American Economic Review 2021, 111(10): 3328–3375
https://doi.org/10.1257/aer.20191277

The Effect of High-Tech Clusters on the Productivity of Top Inventors†

By Enrico Moretti*

The high-tech sector is concentrated in a small number of cities. The ten largest clusters in computer science, semiconductors, and biology account for 69 percent, 77 percent, and 59 percent of all US inventors, respectively. Using longitudinal data on 109,846 inventors, I find that geographical agglomeration results in significant productivity gains. When an inventor moves to a city with a large cluster of inventors in the same field, she experiences a sizable increase in the number and quality of patents produced. The presence of significant productivity externalities implies that the agglomeration of inventors generates large gains in the aggregate amount of innovation produced in the United States. (*JEL* D62, J24, L60, O31, 034, R32)

> *We know from ordinary experience that there are group interactions that are central to individual productivity. We know this kind of external effect is common to all the arts and sciences—the 'creative professions'*
>
> (Lucas 1988)

Firms in the innovation sector display a strong tendency to cluster geographically by research field (Carlino et al. 2012). Prominent examples include the internet and software clusters in Seattle anchored by Amazon and Microsoft, respectively; the medical research and biotech clusters in Boston; the software and telecommunication clusters in Austin; the Raleigh-Durham Research Triangle, with its large concentration of pharmaceutical firms; the nascent autonomous vehicles cluster in Pittsburgh; and the biotech cluster and medical devices clusters in San Diego. The San Francisco-Silicon Valley region has the largest agglomeration of innovative firms in the United States, with important clusters in most research fields.

The geographic concentration of high-tech sectors is not just a curiosity—it has important implications for cities and states. The presence of a high-tech sector has been shown to be a key driver of local economic growth as innovation-oriented industries have taken on larger roles (Glaeser and Saiz 2004, Buera and Kaboski 2012). In the period 1980 to 2010, mean wages and mean income in cities with

* Department of Economics, UC Berkeley; NBER and CEPR (email: moretti@berkeley.edu). Thomas Lemieux was the coeditor for this article. I thank Don Davis, Jorge De La Roca, Gilles Duranton, Bronwyn Hall, Evan Rose, John Van Reenen, Jose Vasquez, Heidi Williams, Paolo Zacchia, the referees, and seminar participants at Berkeley, Columbia, Georgetown, LSE, MIT, NABE, NYU, Penn, Torino, San Francisco Fed, UCL, and USC for helpful comments and suggestions.

† Go to https://doi.org/10.1257/aer.20191277 to visit the article page for additional materials and author disclosure statement.

large high-tech clusters have increased significantly more than in cities without high-tech clusters (Moretti 2012). Thus, it is probably not surprising that competition has emerged among localities to attract high-tech employers, with states and counties offering increasingly generous subsidies designed to spur high-tech clusters in their jurisdiction. The prospect of winning Amazon's second headquarters recently generated intense competition among US cities, with some of finalist cities offering incentives as high as \$5 billion. Tesla received incentives worth \$1.3 billion from Nevada in 2014 to locate its new Gigafactory in Reno, while Foxxcom received \$3 billion from Wisconsin to locate its electronic components factory in the state. In biotech, 11 states provide incentives to relocating firms and the average subsidy has grown four-fold since 1990 (Moretti and Wilson 2014). It is rare for high-tech firms to open large new facilities in the United States today without receiving some type of local subsidy, with areas with limited high-tech presence typically offering the most aggressive subsidies.

Quantitatively, the amount of spatial concentration observed in high tech by research field is remarkable. (In this paper, I will use the term "high tech" broadly, to include any firm that produces innovation.) In my patent data, the ten cities with largest number of inventors in "computer science" account for 69.3 percent of all inventors in 2007. In "semiconductors" and "biology and chemistry," the corresponding shares are 77.0 percent and 59.2 percent, respectively. These shares are not declining over time. Indeed, they are larger in 2007 than in 1971, suggesting that despite the diffusion of the internet, Skype, and other forms of cheap communication, high-tech inventors are more geographically concentrated today than they were in the past. A similar amount of concentration is observed in other countries (Duranton et al. 2010, Kerr 2018). Agglomeration in manufacturing, meanwhile, is much lower on average (Ellison and Glaeser 1997).

The agglomeration of innovative activity by research field raises important questions about the economic geography of the high-tech sector, especially since high-tech clusters tend to be located in cities with high labor and real estate costs—cities like San Francisco, Boston, or Seattle—rather than in cities where costs are low. Why do high-tech inventors tend to locate near other inventors in the same field, despite the higher costs? What are the effects on innovation? One view of Silicon Valley-style clusters—which can be traced to Marshall (1890)—is that agglomeration economies make workers inside large clusters more productive. To use Marshall's words, in industrial clusters "the mysteries of the trade become no mysteries." In his seminal paper on economic growth quoted above, Lucas (1988) explicitly posits a *productivity* effect of agglomeration and argues that it is particularly important in the innovation sector and "creative professions." This productivity effect could reflect localized knowledge spillovers or the fact that the quality of worker-firm matches may be better in larger clusters due to labor pooling.

Empirically determining the productivity advantage of Silicon Valley-style clusters is difficult. First, location is endogenous, since workers and firms decide where to locate based on a number of observed and unobserved factors. Comparing the productivity of inventors in large clusters to the productivity of inventors in small clusters may yield biased estimates of agglomeration effects if particularly productive inventors select into large clusters. Second, direct worker-level measures of productivity are rare. Most existing papers on the effect of agglomeration on innovation rely either on wages—under the assumption that they reflect the marginal product of labor—or on analyses of patent citations (Jaffe, Trajtenberg, and Henderson 1993).[1]

In this paper I use longitudinal data on top inventors based on the universe of patents filed in the United States between 1971 to 2007 to investigate two related questions: (i) Are there productivity benefits for inventors who locate in Silicon Valley-style clusters? (ii) What are the aggregate effects of geographical agglomeration on overall innovation in the United States? In my analysis, I define a cluster as city $\times$ research field. I estimate how inventors' productivity varies with the size of the relevant cluster, measured by the number of other inventors in the same city and field, excluding the focal inventor. I use the number of patents produced in a year and the number of subsequent citations these patents receive, as measures of worker-specific productivity. The latter is a measure of the quality of the patents produced. While these measures have some limitations, they are arguably a more direct measure of inventor productivity than wages. I focus on top inventors—defined as those above the ninetieth percentile in the total number of patents—since the panel is longer for this group.

The analysis proceeds in three parts. I first study the experience of inventors in Rochester, New York, where the high-tech cluster declined due to the demise of its main employer, Kodak. Kodak was the market leader in films for cameras and the fifth most prolific patenter in the United States. At its peak in 1996, Kodak employed more than one-half of all the inventors in Rochester. But due to the diffusion of digital photography and the decline of physical film, Kodak stock price and employment collapsed after 1996. Essentially, demand for Kodak's main product evaporated due to a global technology shock. By 2007, the number of Kodak inventors in Rochester had declined by 84 percent.

Kodak's decline had a profound effect on the broader Rochester high-tech cluster. Measured by the number of inventors in all fields, its size declined by 49.2 percent relative to other cities, dragged down by Kodak's downsizing. The shock was large and arguably exogenous, as it was caused by the advent of digital photography and not factors specific to Rochester's local economy. The experience of Rochester therefore offers an interesting case study for testing the hypothesis that high-tech clusters' size affects inventors' productivity. In particular, examining how the sudden collapse of Rochester's high-tech cluster affected inventors *outside Kodak* offers a direct test of the effects of cluster size on productivity. I focus on non-Kodak inventors outside the photography sector, since the photography sector may have been exposed to the same negative demand shock as Kodak. Finding that their productivity did not change following the demise of the local high-tech cluster would cast doubt on the hypothesis that productivity depends on cluster size.

I compare productivity changes in Rochester with changes in other cities between 1996—the year when Kodak stock price peaked—and 2007. Specifically, I compare the within-inventor change in productivity between 1996 and 2007 for non-Kodak inventors who were in Rochester in 1996 (irrespective of their 2007 location) to the within-inventor change for non-Kodak inventors who were in other cities in 1996. I find that the log productivity of non-Kodak inventors in Rochester declined by 0.206 (0.077) relative to other cities. This estimate is not driven by changes in unobserved

[1] Greenstone, Hornbeck, and Moretti (2010) and Moretti (2004) study productivity spillovers on firm-level total factor productivity.

quality of inventors, since I am following the same inventor over time. Thus, following the decline in the Rochester high-tech cluster, non-Kodak inventors in Rochester experienced large productivity losses relative to non-Kodak inventors in other cities. This is consistent with the existence of important productivity spillovers in the high-tech sector stemming from geographical agglomeration.

In the second part of the paper, I present estimates based on all the data in the sample. The sample includes 109,846 inventors observed between 1971 and 2007, located in 895 clusters (179 cities $\times$ 5 research fields). I regress inventor's log productivity on log cluster size, conditioning on inventor, firm, city $\times$ year effects, as well as other controls.

I find that when an inventor moves to a larger cluster, she experiences significant increases in the number of patents produced in a year and their quality. Models that include five leads and lags allow me to estimate the dynamic response to a change in cluster size and reveal that the productivity increase follows the move, and does not precede it. There is no evidence that future values of treatment affect current productivity. I propose an instrumental variable (IV) based on the geographical structure of firms with laboratories in multiple cities. The instrument uses the expansion of local firms outside the cluster to predict changes in local cluster size. Two-stage least squares (2SLS) models confirm that increases in cluster size cause increases in productivity.

I also find that inventors who move to larger clusters tend to cite more patents per patent created than otherwise similar inventors in smaller clusters, and this is particularly true of local patents. Since citations made by an inventor are arguably a sign that the inventor knows about a specific innovation, this finding suggests that moving to a large clusters results in more knowledge of existing innovations, especially the local ones. This is consistent with the notion that larger clusters allow for knowledge and ideas to spread more efficiently, as documented by Saxenian (1994), who describes how in Silicon Valley ideas flow very fluidly between innovators and this fosters their creativity.

The elasticity of number of patents in a year with respect to cluster size is 0.0676 (0.0139). The estimated elasticity implies that a computer scientist moving from the median cluster in computer science (Gainesville, Florida) to the cluster at the seventy-fifth percentile of size (Richmond, Virgina) would experience a 12.0 percent increase in productivity, holding constant the inventor and the firm. In biology and chemistry, a move from the median cluster (Boise, Idaho) to the seventy-fifth percentile cluster (State College, Pennsylvania) is associated with a productivity gain of 8.4 percent, holding constant the inventor and the firm.

My estimates can be used to quantify the spillover effects that a specific firm generates within a cluster. The size of such an externality varies enormously across firms. The externality generated by the average firm in the average city is 0.3 percent and 0.24 percent in computer science and biology and chemistry, respectively. But it is much larger for firms that account for a large number of inventors in the local cluster. For example, the productivity of non-Microsoft computer scientists in Seattle is estimated to be 8.06 percent higher because of the presence of Microsoft in the local computer science cluster. This large external effect reflects Microsoft's remarkable size in this field in Seattle. Having estimates of the productivity spillover that a specific firm generates in a specific cluster may prove useful to local and state governments that offer subsidies to attract high-tech firms to their jurisdiction.

In the final part of the paper, I seek to quantify the macroeconomic benefits of agglomeration for the United States as a whole. I use the estimated elasticity of productivity with respect to cluster size to ask how much geographical clustering contributes to the overall production of patents in the United States. In particular, is the total number of patents produced each year in the country made larger by the fact that inventors in each field concentrate in a handful of locations, compared to the case where inventors are spread more equally across locations?

I compare the observed total number of patents produced annually in each field in the United States to the number of patents that would be produced if inventor quality and firm quality did not change but some inventors were spatially reallocated from large clusters to small clusters up to the point where clusters size within each field is equalized across cities. Because of the effect of agglomeration on productivity, such spatial redistribution would increase the productivity of inventors in clusters smaller than average and lower the productivity of inventors in clusters larger than average. My estimate of the elasticity of productivity with respect to cluster size implies that the average productivity of computer scientists in the San Francisco-Silicon Valley region, for example, would be 22.76 percent lower than the observed productivity in 2007 because the size of the San Francisco-Silicon Valley computer science cluster would be made significantly smaller. The corresponding losses for New York, Seattle, Austin, and Boston would be −17.81 percent, −16.52 percent, −14.76 percent, and −13.45 percent, respectively. On the other hand, the average productivity of computer scientists in Kansas City would be 2.66 percent higher than the observed productivity in 2007, because the Kansas City computer science cluster is smaller than average. The corresponding gains for Omaha; Portland, Maine; and Memphis would be 13.42 percent; 17.76 percent; and 23.36 percent, respectively.

On net, the magnitude of the aggregate effect for the country as a whole of such spatial redistribution depends on the relative magnitude of the gains in small clusters compared to the losses in large clusters. Empirically, I find significant aggregate efficiency gains from clustering. The total number of patents created in the United States in computer science would be 13.34 percent lower in 2007 if computer scientists were uniformly distributed across cities. The corresponding losses in biology and chemistry, semiconductors, other engineering, and other sciences would be −10.06 percent, −14.83 percent, −7.71 percent, and −9.75 percent, respectively. The change in the total number of patents in the United States would be −11.20 percent.

Thus while the spatial clustering of high-tech industries may exacerbate earning inequality across US communities, it is also important for overall production of innovation in the United States. The unequal distribution of earnings growth across cities is a source of policy concern and has spurred a wealth of policy proposals to help struggling cities and regions, especially those in the Rust Belt (Gruber and Johnson 2019). Policies that seek to attract high-tech investment to communities that do not have any—like the incentives offered by some local and state governments—might help reduce earning differences but are costly in the aggregate, creating a classic case of an equity-efficiency trade-off.

This study relates to an extensive literature on the link between agglomeration and innovation (Audretsch and Feldman 1996). The local nature of knowledge flows and productivity spillovers is frequently noted in the literature. For example, Kantor and Whalley (2014 and 2019) find significant spillovers from academic research and development (R&D) on local firms. Helmers and Overman (2017) find that the establishment of the Diamond Light Source synchrotron in the United Kingdom induced a clustering of related research in its geographic proximity and raised research output within a 25 kilometer radius. Bloom, Schankerman, and Van Reenen (2013) find evidence of large spillovers from R&D. See Carlino and Kerr (2015) for a comprehensive survey of this literature.[2] Not all papers in this literature find evidence of positive spillovers. See, for example, Waldinger (2012), Moser, Voena, and Waldinger (2014), and Azoulay, Graff Zivin, and Wang (2010). This paper is also part of a much larger literature on agglomeration economies in the broader economy (not just the high-tech sector), where most of the focus has been on measures of size at the city level, rather than cluster level.[3]

The remainder of the paper is organized as follows. Section I describes the data. Section II reports the estimates based on the Rochester case study. Sections III and IV report the estimates based on the full sample. Section V discusses the aggregate implications. Section VI concludes.

I. Data

I use data on the universe of US patents filed between 1971 and 2007 and ultimately granted. Unsuccessful applications are not included. The source of the data is the Connecting Outcome Measures in Entrepreneurship, Technology, and Science (COMETS) patent database (Zucker and Darby 2014). It is the same data used in a recent paper by Moretti and Wilson (2017).

A. *Patents*, *Inventors*, *and Their Location*

The data include 4,229,809 patents; 90.9 percent of patents in the sample are filed by inventors employed in private firms and 4.1 percent by inventors employed by universities, with the remainder are filed by inventors working in national labs, government, nonprofit institutions, or for themselves. Online Appendix Table A.1 shows the names of the 25 patent assignees—private or public—with the largest number of patents in my sample period. IBM is by far the most prolific patenter, with 155,790 patents filed between 1971 and 2007, followed by General Electric, Microsoft, Intel, and Kodak with 69,051, 43,556, 42,085, and 41,538 patents filed respectively. The US Navy and the University of California are the only noncorporate organizations in the table.[4]

[2]Guzman (2019) finds that startups that move to Silicon Valley patent more and introduce more products. Additional examples include, but are not limited to, Duranton and Overman (2005); Lychagin et al. (2016); Bosquet and Combes (2017); Acemoglu, Akcigit, and Kerr (2016); Akcigit et al. (2018); and Zacchia (2018).

[3]For surveys, see Behrens and Robert-Nicoud (2015), Combes and Gobillon (2015), Rosenthal and Strange (2004, 2006, and 2008), and Duranton and Kerr (2018).

[4]In the COMETS data, patent assignees (organizations) are identified by a unique code. Different recorded names that standardize to the same name have been assigned the same code. For example: IBM, IBM Corp., IBM Corporation have the same code FI2651. Moreover, organization names were hand cleaned to determine the variant names that a certain organization uses. For example: International Business Machines Corporation has been assigned the code FI2651 as well. Likewise, UCLA, University of California Los Angeles, and Univ Calif Los Angeles have all been assigned the same code.

The geographical units of analysis I employ are the Bureau of Economic Analysis's (BEA) "economic areas." There are 179 economic areas in the United States and they cover the entire country. In most cases, "economic areas" are similar to a metropolitan statistical area (MSA). For large areas like the San Francisco Bay Area, Boston, or New York, they tend to be larger than the corresponding MSAs, since they include the entire economic region. For example, the economic area for the San Francisco Bay Area includes the entire area between Santa Rosa to the north and San Jose to the south (Johnson 2004). In the rest of the paper, I will refer to economic areas as "cities."

Each patent is assigned to an economic area based on the inventor's *residential address*. Patenters must report their home address on their patent application and have no economic incentive to misreport it. There is no legal link between where a patent's inventor lives and the taxation of any income generated by the patent for the patent assignee/owner. In many cases both the inventor's residential address and the assignee address (typically the company that first owned the patent) are available. I do not use the latter because it may not reflect the location where the research was conducted, as in many cases it is the address of the corporate headquarters and not the R&D facility.[5]

In the COMETS data, patents are assigned to one of five main "research fields" and 579 "technology classes." The five research fields are semiconductors, integrated circuits, and superconductors (which accounts for 3.8 percent of all patents—for brevity, in the rest of the paper, I will refer to this field as "semiconductors"); computing and information technology (12.6 percent—for brevity: "computer science"); biology, chemistry, and medicine (22.3 percent—for brevity: "biology and chemistry"); "other engineering" (52.7 percent); and "other science" (8.4 percent). Technology classes are more detailed. Examples include "hybrid electric vehicles," "nanotechnology," "X-ray or gamma ray systems or devices," "exercise devices," and "electrical computers and digital processing systems: memory."

