THE EFFECT OF HIGH-TECH CLUSTERS ON THE PRODUCTIVITY OF TOP
INVENTORS

Enrico Moretti

The Effect of High-Tech Clusters on the Productivity of Top Inventors
Enrico Moretti
NBER Working Paper No. 26270
September 2019, Revised February 2021
JEL No. J01,R11

ABSTRACT

The high-tech sector is increasingly concentrated in a small number of expensive cities, with the top ten cities in "Computer Science", "Semiconductors" and "Biology and Chemistry", accounting for 70%, 79% and 59% of inventors, respectively. Why do inventors tend to locate near other inventors in the same field, despite the higher costs? I use longitudinal data on top inventors based on the universe of US patents 1971 - 2007 to quantify the productivity advantages of Silicon-Valley style clusters and their implications for the overall production of patents in the US. I relate the number of patents produced by an inventor in a year to the size of the local cluster, defined as a city × research field × year. I first study the experience of Rochester NY, whose high-tech cluster declined due to the demise of its main employer, Kodak. Due to the growth of digital photography, Kodak employment collapsed after 1996, resulting in a 49.2% decline in the size of the Rochester high-tech cluster. I test whether the change in cluster size affected the productivity of inventors outside Kodak and the photography sector. I find that between 1996 and 2007 the productivity of non-Kodak inventors in Rochester declined by 20.6% relative to inventors in other cities, conditional on inventor fixed effects. In the second part of the paper, I turn to estimates based on all the data in the sample. I find that when an inventor moves to a larger cluster she experiences significant increases in the number of patents produced and the number of citations received. Conditional on inventor and firm effects, the elasticity of number of patents produced with respect to cluster size is 0.0662 (0.0138). The productivity increase follows the move and there is no evidence of an effect in the years leading up to the move. IV estimates based on the geographical structure of firms with laboratories in multiple cities are statistically similar to OLS estimates. In the final part of the paper, I use the estimated elasticity of productivity with respect to cluster size to quantify the aggregate effects of geographical agglomeration on the overall production of patents in the US. I find macroeconomic benefits of clustering for the US as a whole. In a counterfactual scenario where the quality of U.S. inventors is held constant but their geographical location is changed so that all cities have the same number of inventors in each field, inventor productivity would increase in small clusters and decline in large clusters. On net, the overall number of patents produced in the US in a year would be 11.07% smaller.

Enrico Moretti
University of California, Berkeley
Department of Economics
549 Evans Hall
Berkeley, CA 94720-3880
and CEPR
and also NBER
moretti@econ.berkeley.edu

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)

1   Introduction

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 US, 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-Saiz, 2004; Buera-Kaboski, 2012). In the period 1980 to 2010, mean wages and mean income in cities with large high-tech clusters have increased significantly more than in cities without high-tech clusters (Moretti, 2012; Diamond, 2016a). 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, 2013). 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 10 cities with largest number of inventors in "Computer Science" account for 70% of all inventors in 2007. In "Semiconductors" and "Biology and Chemistry", the corresponding shares are 79% and 59%, respectively. These shares are not declining over time. Indeed, they are larger in 2007 than in 1971, suggesting that despite the diffusion of 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 US between 1971 to 2007 to investigate two related questions: (1) Are there productivity benefits for inventors who locate in Silicon Valley-style clusters? (2) What are the aggregate effects of geographical agglomeration on overall innovation in the US? In my analysis, I define a cluster as city × research field × year. I estimate how inventors' productivity varies with the size of the relevant cluster, measured by the number of other inventors in the same city, field, and year excluding the focal inventor. As measures of worker-specific productivity, I use the number of patents produced in a year and the number of subsequent citations these patents receive. 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 90th percentile in the total number of patents over the sample period—since the panel is longer for this group.

The analysis proceeds in three parts. I first study the experience of inventors in Rochester NY, 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 US. At its peak in 1996, Kodak employed more than 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 83.7%.

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% 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 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 to 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

---

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

in unobserved quality of inventors and it is robust to controlling for field × city and field × year effects. 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 823,359 inventors observed between 1971 and 2007, located in 32,220 clusters (179 cities × 5 research fields × 36 years). I regress inventor's log productivity on log cluster size, conditioning on inventor, firm, city × 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 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. 2SLS models yield an elasticity that is statistically similar to the OLS elasticity.

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 0.0662 (0.0138). The estimated elasticity implies that a computer scientist moving from the median cluster in computer science (Gainesville, FL) to the cluster at the 75th percentile of size (Richmond, VA) would experience a 12.0% increase in productivity, holding constant the inventor and the firm. In biology and chemistry, a move from the median cluster—Boise, ID—to the 75th percentile cluster—State College, PA—is associated with a productivity gain of 8.4%, 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 externality varies enormously across firms. The externality generated by the average firm in the average city is 0.3% and 0.24% 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 scientist in Seattle is estimated to be 8.06% higher because of the presence of Microsoft in the local computer science cluster. This large external effect reflects Microsoft 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 macro-economic benefits of agglomeration for the US 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 US. 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 US 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 22.52% lower than the observed productivity in 2007 because the size of the San Francisco-Silicon Valley Computer Science would be made significantly smaller. The corresponding losses for New York, Seattle, Austin, and Boston would be -16.91%, -16.47%, -14,56%, and -13.30%, respectively. On the other hand, the average productivity of computer scientists in Kansas City would be 3.18% higher than the observed productivity in 2007, because the Kansas City Compute Science cluster is smaller than average. The corresponding gains for Omaha, Portland ME, and Memphis would be 13.23%, 17.50%, and 23.02%, 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 US in Computer Science would be 13.18% 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 -9.92%, -14.68%, -7.62%, and -9.64%, respectively. The change in the total number of patents in the US would be -11.07%.

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 US. 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 forthcoming) find significant spillovers from academic R&D on local firms. Helmers and Overman (2017) find that the establishment of the Diamond Light Source synchrotron in the UK 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 et al. (2014); and Azoulay et al. (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 2 describes the data. Section 3 reports the estimates based on the Rochester case study. Section 4 and 5 report the estimates based on the full sample. Section 6 discusses the aggregate implications. Section 7 concludes.

