuckerman, E.W. (2003) "On Networks and Markets by Rauch and Casella, eds.," *Journal of Economic Literature*, 61, 545-565.