I aggregate the patent-level data to the inventor-year-level data by counting the number of patents created by an inventor in a year and the number of subsequent citations received by those patents, where year is defined as year of the patent *application*, not the year when the patent is granted. For citations, year is defined as the year of the cited patent's application, not the year of the citing patent's application. If a patent has multiple inventors, I assign equally weighted fractions of the patent and citations to each of its inventors. For example, if a patent has four inventors, each inventor is credited with one-quarter of a patent and a quarter of the subsequent citations. If an inventor has multiple patents with multiple addresses in a single year, I use the modal city for that inventor-year pair. If an inventor has patents in more than one research field or technology class in an year, I use the modal research field or technology class for that inventor-year pair. In case of ties (for example, if there are two cities or research fields or technology classes with the same frequency in a given inventor-year pair), I pick randomly.

In the main analysis of the paper, I focus on the productivity of star inventors, defined as those who are at or above the ninetieth percentile in the total number of

[5] It is possible in principle that an inventor's home lies outside a cluster while his professional work takes place inside a cluster. However, given the large size of BEA economic areas, this is unlikely to be a common problem.

patents over the sample period. The main motivation for focusing on stars is the length of the panel that they provide: They are in the sample for an average 7.01 years. The ninetieth cutoff is arbitrary, but I also show results for top 0.5 percent, 1 percent, 5 percent, 25 percent, and the full sample.

The sample of star inventors includes 109,846 inventors and 932,059 inventors $\times$ years observations with nonmissing information on field, class, and city. Of these, 823,375 also have nonmissing employer (patent assignee) identifier. The mean number of patents per year in the sample is 1.08 patents. The tenth, twenty-fifth percentile, median, seventy-fifth, and ninetieth percentiles are 0.25, 0.5, 1, 1.2, and 2 respectively.

B. *Clusters*

Clusters are defined as the combination of city $\times$ research field. Recall that there are 179 cities and 5 research fields, observed over 36 years. Cluster size in a given year is defined as number of inventors of any productivity level (not just stars) in a city $\times$ field pair, *excluding the focal inventor*, as a share of all inventors in that field and year.

For example, the top panel in Table 1 shows the largest clusters in computer science in 2007. The San Jose-San Francisco-Oakland region is by far the largest cluster in this field, accounting for 26.1 percent of all inventors in the field. Second on the list is the New York region, which accounts for 9.2 percent of all the inventors in the field. Seattle, Austin, and Boston follow, with cluster sizes equal to 8.2 percent, 6.0 percent, and 4.7 percent, respectively. The geographical concentration of inventors in computer science is very pronounced. The top ten cities account for 69.3 percent of all inventors in the field in 2007. The ratio between the size of the largest cluster—San Jose-San Francisco-Oakland—and median cluster—Gainesville, Florida—is 866.8, indicating that the Bay Area had 866.8 times the number of inventors in computer science than Gainesville did in 2007. The ratios between the ninety-ninth and median and the ninety-fifth percentile and median are equal to 306.5 and 79.5, respectively.

The middle panel is for biology and chemistry. The top five clusters are New York (11.3 percent), San Jose-San Francisco-Oakland (11.1 percent), Boston (6.9 percent), Philadelphia (6.4 percent), and Los Angeles (5.9 percent). With the total share of top ten cities equal to 59.2 percent, the overall concentration in this research field is lower than in computer science, but it remains quite pronounced. The maximum/median and the ninety-ninth percentile/median ratios are 126.2 and 124.4, respectively.

The bottom panel shows the largest clusters in the semiconductors field. The top five clusters are San Jose-San Francisco-Oakland (25.2 percent), New York (15.2 percent), Los Angeles (6.2 percent), Dallas (5.0 percent), and Phoenix (4.8 percent). The overall concentration in this research field is even higher than in computer science, with the share of top ten cities equal to 77.0 percent. Both the maximum/median ratio and the ninety-ninth percentile/median ratio are undefined since the median city had no inventors in this field in 2007. The tenth, twenty-fifth percentiles, median, seventy-fifth, and ninetieth percentiles across all fields and years are 0.00356, 0.01106, 0.02951, 0.06332, and 0.10786, respectively.

Table 1—Largest Clusters in Computer Science, Biology and Chemistry, and Semiconductors, 2007

	Size
Panel A. Computer science	
San Jose-San Francisco-Oakland, CA	0.261
New York-Newark-Bridgeport, NY-NJ-CT-PA	0.092
Seattle-Tacoma-Olympia, WA	0.082
Austin-Round Rock, TX	0.060
Boston-Worcester-Manchester, MA-NH	0.047
Los Angeles-Long Beach-Riverside, CA	0.039
Minneapolis-St. Paul-St. Cloud, MN-WI	0.034
Raleigh-Durham-Cary, NC	0.028
Denver-Aurora-Boulder, CO	0.023
San Diego-Carlsbad-San Marcos, CA	0.023
Portland-Vancouver-Beaverton, OR-WA	0.022
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	0.019
Dallas-Fort Worth, TX	0.015
Chicago-Naperville-Michigan City, IL-IN-WI	0.015
Panel B. Biology and chemistry	
New York-Newark-Bridgeport, NY-NJ-CT-PA	0.113
San Jose-San Francisco-Oakland, CA	0.111
Boston-Worcester-Manchester, MA-NH	0.069
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	0.064
Los Angeles-Long Beach-Riverside, CA	0.059
San Diego-Carlsbad-San Marcos, CA	0.045
Minneapolis-St. Paul-St. Cloud, MN-WI	0.038
Houston-Baytown-Huntsville, TX	0.031
Chicago-Naperville-Michigan City, IL-IN-WI	0.031
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	0.028
Raleigh-Durham-Cary, NC	0.019
Seattle-Tacoma-Olympia, WA	0.019
Indianapolis-Anderson-Columbus, IN	0.015
Cincinnati-Middletown-Wilmington, OH-KY-IN	0.015
Panel C. Semiconductors	
San Jose-San Francisco-Oakland, CA	0.252
New York-Newark-Bridgeport, NY-NJ-CT-PA	0.152
Los Angeles-Long Beach-Riverside, CA	0.062
Dallas-Fort Worth, TX	0.050
Phoenix-Mesa-Scottsdale, AZ	0.048
Boise City-Nampa, ID	0.047
Portland-Vancouver-Beaverton, OR-WA	0.045
Austin-Round Rock, TX	0.039
Burlington-South Burlington, VT	0.036
Boston-Worcester-Manchester, MA-NH	0.034
Albany-Schenectady-Amsterdam, NY	0.022
Minneapolis-St. Paul-St. Cloud, MN-WI	0.015
San Diego-Carlsbad-San Marcos, CA	0.014
Raleigh-Durham-Cary, NC	0.014

Note: Cluster size is defined as number of inventors in a city $\times$ field $\times$ year, excluding the focal inventor, as a share of all inventors in field $\times$ year.

Remarkably, the amount of spatial agglomeration in these fields has not been declining overtime. Figure 1 shows how the share of inventors in top ten cities has changed over time in computer science (top), biology and chemistry (center), and semiconductors (bottom). The share in computer science has increased monotonically. The share in semiconductors and biology and chemistry has also tended to increase, but not monotonically. Overall, the geographical concentration

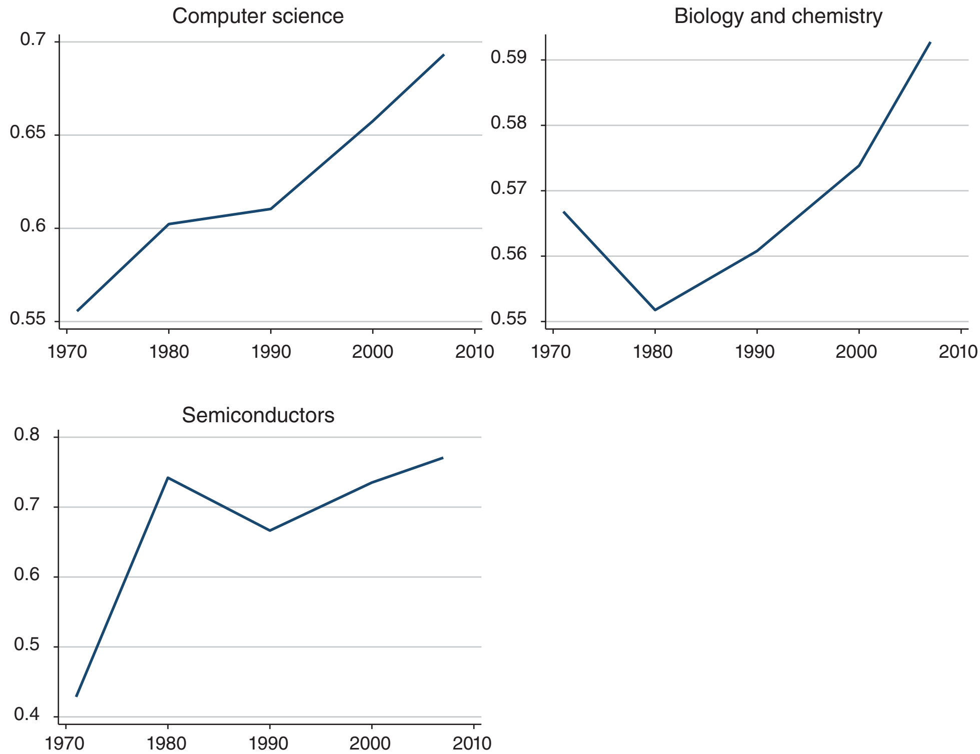

Figure 1. Share of Top Ten Cities over Time

of inventors at any moment in time is very high and it has been generally increasing over time since the 1970s.[6]

C. *Data Limitations*

Patent count is an imperfect measure of the amount of innovation created by an inventor. While some patents represent important innovations in a field, others represent trivial innovation. In the extreme, some patents do not represent any innovation and are filed for defensive purposes only. To account for these limitations, I use subsequent citations to measure patent quality.[7]

Not all innovation is patented and the fraction of innovation that is patented varies across fields and technology classes. Cohen, Nelson, and Walsh (2000) report that firms in the chemicals, drugs, mineral products, and medical equipment 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

[6] Alternative definitions of cluster size are possible. One could measure size of a cluster not based on the number of inventors in a city-field pair, but based on the number of *firms* (patent assignees). Based on this measure, the picture that emerges is also one of significantly concentrated patenting activity. Results available on request.

[7] Williams (2013) is an example of a paper that measures scientific research and product development directly rather than via patent activity.

on fewer than 15 percent of their product innovations. Moreover, the forms of technologies that are patented change over time. For example, the rate of innovation in software appears to accelerate in the 1990s, but this reflects at least in part changes in the legal practice of patenting software (Carlino and Kerr 2015). Lerner and Seru (2017) highlight the importance of controlling for the sectoral composition of inventive activity. For this reason, my models include research field $\times$ year and technology class $\times$ year effects. The assumption is that the fraction of innovation that is patented does not vary across cities within a field $\times$ year pair and technology class $\times$ year pair.

A concern is the possibility that the propensity to patent a given innovation may vary geographically as a function of cluster size, even within a field $\times$ year pair and technology class $\times$ year pair. This could be the case, for example, if firms located in larger clusters are more likely to patent a given innovation than firms in the same field and class located in smaller clusters, possibly due to a higher concern of being scooped by competitors. However, this possibility would imply that marginal innovations patented in large clusters are less valuable than the ones patented in small clusters, everything else equal. As I discuss below, this is inconsistent with the evidence on patent citations. Models based on number of citations per patent show a positive association between patent quality and cluster size.

A second limitation of the data stems from the fact that not every inventor applies for a patent every year, so I don't observe the productivity and location of every inventor in every year. This generates sample selection, as inventors are in the sample only when they patent. This selection problem is conceptually similar to the sample selection that exists in wage data due to the fact that wages are observed only in years when a worker is employed. To minimize this problem I focus my analysis on the population of star inventors. Star inventors are by construction prolific patenters and the typical individual is observed patenting in most years (over the period in which they patent at all).

Since I do not observe inventors in years when their number of patents is zero, a regression of number of patents on cluster size quantifies the effect of cluster size on inventor productivity *given a patent* (intensive margin), but it misses the effect of cluster size on probability of patenting (extensive margin). If the intensive and extensive margin effects go in the same direction, my estimates should be interpreted as a lower bound of total effect of cluster size on patenting.

To empirically probe the direction and magnitude of the bias, I provide two pieces of evidence in Section IVC. First, I add some of the missing zeros by interpolating the data when one missing year is immediately preceded and followed by nonmissing years. That is, if an inventor is observed in years $t-1$ and $t+1$, but is missing in year t, I assign her zero patents in t and I assign her to the cluster in which she was located at $t-1$. Estimates in the sample that includes these interpolated zeros are larger than the baseline estimates that do not include interpolation, because they include (part of) the extensive margin. If I interpolate data when two missing years are immediately preceded and followed by nonmissing years, estimates are even larger.

Second, I test whether my estimates are sensitive to different definition of the temporal unit of analysis. Specifically, I re-estimate my models using samples where inventor productivity is measured over one month, two month, three month, six month, one year (the baseline), two year, and three year periods. When the temporal unit of analysis is short (one or two months), the problem of missing zeros and the resulting downward bias should be particularly pronounced. In the extreme, if one could measure patent creation second by second, virtually all of the inventor-second pairs would be missing. By contrast, when the temporal unit of analysis is long (two or three years), the problem of the missing zeros and the downward bias should be less pronounced. In the extreme, if I were to use a temporal unit that includes all the years in the sample, there would be no selection and both the intensive margin and extensive margin would be reflected in my estimates. Empirically, I find that my estimates are increasing with the length of the unit of analysis. Overall, the evidence indicates that my baseline estimates should be interpreted as a lower bound of the total effect of cluster size on patenting.

II. The Effect of the Demise of Kodak on Inventors in the Rochester High-Tech Cluster

One way to credibly quantify the productivity advantages of large Silicon-Valley style clusters relative to small clusters is to isolate shocks to high-tech clusters that are exogenous to the local economy. The ideal shock is one that significantly affects the size of a local high-tech cluster—either positively or negatively—and at the same time is initially uncorrelated with the determinants of productivity of local inventors. The change caused by the shock in the productivity of inventors located in the affected cluster is potentially informative of the importance of localized agglomeration economies in the high-tech sector.

The rise and fall of some US tech clusters offer interesting case studies. In this section, I focus on the experience of Rochester, New York, over the past 25 years. It represents a useful case study because of the large, arguably exogenous change in the size of its high-tech cluster due to the demise of its main high-tech employer, Kodak. Rochester is not an isolated case. In the history of the US innovation sector, there are several examples of high-tech clusters that are born or die due to idiosyncratic, firm specific shocks. The rise of the software cluster in Seattle is a prominent example. The initial seed for the Seattle software cluster was Microsoft's relocation from Albuquerque, New Mexico, where it was founded. At the time, Seattle had a very limited software sector and the move can be considered largely exogenous, since it was motivated by personal reasons on the part of the company cofounders rather than business reasons or the conditions of the local economy.[8] Since Microsoft's relocation, Seattle has become one of the largest software clusters in the United States. The Seattle experience exemplifies a common way in which US high-tech clusters tend to emerge, namely through the growth of a local firm that becomes the seed around which a cluster agglomerates (Kerr 2010, Moretti 2012).[9]

[8] The key motivation for the move was the desire of Bill Gates and Paul Allen to be close to their families. At the time, the Albuquerque software sector was significantly more developed than the Seattle one. Indeed, Microsoft was founded in Albuquerque because its main clients were there. For a software start-up, there was no obvious business reason to choose Seattle over Albuquerque in 1979 (Moretti 2012).

[9] Kerr (2010) finds that patenting growth is significantly higher in cities where breakthrough inventions are located. Consistent with this notion, Duranton (2007) proposes a model of industry migration in which centers of innovation are dictated by where frontier inventions occur.

In a similar way, the birth of the biotech sector in San Diego can be traced to the serendipitous presence of star biologists at the University of California San Diego in the mid-1970s who founded some of the early local biotech firms (Zucker and Darby 1996). Austin had very little high tech in the early 1980s when Michael Dell started Dell Computers in his dorm at the University of Texas. The company is often credited to be the initial seed for the emergence of the Austin tech cluster. Relative to Seattle, San Diego, and Austin, the Rochester case study has the advantage that the effect of the shock on the local economy was particularly sudden.

In the 1980s and the 1990s, Kodak was the leading producer of films for cameras. The firm was one of the most prolific patenters in the United States. Indeed, as we saw in online Appendix Table A.1, Kodak had the fifth largest number of patents filed between 1971 and 2007. Kodak's patents are not concentrated in one research field, but span all five fields. Kodak was headquartered in Rochester and it was the largest patenter in the city by a vast margin. At its peak in 1996, Kodak accounted for 49 percent of all the inventors in Rochester. But after 1996, Kodak entered a period of dramatic decline caused by the diffusion of digital photography and the collapse in the demand for physical film (Dickinson 2017). In essence, Kodak experienced a large negative shock caused by a technological innovation that made its main product obsolete.

Kodak's demise can be seen in the top panel in Figure 2, which shows the firm's stock price since 1990. The price grew in the first half of the 1990s, reached a peak in 1996 and then began declining, as digital photography spread and the market for physical films shrank. Between 1996 and 2007, Kodak's stock price declined by 82 percent. Shrinking product demand led to less investment in R&D. The bottom panel in Figure 2 shows that the number of Kodak inventors in Rochester reached a peak in 1997, with 1,254 inventors, and then began a steep decline. By 2007 there were only 204 Kodak inventors in Rochester. The size of the overall Rochester high-tech cluster measured by the number of inventors in all fields declined by 50.5 percent in this period, dragged down by the collapse of Kodak.

The experience of inventors in Rochester can be used to test the hypothesis that the size of a high-tech cluster affects inventor productivity. Specifically, I test whether the change in the size of the Rochester high-tech cluster caused by Kodak's demise affected the productivity of inventors *outside Kodak*. The shock was large and arguably exogenous, as it had little to do with what was happening to the Rochester local economy. Finding that the productivity of inventors outside Kodak did not change following the large negative shock to Rochester high-tech cluster would cast doubt on importance of productivity benefits stemming from cluster size.