---

[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, Pinkse, Slade, and Van Reenen (2016); Bosquet and Combes (2017); and Acemoglu, Akcigit, and Kerr (2016); Akcigit et al (2018); Zacchia (2018).

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

2 Data

I use data on the universe of U.S. patents filed between 1971 and 2007 and ultimately granted. Unsuccessful applications are not included. The source of the data is the COMETS patent database, described in detail by Zucker, Darby, and Fong (2014). It is the same data used in a recent paper by Moretti and Wilson (2017).

2.1 Patents, Inventors, and Their Location

The data include 4,229,818 patents. 91.5% of patents in the sample are filed by inventors employed in private firms, 4.1% by inventors employed by universities, with the remainder are filed by inventors working in national labs, government, nonprofit institutions, or for themselves. 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 155,708 patents filed between 1971 and 2007, followed by General Electric, Microsoft, Intel and Kodak with 69,019, 43,550, 42,087 and 41,553 patents filed respectively. The U.S. Navy and the University of California are the only non-corporate organizations in the Table.[4]

The geographical units of analysis I employ are the Bureau of Economic Analysis's (BEA) "Economic Areas." There are 179 Economic Areas in the US and they cover the entire country. In most cases, "Economic Areas" are similar to an 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. 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 4.25% of all patents—for brevity, in the rest of the paper, I will refer to this field as "semiconductors"); Computing and Information Technology (8.75%—for brevity: "computer science"); Biology, Chemistry and Medicine (23.15%—for brevity: "biology and chemistry"); "Other Engineering" (54.26%); and "Other Science" (9.59%). 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

---

[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, Univ Calif Los Angeles have all been assigned the same code.

[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.

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 follow Moretti and Wilson (2017) and focus on the productivity of star inventors, defined as those who are at or above the ninetieth percentile in the total number of 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%, 1%, 5%, 25% and the full sample.

The sample of star inventors includes 109,846 inventors and 932,008 inventors x years observations with non missing information on field, class, and city. Of these, 823,359 also have non-missing 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 .25, .5, 1, 1.2 and 2 respectively.

2.2 Clusters

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

For example, the top panel in Table 1 shows the largest clusters in the computer science in 2007. The San Jose-San Francisco-Oakland region is by far the largest cluster in this field, accounting for 26.1% of all inventors in the field. Second on the list is the New York region, which accounts for 9.1% of all the inventors in the field. Seattle, Austin, and Boston follow, with cluster sizes equal to 8.3%, 5.9%, and 4.7%, respectively. The geographical concentration of inventors in computer science is very pronounced. The top ten cities account for 69.33% of all inventors in the field in 2007. The ratio between the the size of the largest cluster—San Jose-San Francisco-Oakland—and median cluster—Gainesville, FL—is 865.5, indicating that the Bay Area had 865.5 times the number of inventors in computer science than Gainesville did in 2007. The ratios between the 99th and median and 95th percentiles and median are equal to 300.9 and 78.6, respectively.

The middle panel is for biology and chemistry. The top five clusters are New York (11.3%), San Jose-San Francisco-Oakland (11.1%), Boston (6.9%), Philadelphia (6.3%) and Los Angeles (5.9%). With the total share of top ten cities equal to 59.05%, the overall concentration in this research field is lower than in computer science, but it remains quite pronounced. The max / median and the 99th percentile / median ratios are 126.8 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 (25.5%), New York (14.9%), Los Angeles (6.1%), Dallas (5.0%) and Phoenix (4.9%). The overall concentration in this research field is even higher than in computer science, with the share of top ten cities equal to 77.16%. Both the max / median ratio and the 99th percentile / median ratio are undefined since the median city had no inventors in this field in 2007. The tenth, twenty-fifth percentile, median, seventy-fifth and ninetieth percentiles across all fields and years are 0.00356, 0.01106, 0.02951, 0.06332, 0.10786, respectively.

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

2.3 Data Limitations

(A) There are two main limitations of the data. First, patent count is an imperfect measure of the amount of innovation created by an inventor. Not all patents are equal, in the sense that not all patents embody the same amount of innovation. 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. In particular, I estimate models where the dependent variable is an inventor's number of citations per year or an inventor's number of citations per patent per year.[7]

In addition, not all innovation is patented and the fraction of innovation that is patented varies across fields and technology classes. Cohen et al. (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 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 1990's, 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.[8] 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.

(B) 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

---

[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, 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.

[7]Bernstein McQuade and Townsend (2018) and Akcigit et al (2018) also use number of patents and citations to measure inventor productivity. On the other hand, Williams (2013) is an example of a paper that measures scientific research and product development directly rather than via patent activity.

[8]It could also happens because high tech clusters tend to have high tech lawyers who help in the patenting activity, which reduces the cost of obtaining advice.

at all).

Since I do not observe inventors in years when their number of patents is 0's, 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 three pieces of evidence in Section 5.4 below. First, I add some of the missing zeros by interpolating the data when one missing year is immediately preceded and followed by non-missing years. That is, if an inventor is observed in years $t-1$ and $t+1$, but is missing in year t, I assign her 0 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 non-missing 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 1 month, 2 month, 3 month, 6 month, 1 year (the baseline), 2 year, and 3 year periods. When the temporal unit of analysis is short (1 or 2 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 (2 or 3 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.

Third, I find that estimates are larger if more stringent definitions of star patenters are adopted. In particular, estimates for the top 0.5%, 1%, and 5% of inventors are larger than the baseline estimates that use the top 10%. In turn, the baseline estimates are larger than estimates for the top 25% and for the full population of inventors. This is due to the fact that the more stringent the definition of star patenters the longer an inventor is in the data and therefore the smaller is the number of missing 0's.

Overall, the evidence indicates that my baseline estimates miss part of the extensive margin and therefore should be interpreted as a lower bound of the total effect of cluster size on patenting. The problem is smaller in specifications where I focus on the subset of inventors who are in the sample for several years. For example, in estimating the dynamic response to changes in cluster size induced by a move, I will focus on the subset of inventors who are observed for 11 consecutive years (including 5 year before the move, the year of the move and 5 years after the move). By construction, this sample has no missing 0's in the 11 years under consideration. Estimates based on this sample reflect both the intensive and extensive margins and therefore are larger in magnitude.