Baseline Estimates.—Figure 3 shows visually what happened to the mean productivity of *non-Kodak* patenters in Rochester in the period 1990–2007. Specifically, it plots the mean log productivity of inventors who do not work for Kodak, after controlling only for research field dummies.[10] The vertical red line marks 1996, which was the peak of Kodak's stock price. While log mean productivity appears generally flat before 1996, it declines after 1996. Excluding Kodak, the main patent

[10] In practice, I plot the mean residual from a regression of log productivity on research field dummies.

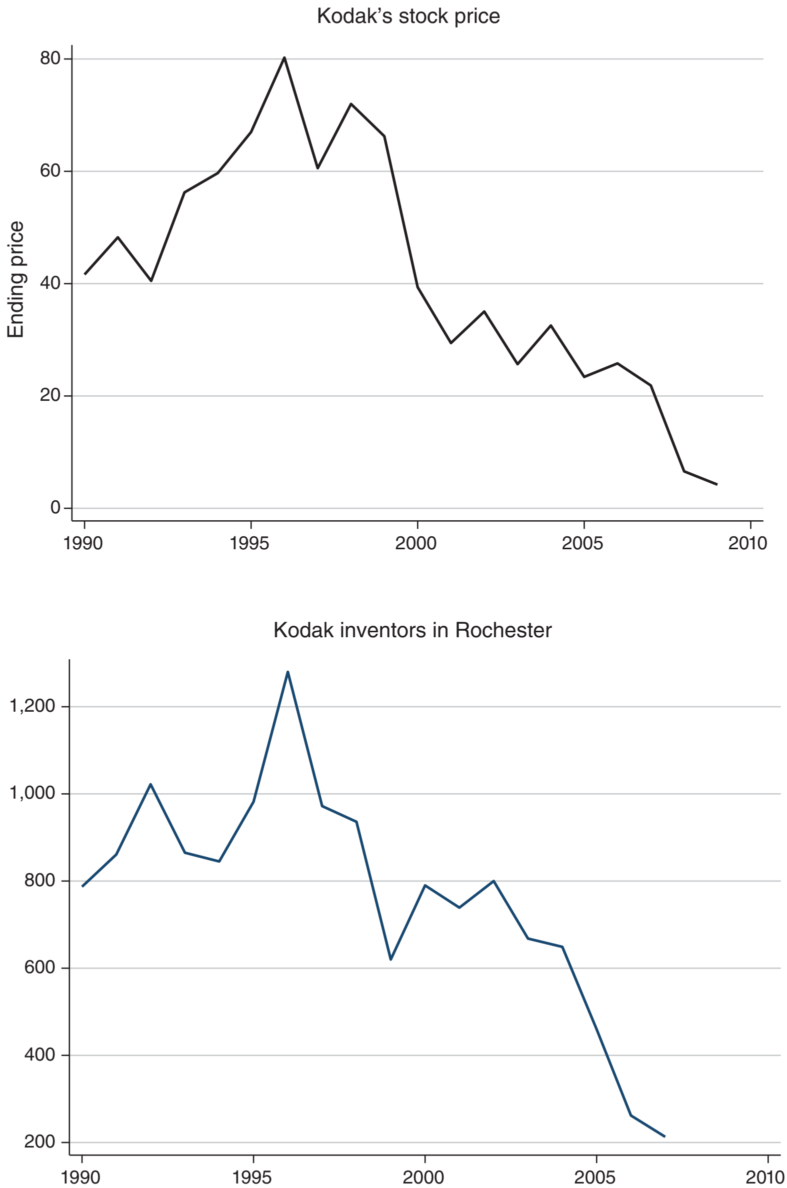

Figure 2. Kodak's Decline

producing firms in Rochester in 1996 were Coring (99 patents), Bausch Lomb (42), General Motors (41), Tenneco Chemical (38), Mobil Oil (33), Johnson and Johnson (28), Osram Sylvania (24), PSC (20), and IBM (16).

Table 2 presents corresponding difference-in-difference estimates comparing the 1996–2007 change experienced by non-Kodak inventors in Rochester to the change experienced by non-Kodak inventors in other cities. The two panels in the table differ because panel A is based on a cross-sectional comparison, while panel B uses the longitudinal nature of the data and focuses on within-inventor productivity change for a fixed set of inventors. (Since in 1996 Kodak had a presence in

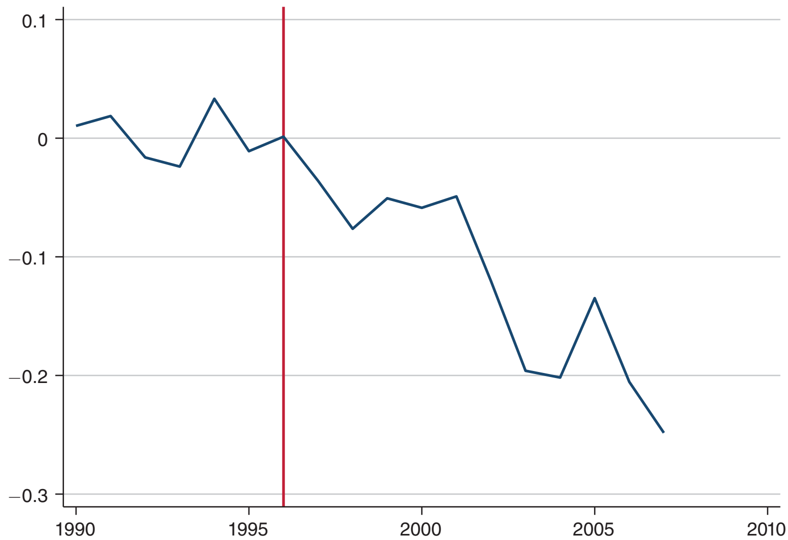

Figure 3. Average Inventor Productivity in Rochester outside Kodak

Note: Controls include research field dummies.

all five main research fields, no field can be considered unaffected by its demise. In the next section, where I use all cities and years, I will be able to identify field-specific changes in cluster size.) In both panels, the sample includes years 1996 and 2007. I use 1996 as the initial year because it is the year when Kodak stock prices peaked. I use 2007 as the final year because it is the last year available in the data.

The level of observation in the regressions is inventor-year. I drop all Kodak inventors from the analysis, whether in Rochester or in other cities. I also drop inventors in the photography sector, identified as those with patents in at least one of the following technology classes: "396, photography;" or "399, electro-photography," irrespective of their location. They are dropped because although they do not work for Kodak, they may be directly affected by the same negative shock. Thus, they may experience a productivity loss not due to agglomeration economies, but due to a decline in product demand for traditional photographic film.

In column 1 of panel A, inventor log productivity is regressed on a 2007 dummy, a Rochester dummy, and the interaction. In this table, standard errors are clustered by city. The coefficient on the Rochester dummy indicates that in 1996, the mean productivity of non-Kodak inventors in Rochester was not statistically different from the productivity in other cities. The coefficient of interest is the one on the interaction, which indicates that the mean number of patents filed in a year by non-Kodak inventors in Rochester declined by 6.41 percent between 1996 and 2007 relative to other cities. In columns 2 to 4, I add field, field $\times$ year, and field $\times$ city effects as controls. The coefficient on the interaction becomes more negative and suggests a productivity loss in Rochester between 6.73 percent and 9.16 percent relative to other cities, depending on the controls. For robustness, I estimated an additional

Table 2—Difference-in-Difference Estimates: 1996–2007 Productivity Change of Non-Kodak Inventors in Rochester Compared to Other Cities

					Weighted
	(1)	(2)	(3)	(4)	(5)
Panel A					
Rochester × 2007	−0.0641	−0.0673	−0.0805	−0.0916	−0.0947
	(0.00757)	(0.00674)	(0.00631)	(0.00665)	(0.00860)
Rochester	−0.0148	−0.0364	−0.0317		
	(0.0105)	(0.0101)	(0.00987)		
2007	−0.190	−0.189			
	(0.00757)	(0.00713)			
Observations	194,120	194,120	194,120	194,120	193,331
Field		Yes	Yes	Yes	Yes
Field × year			Yes	Yes	Yes
Field × city				Yes	Yes
Panel B. Within-inventor difference-in-difference estimates					
Rochester × 2007	−0.206	−0.222	−0.257	−0.309	−0.363
	(0.0772)	(0.0782)	(0.0825)	(0.0609)	(0.0387)
2007	−0.205	−0.206			
	(0.0190)	(0.0193)			
Observations	16,430	16,430	16,430	16,430	16,379
Inventor	Yes	Yes	Yes	Yes	Yes
Field		Yes	Yes	Yes	Yes
Field × year			Yes	Yes	Yes
Field × city				Yes	Yes

Notes: Each column within a panel is a separate regression. The level of observation in the regressions is inventor-year. The dependent variable is log of number of patents filed in a year. The sample in panel A includes non-Kodak inventors excluding the "photography" or "electrophotography" technology classes. The sample in panel B only includes the subset observed both in 1996 and 2007. Weights in column 5 are based on a vector of observable city characteristics measured before the shock to Kodak: 1990 city population; 1990 mean household income; 1996 mean inventor productivity; 1996 share of patents in each research field; 1990 share of non-White residents; 1990 city total employment; and 1990 city industry mix defined as share of employment in manufacturing, trade, construction, and agriculture. The weight for inventors who are not in Rochester is $1/(1-p)$. Standard errors are clustered by city, in parentheses.

model where I drop Xerox inventors, both in Rochester and other cities and find a similar estimate.[11]

The estimates in columns 1 to 4 compare productivity changes of inventors in Rochester to productivity changes of inventors in all other US cities. One reasonable concern is that not all other US cities offer a compelling counterfactual, since not all cities are similar to Rochester before the shock. For example, before 1996, Rochester had a larger population than the median city, and a higher median family income. The main research field (outside Kodak) in 1996 was "other engineering," which accounted for 67 percent of all inventors, a much larger share than the typical city. In column 5, I report weighted estimates, with weights reflecting how close the focal inventor city is to Rochester based on a vector of observable city characteristics measured before the Rochester shock: 1990 city population; 1990 mean household

[11] Xerox had an important presence in Rochester and experienced negative shocks in this period. It is not clear whether one should consider the Xerox shock in Rochester as exogenous or as an endogenous effect of Kodak's decline.

income; 1996 mean inventor productivity; 1996 share of patents in each research field; 1990 share of non-White residents; 1990 city total employment:, and 1990 city industry mix defined as share of employment in manufacturing, trade, construction, and agriculture.[12] The entry in column 5 puts more weight on inventors in cities that were more similar to Rochester and is similar to the entry in column 4.[13]

Within-Inventor Estimates.—Since the estimates in panel A are based on cross-sectional comparisons, one obvious concern is the possibility of unobserved quality changes due to inventor selection. In particular, it is possible that the quality of non-Kodak inventors in Rochester in 1996 is not the same as the quality in 2007, although the direction of the bias is a priori unclear. On the one hand, it is possible that the decline in the local high-tech cluster induced some of the best non-Kodak inventors to leave Rochester. On the other, it is also possible that Kodak's downsizing allowed other firms in Rochester to hire good inventors laid off by the firm. If former Kodak inventors hired by non-Kodak employers are on average better than incumbent inventors, the average unobserved ability of inventors working outside Kodak may increase.

In panel B, I present estimates that use the longitudinal structure of the data. I use a balanced panel made up of the inventors who are observed both in 1996 and in 2007 and I add inventor fixed effects to my models. Thus, panel B reports estimates where I compare within-inventor productivity changes in Rochester to within-inventor productivity changes in other cities for a fixed set of inventors. The sample includes 8,215 inventors observed twice, for a total sample size of 16,430, as shown in the bottom row.

I assign Rochester status to an inventor based on her 1996 location. For a given inventor, the Rochester dummy is set equal to one if the inventor is in Rochester in 1996. That is, I compare within-inventor productivity changes of inventors who in 1996 were located in Rochester to within-inventor productivity changes of inventors who in 1996 were located in other cities, *irrespective of their 2007 location*. This specification is preferred to one where the Rochester dummy is set equal to one if the inventor is in Rochester both in 1996 and in 2007, because 2007 location is potentially endogenous. It estimates the effect of the shock to the Rochester high-tech cluster on inventors who were there at the beginning of the shock, allowing inventors to optimally choose their 2007 location.

In column 1 the only set of controls used is inventor fixed effects. The difference-in-difference estimate in the first row suggests that inventors who were in Rochester in 1996 experienced a change in productivity that was 20.6 percent

[12] The weight for inventors who are not in Rochester is $1/(1-p)$, where p is estimated using a logit. Data are from Moretti (2021a).

[13] A more extreme option is to include in the sample only cities with characteristics similar to Rochester. This is equivalent to putting zero weight on some cities. I re-estimated the models in columns 1 to 4 including a subset of cities that are more similar to Rochester than the full sample of cities based on city population in 1990, mean household income in 1990, mean inventor productivity in 1996 (excluding Kodak) and share of patents in Rochester's main research field in 1996 (excluding Kodak). I dropped cities with below median population in 1990, below median household income in 1990, 1996 mean inventor productivity that was 10 percent larger or 10 percent smaller than 1996 mean inventor productivity in Rochester (excluding Kodak) and with a presence in "other engineering" research field below 40 percent. Results are generally similar to the ones presented in the table (available on request).

more negative than inventors who were not in Rochester in 1996, after controlling for inventor fixed effects. When I add field, field $\times$ year, and field $\times$ city effects as controls in columns 2, 3, and 4, the coefficient on the interaction becomes more negative and suggests a productivity loss in Rochester between 22.2 percent and 30.9 percent relative to other cities. Weighted estimates in column 5 are even more negative. A comparison with the corresponding cross-sectional estimates in panel A indicates that within-inventor estimates are more negative than cross-sectional estimates, suggesting that unobserved quality of non-Kodak inventors biases cross-sectional estimates toward zero. This is consistent with the possibility that some of the best Kodak inventors laid off by Kodak are hired by other Rochester employers.[14]

Kodak Suppliers.—Product demand faced by Kodak's suppliers likely declined following Kodak demise. If the decline in product demand resulted in lower inventor productivity at Rochester suppliers, it could in principle explain part of the decline in inventor productivity in Rochester after Kodak decline. In this case, one would want to exclude the effect on supplier productivity from the estimate of the productivity losses in Rochester, since it reflects a product demand shock, not productivity spillovers. On the other hand, it is also possible that the decline in supplier productivity—if it took place—was the endogenous effect of the demise of Kodak. Geographical agglomeration of specialized suppliers is one of the mechanisms that the literature on agglomeration effects has identified as possible explanations of productivity spillovers. In this case, one would want to include the effect on supplier productivity in the estimate of the productivity losses in Rochester.

It should be noted that suppliers with patents in the "photography" and "electrophotography" classes—probably a sizable share of Kodak suppliers who engage in patenting—are already excluded from the sample used to estimate Table 2. In order to assess how robust my estimates are to excluding remaining Kodak's suppliers, I use input-output tables (BEA 2012) to identify likely Kodak suppliers outside "photography" and "electrophotography." In particular, I use input-output tables at the six-digit level to identify how much each six-digit industry sells to North American Industry Classification System (NAICS) 333316 (photographic and photocopying equipment manufacturing), which is the one that Kodak likely belongs to. I then re-estimate Table 2 excluding inventors in technology classes that have a positive amount of expected sales to NAICS 333316.[15] Estimates corresponding to column 4 of Table 2 are -0.0633 (0.0139) and -0.459 (0.1210). Thus, good flows are unlikely to be the main driver of the findings.

Overall, I conclude that following the decline in the Rochester high-tech cluster, non-Kodak inventors in Rochester experienced large productivity losses relative to non-Kodak inventors in other cities. While this is by no means direct proof of productivity spillovers, this finding is consistent with the hypothesis that cluster size affects inventor productivity.

[14] In addition, estimates in panel B are based on inventors who are in the sample both in 1996 and 2007. Estimates are larger for inventors who are in the sample for long periods of time, as I discussed in Section IC above.

[15] To link patents to industries, I used the crosswalk by Zolas (2016): for each parent class, the crosswalk lists multiple six-digit industries, with assigned probabilities. Based on the probabilities, I computed the expected value of inputs used by Kodak for each technology class.

III. Inventor Productivity and Cluster Size

The Rochester analysis focuses on a specific shock to one city. It has the advantage that the source of variation in cluster size is clear. It has the disadvantage that it is a case study based on the experience of only one community. I now turn to estimates of the relationship between inventor productivity and cluster size based on all cities, fields, and years in my sample. I assume that the log of productivity of an inventor in a cluster depends upon her skills, location fundamentals, agglomeration effects, and an idiosyncratic component:

$$(1) \qquad \ln y_{ijfkct} \,=\, \alpha \ln S_{-ifct} + d_{cf} + d_{ck} + d_{ft} + d_{kt} + d_{ct} + d_i + d_j + u_{ijfkct},$$

where y_{ijfkct} is the number of patents produced (or citations received) in year t by inventor i working in firm j in research field f, technology class k, and located in city c; S_{-ifct} is the size of the cluster in the relevant field, city, and year, excluding inventor i; d_{cf} and d_{ck} are city $\times$ field and city $\times$ class effects included in order to absorb time invariant factors that may make specific fields or specific technology classes particularly productive in some locations; d_{ft} and d_{kt} are field $\times$ year and technology class $\times$ year effects and are included to absorb nationwide time-varying technological and sectoral changes; d_{ct} represents city $\times$ year effects and is included to absorb citywide changes in the determinants of productivity shared by all fields in a location; and d_i and d_j are inventor and firm effects. To account for possible serial correlations in the residual, I report standard errors clustered by city $\times$ research field.

If cluster size raises inventor productivity due to productivity spillovers, α should be greater than zero. In the absence of spillovers α should be zero. Note, α is identified by inventors who move across cities or across fields and by changes over time in clusters size in a given city and field pair. For ordinary least squares (OLS) to identify α, the unobserved determinants of productivity u_{ijfkct} need to be orthogonal to S_{-ifct}. There are two identification concerns: sorting (or endogenous quality of labor), and simultaneity (or endogenous quantity of labor).