3   The Effect of the Demise of Kodak on Inventor Productivity 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 initially 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, NY 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, NM, where it was founded, to Seattle in 1979. 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 parts of the company co-founders rather than business reasons or the conditions of the local economy.[9] Since Microsoft relocation, Seattle has become one of the largest software clusters in the US. 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).[10] In a similar way, the birth of the biotech sector in San Diego can be traced to the serendipitous presence of star biologists at UCSD in the mid-1970's 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 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 US. Indeed, as we saw in Table A.1, Kodak had the 5th largest number of patents filed between 1971 and 2007. Kodak 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 51.06% 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 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 stock price declined by 82%. 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 1254 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% 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

---

[9]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).

[10]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.

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.[11] 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 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).

Tables 2 presents corresponding differences-in-differences 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 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, Electrophotography", 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 and the next 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 6.58% between 1996 and 2007 relative to other cities. In columns 2 to 4, I add field, field × year and field × city effects as controls. The coefficient on the interaction becomes more negative and suggests a productivity loss in Rochester between 6.96% and 9.84% relative to other cities, depending on the controls. For robustness, I estimated an additional model where I drop Xerox inventors, both in Rochester and other cities and find a similar estimate.[12]

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, a higher median family income, a slightly higher mean inventor productivity. The main research field (outside Kodak) in 1996 was "Other Engineering", which accounted for 67% of all inventors,

---

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

[12]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. After dropping Xerox, the sample size is 193,194 and the difference-in-difference estimate is -0.0760 (0.0059).

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 income, 1996 mean inventor productivity, 1996 share of patents in each research fields, 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.[13] 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.[14]

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 8213 inventors observed twice, for a total sample size of 16,426, 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 1 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 1 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% more negative than inventors who were not in Rochester in 1996, after controlling for inventor fixed effects.[15] When I add field, field × year, and field × city effects as controls in columns 2, 3 and 4, the coefficient on the interaction become more negative and suggests a productivity loss in Rochester between 22.0% and 32.1% relative to other cities. Weighted estimates in column 5 are 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. In

---

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

[14]A more extreme option is to include in the sample only cities with characteristics similar to Rochester. This is equivalent to putting 0 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%. Results are generally similar to the ones presented in the table (available on request).

[15]The average change for inventors who were in Rochester in 1996 is -.412 (0.0153)—as can be seen by adding the coefficients in row 1 and 2—significantly more negative than the decline for inventors who were not in Rochester in 1996, which is -.206 (0.0190).

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 2.3 above.[16]

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.

In order to assess how robust my estimates are to excluding remaining Kodak's suppliers, I contacted Kodak and requested a list with the names of their suppliers in 1996 with the intent to match their names with patent data, but failed to obtain the data from the company. 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. I used input-output tables to identify likely Kodak suppliers outside "Photography" and "Electrophotography". In particular, I use input-output tables at the 6 digit level to identify how much each 6 digit industry sells to NAICS 333316 (Photographic and photocopying equipment manufacturing), which is the one that Kodak likely belongs to. I then re-estimated Table 2 excluding inventors in technology classes that have a positive amount of expected sales to NAICS 333316.[17] Estimates corresponding to column 4 of Table 2 are -.0823 (.0131) and -.3771 (.1475), respectively. I conclude that 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.

4 Inventor Productivity and Cluster Size

The Rochester case study focuses on a specific shock to clusters size in one city. 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:

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

---

[16]I have estimated similar models where the Rochester dummy is set equal to 1 if the inventor is in Rochester both in 1996 and in 2007. This specification estimates the effect of the shock to Rochester on inventors who were there at the beginning of the shock and are still there in 2007. It is less ideal than the specification in the Table, since the 2007 location is endogenous. The difference in difference estimates are between -0.313 (0.0189) and -0.399 (0.0233) and are available upon request.

[17]To link patents to industries, I used the crosswalk created by Lybbert and Zolas (2014): for each parent class, the crosswalk lists multiple 6-digit industries, with assigned probabilities. Based on the probabilities, I computed the expected value of inputs used by Kodak for each technology class.

where y_{ijfct} 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} represent city $\times$ year effects and are included to absorb city-wide 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.

Equation 1 assumes that agglomeration effects vary as a function of cluster size. 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. 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 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—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 (2016b) are also absorbed by d_{ct}. City-wide productive amenities—i.e., factors that affect the 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 or 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 to cluster size productivity increases that arise because of sorting. 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 city-wide—for example: a new airport, or other form of improved infrastructure—and therefore are controlled for in equation 1. But the city × 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.

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 $city x year$ dummies such as the one in Equation 1 yields an estimate of α that is numerically identical to the one would obtain if the independent variable is log density defined as number of inventors per-square mile: $\ln(S_{-ifct}/A_c)$, where A_c is the area of the city.[18]

4.1 Baseline Estimates

Figure 4 shows a binned scatter plot 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 (0.008).

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: 0.0514 (0.0081). In columns 2 and 3, I add dummies for the interaction 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 interaction 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 0.0910 (0.00980). A comparison with column 5 indicates that the within-inventor estimate is significantly 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.[19]

In column 7, I add dummies for the interaction of city × year in order to absorb city-wide shocks to local productive amenities and selection driven by shocks to local consumption amenities. The coefficient drops to 0.0514

---

[18]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: $\ln(S_{-ifct}/A_{fc})$, where A_{fc} is the area of the cluster.

[19]One concern is the possibility 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.

(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 city-wide 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 932,008 to 823,359. The coefficient is 0.0662 (0.0138).

The elasticity of productivity with respect to cluster size estimated in column 8 indicates that a 10% increase in cluster size is associated with a 0.66% 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, FL—to the cluster at the 75th percentile—Richmond, VA. Based on the coefficient in column 8, such inventor would experience a 12.0% 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, ID—to the 75th percentile cluster—State College, PA—would be associated with a productivity gain of 8.4%, 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.[20]

The magnitude of the implied agglomeration economies is in line with existing estimates in the literature on agglomeration economies. A meta-analysis of 34 different studies (Melo et al., 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.[21]

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. Appendix Table A.2 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 unchanged.[22] 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%. This large external effect reflects Microsoft remarkable size in the Seattle computer science cluster. With 1389 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%). 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%. The external effect of Dell in Austin, Texas Instruments in Dallas, Caterpillar in Peoria, and Motorola in Chicago are, respectively, 0.61%, 1.84%, 10.47%, 0.88%. The estimated external effect of Hewlett-Packard in San Francisco-Silicon Valley is only 0.07%. The last row in

---

[20]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. Appendix Table A.3 shows estimates of the effect by research field. These models do not include city x year effects to avoid multicollinearity. (The corresponding elasticity for the full sample is .1081 (.0119).) The elasticity is largest for Semiconductors (0.258), and Computer Science (0.194), and it is smallest for Other Science (0.076).