The first concern stems from the fact that inventors *choose* their cluster. An inventor's choice of location likely depends on a number of factors, including earnings, cost of living, local amenities, and idiosyncratic preferences for specific locations. Some forms of sorting do not violate the orthogonality condition. For example, it is possible that highly productive scientists tend to locate in large clusters. Sorting based on the permanent component of productivity does not introduce bias in OLS estimates, since equation (1) conditions on the inventor fixed effects. The equation also controls for firm effects, and thus is identified by within firm differences in cluster size across inventors.

Sorting driven by local amenities also does not violate the orthogonality condition. Dummies for the interaction of city $\times$ year (i.e., d_{ct}) absorb all characteristics of a city that may affect its attractiveness, both permanent and time-varying, including cultural amenities, restaurants, entertainment, school quality, crime, congestion, costs of living, and local taxes. Endogenous changes in local amenities of the type discussed by Diamond (2016) are also absorbed by d_{ct}. Citywide productive amenities—i.e., factors that affect the productivity of all inventors, irrespective of field, such as local infrastructure, productive public goods and regulations—are also absorbed. Identification relies on the fact that there are multiple fields within each city.

Sorting due to time-invariant factors that shift the productivity of scientists in specific fields and locations is also accounted for. As an example, consider cases where particularly productive computer scientists are attracted to the San-Francisco Silicon Valley area by the proximity to top engineering departments in Stanford University or the University of California, Berkeley, or, alternatively, where the engineering departments of Stanford and Berkeley produce top computer scientists who after graduation tend to stay in the area. City $\times$ field effects control for this type of productivity differences, to the extent that they are time invariant.

Sorting into large clusters caused by time-varying unobserved productivity shocks, on the other hand, is a potentially important concern. Consider, for example, the case of inventors in a small cluster whom employers in large clusters expect to become more productive in the future. These expectations are of course not observed in my data. If employers in large clusters systematically hire promising inventors from small clusters, equation (1) will overestimate the effect of cluster size, because it would attribute productivity increases that arise because of sorting to cluster size. In this case, productivity might be increasing even before the move. To assess the importance of this type of sorting, I study the timing of productivity changes relative to changes in cluster size. If the model is properly specified, future values of treatment should not affect current outcomes.

A second identification concern is simultaneity: the existence of unobserved time-varying productivity shocks at the city-field level that attract more inventors to a city field. In practice, some of the localized shocks that affect firms productivity are likely to be citywide—for example, a new airport, or other form of improved infrastructure—and therefore are controlled for in equation (1). But the city $\times$ year effects do not control for time-varying productivity shocks that are both city and field specific. One example might be changes in local subsidies, to the extent that they target specific fields. This could be the case if city or county adopts subsidies for, say, biotech firms, causing an increase in the size of the local biotech cluster. If subsidies directly affect biotech firms' productivity, equation (1) will overestimate the effect of cluster size. This could happen if subsidies allow local biotech firms to buy equipment that they could not afford in the absence of a subsidy. (There is, of course, no bias if subsidies affect productivity only indirectly by increasing cluster size—which is the case if subsidies increase the number of local inventor and that in turns affects inventor productivity.) To assess the importance of this type of bias, I use two alternative instrumental variables (IVs).

In some models of agglomeration economies, productivity spillovers are assumed to depend on density of a cluster, rather than its absolute size. In interpreting the parameter α, one should keep in mind that it identifies the effect of cluster size on productivity, holding constant the land mass of the relevant area. Thus, the estimated effect can be interpreted as the effect of cluster density on productivity. To see this, consider that a log-log model with city $\times$ year dummies such as the one in equation (1) yields an estimate of α that is numerically identical to that which one would obtain if the independent variable is log density defined as number of inventors per square mile: $\ln\left(S_{-ifct}/A_c\right)$, where A_c is the area of the city.[16]

[16]The same conclusion applies to the case where density is defined as number of inventors divided by the area occupied by a specific field within a city: $\ln\left(S_{-ifct}/A_{fc}\right)$, where A_{fc} is the area of the cluster.

A. *Baseline Estimates*

Figure 4 shows a binned scatterplot of the correlation between log number of patents and log cluster size, conditioning only on year effects, research field effects and city effects. The positive slopes indicate that larger clusters appear to be associated with a higher number of patents generated per year. The estimated slope is 0.053 (0.002).

Table 3 reports estimates of variants of equation (1). Column 1 reports estimates from a model that includes year, research field, technology class, and city effects. Thus, the controls are the same as those used in Figure 4, with the addition of technology class effects. The estimated coefficient is similar: 0.0518 (0.0081). In columns 2 and 3, I add dummies for the interactions of city × field and city × class. While the field and class effects in column 1 absorb nationwide productivity differences across fields and class, the interactions in columns 2 and 3 absorb features of an area that may make specific fields or specific technology classes particularly productive. In columns 4 and 5, I add dummies for the interactions of field × year and class × year in order to control for nationwide technological shocks. In column 6, I add inventor fixed effects. This specification absorbs time-invariant quality differences between inventors. The coefficient is 0.0923 (0.00990). A comparison with column 5 indicates that the within-inventor estimate is larger. This is surprising, because it suggests that conditional on the controls included in the model in column 5, larger clusters attract inventors with lower mean unobserved quality.[17]

In column 7, I add dummies for the interaction of city × year in order to absorb citywide shocks to local productive amenities and selection driven by shocks to local consumption amenities. The coefficient drops to 0.0545 (0.0116), significantly lower than the one in column 6. This suggests that during my sample period, large clusters tend to experience on average more positive citywide productivity shocks than small clusters; or that changes in local amenities in large clusters tend to attract more productive inventors compared to small clusters; or both. Finally, in the last column I add firm effects. Firm identifiers are not available for all patents, since not all inventors are employed by firms and as a consequence sample size drops from 932,059 to 823,375. The coefficient is 0.0676 (0.0139).

The elasticity of productivity with respect to cluster size estimated in column 8 indicates that a 10 percent increase in cluster size is associated with a 0.67 percent increase the number of patents produced by a scientist in a year. To help interpret the magnitude of the estimated effect, consider an inventor in the computer science field who in 2007 moves from the median cluster—Gainesville, Florida—to the cluster at the seventy-fifth percentile—Richmond, Virgina. Based on the coefficient in column 8, such an inventor would experience a 12.0 percent increase in the number of patents produced in a year, holding constant the inventor and the firm. In biology and chemistry, a move from the median cluster—Boise, Idaho—to the seventy-fifth percentile cluster—State College, Pennsylvania—would be associated with a productivity gain

[17] It is possible that moving lowers the probability of patenting—either because it is distracting and time consuming, or because it coincides with a job spell with a new employer and the inventor may be restricted from using intellectual property created while working for a previous employer. This could lead me to underestimate productivity following a move, but there is no obvious reason to expect that this effect is different for those moving from small to large clusters compared to those moving from large to small clusters.

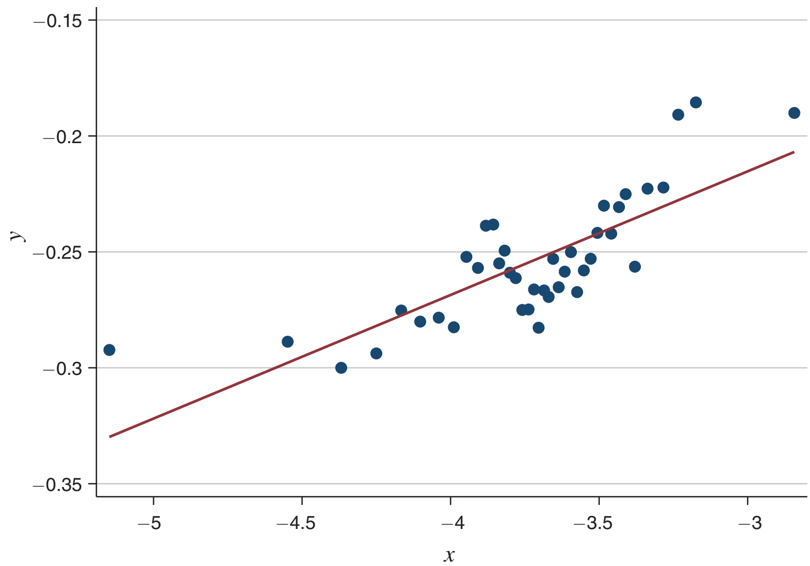

FIGURE 4. AVERAGE LOG NUMBER OF PATENTS PER INVENTOR PER YEAR AND LOG CLUSTER SIZE: ALL YEARS AND FIELDS

Notes: The slope is 0.053 (0.008). Controls include dummies for year, field, and city.

of 8.4 percent, holding constant the inventor and the firm. The estimated productivity gain is smaller than in computer science because the difference in cluster size is smaller.[18] The magnitude of the implied agglomeration economies is in line with existing estimates in the literature on agglomeration economies.[19]

One can use the estimated elasticity to quantify the externality that a given firm generates in a given cluster. The externality reflects the impact that a specific firm is estimated to have on the productivity of scientists in other firms in the same cluster in a given year. Online Appendix Table A.3 shows examples of estimated firm-specific productivity spillovers for selected firms and clusters in 2007. The estimate for a given firm and cluster quantifies the percent gain in the productivity of local scientists in other firms due to the presence of the firm in the cluster relative to the case where the firm was not present in the cluster and everything else in the cluster was

[18] The baseline model in equation (1) assumes that the productivity spillover is the same across fields. This assumption has the advantage of being parsimonious. In addition, it allows me to use within city-year variation in cluster size, since in each city and year there are five research fields. Online Appendix Table A.2 shows estimates of the effect by research field. These models do not include city × year effects to avoid multicollinearity. (The corresponding elasticity for the full sample is 0.1081 (0.0119).) The elasticity is largest for semiconductors (0.262), and computer science (0.187), and it is smallest for other science (0.076).

[19] A meta-analysis of 34 different studies (Melo, Graham, and Noland 2009) indicates that my estimated elasticity is below the middle of the distribution of existing estimates but within the range of elasticities reported in several recent studies. For example, Henderson (2003) obtains an elasticity of productivity with respect to density of 0.01–0.08. Estimates for France in Combes et al. (2010) and Combes et al. (2012) imply elasticities of 0.029 and 0.032, respectively. Kline and Moretti (2014a) find an elasticity of 0.2 for US manufacturing. At the other extreme, Greenstone, Hornbeck, and Moretti's (2010) estimates imply an elasticity in the range 1.25–3.1. Of course, part of the variation in these estimates is due to the fact that the models, data, time periods and industries used in the studies are vastly different. The elasticity estimated in this paper is also consistent with estimates based on wages (for example, Ciccone and Hall 1996; Ciccone and Peri 2005; Combes, Duranton, and Gobillon 2008; Rosenthal and Strange 2008).

Table 3—Effect of Cluster Size on Inventor Productivity: Baseline Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log size	0.0518	0.0762	0.0881	0.0907	0.0677	0.0923	0.0545	0.0676
	(0.00815)	(0.0167)	(0.0187)	(0.00926)	(0.00862)	(0.00990)	(0.0116)	(0.0139)
Observations	932,059	932,059	932,059	932,059	932,059	932,059	932,059	823,375
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class			Yes	Yes	Yes	Yes	Yes	Yes
Field × year				Yes	Yes	Yes	Yes	Yes
Class × year					Yes	Yes	Yes	Yes
Inventor						Yes	Yes	Yes
City × year							Yes	Yes
Firm								Yes

Notes: Each column is a separate regression. The level of observation in the regressions is inventor-year. The dependent variable is log of number of patents filed in a year. The model estimated is equation (1). Standard errors are clustered by city × research field.

unchanged.[20] The top panel is for computer science. The entry in the top row indicates that Microsoft productivity spillover on the Seattle computer science cluster is estimated to be 8.06 percent. This large external effect reflects Microsoft's remarkable size in the Seattle computer science cluster. With 1,389 inventors in computer science in Seattle in 2007, Microsoft was the largest firm in the cluster by a vast margin. IBM's spillover effect on Minneapolis is estimated to be very large as well (4.15 percent). Although IBM's headquarter is not in Minneapolis, the firm has a large R&D presence in the city. On the other hand, Cisco's spillover effect on the San Francisco-Silicon Valley cluster is only 0.2 percent. The external effect of Dell in Austin, Texas Instruments in Dallas, Caterpillar in Peoria, and Motorola in Chicago are, respectively, 0.61 percent, 1.84 percent, 10.47 percent, 0.88 percent. The estimated external effect of Hewlett-Packard in San Francisco-Silicon Valley is only 0.07 percent. The last row in the panel indicates that the spillover effect generated by the average firm in the average city in computer science was 0.32 percent. Thus, Microsoft and Caterpillar are outliers and the more typical case is closer to that of Cisco or Motorola. The bottom panel shows examples for the biology and chemistry field.[21]

Having estimates of the productivity spillover that a specific firm generates in a specific cluster may prove useful to local and state governments that offer subsidies

[20] The estimate is a function of the number of inventors that the firm has in the cluster relative to the overall number of inventors in the cluster. For a given firm j, field f, and city c, the estimated spillover is obtained as $\hat{\alpha}\Delta S_{-jfct}$ where $\hat{\alpha} = 0.067$ is the estimated elasticity and $\Delta \ln S_{-jfct}$ is the difference in log cluster size with and without a given firm: $\Delta \ln S_{-jfct} = \left[\ln\left(N_{fct}/N_{ft}\right) - \ln\left(N_{jfct}/N_{ft}\right)\right]$, where N_{fct} is the number of scientists in cluster fct, N_{ft} is number of scientists in field f and year t, N_{jfct} is number of scientists in firm j in cluster fct, and $t = 2007$.

[21] Du Pont has a very large presence in Philadelphia, with 986 scientists in 2007. It generates a productivity spillover equal to 1.17 percent. Two examples of firms with a dominant position in their local cluster and a very large estimated productivity spillover are Procter and Gamble in Cincinnati, Ohio, and 3M in Duluth, Minnesota, with productivity spillovers estimated to be 3.47 percent and 8.99 percent, respectively. More typical cases are Bristol-Myers in New York, Amgen in Los Angeles, Chevron in the Bay Area, and Exxon in Washington, DC, with spillovers in the 0.14–0.65 percent range. The spillover for the average firm in the average city reported in the last row is 0.24 percent.

to attract high-tech firms to their jurisdiction or to retain incumbent firms. In theory, one economic rationale for offering subsidies to attract or retain high-tech firms is the existence of localized productivity spillovers. (See Greenstone, Hornbeck, and Moretti 2010, for a discussion of firm specific subsidies and productivity spillovers in manufacturing). Local and state governments interested in offering subsidies proportional to the spillover effects that a firm may generate in their jurisdiction could use an approach similar to the one used in online Appendix Table A.2 together with information on the expected number of local inventors in the firm they are targeting. The estimates in Table A.2 reflect productivity gains scaled in terms of the number of additional patents. Local governments interested in the corresponding dollar value would need to make an assumption on the expected monetary value of the marginal patent generated.

B. *Dynamic Response*

The baseline estimates in Table 3 use all the variation in cluster size observed in the data to estimate the effect on inventor productivity. As discussed above, one can't necessarily interpret the baseline estimates as the causal effect of cluster size on inventor productivity. There are two main identification concerns: sorting and simultaneity. In this subsection and the next one, I present estimates that are useful in establishing the validity of the baseline model.

I begin by studying the dynamic response of productivity following a change in cluster size. Sorting into large clusters of "rising stars" is a potentially important concern. To assess the importance of this type of sorting, I study the timing of productivity changes relative to changes in cluster size. If my model is properly specified, cluster size in the future should have no effect on productivity in the current period, conditional on current cluster size. Finding that cluster size in the future is correlated with current productivity would indicate that inventors with rising productivity systematically move to larger clusters, as it would be the case if firms in large clusters can anticipate productivity growth increases and attract them.

I estimate a version of equation (1) that includes the current cluster size and five leads and five lags:

$$\ln y_{ijfct} = \sum_{s=-5}^{-1} \beta_s \ln S_{-ifc(t+s)} + \beta_0 \ln S_{-ifc(t)} + \sum_{s=1}^{5} \beta_s \ln S_{-ifc(t+s)} + d_{cf} + d_{ck} + d_{ft} + d_{kt} + d_{ct} + d_i + d_j + u_{ijfkct}, \quad (2)$$

where the five leads and five lags ($S_{-ifc(t+s)}$ for $s = -5, \ldots, -1, 1, \ldots, 5$) refer to the cluster where the focal inventor i is at time $t + s$. Note that this is not a standard event study. The coefficients on the lead terms, β_1 through β_5, allow me to determine how an inventor productivity in a given year responds to a future changes in size. The coefficients on the lag terms, β_{-5} through β_{-1}, allow me to examine how a change in cluster size propagates over time and in particular whether the effect is short lived or permanent.

In the top panel of Figure 5, I plot the coefficients β_5 through β_{-5}. The leftmost coefficient, β_5, represents the change in the focal inventor's productivity in response to a change in cluster size five years into the future. The rightmost coefficient, β_{-5}, represents the focal inventor's change in productivity in response to a change in size five years in the past. The dotted lines are 95 percent confidence bands based on standard errors clustered by city $\times$ field. The sample size is only 21,787 because for a inventor to be in this sample, the five leads and lags need to be nonmissing, which implies that only inventors observed in 11 consecutive years are included. In the bottom panel, I present the cumulative estimates corresponding to equation (2) (the impulse response function). Specifically, the figure displays $\mu_n = \beta_5 + \beta_4 + \cdots + \beta_n$ for $n = -5$ through 5, along with an accompanying error band.

To interpret this figure, suppose that a change in cluster size takes place at time $t = 0$. For example, the change could be caused by the focal inventor moving from a small cluster to a large cluster at $t = 0$. The point that is farthest to the left, μ_5, represents the productivity response five years prior to the move. The next point moving to the right, μ_4, is the estimated cumulative productivity response up to four years before the move $(\beta_5 + \beta_4)$. The point μ_{-5} is the cumulative productivity response from five years before the move through five years after the move.

Figure 5 is estimated using all the variation in cluster size in the sample—both variation coming from movers and stayers. It's a useful benchmark because my baseline estimates in Table 3 use both sources of variation. In Figure 6 I replicate the figure using only variation from inventors who move across cities. This specification has the advantage of being more easily interpretable. In particular, I use the subset of inventors who change city once to estimate a variant of equation (2), where $t = 0$ marks the time of the move, and the timing relative to an inventor's move is interacted with the corresponding cluster size. As treatment variable, I use average cluster size before the move and after the move. In other words, cluster size in the years before or after a move is measured as the average cluster size in the years before or after the move, so that within-city variation in cluster size over time is not used to identify the parameters. Thus, this specification is an event study, based on a "pure" movers design solely exploiting variation in cluster size induced by moves.