[21]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 (2014) find an elasticity of 0.2 for US manufacturing. At the other extreme, Greenstone, Hornbeck and Moretti (2010)'s 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 et al.; 2008; Rosenthal and Strange; 2008).

[22]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 $\hat{\alpha} = 0.066$ is the estimated elasticity and $\Delta \ln S_{-jfct}$ is the difference in log cluster size with and without a given firm: $\Delta \ln S_{-jfct} = [\ln(N_{fct}/N_{ft}) - \ln(N_{jfct}/N_{ft})]$, 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 $t = 2007$.

the panel indicates that the spillover effect generated by the average firm in the average city in computer science was 0.32%. 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.[23]

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 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 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 number of additional patents. Local government interested in the corresponding dollar value would need to make an assumption on the expected monetary value of the marginal patent generated.

4.2 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 sub-section 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 raising 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 $s = -5, ..., -1, 1, ..., 5$—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 left most 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

---

[23]Du Pont has a very large presence in Philadelphia, with 986 scientists in 2007. It generates a productivity spillover equal to 1.17%. 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, OH, and 3M in Duluth, MN, with productivity spillovers estimated to be 3.47% and 8.99%, 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% range. The spillover for the average firm in the average city reported in the last row is 0.24%.

rightmost coefficient, β_{-5} represents the focal inventor's change in productivity in response to a change in size 5 years in the past. The dotted lines are a 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 non-missing, 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 $\mu_n = \beta_5 + \beta_4 + ... + \beta_n$ 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 5 years prior to the move. The next point moving to the right, μ_4, is the estimated cumulative productivity response up to 4 years before the move ($\beta_5 + \beta_4$). The point μ_{-5} is the cumulative productivity response from 5 years before the move through 5 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 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 (ICT) and even shorter for software (Griss, 1993; Krasner, 2003; Wu, 2011; Kapoor, 2012; Haran, 2011; Mansfield, 1972). 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 early stage (Cohen 2010). Previous studies have found quick effects on patenting. For example, Hall, Griliches, 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.[24] 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 codes 424 and 514). While the confidence intervals are inevitably larger, there appears to be no evidence of an immediate effect in this case. In fact, we see no effect within five years.

A third feature of Figures 5 and 6 worth highlighting is that the effect appears, at least in part, persistent. In Figure 5, after the initial increase in the first two years, productivity declines 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.[25]

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. (Recall that the average inventor is in my sample for only 7.01 years and not necessarily consecutively.) Estimates of the effect of cluster size are expected to be larger for superstar inventors, as I discussed in Section 2.3. The reason is that inventors who are particularly prolific tend to be in the data for many years, so that all the variation in size is from the intensive margin and none from the extensive margin. I will come back to this point in Section 5.4 below, where I present estimates for increasingly stringent definitions of star inventors and confirm that the effect is largest when I restrict the sample to inventors in the top 1% or 0.5% of lifetime patent count, which account for the vast majority of the sample used in Figures 5 and 6.[26]

4.3 Instrumental Variable Estimates

In this sub-section, I consider models based on two instrumental variables. 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.

(A) Rochester. First, I re-visit the Rochester case study. The difference in difference estimates uncovered in Section 3 can be interpreted as reduced form estimates in a model where the excluded instrument is the Rochester $\times$ 2007. The 2SLS estimates can be obtained by re-scaling 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 -.492 (.128) in column 1 to -.396 (.0455) in column 4. This last coefficient, for example, indicates that in a model

---

[24]On the other hand, some pharmaceutical patents are secondary. That is, they are modifications of the original idea, and have short research times.

[25]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.

[26]To see if inventors who move to larger clusters are more likely to continue patenting, I estimated a model where the dependent variable is a dummy equal to 1 if the inventor is observed patenting after the move. This dependent variable is regressed on a dummy equal to one if the destination cluster is larger than the origin cluster, average productivity of the inventor in the years before the move and controls for the destination city x year and research field x year. This model is estimated on a sample of inventors who move only once, with one observation per inventor. The coefficient on the dummy equal to one if the destination cluster is larger than the origin cluster is .0475 (.0161), indicating that inventors who move to larger clusters are more likely to continue patenting.

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.)

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 0.134 (0.038) in column 1 to .249 (.0398) in column 4.

(B) Firm Spatial Networks. Next, I use the geographical structure of firms with multiple locations to build an instrumental variable that isolates variation in local cluster size that originate 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 but 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 <u>excluding</u> 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 instrumental variable 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 $D_{scf(t-1)}$ is an indicator equal to 1 if firm s has at least 1 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 3 inventors in the computer science field. Between 1996 and 1997 Microsoft is generally expanding its overall R&D investment in the US 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 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 $\times$ year and class $\times$ year effects, identifica-

tion 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:

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

Columns 1-6 of Table 5 report OLS estimates of equation 3. The sample in Panel A includes inventors who are observed in two consecutive years. Standard errors in this table are clustered by city. The OLS estimates range between 0.0144 (0.00376) and 0.0169 (0.00387), 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 on 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.[27]

Columns 7-12 report 2SLS estimates and the corresponding first stage estimates. The F-stats, reported at the bottom of the Table, are between 42.9 and 87.3. The 2SLS estimates range from 0.0291 (0.0177) to 0.0475 (0.0200). The entry in column 12 is equal to 0.0307 (0.0147). Measurement error in first differences models means that part of the reason why IV is larger is that it both corrects for measurement error and endogeneity. While the 2SLS point estimates are larger than the corresponding OLS estimate, they are generally not statistically different from the OLS estimates. The p-values from a Durbin–Wu–Hausman test for the models in columns 1 to 6 are, respectively: 0.36; 0.08; 0.25; 0.22; 0.34; and 0.30.

To asses 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.[28]

The sample in Panel B includes inventors who are observed in two consecutive years and do not change city or field while in the sample. The OLS estimate in column 6 is 0.0415 (0.0090), larger than the estimate in Panel A, and smaller than the estimate from the model in levels. The 2SLS estimates are larger than the ones in Panel A, but display more variability and less precision, probably due to the smaller sample size. The entry in column 12 is equal to 0.124 (0.0433).

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 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.

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 instrumental variable estimates based on the Rochester shock and the ones based on firm spatial networks, taken together, appear to allay concerns about simul-

---

[27]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 0.036 (0.006), larger than the entry in the Table although still smaller than the corresponding estimates in levels.

[28]The coefficients for the first and second stage are -.0460 (.032) and -.0022 (.511), respectively.

taneity.

5 Citations, Heterogeneity, Cross-Field Spillovers and Robustness

5.1 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 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 8 suggest that the elasticities for the overall number of citations and citations per patent are equal to 0.156 (0.043) and 0.089 (0.040), respectively. The latter suggests that inventors in larger clusters produce not just more patents bit also more influential patents compared to otherwise similar inventors in smaller clusters. The last two columns show 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, as opposed to patents created by inventors in a different city.[29]

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.[30] 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—which is probably not surprising given that inventors in larger clusters create more patents—but also of the number of citations per patent. Since citations made by an inventor are arguably a sign that the inventor knows about a specific innovation, this finding 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 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.

---

[29]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 0 to 1. If a patent is cited by subsequent patents that belong to a wide range of fields the measure of "Generality" is close to 1, whereas if most citations are concentrated in a few fields it is close to 0. "Originality" is defined the same way, except that it refers to citations made. Thus, if a patent cites previous patents that belong to a narrow set of technologies the originality score is close to 0, whereas citing patents in a wide range of fields would render a score close to 1.More specifically, Generality is defined as $1 - \sum_k p_{ik}$ where 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 ion request) indicate that inventors in larger clusters do not tend to produce patents that are more original or general than inventors in smaller clusters.

[30]Inventors in my sample cite an average of 37.2 patents per year. Of these, 5.7 citations are to local patents.

5.2 Are The Effect Larger in Larger Clusters or For More Productive Firms?

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

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 may vary 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% increase in size in a large cluster results in productivity gains that are proportionally smaller than a 1% increase in size in a small cluster. This would be the case, for example, is 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 the 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 0.0523 (0.0113) to 0.0571 (0.0137). A test for equality, reported at the bottom of the table, has p-value equal to 0.224, 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 0.094. The elasticities remain quantitatively similar. The smallest is the one for the first quartile, which is equal to 0.0685 (0.0136), while the largest is the one for the fourth quartile, which is equal to 0.0776 (0.0162) 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 Moretti (2014). 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, Duranton, Gobillon, Puga and Roux, 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.[31] 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, but the difference is quantitatively limited. The elasticity estimated for the top quartile—0.0736 (0.141)—is only slightly larger than the elasticity for the bottom quartile—0.0617 (0.0142). Statistically, I can't reject that the four elasticities are identical, as shown by the p-value reported at the bottom of the table. I conclude that there is limited evidence of heterogeneity in the spillover effect as a function of firm productivity.

5.3 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 does 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 existence and magnitude of cross-field spillovers also have important implications for our understanding of the nature and scope of knowledge spillovers.[32]

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 Appendix Table A.5 show 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-fields 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 size, one for each of the main research fields. This allows me to estimate a $5x5$ matrix where each element is the effect of a field on another field. The matrix is shown in columns 2 to 6.[33] In each column, the sample includes inventors in the field specified at the top. Entries in the first row show the effect of cluster size in a field on productivity in that field, while entries in rows 3 to 7 show the cross fields effects. Empirically, it seems that most of the cross-field effects are not statistically different from zero. One exception is the Biology and Chemistry field in column 2. The productivity of inventors in this field depends not only by the size of the local Biology and Chemistry cluster, but also by the size of the local Computer Science cluster. The cross-field effect is 0.039 (0.0147), about a fifth of the own-field effect. A second exception is the Semiconductors field in column 6. The productivity of inventors in this field depends on size of the local Computer Science cluster and the size of the Other Science cluster. The magnitude of the cross-fields effects is comparable to the own field effect, suggesting that Semiconductor scientist productivity is highly sensitive to the presence of scientists in these two other fields.

Finally, I 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. In

---

[31]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.

[32]More narrowly: If the productivity of scientists depends on cluster size both in their own field and in other fields, omitting cluster size in other fields may result in omitted variable bias. 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%.

[33]Models in this Table do not include $city * year$ effects to avoid multicollinearity.

particular, 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%. The coefficient on cluster size is .064 (.0136), and the one on the interaction is -.0125 (.0039). 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.

5.4   Robustness

Sample Selection. In Section 2.3, I noted that when inventors don't apply for a patent in a given year, they are missing from my data because their location in 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 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 a lower bound of total effect of cluster size on patenting. Here I present three pieces of evidence intended to empirically probe the direction and magnitude of the sample selection bias.

(A) 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 reflects not only the 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 non-missing 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 860,753 observations, indicating that 37,394 observations are interpolated.[34] In column 3, I interpolate the data when 2 missing years are immediately preceded and followed by non-missing 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 873,331 observations.

The estimated coefficients in column 2 are larger than the corresponding coefficients in column 1. Intuitively, by adding some of the missing 0's, 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.

(B) Appendix Table A.7 reports estimates based on different cutoff 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. In the top row, I report the average number of years that inventors are in the sample. It is clear that the more stringent the criterion, the larger the number of years. At one extreme, column 1 indicates that if all inventors are included, the average inventors in present in the sample only 2.1 years. At the other extreme, if only the top 0.5% of inventors are used, the mean inventor is in the sample for an average of 13.60.