Three features of Figure 5 and 6 are worth highlighting. First, the lead terms β_1 through β_5 test for whether future values of treatment affect current productivity. If my estimates reflect a true productivity spillover, and not spurious correlation, then cluster size in the future should have no effect on inventor productivity in the current year, conditional on current cluster size. Finding positive coefficients on the lead terms would cast doubt on the causal interpretation of my estimates, because it would suggest that the productivity of inventors today depends on cluster size they will be exposed to in future years. In practice, there appears to be little evidence in Figures 5 and 6 that future values of treatment affect current outcomes. In both cases, I cannot reject that any of the coefficients on the lead terms are equal to zero. This is reassuring, because it is inconsistent with the possibility of selection on unobservables discussed above, in which promising inventors in small clusters who are experiencing productivity gains tend to be systematically hired by employers in large clusters.

Second, the estimates reveal that in the year when cluster size changes, there is an immediate rise in the focal worker's productivity. The increase in productivity may appear surprisingly fast. Recall, however, that I use the date of patent *filing*, not the

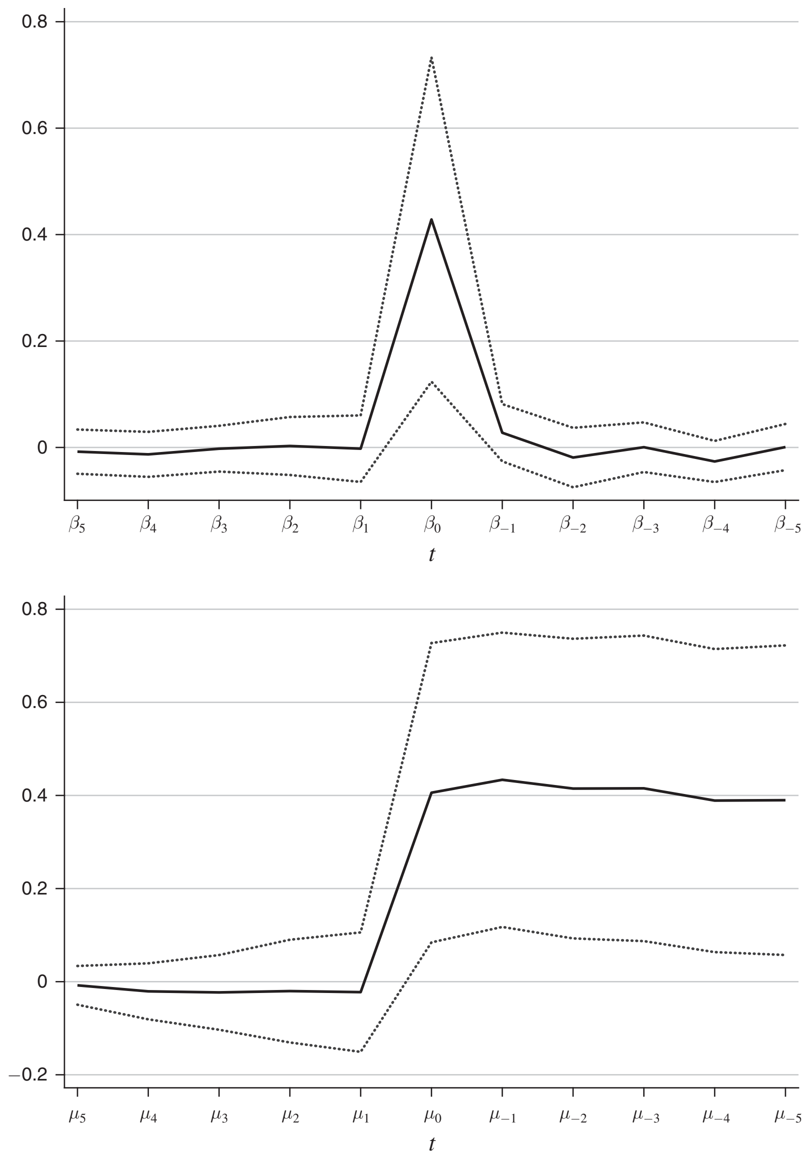

Figure 5. Dynamic Response Following a Change in Cluster Size

Notes: This figure is based on equation (2) in the text. In the top panel I plot the estimated β coefficients in equation (2) on the lag and lead terms. For example, β_5 is the coefficient on the fifth lead term. In the bottom panel, I plot the cumulative response, where the μ terms are defined as $\mu_n = \beta_5 + \beta_4 + \cdots + \beta_n$ for $n = -5$ through 5. Only inventors who are in the data for 11 consecutive years are included. 21,787 observations. Standard errors are clustered by city $\times$ research field.

date when the patent is granted. Griliches (1998) points out that in many fields the timing of patent application and R&D are close, often measured in months or even weeks. Indeed, industry studies report average length of R&D projects of less than 12 months for semiconductors, 3 to 6 months for information and communication technologies, and even shorter for software (Griss 1993, Krasner 2003, Wu 2011, Kapoor 2012, Haran 2011, Mansfield et al. 1971). In addition, there is evidence that in some cases, patents are applied for not at the end of the R&D process but at an

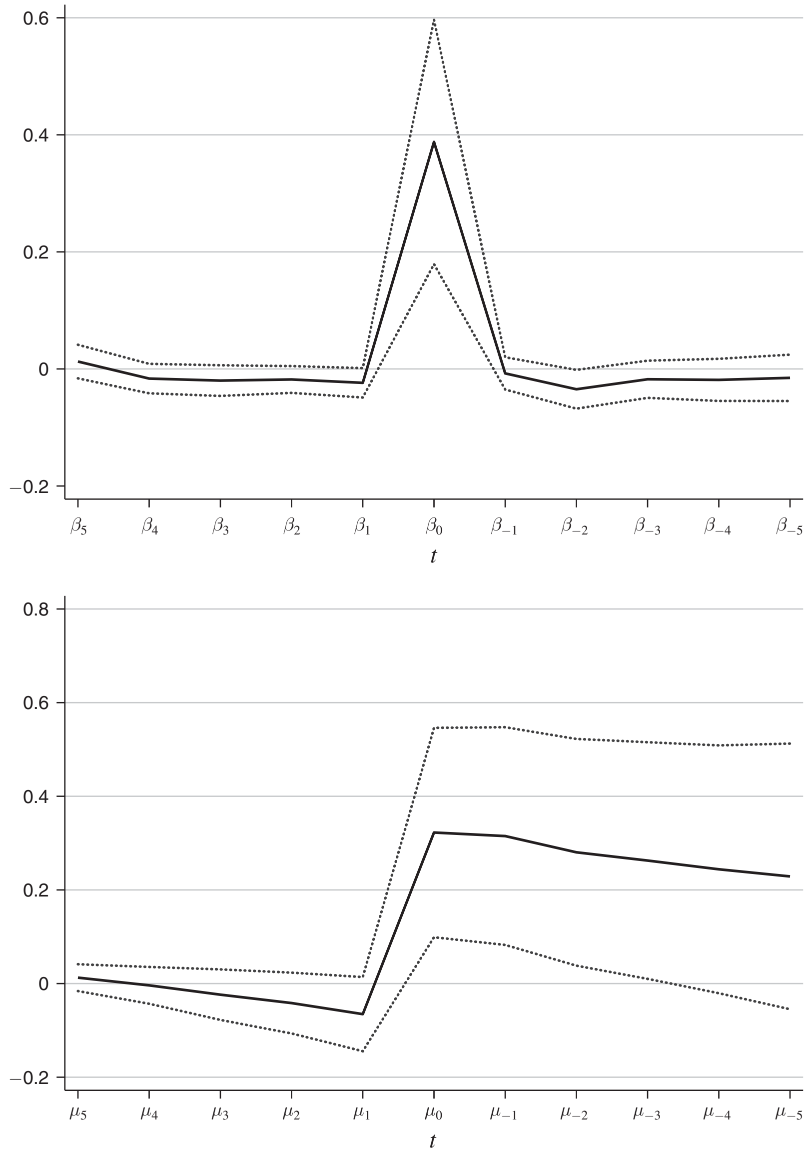

Figure 6. Dynamic Response Following a Change in Cluster Size—Movers

Notes: In the top panel I plot the estimated β coefficients on the lag and lead terms. In the bottom panel, I plot the cumulative response, where the μ terms are defined as: $\mu_n = \beta_5 + \beta_4 + \cdots + \beta_n$ for $n = -5$ through 5. Only inventors who are in the data for 11 consecutive years and move across cities are included. Standard errors are clustered by city $\times$ research field.

early stage (Cohen 2010). Previous studies have found quick effects on online patenting. For example, Hall, Griliches, and Hausman (1986) find the effect of R&D expenditures on patenting to be immediate.

In principle, one exception should be pharmaceutical R&D, where it likely takes several years and sometimes decades to invent and patent new drugs. Online Appendix Figure A.1 replicates Figure 5 limiting the sample to inventors working in technology classes for drug development ("drug, bio-affecting and body treating composition;" class code 424). While the confidence intervals are inevitably much larger, there appears to be no evidence of an immediate effect in this case.

A third feature of Figures 5 and 6 worth highlighting is that the effect appears, in large part, persistent. In Figure 5, after the initial increase in the first two years, productivity declines slightly over the next three years, as inferred from the downward drifting cumulative response in the bottom panel, but most of the effect persists. In the case of movers in Figure 6, there is more decline, but even here most of the effect persists five years after the move.[22]

From the quantitative point of view, the estimated effects in Figures 5 and 6 appear larger than the baseline results in Table 3. The reason is that these figures are based on a sample that includes inventors observed for 11 consecutive years. By construction, this sample includes only the most prolific inventors. The effect of cluster size is particularly large for superstar inventors, as I show in Section IVB. There, I present estimates for increasingly stringent definitions of star inventors and find that the effect is largest when I restrict the sample to inventors in the top 1 percent or 0.5 percent of lifetime patent count, which accounts for the vast majority of the sample used in Figures 5 and 6.

C. *Instrumental Variable Estimates*

In this subsection, I consider models based on two IVs. While the baseline estimates in Table 3 use all the variation in cluster size observed in the data to estimate the effect on productivity, the IV estimates isolate variation in cluster size that comes from specific and arguably exogenous sources. This approach is useful to deal with simultaneity, namely the possible existence of unobserved time-varying productivity shocks at the city-field level that are correlated with variation in cluster size.

Rochester.—First, I revisit the Rochester case study. The difference-in-difference estimates uncovered in Section II can be interpreted as reduced form estimates in a model where the excluded instrument is the Rochester $\times$ 2007. The two-stage least squares (2SLS) estimates can be obtained by rescaling the reduced form estimates by the relevant first stage estimates. The identifying assumption is that the change in the number of inventors in Rochester caused by the demise of Kodak after 1996 is uncorrelated with unobserved productivity shocks of non-Kodak inventors outside the photography sector, conditional on controls.

The first stage estimates are presented in the top panel of Table 4. They are obtained by estimating models similar to those in Table 2 (panel A), where the dependent variable is the log of cluster size. Unweighted estimates range from -0.491 (0.128) in column 1 to -0.396 (0.0460) in column 4. This last coefficient, for example, indicates that in a model that conditions on field $\times$ year and field $\times$ city effects, the cluster size in Rochester declined by 39.6 percent relative to other cities. (As discussed above, this is the mean across the five fields in Rochester.)

[22] Glaeser and Mare (2001) find that the effect of city size on wages manifests itself over time. De La Roca and Puga (2017) find that Spanish workers in bigger cities obtain an immediate static wage premium and also accumulate valuable experience over time. They discus the biases that arise if the benefits of bigger clusters take time to realize.

2SLS estimates are shown in the bottom panel. They are the ratio of the reduced form coefficients (Table 2) and the corresponding first stage coefficients. The unweighted 2SLS estimates of the elasticity of inventor productivity with respect to cluster size range from 0.131 (0.037) in column 1 to 0.232 (0.0393) in column 4.

Firm Spatial Networks.—Next, I use the geographical structure of firms with multiple locations to build an IV that isolates variation in local cluster size that originates elsewhere. The idea is that changes over time in the number of inventors employed in other cities by firms other than the focal inventor's firm but that have a presence in the focal inventor's city and field are predictive of changes in the local cluster size and are unlikely to be systematically correlated with unobserved shocks to the focal inventor's productivity. Specifically, let $N_{jf(-c)t}$ be the number of inventors that firm j has in field f, year t in all the cities excluding city c, so that $\Delta N_{jf(-c)t} = N_{jf(-c)t} - N_{jf(-c)(t-1)}$ is the change between year $t-1$ and t. The IV for inventors in firm j in cluster fct is defined as

$$IV_{jfct} = \sum_{s \neq j} D_{sfc(t-1)} \frac{\Delta N_{sf(-c)t}}{\Delta N_{ft}},$$

where $D_{sfc(t-1)}$ is an indicator equal to one if firm s has at least one inventor in city c in field f in year $t-1$, and ΔN_{ft} is the nationwide change in inventors in the field. Note that the summation is across all firms that have a presence in the city *excluding* the focal firm j.

Identification comes from changes over time in the number of inventors employed in other cities by local firms besides the focal inventor's own firm. To see the intuition, consider as an example Boston and Minneapolis. In 1996, Microsoft has a presence in Boston, with three inventors in the computer science field. Between 1996 and 1997 Microsoft is expanding its overall R&D investment in the United States and the total number of its inventors in all cities outside Boston is growing. Now take a computer scientist in Boston employed by a firm other than Microsoft. The first stage captures whether the size of the cluster that this inventor is exposed to in Boston increases as the number of Microsoft computer science inventors outside Boston increases between 1996 and 1997. The idea is that the growth of Microsoft inventors outside Boston may affect the size of the cluster that non-Microsoft inventors in Boston are exposed to but it is arguably not correlated with productivity shocks of non-Microsoft inventors in Boston, since it is driven by Microsoft's growth elsewhere.

By contrast, consider Minneapolis in 1996, where Kodak had a presence, with two inventors in the computer science field. Between 1996 and 1997, Kodak's total number of inventors in all cities outside Minneapolis is declining, since the firm is struggling due to rising competition from digital photography. For a computer scientist in Minneapolis employed by a firm other than Kodak, the instrument will likely capture a decrease in cluster size. The decrease arguably reflects Kodak's overall demise rather than unobserved productivity shocks in Minneapolis. (In practice, the instrument is not based only on one firm per city, but instead reflects the sum across all firms that have a presence there. It also includes all years, not just 1996 and 1997).

More explicitly, the assumption is that for focal inventor i, variation in the number of inventors in i's field who are located outside i's city and work for firms other than i's firm that have a presence in i's city is orthogonal to unobserved factors

Table 4—First Stage and 2SLS Estimates: Non-Kodak Inventors in Rochester Compared to Other Cities

					Weighted
	(1)	(2)	(3)	(4)	(5)
First stage					
Rochester × 2007	−0.491	−0.462	−0.449	−0.396	−0.385
	(0.128)	(0.113)	(0.117)	(0.0460)	(0.0466)
F-statistic	14.63	16.59	14.83	73.89	68.19
Second stage					
Cluster size	0.131	0.145	0.179	0.232	0.247
	(0.0372)	(0.0372)	(0.0501)	(0.0393)	(0.0469)
Rochester	Yes	Yes	Yes	Yes	Yes
2007	Yes	Yes	Yes	Yes	Yes
Field		Yes	Yes	Yes	Yes
Field × year			Yes	Yes	Yes
Field × city				Yes	Yes

Notes: Each entry is a separate regression. The dependent variable in the first stage is log cluster size. The dependent variable in the second stage is log number of patents in a year. Entries in the second stage are the ratio of reduced form estimates in the first row of Table 2 divided by the corresponding first stage estimates of this table. The level of observation in the regressions is inventor-year. The sample includes all non-Kodak inventors in the years 1996 and 2007, excluding those employed in the photography sector. Standard errors are clustered by city, in parentheses.

that affect i's productivity, after conditioning on covariates. In the Boston example, the assumption is that conditional on the controls, unobserved productivity shocks experienced by an inventor in computer science in Boston who is not employed by Microsoft are orthogonal to changes in the number of computer scientists who work for Microsoft outside Boston. Since the econometric model includes field × year and class × year effects, identification is not driven by the sectoral mix of employers in a city. Rather, it is driven by the identity of the firms that exist in a cluster at $t-1$, other than the one where the focal inventor is employed, and by changes in their number of inventors outside the focal inventor cluster.

One limitation is that this instrument predicts changes in cluster size, not its level. I estimate a version of equation (1) in first differences, precluding a direct comparison with the baseline models:

$$(3) \qquad \Delta \ln y_{ijfkct} = \alpha \Delta \ln S_{-ifct} + d_t + d_f + d_k + d_j + d_{ft} + d_{kt} + u_{ijfkct}.$$

The top panel of Table 5 reports OLS estimates of equation (3). The sample includes inventors who are observed in two consecutive years. Standard errors in this table are clustered by city. The OLS estimates range between 0.0141 (0.00394) and 0.0164 (0.00397), depending on the set of controls, and are smaller that the corresponding models in levels in Table 3. This is in part due to the fact that relative to fixed effects models, models in first differences magnify measurement error biases (Griliches and Hausman 1986). Moreover, models in first differences only estimate the contemporaneous effect of cluster size. First differences models that include the contemporaneous change in log size and lagged changes yield larger long run estimates.[23]

[23] For example, in a model that has the contemporaneous change in log size and two lagged changes, and all the same controls as column 6, the sum of the three coefficients is 0.031 (0.008), larger than the entry in the table although still smaller than the corresponding estimates in levels.

Table 5—Models in Differences: Effect of Changes in Cluster Size on Changes in Inventor Productivity: OLS and IV Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. OLS						
Δlog size	0.0141	0.0145	0.0153	0.0164	0.0162	0.0159
	(0.00394)	(0.00392)	(0.00376)	(0.00397)	(0.00392)	(0.00385)
Panel B. 2SLS						
Δlog size	0.0422	0.0630	0.0502	0.0496	0.0502	0.0491
	(0.0186)	(0.0211)	(0.0189)	(0.0131)	(0.0137)	(0.0144)
First stage	1.109	1.076	1.096	1.431	1.475	1.488
	(0.151)	(0.170)	(0.167)	(0.214)	(0.189)	(0.185)
F-statistic	53.8	40.2	43.0	44.5	60.8	64.2
Observations	419,596	419,596	419,565	405,111	405,111	403,955
Year	Yes	Yes	Yes	Yes	Yes	Yes
Field		Yes	Yes	Yes	Yes	Yes
Class			Yes	Yes	Yes	Yes
Firm				Yes	Yes	Yes
Field × year					Yes	Yes
Class × year						Yes

Notes: Each entry is a separate regression. Dependent variable is the change in the log number of patents in a year. The model estimated is equation (3). The IV for workers in firm j in cluster fct is defined as $IV_{jfct} = \sum_{s \neq j} D_{sfc(t-1)} \left(\Delta N_{sf(-c)t} / \Delta N_{ft}\right)$ where $D_{jcf(t-1)}$ is an indicator equal to one if firm j has at least one inventor in city c in field f in year $t-1$. $N_{jf(-c)t}$ is the number of inventors that firm j has in field f, year t in all the cities excluding city c; and $\Delta N_{jf(-c)t} = N_{jf(-c)t} - N_{jf(-c)(t-1)}$ is the change in $N_{jf(-c)t}$ between time $(t-1)$ and t. Standard errors are clustered by city.

The bottom panel reports 2SLS estimates and the corresponding first stage estimates. The F-statistics are between 40.2 and 64.2. The 2SLS estimates range from 0.0422 (0.0186) to 0.0630 (0.0211). The entry in column 6 is equal to 0.0491 (0.0144). The reason why IV is larger than OLS is that it corrects both for measurement error and endogeneity.

To assess the validity of the instrument, I re-estimated both the first stage and the second stage using as IV the one-year lead. The idea is that a future shock to cluster size predicted by a future change in the IV should have no effect on productivity or cluster size in the current period. In both cases, I find statistically insignificant coefficients, suggesting that future values of the IV do not predict changes in current cluster size or changes in current productivity.[24]

By construction, variation in the instrument comes from firms that have a presence in more than one city. Variation in the focal inventor's own firm is excluded to minimize the likelihood that unobserved shocks to the focal inventor might be correlated with the instrument. To push the leave-out logic further, I re-estimate my models using the same instrument but excluding from the regressions firms that have a presence in more than one city. In particular, estimates in online Appendix Table A.4 are based on the sample of inventors who work in firms that in every year in which they appear in the data are present in only one city. The IV estimates are larger, but less precisely estimated.

[24] The coefficients for the first and second stage are −0.039 (0.032) and −0.229 (0.609), respectively.

While I can't completely rule out the possible existence of unobserved time-varying productivity shocks at the city-field level that are correlated with variation in cluster size, the IV estimates based on the Rochester shock and the ones based on firm spatial networks, taken together, appear to allay concerns about simultaneity.

IV. Citations, Heterogeneity, and Robustness

A. *Citations Received and Citations Made*

To understand the effect of cluster size on the quality of inventor output, not just its quantity, I now focus on patent citations. The dependent variable in panel A of Table 6 is the log number of subsequent patents that cite any patent filed by inventor i in year t. The citing patents include any patent filed between year t and the end of the sample, not just patents filed in t. The dependent variable in panel B is the log of number of subsequent patents that cite patents filed by inventor i in year t divided by the number of patents filed by inventor i in year t. The former is a measure of the overall impact of patents produced by an inventor in a given year, while the latter is a measure of the mean quality of patents filed by an inventor in a given year. Estimates in column 6 suggest that the elasticities for the overall number of citations and citations per patent are equal to 0.160 (0.043) and 0.092 (0.041), respectively. This means that inventors in larger clusters produce not just more patents but also more influential patents compared to otherwise similar inventors in smaller clusters. Estimates in the last two columns indicate that the increase in citations is largely driven by an increase in local citations, defined as citations coming from patents created by inventors in the same city as the focal inventor (column 7), as opposed to patents created by inventors in a different city (column 8).[25]

To shed some light on the channels that may lead to the productivity gains enjoyed by scientists in larger clusters, I now turn to an analysis of citations *made* by focal inventors—namely citations of previous patents included in the focal inventor patents. The dependent variable in columns 1 and 2 in Table 7 is the log of the total number of citations by the focal inventor in the focal year, and the log of the total number of citations per patent filed, respectively. Entries indicate that inventors in larger clusters tend to cite more than inventors in smaller clusters. This is true both of the overall number of citations made (column 1)—which is probably not surprising given that inventors in larger clusters create more patents—but also of the number of citations per patent (column 2). Since citations made by an inventor are arguably a sign that the inventor knows about a specific innovation, this finding

[25] I have also estimated the effect of cluster size on measures of patent "generality" and "originality" based on citations. The measures, first proposed by Trajtenberg, Jaffe, and Henderson (1997), are between zero and one. If a patent is cited by subsequent patents that belong to a wide range of fields the measure of "generality" is close to one, whereas if most citations are concentrated in a few fields it is close to zero. "Originality" is defined the same way, except that it refers to citations made (Hall, Jaffe, and Trajtenberg 2001). Thus, if a patent cites previous patents that belong to a narrow set of technologies the originality score is close to zero, whereas citing patents in a wide range of fields would render a score close to one. More specifically, generality is defined as $1 - \sum_k p_{ik}$ where p_{ik} denotes the percentage of citations received by patent i that belong to patent class k. The sum is the Herfindahl concentration index. "Originality" has the same definition, except that p_{ik} denotes the percentage of citations made by patent i that belong to patent class k. The estimates (available on request) indicate that inventors in larger clusters do not tend to produce patents that are more original or general than inventors in smaller clusters.

Table 6—Effect of Cluster Size on Inventor Productivity: Patent Citations Received by Focal Inventor

	(1)	(2)	(3)	(4)	(5)	(6)	Same city (7)	Different city (8)
Panel A. Dependent variable: number of citations								
log size	0.195	0.205	0.230	0.209	0.168	0.160	0.193	0.0194
	(0.0938)	(0.0448)	(0.0392)	(0.0297)	(0.0419)	(0.0438)	(0.0786)	(0.0302)
Panel B. Dependent variable: number of citations per patent								
log size	0.107	0.114	0.162	0.117	0.114	0.0927	0.126	−0.0476
	(0.0812)	(0.0435)	(0.0372)	(0.0272)	(0.0396)	(0.0411)	(0.0740)	(0.0283)
Observations	932,059	932,059	932,059	932,059	932,059	823,375	730,283	730,283
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes	Yes	Yes
City × year					Yes	Yes	Yes	Yes
Firm						Yes	Yes	Yes

Notes: The dependent variable is log of patent citations (panel A) or log of citations per patent (panel B). In particular, panel A shows estimates where the dependent variable is the log number of patents that cite any patent filed by inventor i in year t, where the citing patent may be filed in any year between t and the end of the sample. Panel B shows estimates where the dependent variable is the log of number of subsequent patents that cite patents filed by inventor i in year t divided by the number of patents filed by inventor i in year t. The model estimated is equation (1). Column 7 is for citations by inventors located in the same city as the focal inventor. Column 8 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

is consistent with the idea that scientists in larger clusters have more knowledge of existing innovations than otherwise similar scientists in smaller clusters, possibly because larger clusters allow for more knowledge diffusion than smaller clusters.

The dependent variable in column 3 is the share of citations made by the focal inventor to inventors located in the same city. The entry indicates that larger clusters are associated with a larger share of local citations, suggesting that scientists in larger clusters not only have more overall knowledge of existing innovations, but that this is particularly true of local innovations. This is consistent with the notion that larger clusters allow for knowledge and ideas to spread more efficiently, as suggested, among others, by Saxenian (1994), who describes how in Silicon Valley ideas flow very fluidly between innovators and this fosters their creativity.

The dependent variable in column 4 is the share of citations from the focal inventor to inventors in the same field. The estimated elasticity is only marginally significant, precluding definitive conclusions.

B. *Is the Effect Larger for Larger Clusters or More Productive Firms or More Productive Inventors?*

I now examine three potentially important sources of heterogeneity in the magnitude of the spillover effect.

Table 7—Citations by Focal Inventor

	Citations (1)	Citations per patent (2)	Share same city (3)	Share same field (4)
log size	0.158	0.0899	0.00752	0.00553
	(0.0158)	(0.0137)	(0.00268)	(0.00330)
Observations	810,495	810,495	810,495	810,495

Notes: Each column is a separate regression. The level of observation in the regressions is inventor-year. Standard errors are clustered by city × research field. Models include year, city, field, class, city × field, city × class, field × year, class × year, inventor, city × year, and firm effects.

Estimates by Cluster Size.—My baseline estimates are based on a log-log specification that assumes a constant elasticity. This is a natural starting point, but in reality, it is possible that the elasticity varies depending on cluster size. On the one hand, it is in principle possible that the elasticity in large clusters is lower than the elasticity in small clusters, so that a 1 percent increase in size in a large cluster results in productivity gains that are proportionally smaller than a 1 percent increase in size in a small cluster. This would be the case, for example, if size increases above a certain threshold yield limited productivity benefits. On the other hand, the opposite could also be true—namely, that the elasticity grows with size. This could happen, for example, if a few extra inventors in a tiny cluster have limited productivity benefits, and agglomeration economies begin to materialize only above a certain size. It's even possible that both cases are true, and that the relationship between log productivity and log cluster size is best represented by an S curve, like many network models with social interaction would suggest. Ultimately, this is an empirical question.

In Figure 4, the relationship between log productivity and log cluster size appears visually slightly convex, after controlling for a limited set of controls, more consistent with an elasticity that grows with cluster size. To test more formally whether the estimated elasticity is constant across clusters of different size, panel A in Table 8 shows estimates where I allow the coefficient on log size to vary across size quartiles and the vector of controls includes all the controls that are included in the baseline model. The coefficients in the table give the elasticity of productivity with respect to size for the relevant size. This specification allows me to determine whether the effect of log cluster size on log productivity in large clusters is different from the effect of log cluster size on log productivity in small clusters.

Estimates in column 1 condition on all the effects excluding firm effects. They appear to quantitatively similar, ranging from 0.0556 (0.0114) to 0.0633 (0.0137). A test for equality, reported at the bottom of the table, has p-value equal to 0.157, indicating that the coefficients are statistically not different from each other. In column 2, I add firm effects. Here, the p-value for the test for equality is 0.080. The elasticities remain quantitatively similar. The smallest is the one for the first quartile, which is equal to 0.0702 (0.0137), while the largest is the one for the fourth quartile, which is equal to 0.0821 (0.0163) and the relationship is not monotonic.

Overall, I conclude that there is limited evidence of large heterogeneity in elasticities. This is consistent with the findings for the manufacturing sector by Kline and

Table 8—Heterogeneity in Elasticity by Cluster Size and by Firm Productivity

	(1)	(2)
Panel A. Heterogeneity by cluster size		
First quartile (smallest)	0.0556	0.0702
	(0.0114)	(0.0137)
Second quartile	0.0596	0.0760
	(0.0120)	(0.0143)
Third quartile	0.0574	0.0743
	(0.0123)	(0.0148)
Fourth quartile (largest)	0.0633	0.0821
	(0.0137)	(0.0163)
Observations	932,059	823,375
p-value equal elasticities	0.157	0.080
Firm		Yes
Panel B. Heterogeneity by firm productivity		
First quartile (least productive)	0.0539	0.0583
	(0.0122)	(0.0142)
Second quartile	0.0568	0.0731
	(0.0117)	(0.0140)
Third quartile	0.0534	0.0713
	(0.0116)	(0.0145)
Fourth quartile (most productive)	0.0558	0.0744
	(0.0118)	(0.0141)
Observations	932,059	823,375
p-value equal elasticities	0.785	0.003
Firm		Yes

Notes: Dependent variable is log number of patents in a year. Each column within a panel is a separate regression. Standard errors are clustered by city × research field. Models include year, city, field, class, city × field, city × class, field × year, class × year, inventor, and city × year effects. Column 2 also includes firm effects.

Moretti (2014a). They find that the elasticity of manufacturing productivity with respect to density of labor in a county is similar in counties with low density and high density.

Estimates by Firm Productivity.—A separate question is whether the elasticity of productivity with respect to cluster size depends on firm productivity. It's in principle possible that the most productive firms are the ones most able to take advantage of productivity spillovers. Indeed, recent contributions on firms sorting and agglomeration (Gaubert 2018, Combes et al. 2012) are built on an assumption that the most highly productive firms gain the most from agglomeration. On the other hand, it is also possible that the opposite is true, namely that productivity spillovers are quantitatively more important for low productivity firms. This would be the case, for example, if inventors in low productivity firms are the ones with the most to gain from proximity to others inventors.

Panel B in Table 8 tests whether estimates of the effect of cluster size on productivity depend on firm productivity. Firms are divided in four quartiles, based on their average productivity, and a separate elasticity is estimated for each quartile by interacting cluster size with four indicators for each quartile.[26] Estimates in column 1 indicate that the spillover effect is very similar across quartiles. When firm fixed effects are added in column 2, the point estimates show a weak relationship with firm productivity. Statistically, I can reject that the four elasticities are identical, as shown by the p-value reported at the bottom of the table. But quantitatively, the difference is limited. The elasticity estimated for the top quartile, 0.0744 (0.141), is only slightly larger than the elasticity for the bottom quartile, 0.0583 (0.0142).

I conclude that there is some evidence of differences in the spillover effect as a function of firm productivity, but the differences are economically small.

Estimates by Inventor Productivity.—Does the elasticity of productivity with respect to cluster size depend on inventor productivity? On the one hand, it's possible that the most productive inventors are the ones most able to take advantage of productivity spillovers. On the other hand, it is also possible that the opposite is true, and that relatively weaker inventors are the ones with the most to gain from proximity to others inventors.

Table 9 reports estimates based on different cutoffs for the definition of star inventors. Column 1 includes all inventors, while columns 2 to 6 include increasingly stringent definitions of star inventors based on the total number of patents over their lifetime. The estimated coefficients indicate that estimates are larger if more stringent definitions of stars are adopted. In particular, the estimate for the top 0.5 percent of inventors (column 6) is larger than the estimate for the top 1 percent (column 5), which in turn is larger than the estimate for the top 5 percent (column 4), and so on. The estimate for the full population of inventors (column 1) is the smallest and it is indistinguishable from zero.

I conclude that the most productive inventors are the ones who benefit the most from agglomeration economies. This suggests an interesting form of complementarity between inventor quality and cluster size.

C. *Robustness*

Cross-Field Spillovers.—The baseline model in equation (1) assumes that the productivity spillovers enjoyed by scientists in a given research field only depend on the size of the cluster in that field, and do not depend on the size of clusters in other fields. In reality, cross-field spillovers could be empirically important if scientists in a city benefit from spillovers stemming not only from their own research field, but also from other research fields. It is not hard to imagine that researchers in semiconductors, for example, might benefit from learning about new discoveries in computer science. Strong cross-field spillovers may help explaining the tendency of firms belonging to different parts of the high-tech sector to locate near each other, as observed in cities like San Francisco, Boston and Seattle. The

[26] Firm productivity is measured as mean inventor productivity by firm and research field across all years. Thus, if a firm has inventors working in different fields, there are different measures of productivity depending on the field.

Table 9—Alternative Definitions of Star Inventors

	All (1)	Top 25 percent (2)	Top 10 percent (baseline) (3)	Top 5 percent (4)	Top 1 percent (5)	Top 0.5 percent (6)
log size	0.0182	0.0371	0.0669	0.102	0.250	0.274
	(0.0117)	(0.0116)	(0.0139)	(0.0176)	(0.0403)	(0.0424)
Observations	2,360,153	1,349,320	826,761	528,833	156,051	132,445

Notes: Dependent variable is log number of patents in a year. Baseline is the entry in column 8 of Table 3. Models include year, city, field, class, city × field, city × class, field × year, class × year, inventor, city × year, and firm effects. Standard errors are clustered by city × research field.

existence and magnitude of cross-field spillovers also have important implications for our understanding of the nature and scope of knowledge spillovers.[27]

To assess how important cross-field spillovers are, I estimate a model that includes not only cluster size based on the focal inventor's own research field, S_{-ifct}, but also the mean cluster size of the other four fields in the relevant city and year. Column 1 in online Appendix Table A.5 shows that the coefficient on the mean cluster size of the other fields is not statistically different from zero, suggesting that on average cross-fields spillovers are not very important.

It is still possible that cross-field spillovers are important for specific pairs of fields. In columns 2 to 6, I estimate a more general version of equation (1) that includes five cluster sizes, one for each of the main research fields. This allows me to estimate a 5 × 5 matrix where each off-diagonal element is the effect of a field on another field. The diagonal elements are the own-field effect. The matrix is shown in columns 2 to 6.[28] Empirically, most of the cross-field effects are not statistically different from zero. One exception is the semiconductor field in column 6. The productivity of inventors in this field depends not only on the size of the local semiconductor cluster, but also on the size of the local clusters in computer science and other science. The cross-field effects are 0.138 (0.0650) and 0.196 (0.0585)—large compared with the own-field effect estimate of 0.218 (0.0644)—suggesting that semiconductor scientist productivity is highly sensitive to the presence of scientists in these two other fields.[29]

Sample Selection.—In Section IC, I noted that when inventors don't apply for a patent in a given year, they are missing from my data because their location is unknown. I argued that a regression of number of patents on cluster size captures the intensive margin—namely the effect of cluster size on number of patents, given a

[27] Patent citations seem to suggest that cross-field spillovers may be important: the fraction of patents cited by scientists in my sample that are in one of the four fields different from their own is 53.3 percent.

[28] Models in this table do not include city × year effects to avoid multicollinearity.

[29] To test whether the productivity spillovers enjoyed by an inventor depend on her intellectual linkages to research fields other than her own, as measured by an inventor's propensity to cite patents outside her own field, I estimate the baseline model augmented by the interaction of cluster size with an indicator for whether the focal scientist share of citations to her own field is above 90 percent. The coefficient on the interaction is negative, indicating that scientists who tend to cite patents only in their own field enjoy smaller productivity spillovers, possibly because they are less open to learning about other fields.

positive number of patents—but misses the extensive margin—namely the effect of size on the probability of patenting. As a consequence, the elasticities estimated so far should be interpreted as a lower bound of total effect of cluster size on patenting. Here I present two pieces of evidence intended to empirically probe the direction and magnitude of the sample selection bias.