I expect the estimated coefficient to grow as we move from left to right, since the problem of sample selection and

---

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

the downward bias should be more pronounced in columns to the left of the Table, and less and less pronounced the more we move to the right. Intuitively, the longer an inventor is in the data, the smaller is the missing extensive margin. The coefficients confirm that estimates are larger if more stringent definitions of stars are adopted. In particular, the estimate for the top 0.5% of inventors (column 6) is larger than estimate for the top 1% (column 5), which in turn is larger than the estimate for the top 5% (column 4), and so on. The estimate for the full population of inventors (column 1) is the smallest and it is indistinguishable from zero.

(C) Appendix Table A.8 reports estimates obtained with different temporal units of analysis. Specifically, I re-estimate my baseline models using measures of productivity defined over 1 month, 2 months, 3 months, 6 months, 2 years and 3 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 was to measure productivity second by second, very few inventor-second pair would be non-missing and the selection bias would be large. By contrast, when the temporal unit of analysis is long (2 or 3 years), I expect the problem of sample selection and the downward bias to be less pronounced. In the extreme, if I was 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 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 0. In column 4 it is positive, although 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% of lifetime patent count. It shows that even for the most prolific inventors the problem of the missing 0's is pronounced when the unit of time is only 1 month, since even the most prolific inventors rarely have patents every single month. As expected, the problem declines the longer the temporal unit. A comparison of Panels A and B shows that for a given unit of time, the coefficients is significantly larger for the most prolific inventors, since they are present in the data more frequently and have fewer missing 0's.

Overall, findings in Appendix Tables A.6 to A.8 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 (Appendix Table A.6) to include some of the 0's, the estimated effects increase precisely because part of the extensive margin gets included in the estimates. When I focus on the very top inventors (top 5%, top 1% or top 0.5%), as in Appendix Table A.7, the estimated effect increases precisely because these prolific inventors are in the data almost continuously over several years, and most of the variation in the dependent variable is from the intensive margin. By contrast, the baseline sample includes many inventors who have either 0 or 1 patent in a typical year. For this group of less prolific inventors, there is no intensive margin variation. Their presence in the baseline sample mechanically reduces the estimated baseline effect. But even superstar inventors rarely have several patents in a given month, or a 2-3 month period. Thus, when I focus on short units of time (Appendix Table A.8), it is not surprising that my estimated effects are small. 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.

Common Shocks. A related concern is the presence of common shocks— namely shocks that affect both the measured number of patents filed in a year and measured number of inventors in that year. This would be the case, for example, if an earthquake or a hurricane in a city lowers the number of patents filed by local inventors in a year and also the measured number of local inventors in that year. In Appendix Table A.9, I probe the extent of the problem. Column 1 reproduces the baseline estimate (Table 3, column 8). Columns 2 and 3 show that using cluster

size lagged by one year or two years, the estimated coefficients are 0.0799 (0.0103) and 0.0531 (0.0106), respectively. To be clear: if the focal inventor is in cluster c in year t, lagged cluster size here is measured as the size of cluster c in year $t-1$ (in column 2) and $t-2$ (in column 3), irrespective of the inventor's location at $t-1$ and $t-2$. This is conceptually different from the model in equation 2 where lagged cluster size refers to the cluster where the focal inventor is at time $t-1$ or $t-2$. In column 4, I use the average cluster size over the the last 3 years. More precisely, if the focal inventor is in cluster c in year t, cluster size is measured as the average size of cluster c in year t, $t-1$ and $t-2$, irrespective of the inventor's location at t, $t-1$ and $t-2$. The coefficient increases to 0.100 (0.0150). In columns 5 and 6, I use the min and max cluster size over the the last 3 years. If the focal inventor is in cluster c in year t, cluster size is measured as the min or max size of cluster c in years t, $t-1$ and $t-2$. The coefficients are 0.0637 (0.0114) and 0.105 (0.015).

Overall, I conclude that there is limited evidence of a positive bias from common shocks. If anything, using lagged measures of cluster size or average measures leads to larger estimates of the productivity effect in most specifications.

Quality of the Cluster. It is possible that productivity spillovers depend not just of the overall size of a cluster, but also on the quality of the inventors in that cluster.[35] 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 Appendix Table A.10 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 3. 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 5. Compared with the baseline elasticity in Table 3, column 8, the elastcities 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 Appendix Table A.10, I control for a quadratic in team size. The coefficient is 0.118 (0.0118), significantly larger than the corresponding baseline coefficient, suggesting that teams size is negatively correlated with inventor 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).[36] The positive coefficient on the interaction in column 7 indicates that the spillover effect is larger for larger teams. The difference is statistically significant, but economically not enormous. The elasticity for inventors in teams larger than the median

---

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

[36] 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 1 if the focal inventor's share of solo patents in the relevant year is above .9.

is 14% larger than the elasticity for inventors in teams smaller than the median. Similarly, the negative coefficient on the interaction in column 8 indicates that the spillover effect is smaller for solo inventors.

6   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 patents in the US. Put differently, 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 macro-economic benefits of agglomeration for the US as a whole. I seek to estimate what would happen to the total number of patents produced annually in each field in the US 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. 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.

6.1   How Spatial Agglomeration Affects the Aggregate Number of Patents

To see how spatial agglomeration might affect the aggregate number of patents produced in the US, 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 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: $S_A = S_B = \frac{S}{2}$. 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 $Y^2 = Sg(\frac{S}{2})$. The change in the aggregate output is

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

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: $[g(\frac{S}{2}) - g(S_A)] < 0$. By contrast, the second

term, which measures the change in number of patents in B, is positive because B has become larger and therefore $[g(\frac{S}{2}) - g(S_B)] > 0$.

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 (2014).[37]

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:

$$Y_{ft}^2 - Y_{ft}^1 = \sum_c S_{cft}[g(\overline{S_{ft}}) - g(S_{cft})] \tag{5}$$

where $\overline{S_{ft}}$ 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 $[g(\overline{S_{ft}}) - g(S_{cft})]$ can be quantified empirically as $[g(\overline{S_{ft}}) - g(S_{cft})] = \overline{S_{ft}}^{\hat{\alpha}} - S_{cft}^{\hat{\alpha}}$ 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.[38]

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.[39]

---

[37]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 IV.B of Kline and Moretti (2014) 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 cities, 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 homogenous tastes for location and perfect mobility, Kline and Moretti (2014) 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 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 Gottlieb (2008) and Gaubert (2018).