Online Appendix Table A.6 shows what happens when I use interpolation to impute some of the inventor-year pairs that are missing due to lack of patenting by an inventor in a given year. I expect interpolation to result in larger estimates. By including some of inventor-year pairs when an inventor does not patent, estimates in the interpolated sample reflect not only the effect of size on number of patents (intensive margin), but also part of the effect of cluster size on the probability of patenting (extensive margin).

Since the dependent variable is in logs, and log(0) is undefined, I set it equal to the inverse hyperbolic sine (panel A) or log(number of patents + 1) (panel B). Column 1 uses the baseline sample. In column 2, I interpolate the data when one missing year is immediately preceded and followed by nonmissing years. If a scientist is missing in year t but observed in years $t - 1$ and $t + 1$, she is assigned to the cluster in which she is at $t - 1$. For example, this would be the case of an inventor observed patenting in 2003 and 2005, and not observed in 2004 due to the lack of patenting. In this case, I assume that the inventor is located in 2004 in the same cluster as the one where I observe her in 2003. The sample increases to 860,806 observations, indicating that 37,431 observations are interpolated.[30] In column 3, I interpolate the data when two missing years are immediately preceded and followed by nonmissing years. For example, this would be the case of an inventor observed patenting in 2003 and 2006, and not observed in 2004 and 2005. In this case, I assume that the inventor is located in 2004 and 2005 in the same cluster as the one in 2003. The sample increases to 873,346 observations.

The estimated coefficients in column 2 are larger than the corresponding coefficients in column 1. Intuitively, by adding some of the missing zeros, the estimate in column 2 reflects not only the intensive margin, but also part of the effect of the extensive margin. The coefficients in column 3 are even larger because an even larger share of the extensive margin is captured.

Online Appendix Table A.7 reports estimates obtained with different temporal units of analysis. Specifically, I re-estimate my baseline models using measures of productivity defined over one month, two months, three months, six months, two years, and three years. I expect that when the temporal unit of analysis is short (months), the problem of sample selection and the downward bias are more pronounced. In the extreme, if I were to measure productivity second by second, very few inventor-second pairs would be nonmissing and the selection bias would be large. By contrast, when the temporal unit of analysis is long (two or three years), I expect the problem of sample selection and the downward bias to be less pronounced. In the extreme, if I were to have just one observation per inventor with productivity defined as the number of patents created in all the years in the sample, there would be no selection and both the intensive margin and extensive margin

[30] When I do this interpolation, I do not change the measure of cluster size on the right-hand side.

would be reflected in my estimates. Empirically, panel A confirms that the estimated coefficient is monotonically increasing with the length of the unit of analysis. In columns 1 and 2, it is negative. In column 3 it is close to zero. In column 4 it is positive, although one-half of the size of the baseline coefficient. In columns 6 and 7, the estimated coefficient is significantly larger than the baseline coefficient. Panel B repeats the exercise for inventors in the top 1 percent of lifetime patent count. It shows that even for the most prolific inventors the problem of the missing zeros is pronounced when the unit of time is only one month, since even the most prolific inventors rarely have patents every single month. As expected, the problem declines the longer the temporal unit.

Overall, findings in online Appendix Tables A.6 and A.7 confirm that the baseline estimates capture only part of the overall effect—namely the intensive margin—and that including the extensive margin leads to larger estimates. When I interpolate the data to include some of the zeros, the estimated effects increase precisely because part of the extensive margin gets included in the estimates. Similarly, when I focus on long units of time, my estimated effects grow because part of the extensive margin gets included in the estimates. I conclude that the baseline estimates in Table 3 should be interpreted as a lower bound of the true effect of cluster size on productivity.

Quality of the Cluster.—It is possible that productivity spillovers depend not just on the overall size of a cluster, but also on the quality of the inventors in that cluster.[31] To investigate this possibility, I estimate models where the size of cluster is defined as a function not just of the number of inventors, but also of their quality, where quality is measured by the number of patents created, or the number of citations received. In column 1 of online Appendix Table A.8, cluster size is measured as the weighted sum of inventors in a given city-field-year cell, with weights reflecting the lifetime number of patents of each inventor. In column 2, cluster size is measured as the number of inventors with a lifetime patent count above three. In column 3, cluster size is measured as the weighted sum of inventors in a given city-field-year cell, with weights reflecting the lifetime number of patent citations received. Finally, in column 4, cluster size is measured as the number of inventors with a lifetime patent citation received count above five. Compared with the baseline elasticity in Table 3 column 8, the elasticities in columns 1 to 4 are about twice as large, suggesting that local spillovers from high quality scientists have a significantly larger impact on productivity than spillovers from the average scientist.

Teams.—Increasingly, innovation is created by teams of inventors working together. If larger teams are both more productive and more likely to be in larger clusters, team size could be an important omitted variable. (Note that there is no mechanical relationship between team size and productivity, since my measure of productivity is already adjusted for team size: in case of patents with multiple inventors, each of them receives a fraction of the patent.) In my data, I define teams as a group of inventors whose names are on the same patent. In column 5 of online Appendix

[31] For example, Iaria, Schwarz, and Waldinger (2018) show that access to the very upper tail of scientists is crucial for scientific output.

Table A.8, I control for a quadratic in team size. The coefficient is 0.118 (0.0118), significantly larger than the corresponding baseline coefficient, suggesting that team size is negatively correlated with inventors' unobserved productivity determinants. In column 6, cluster size is defined excluding all members of the focal inventor's team. The estimated elasticity is larger than the baseline elasticity, confirming that if anything, omitting team size biases the results downward. In the last 2 columns, I ask whether the spillover effect is larger for larger teams. For each inventor, I compute the mean team size in the relevant year. I estimate models where cluster size is interacted with an indicator for teams with size above median (column 7) or an indicator for solo inventors (column 8).[32] The positive coefficient on the interaction in column 7 and the negative coefficient on the interaction in column 8 indicate that the spillover effect is larger for larger teams and smaller for solo inventors

V. Implications of Agglomeration for the Aggregate Production of Innovation

I found that inventors tend to cluster geographically in a small number of areas and that inventors who locate in large clusters enjoy productivity gains relative to those in small clusters. A natural question is therefore how much geographical clustering contributes to the overall production of innovation in the United States. Specifically, is the total number of patents produced each year in the country made larger by the fact that inventors in each field concentrate in a handful of locations, compared to the case where inventors are spread more equally across locations?

In this section, I use my estimates of the elasticity of inventor productivity with respect to cluster size to quantify the macroeconomic benefits of agglomeration for the United States as a whole. I seek to estimate what would happen to the total number of patents produced annually in each field in the United States if inventor quality and firm quality did not change but some inventors were spatially reallocated from large clusters to small clusters up to the point where cluster size within each field is equalized across cities. In the presence of productivity spillovers, one would expect that such spatial redistribution would increase the productivity of inventors in clusters smaller than average and lower the productivity of inventors in clusters larger than average. On net, the magnitude of the aggregate effect for the country as a whole of such spatial redistribution depends on the relative magnitude of the gains in smaller clusters compared to the losses in larger clusters. In turn, this depends on the strength of agglomeration economies and how unequal is the initial spatial distribution.

Complete equalization of cluster size is of course an extreme counterfactual. It is intended to be a useful benchmark to assess the aggregate benefits of agglomeration, rather than a specific policy objective. At the end, I provide an additional estimate based on partial equalization.

[32] More precisely, for each inventor, I computed the share of solo patents in the relevant year, defined as patents with only one inventor. Column 8 is based on an indicator equal to one if the focal inventor's share of solo patents in the relevant year is above 0.9.

A. *How Spatial Agglomeration Affects the Aggregate Number of Patents*

To see how spatial agglomeration might affect the aggregate number of patents produced in the United States, consider the following simplified example. Assume there are only two clusters, A and B, and cluster A is initially larger: $S_A > S_B$. Assume that inventor i's output is only a function of cluster size S_c: $y_{ic} = g(S_c)$, with $g' > 0$. (In equation (1) there are of course many other terms that affect an inventor's productivity which I ignore in this example to keep notation as simple as possible.) Aggregate output in this economy is the sum of output in location A and location B: $Y^1 = S_A g(S_A) + S_B g(S_B)$, where the total number of patents produced in a location is simply the product of inventor output times the number of inventors in that location.

Consider a counterfactual where some inventors are moved from cluster A to B so that the number of inventors is equalized: $S_A = S_B = S/2$. The total number of inventors in the economy, S, does not change, but cluster A becomes smaller and cluster B larger. This could happen, for example, through the provision of subsidies in B. Aggregate output in this counterfactual is $Y^2 = Sg(S/2)$. The change in the aggregate output is

$$Y^2 - Y^1 = S_A\left[g\left(\frac{S}{2}\right) - g(S_A)\right] + S_B\left[g\left(\frac{S}{2}\right) - g(S_B)\right]. \tag{4}$$

The first term is the change in number of patents in A: it is the product of the change in inventor productivity in A times the initial number of inventors in A. This term is negative because cluster A has become smaller, and as a consequence the change in inventor productivity is negative: $\left[g(S/2) - g(S_A)\right] < 0$. By contrast, the second term, which measures the change in number of patents in B, is positive because B has become larger and therefore $\left[g(S/2) - g(S_B)\right] > 0$.

The effect of redistribution on aggregate output depends on the magnitude of the output losses in the cluster that was initially larger, A, and the output gains in the cluster that was initially smaller, B. Equation (4) clarifies that the aggregate effect depends on the change in inventor productivity in each cluster weighted by the initial cluster size. Intuitively, a certain change in inventor productivity in a cluster that is initially large has a larger aggregate impact on total number of patents produced than the same change affecting a cluster that is initially small. Although it may not be immediately obvious, equation (4) is analogous to the expression for aggregate effects derived by Kline and Moretti (2014a).[33]

[33] To see the analogy, rewrite equation (4) as $Y^2 - Y^1 = -g(S_A)\alpha_A + g(S_B)\alpha_B = -(Y_A/S_A)\alpha_A + (Y_B/S_B)\alpha_B$ where Y_c is the total number of patents produced in a cluster and α_c is the elasticity, which is allowed to vary across locations. This is equivalent to the expression in Section IVB of Kline and Moretti (2014a) for the effect of redistribution from one county to another. They write their expression in terms of elasticity of output with respect to density and show that under the assumption of perfect mobility and homogeneous tastes for locations, when both the elasticity and per worker output are the same in all locations, reallocating workers has no aggregate effects, as the benefits in the areas that gain activity are identical to the costs in areas that lose it. In this paper, elasticity of agglomeration is found to be constant across clusters of different size, but per inventor productivity is not assumed to be the same in all cities. This is the reason why spatial redistribution is found to have aggregate effects even with constant elasticity. In a model with homogeneous tastes for location and perfect mobility, Kline and Moretti (2014a) argue that the welfare effects are the same as the output effects. Put differently, if wage differences are only a function of local amenities, cities with low wages are larger and have better amenities than cities with high wages. Thus, redistributing workers from larger to smaller cities implies a utility loss, as more workers end up living in less

Extending equation (4) to the case of many cities, fields and years, the difference between counterfactual and observed aggregate number of patents in field *f* and year *t*, $Y_{ft}^2 - Y_{ft}^1$, can be estimated by summing across all cities the estimated change in inventor productivity in each cluster multiplied by the relevant cluster size:

$$(5) \qquad Y_{ft}^2 - Y_{ft}^1 = \sum_c S_{cft}\left[g\left(\bar{S}_{ft}\right) - g\left(S_{cft}\right)\right],$$

where $\bar{S}_{ft}$ is the average cluster size in the relevant field and year across all cities; S_{cft} is the actual cluster size; and the change in inventor productivity $\left[g\left(\bar{S}_{ft}\right) - g\left(S_{cft}\right)\right]$ can be quantified empirically as $\left[g\left(\bar{S}_{ft}\right) - g\left(S_{cft}\right)\right] = \bar{S}_{ft}^{\hat{\alpha}} - S_{cft}^{\hat{\alpha}}$, where $\hat{\alpha} = 0.0662$. The reason is that equation (1) assumes that $\ln g(S) = \alpha S$, so that $g(S) = S^{\alpha}$, and the elasticity of productivity with respect to cluster size was estimated to be $\hat{\alpha} = 0.0662$ in the baseline estimates in column 8 of Table 3. The elasticity α was found to be constant across clusters of different sizes.[34]

I stress that my analysis is deliberately a partial equilibrium analysis. In the counterfactual, only cluster size changes, while everything else in the economy is kept unchanged. In reality it is possible that changes in cluster size induce general equilibrium effects. Examples of general equilibrium effects may include lower congestion and housing prices in cities that currently have large high-tech clusters, and higher congestion and housing prices in cities that currently do not have large high-tech clusters. While the question of how these changes sum up in the aggregate is an interesting one, it is well outside the scope of this paper. In thinking about the possible magnitude of these general equilibrium effects, it is important to keep in mind that in my counterfactual, I equalize cluster size, not city size. The implied changes in congestion and land prices are likely to be considerably smaller than the changes that would be caused by equalizing city size.[35]

B. *Estimates of the Effect of Agglomeration on the Aggregate Number of Patents*

To gain a concrete idea of which cities may gain and which may lose in the counterfactual scenario, Table 10 reports examples of the estimated effect of size equalization on mean inventor productivity for computer science in 2007.[36] My estimate of $\hat{\alpha} = 0.0662$ from column 8 of Table 3 implies that under size equalization, clusters that are currently large would lose productivity. For example, the average productivity of computer scientists in the San Francisco-Silicon Valley region would be 22.76 percent lower than the observed productivity in 2007. This productivity

desirable locations. This is not true in a more general setting with idiosyncratic preferences for location. When labor supply to a locality is upward sloping, equilibrium wages will reflect both local amenities and local productivity. See also Glaeser and Gottleib (2008) and Gaubert (2018).

[34] A log-log model with $\alpha < 1$ implies that the relationship between number of patents and cluster size *in levels* is concave. The magnitude of coefficient α governs the degree of concavity: the smaller the coefficient, the more concave the function in levels. My estimate of α equal to 0.0662 points to a very concave function.

[35] Moreover, what matters is how the general equilibrium effects sum up in the aggregate. For each city that in the counterfactual experiences a decline in congestion and prices, there is a city that experiences an increase. While the former does not need to be equal to the latter, in the aggregate, part of the change among "winners" will be offset by changes among "losers."

[36] Specifically, for a city *c*, I am showing $\left[g\left(\bar{S}_{ft}\right) - g\left(S_{cft}\right)\right]$ estimated as $\bar{S}_{ft}^{\hat{\alpha}} - S_{cft}^{\hat{\alpha}}$ where $f =$ "computer science" and $t = 2007$.

loss stems from the fact that the size of the San Francisco-Silicon Valley cluster is larger than the average, so that in the counterfactual the cluster is made smaller. The corresponding figures for other large clusters like New York, Seattle, Austin, and Boston are −17.81 percent, −16.52 percent, −14.76 percent, and −13.45 percent.

On the other hand, the bottom panel shows that the average productivity of computer scientists in clusters that are originally below average increase, since their counterfactual size is larger. Many of these clusters are in the South and the Midwest. For example, the average productivity of computer scientists in Kansas City would be 2.66 percent higher than the observed productivity in 2007. The corresponding figures for Omaha, Nebraska; Portland, Maine; Memphis, Tennessee; and New Orleans, Louisiana are 13.42 percent; 17.76 percent; 23.36 percent; and 35.36 percent.

Table 11 reports the estimated effect of size equalization on inventor productivity for all fields in 2007 by initial cluster size. Unsurprisingly, large clusters experience declines in per inventor productivity, while small clusters experience gains in per inventor productivity. Clusters in the bottom quartile of the size distribution gain on average 27.53 percent in per inventor productivity. Clusters with sizes between the twenty-fifth percentile and the median gain 18.49 percent. Since the mean is above the median in most fields, clusters with size between the median percentile and the seventy-fifth percentile also gain, although only 9.41 percent. By contrast, clusters above the mean lose productivity. The changes in per inventor productivity for clusters with size between seventy-fifth and ninetieth percentile, ninetieth and ninety-fifth percentile, and ninety-sixth and one hundredth percentile are, −0.50 percent, −8.18 percent, and −14.73 percent, respectively.

Overall, Tables 10 and 11 indicate that small clusters gain productivity and large clusters lose. The question is what happens in the aggregate. Using equation (5), I estimate the aggregate effect of equalizing cluster size in each research field on the total number of patents produced in the United States in that field. Column 1 of Table 12 shows estimates for 2007, by field. The total number of patents created in the United States in computer science would be 13.34 percent lower in 2007 if computer scientists were uniformly distributed across cities. The losses in biology and chemistry, semiconductors, other engineering, and other science would be −10.06 percent, −14.83 percent, −7.71 percent, and −9.75 percent, respectively. The last row shows the total effect across all fields. The change in the total number of patents produced in the United States in 2007 is −11.20 percent.

Estimates in column 1 are based on a specific functional form assumption, namely that the relationship between log inventor productivity and log cluster size is linear. In Section IVB above, I tested this assumption and concluded that there is limited evidence of large departures from a log-log specification. My most reliable estimates of any curvature in the relationship between log inventor productivity and log cluster size (in column 2 of Table 8, panel A) indicate that the effect may be slightly larger for larger clusters, although the amount of heterogeneity is small. To assess the sensitivity of my aggregate findings to departures from the functional form assumption, I re-estimated the aggregate effects using the parameters in Table 8, panel A, column 2. The aggregate losses from equalization become slightly larger, 12.35 percent for all fields, since a larger elasticity

Table 10—Some Examples of the Effect of Cluster Size Equalization on Inventor Productivity: Computer Science in 2007

	Percent change
Examples of losers	
San Jose-San Francisco-Oakland, CA	−22.76
New York-Newark-Bridgeport, NY-NJ-CT-PA	−17.81
Seattle-Tacoma-Olympia, WA	−16.52
Austin-Round Rock, TX	−14.76
Boston-Worcester-Manchester, MA-NH	−13.45
Minneapolis-St. Paul-St. Cloud, MN-WI	−11.48
Raleigh-Durham-Cary, NC	−10.42
San Diego-Carlsbad-San Marcos, CA	−9.10
Portland-Vancouver-Beaverton, OR-WA	−8.84
Pittsburgh-New Castle, PA	−2.64
Boise City-Nampa, ID	−2.54
Examples of winners	
Miami-Fort Lauderdale-Miami Beach, FL	1.36
Kansas City-Overland Park-Kansas City, MO-KS	2.66
Buffalo-Niagara-Cattaraugus, NY	6.88
Omaha-Council Bluffs-Fremont, NE-IA	13.42
Des Moines-Newton-Pella, IA	14.60
Portland-Lewiston-South Portland, ME	17.76
Scranton-Wilkes-Barre, PA	21.53
Toledo-Fremont, OH	23.36
Memphis, TN-MS-AR	23.36
Oklahoma City-Shawnee, OK	25.76
New Orleans-Metairie-Bogalusa, LA	35.36

Note: Entries are estimates of the percent difference between mean inventor productivity in the counterfactual scenario and observed productivity.