[38]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.

[39]Moreover, what matters is how the general equilibrium effects sum up in the aggregate. For each city that in the counterfactual experience a decline in congestion and prices there is a city that experience 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".

6.2 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 9 reports examples of the estimated effect of size equalization on mean inventor productivity for computer science in 2007.[40] My estimate of $\hat{\alpha} = 0.0662$ from column 8 of Table 3 implies that under size equalization, the average productivity of computer scientists in the San Francisco-Silicon Valley region would be 22.52% lower than the observed productivity in 2007. This productivity 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 -16.91%, -16.47%, -14.56% and -13.30%.

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. For example, the average productivity of computer scientists in Kansas City would be 3.18% higher than the observed productivity in 2007. The corresponding figures for Omaha, NE, Portland ME, Memphis, TN and New Orleans are 13.23%, 17.50%, 23.02% and 34.85%.

Table 10 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 27.22% in per inventor productivity. Clusters with sizes between the 25th percentile and the median gain 18.22%. Since the mean is above the median in most fields, clusters with size between the median percentile and the 75th percentile also gain, although only 9.29% By contrast, clusters above the mean lose productivity. The changes in per inventor productivity for clusters with size between 75th and 90th percentile, 90th and 95th percetile, and 96th and 100th percetile are, -0.50%, -8.09% and -14.56%, respectively.

Overall, Tables 9 and 10 indicate that small clusters gain productivity and large cluster lose. The question is what happens in the aggregate. Using equation 5, I estimate the the aggregate effect of equalizing cluster size in each research field on the total number of patents produced in the US in that field. Column 1 Table 11 shows estimates for 2007, by field. The total number of patents created in the US in Computer Science would be 13.18% 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 -9.92%, -14.68%, -7.62%, and -9.64%, respectively. The last row shows the total effect across all fields. The change in the total number of patents produced in the US in 2007 is -11.07%.

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 5.2 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, since a larger elasticity in larger clusters implies that redistribution away from larger clusters is more costly in the aggregate.[41]

Overall, based on estimates in Table 11, I conclude that the aggregate productivity gains from agglomeration are large in the US.

---

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

[41]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 $1/2$ 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%.

7 Conclusions

One of the most remarkable and consequential aspects of the economic geography of the US 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 70%, 79% and 59% 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 0.0662 (0.0138). 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 US as a whole from agglomeration of inventors. My estimates suggest that the overall number of patents created in the US in a given year is 11.07% 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 US.[42] 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.

---

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

Bibliography

- Acemoglu, Daron, Ufuk Akcigit, and William Kerr (2016) "Innovation Network." Proceedings of the National Academy of Sciences 113:41, 11483-11488.

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

- Audretsch, D. B. and M. P. Feldman. (1996): 'RD Spillovers and the Geography of Innovation and Production,' American Economic Review, 86(3): 630-640.

- Azoulay, Pierre, Joshua Graff Zivin and Jialan Wang. (2010) "Superstar Extinction." Quarterly Journal of Economics Vol. 125, No. 2 : 549-589.

- Behrens, K., and F. Robert-Nicoud (2015). "Agglomeration Theory with Heterogeneous Agents", G. Duranton, J. V. Henderson and W. Strange (eds.), Handbook in Regional and Urban Economics, Amsterdam (Holland), Elsevier Press.

- Bernstein, Shai, Timothy McQuade, Richard R. Townsend (2018) "Do Household Wealth Shocks Affect Productivity? Evidence from Innovative Workers During the Great Recession" NBER Working Paper No. 24011

- Bloom, Nicholas, Mark Schankerman, and John Van Reenen. (2013) "Technology Spillovers and Product Market Rivalry." Econometrica Vol. 81, No. 4: 1347–1393

- Bosquet Clément and Pierre-Philippe Combes (2017) Do large departments make academics more productive? Sorting and agglomeration economies in research. Journal of Urban Economics, 101, 27–44.

- Buera, F. and J. Kaboski (2012). "The Rise of the Service Economy", American Economic Review, 102(6), 2540-2569.

- Carlino, G., Carr, J., Hunt, R., Smith, T. (2012). "The agglomeration of R&D labs." Federal Reserve Bank of Philadelphia Working Paper 12-22

- Carlino, Gerald and William R. Kerr (2015) "Agglomeration and Innovation", Handbook of Regional and Urban Economics, Volume 5, Pages 349-404

- Ciccone and Hall (1996) "Productivity and Density of Economic Activity", American Economic Review, vol. 86, no. 1, pp. 54-70.

- Ciccone Antonio and Giovanni Peri (2005) "Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950–1990" The Review of Economics and Statistics 87:4, 652-663

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

- Cohen, W., Nelson, R., Walsh, J. (2000). "Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not)." NBER Working Paper 7552.

- Combes, P.P., G. Duranton, and L.Gobillon (2008). "Spatial Wage Disparities: Sorting Matters!" Journal of Urban Economics 63(2), 723-742.

- 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.

- Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux. (2010) "Estimating Agglomeration Effects With History, Geology, and Worker Fixed-Effects," in Agglomeration Economics, Vo l. 1, ed. by E. L. Glaeser. Chicago, IL: Chicago University Press 15-65.

- Combes, P.P., and L. Gobillon (2015). "The Empirics of Agglomeration Economies" in G. Duranton, J. V. Henderson and W. Strange (eds.), Handbook in Regional and Urban Economics, Volume 5, Amsterdam (Holland), Elsevier Press.

- De La Roca, Jorge and Diego Puga (2017) "Learning by Working in Big Cities" , Review of Economic Studies 84, 106–142

- Diamond, Rebecca (2016a) US Workers' Diverging Locations: Causes and Inequality Consequences Shared Prosperity in America's Communities. University of Pennsylvania Press, Pages 57.

- Diamond, Rebecca (2016b) The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000 American Economic Review. Vol. 106, Issue 3, Pages 479-524.

- Dickinson Mike, (2017) " Kodak's decades of decline" Rochester Business Journal, September 13.