Table 11—Effect of Cluster Size Equalization on Inventor Productivity: All Fields, 2007

	Percent change
0–25th percentile	27.53
25th–50th percentile	18.49
50th–75th percentile	9.41
75th–90th percentile	−0.50
90th–95th percentile	−8.18
95th–100th percentile	−14.73

Note: Entries are the estimated percent difference between inventor productivity in the counterfactual scenario and observed productivity, by initial size of cluster.

in larger clusters implies that redistribution away from larger clusters is more costly in the aggregate.[37]

Overall, based on estimates in Table 12, I conclude that the aggregate productivity gains from agglomeration are large in the United States.

[37] Alternative counterfactuals are possible. For example, I have estimated aggregate losses in the case of "partial equalization," where the number of inventors in city-field pairs with observed number of inventors below (above) the field mean is increased (decreased) by one-half of the difference between that city-field number and the mean. In this counterfactual, spatial inequality in the number of inventors is lower than observed inequality but higher than in the case of full equalization. I estimate that the aggregate loss in the number of patents would be −2.5 percent.

Table 12—Aggregate Effects of Cluster Size Equalization: 2007

	Constant elasticity (percent) (1)	Heterogeneous elasticity (percent) (2)
Computer science	−13.34	−14.54
Biology and chemistry	−10.06	−11.27
Semiconductors	−14.83	−16.05
Other engineering	−7.71	−8.61
Other science	−9.75	−10.93
All fields	−11.20	−12.35

Notes: Entries are the estimated percent difference between total number of patents created in the United States in the counterfactual scenario and observed number of patents created in the United States. Entries in column 1 are computed using a constant elasticity (Table 3, column 8). Entries in column 2 are computed using an elasticity that varies by cluster size (Table 8, panel A, column 2).

VI. Conclusions

One of the most remarkable and consequential aspects of the economic geography of the United States is the strong degree of geographical clustering of the high-tech sector. High-tech firms and workers appear to concentrate in a small number of expensive labor markets, such as San Francisco, New York, Boston, and Seattle, and not in less expensive locations. The top ten clusters in the computer science, semiconductors, and biology and chemistry fields account, respectively, for 69.3 percent, 77.0 percent, and 59.2 percent of all inventors in their field in 2007. These shares were significantly larger in 2007 than in 1971, pointing to increasing geographical agglomeration of inventors.

I find an economically important effect of cluster size on an inventor's productivity. An inventor moving from a small cluster to a large cluster enjoys an increase in annual productivity, as measured by the number of patents produced in a year or number of citations. The estimated elasticity of number of patents produced in a year with respect to cluster size is 0.0676 (0.0139). This estimate reflects the intensive margin of the effect of cluster size on productivity but misses the extensive margin. Therefore, it is likely to be a lower bound of overall effect. I also find aggregate gains for the United States as a whole from agglomeration of inventors. My estimates suggest that the overall number of patents created in the United States in a given year is 11.20 percent larger relative to a counterfactual scenario where all clusters are equalized.

Clustering of the high-tech sector may exacerbate inequality in earnings and income across communities. At the same time it appears to be important for overall production of innovation in the United States.[38] Policies designed to spread innovation across communities, such as place based subsidies that favor areas with little high-tech presence, need to take into account both benefits and costs.

[38] Kline and Moretti (2014b) provide a formal discussion of the equity-efficiency trade off in placed-based policies.

REFERENCES

1stock1.com. 2017. "Table with Kodak Prices." http://www.1stock1.com/1stock1_173.htm (accessed on September 30, 2017).

Acemoglu, Daron, Ufuk Akcigit, and William R. Kerr. 2016. "Innovation Network." *Proceedings of the National Academy of Sciences of the United States of America* 113 (41): 11483–88.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018. "Dancing with the Stars: Innovation Through Interactions." NBER Working Paper 24466.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–40.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics* 125 (2): 549–89.

Behrens, Kristian, and Frédéric Robert-Nicoud. 2015. "Agglomeration Theory with Heterogeneous Agents." In *Handbook of Regional and Urban Economics*, Vol. 5A, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 171–246. Amsterdam: Elsevier.

Bernstein, Shai, Timothy McQuade, and Richard R. Townsend. 2018. "Does Economic Insecurity Affect Employee Innovation? Evidence from Innovative Workers During the Great Recession." NBER Working Paper 24011.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81 (4): 1347–93.

Bosquet, Clément, and Pierre-Philippe Combes. 2017. "Sorting and Agglomeration Economies in French Economics Departments." *Journal of Urban Economics* 101: 27–44.

Buera, Francisco J., and Joseph P. Kaboski. 2012. "The Rise of the Service Economy." *American Economic Review* 102 (6): 2540–69.

Carlino, Gerald A., Robert M. Hunt, Jake K. Carr, and Tony E. Smith. 2012. "The Agglomeration of R&D Labs." Federal Reserve Bank of Philadelphia Working Paper 12-22.

Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Regional and Urban Economics*. Vol. 5A, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 394–404. Amsterdam: Elsevier.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1): 54–70.

Ciccone, Antonio, and Giovanni Peri. 2005. "Long-Run Substitutability between More and Less Educated Workers: Evidence from U.S. States, 1950–1990." *Review of Economics and Statistics* 87 (4): 652–63.

Cohen, Wesley M. 2010. "Fifty Years of Empirical Studies of Innovative Activity and Performance." In *Handbook of the Economics of Innovation*, Vol. 1, edited by Bronwyn H. Hall and Nathan Rosenberg, 129–213. Amsterdam: Elsevier.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)." NBER Working Paper 7552.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63 (2): 723–42.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80 (6): 2543–94.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Effects." In *Agglomeration Economics*, edited by Edward L. Glaeser, 15–65. Chicago: University of Chicago Press.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 5A, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 247–348. Amsterdam: Elsevier.

De La Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." *Review of Economic Studies* 84 (1): 106–42.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524.

Dickinson, Mike. 2017. "Kodak's Decades of Decline." *Rochester Business Journal*. September 13.

Duranton, Gilles. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review* 97 (1): 197–221.

Duranton, Gilles, and William R. Kerr. 2018. "The Logic of Agglomeration." In *The New Oxford Handbook of Economic Geography*, edited by Gordon L. Clark, Maryann P. Feldman, Meric S. Gertler, and Dariusz Wójcik, 347–65. Oxford: Oxford University Press.

Duranton, Gilles, Philippe Martin, Thierry Mayer, and Florian Mayneris. 2010. *The Economics of Clusters: Evidence from France*. Oxford: Oxford University Press.
Duranton, Gilles, and Henry G. Overman. 2005. "Testing for Localization Using Micro-Geographic Data." *Review of Economic Studies* 72 (4): 1077–106.
Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (5): 889–927.
Gaubert, Cecile. 2018. "Firm Sorting and Agglomeration." *American Economic Review* 108 (11): 3117–53.
Glaeser, Edward L., and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity* 39 (1): 155–239.
Glaeser, Edward L., and David C. Mare. 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316–42.
Glaeser, Edward L., and Albert Saiz. 2004. "The Rise of the Skilled City." *Brookings-Wharton Papers on Urban Affairs*: 47–94.
Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.
Griliches, Zvi. 1998. *R&D and Productivity: The Econometric Evidence*. Chicago: University of Chicago Press.
Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (1): 93–118.
Griss, Martin. 1993. "Hewlett-Packard Software Reuse: from Library to Factory." Unpublished.
Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. New York: PublicAffairs.
Guzman, Jorge. 2019. "Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley." Unpublished.
Hall, Bronwyn H., Zvi Griliches, and Jerry A. Hausman. 1986. "Patents and R and D: Is There a Lag?" *International Economic Review* 27 (2): 265–83.
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File." NBER Working Paper 8498. https://www.nber.org/research/data/us-patents.
Haran, Hadas. 2011. "Time to Market Research: Highlights and Key Findings." Unpublished.
Helmers, Christian, and Henry G. Overman. 2017. "My Precious! The Location and Diffusion of Scientific Research: Evidence from the Synchrotron Diamond Light Source." *Economic Journal* 127 (604): 2006–40.
Henderson, J. Vernon. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 53 (1): 1–28.
Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science." *Quarterly Journal of Economics* 133 (2): 927–91.
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.
Johnson, Kenneth P. 2004. "Redefinition of the BEA Economic Areas." https://www.csus.edu/indiv/j/jensena/sfp/ea_desc.htm (accessed on April 3, 2019.)
Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics* 96 (1): 171–88.
Kantor, Shawn, and Alexander Whalley. 2019. "Research Proximity and Productivity: Long-Term Evidence from Agriculture." *Journal of Political Economy* 127 (2): 819–54.
Kapoor, Rahul. 2012. *Managing Complexity and Change in the Semiconductor Industry: Findings from the Wharton-ATREG Industry Study*. Wharton-ATREG.
Kerr, William R. 2010. "Breakthrough Inventions and Migrating Clusters of Innovation." *Journal of Urban Economics* 67 (1): 46–60.
Kerr, William R. 2018. "Navigating Talent Hot Spots." *Harvard Business Review* 96 (5): 80–86.
Kerr, William R., and Scott Duke Kominers. 2015. "Agglomerative Forces and Cluster Shapes." *Review of Economics and Statistics* 97 (4): 877–99.
Kline, Patrick, and Enrico Moretti. 2014a. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *Quarterly Journal of Economics* 129 (1): 275–331.
Kline, Patrick, and Enrico Moretti. 2014b. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." In *Annual Review of Economics*, Vol. 6, edited by Kenneth J. Arrow and Timothy F. Bresnahan, 629–62. Palo Alto, CA: Annual Reviews.

Krasner, Jerry. 2003. "Total Cost of Development: A Comprehensive Cost Estimation Framework for Evaluating Embedded Development Platforms." Unpublished.

Lerner, Josh, and Amit Seru. 2017. "The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond." Unpublished.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.

Lychagin, Sergey, Joris Pinkse, Margaret E. Slade, and John Van Reenen. 2016. "Spillovers in Space: Does Geography Matter?" *Journal of Industrial Economics* 64 (2): 295–335.

Mansfield, Edwin, John Rapoport, Jerome Schnee, Samuel Wagner, and Michael Hamburger. 1971. *Research and Innovation in the Modern Corporation*. London: Palgrave Macmillan.

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

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. "Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics* 39 (3) 332–42.

Moretti, Enrico. 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94 (3): 656–90.

Moretti, Enrico. 2012. *The New Geography of Jobs*. Boston: Houghton Mifflin Harcourt.

Moretti, Enrico. 2021a. "Replication Data for 'Local Economic Development, Agglomeration Economics, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority.'" Harvard Dataverse [distributor]. https://doi.org/10.7910/DVN/G1UYL0.

Moretti, Enrico. 2021b. "County - Economic Area Crosswalk." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E140581V1.

Moretti, Enrico. 2021c. "Replication Data for: The Effect of High-Tech Clusters on the Productivity of Top Inventors." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E140662V1.

Moretti, Enrico, and Daniel J. Wilson. 2014. "State Incentives for Innovation, Star Scientists and Jobs: Evidence from Biotech." *Journal of Urban Economics* 79: 20–38.

Moretti, Enrico, and Daniel J. Wilson. 2017. "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists." *American Economic Review* 107 (7): 1858–903.

Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish Émigrés and US Invention." *American Economic Review* 104 (10): 3222–55.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics Volume 4: Cities and Geography*, edited by J. Vernon Henderson and Jacques-Francois Thisse, 2119–71. Amsterdam: Elsevier.

Rosenthal, Stuart S., and William Strange. 2006. "The Micro-Empirics of Agglomeration Economics." In *A Companion to Urban Economics*, edited by Richard J. Arnott and Daniel P. McMillen, 7–23. Malden, MA: Blackwell Publishing.

Rosenthal, Stuart S., and William C. Strange. 2008. "The Attenuation of Human Capital Spillovers." *Journal of Urban Economics* 64 (2): 373–89.

Saxenian, AnnaLee. 2014. *Regional Advantage Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Trajtenberg, M., R. Henderson, and A. Jaffe. 1997. "University versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation & New Technology* 5: 19–50.

US Bureau of Economic Analysis (BEA). 2012. "2007, 2012: 405 Industries." https://www.bea.gov/industry/input-output-accounts-data (accessed September 1, 2019).

Waldinger, Fabian. 2012. "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany." *Review of Economic Studies* 79 (2): 838–61.

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy* 121 (1): 1–27.

Wu, Mary. 2011. "Agile Method to Improve Delivery of Large-Scale Software Projects." Unpublished.

Zacchia, Paolo. 2018. "Benefiting Colleagues but Not the City: Localized Effects from the Relocation of Superstar Inventors." *Research Policy* 47 (5): 992–1005.

Zolas, Nikolas. 2016. "United States Patent Classification (USPC) Crosswalks – Version 1610." https://sites.google.com/site/nikolaszolas/PatentCrosswalk.

Zucker, Lynne G., and Michael R. Darby. 1996. "Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry." *Proceedings of the National Academy of Sciences* 93 (23): 12709–16.

Zucker, Lynne G., and Michael R. Darby. 2014. "Connecting Outcome Measures in Entrepreneurship Technology and Science (COMETS) Database, Public-Use Version COMETS 2.0." https://kauffman.org/entrepreneurship/research/comets/.

This article has been cited by:

1. Wenyue Cui, Jie Tang. 2023. Innovation convergence clubs and their driving factors within urban agglomeration. *Economic Modelling* **121**, 106199. [Crossref]
2. Xun Bian, Erik Devos, Zifeng Feng. 2023. Commercial Real Estate Returns and Innovation. *Journal of Real Estate Research* **59**, 1-20. [Crossref]
3. Chenyang Yu, Hongyu Long, Xiang Zhang, Chenglin Tu, Yuanfang Tan, Yu Zhou, Chuanxiang Zang. 2023. Regional integration and city-level energy efficiency: Evidence from China. *Sustainable Cities and Society* **88**, 104285. [Crossref]
4. Muniba Muniba, Baorong Yu. 2023. Does Innovative City Pilot Policy Stimulate the Chinese Regional Innovation: An Application of DID Model. *International Journal of Environmental Research and Public Health* **20**:2, 1245. [Crossref]
5. Zhongxin Ma, Fenglan Wu. 2023. Smart City, Digitalization and CO2 Emissions: Evidence from 353 Cities in China. *Sustainability* **15**:1, 225. [Crossref]
6. Breandán Ó hUallacháin. 2022. Taking Organizations Seriously in Understanding the Geography of Invention. *International Regional Science Review* **5**, 016001762211458. [Crossref]
7. Dori A. Cross, Julia Adler-Milstein, A. Jay Holmgren. Management Opportunities and Challenges After Achieving Widespread Health System Digitization 67-87. [Crossref]
8. Jing Tan, Yaqiao Zhang, Hui Cao. 2022. The FDI-spawned technological spillover effects on innovation quality of local enterprises: evidence from industrial firms and the patents in China. *Applied Economics* **3**, 1-16. [Crossref]
9. Felix Ehrenfried, Thomas A. Fackler, Valentin Lindlacher. 2022. New region, new chances: does moving regionally for university shape later job mobility?. *Regional Studies* **110**, 1-15. [Crossref]
10. Siyu Huang, Yi Shi, Qinghua Chen, Xiaomeng Li. 2022. The growth path of high-tech industries: Statistical laws and evolution demands. *Physica A: Statistical Mechanics and its Applications* **603**, 127719. [Crossref]
11. Lu Dai, Jiajun Zhang, Shougui Luo. 2022. Effective R&D capital and total factor productivity: Evidence using spatial panel data models. *Technological Forecasting and Social Change* **183**, 121886. [Crossref]
12. Farhat Chowdhury, Albert N Link, Martijn van Hasselt. 2022. The spatial distribution of public support for AI research. *Science and Public Policy* **49**:4, 573-579. [Crossref]
13. Wolfgang Dauth, Sebastian Findeisen, Enrico Moretti, Jens Suedekum. 2022. Matching in Cities. *Journal of the European Economic Association* **20**:4, 1478-1521. [Crossref]
14. Harald Bathelt, Pengfei Li. 2022. The interplay between location and strategy in a turbulent age. *Global Strategy Journal* **12**:3, 451-471. [Crossref]
15. Magnus Henrekson, Anders Kärnä, Tino Sanandaji. 2022. Schumpeterian entrepreneurship: coveted by policymakers but impervious to top-down policymaking. *Journal of Evolutionary Economics* **32**:3, 867-890. [Crossref]
16. Farhat Chowdhury, Albert N. Link, Martijn van Hasselt. 2022. Public support for research in artificial intelligence: a descriptive study of U.S. Department of Defense SBIR Projects. *The Journal of Technology Transfer* **47**:3, 762-774. [Crossref]
17. Nils M. Denter, Lukas Jan Aaldering, Huseyin Caferoglu. 2022. Forecasting future bigrams and promising patents: introducing text-based link prediction. *foresight* **241**. . [Crossref]
18. Hector Rocha, David B. Audretsch. 2022. Entrepreneurial ecosystems, regional clusters, and industrial districts: Historical transformations or rhetorical devices?. *The Journal of Technology Transfer* **49**. . [Crossref]

19. Jessica Bai, Shai Bernstein, Abhishek Dev, Josh Lerner. 2021. Public Entrepreneurial Finance around the Globe. *SSRN Electronic Journal* **49**. . [Crossref]
20. Daniel P. Gross, Bhaven N. Sampat. 2020. Inventing the Endless Frontier: The Effects of the World War II Research Effort on Post-War Innovation. *SSRN Electronic Journal* **6**. . [Crossref]