- Duranton, G., (2007) "Urban evolutions: The fast, the slow, and the still", The American Economic Review 97:1, 197-221

- Duranton Gilles and Henry G. Overman, (2005). "Testing for Localization Using Micro-Geographic Data," Review of Economic Studies, Oxford University Press, vol. 72(4), pages 1077-1106.

- Duranton Gilles, Philippe Martin, Thierry Mayer, and Florian Mayneris (2010) "The economics of clusters: Evidence from France", Oxford University Press.

- Duranton Gilles and William R. Kerr (2018) "The logic of agglomeration" New Oxford Handbook of Economic Geography, Clark, Feldman, Gertler and Wojcik, (eds), Oxford University Press, Oxford

- Ellison Glenn and Edward L. Glaeser (1997) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach, Journal of Political Economy, Vol. 105, no. 5: 889-927.

- Gaubert, Cecile (2018) Firm Sorting and Agglomeration. American Economic Review

- Glaeser, Edward and Joshua Gottlieb. (2008) "The Economics of Place-Making Policies." Brookings Papers on Economic Activity 2 155-239.

- Glaeser, Edward L. and David C. Mare. (2001) "Cities And Skills," Journal of Labor Economics, v19(2,Apr), 316-342.

- Glaeser, E., A. Saiz, (2004). The Rise of the Skilled City. Brookings-Wharton Papers on Urban Affairs, 47-105.

- 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: 536-598.

- Griliches, Zvi and Jerry A. Hausman, "Errors in variables in panel data" Journal of Econometrics, Volume 31, Issue 1, 1986, Pages 93-118,

- Griliches, Zvi (1998) "R&D and Productivity: The Econometric Evidence". Chicago: University of Chicago Press

- Griss, Martin (1993) "Hewlett-Packard Software Reuse: from Library to Factory"

- Gruber, Jonathan and Simon Johnson (2019) "Jump-Starting America How Breakthrough Science Can Revive Economic Growth and the American Dream", Public Affairs.

- Guzman Gorge (2019) "Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley", mimeo.

- Hall BH, Griliches Z, Hausman JA (1986) "Patents and R&D: Is there a lag?" International Economic Review 27: 265-283

- Haran, Hadas (2011) "Time to Market Research: Highlights and Key Findings"

- Helmers, Christian and Henry G. Overman, (2017) "My precious! The location and diffusion of scientific research: evidence from the Synchrotron Diamond Light Source," The Economic Journal, Volume 127, Issue 604

- Henderson, Vernon (2003) "Marshall's Scale Economies" Journal of Urban Economics, 53, 1-28

- Jaffe, Adam B., Manuel Trajtenberg, Rebecca Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," The Quarterly Journal of Economics, Volume 108, Issue 3

- Kantor Shawn and Alex Whalley (2014) "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks" Review of Economics and Statistics, March, 96(1): 171-188.

- Kantor Shawn and Alex Whalley (forthcoming) "Research Proximity and Productivity: Long-Term Evidence From Agriculture" Journal of Political Economy

- Kapoor, Rahul (2012) "Managing Complexity and Change in the Semiconductor Industry" Findings From the Wharton-Atreg Industry Study 5, 14

- Kerr, William R. (2010) "Breakthrough inventions and migrating clusters of innovation," Journal of Urban Economics, Elsevier, vol. 67(1), pages 46-60, January

- Kerr, William R. (2018) "Navigating Talent Hot Spots." Harvard Business Review 96, no. 5: 80–86.

- Kerr, William and Scott Kominers (2015) "Agglomerative Forces and Cluster Shapes," Review of Economics and Statistics 97:4 (October), 877-899.

- Kline, Pat and Enrico Moretti (2014) "Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority", Quarterly Journal of Economics, 129(1).

- Kline, Pat and Enrico Moretti (2014b) "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs", Annual Review of Economics, 6.

- Krasner, Jerry (2003) "Total Cost of Development: A Comprehensive Cost Estimation Framework for Evaluating Embedded Development Platforms"

- Lerner, Josh and Seru, Amit. (2017). The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond. SSRN Electronic Journal.

- Lucas, R.E. (1988). "On the mechanics of economic development." Journal of Monetary Economics, 22, 3-42.

- Lychagin, Sergey, Joris Pinkse, Margaret Slade, and John Van Reenen. (2016) "Spillovers in Space: Does Geography Matter?" Journal of Industrial Economics Vol. 64, No. 2: 295-335.

- Mansfield (1972) Edwin "Research and Innovation in the Modern Corporation" Norton Company

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

- Melo, P. C., D. J. Graham, and R. B. Noland. (2009) "A Meta-Analysis of Estimates of Urban Agglomeration Economies," Regional Science and Urban Economics, 39 (3) 332-342.

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

- Moretti, Enrico (2012) "The New Geography of Jobs", Houghton Mifflin Harcourt.

- Moretti, Enrico and Daniel Wilson (2013) "State Incentives for Innovation, Star Scientists and Jobs: Evidence from Biotech", The Journal of Urban Economics, 79.

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

- Moser, Petra, Alessandra Voena, and Fabian Waldinger (2014) "German Jewish Émigrés and US Invention" American Economic Review, 104(10): 3222–3255

- Rosenthal, S.S and W.C. Strange (2004). "Evidence on the Nature and Sources of Agglomeration Economies", in Henderson, J.V. and Thisse, J.-F. (Eds.), Handbook of Urban and Regional Economics, Volume 4, Amsterdam: Elsevier, 2119-2172.

- Rosenthal, Stuart S. and William Strange (2006). "The Micro-Empirics of Agglomeration Economies," in A Companion to Urban Economics, Daniel P. McMillen and Richard Arnott (eds.), pp 7-23, Blackwell

- Rosenthal, Stuart S. and William Strange (2008). "The Attenuation of Human Capital Spillovers," Journal of Urban Economics, 64(2), 373-389.

- Waldinger, Fabian (2012) Peer Effects in Science – Evidence from the Dismissal of Scientists in Nazi Germany The Review of Economic Studies, vol. 79, no. 2, pp. 838-861

- Williams Heidi (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"

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

  Zucker, LynneG and MichaelR. Darby "Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnologyindustry" Proceedings of the National Academy of Sciences Nov 1996, 93 (23) 12709-12716

- Zucker, Lynne G., Michael R. Darby, Jason Fong (2014) Communitywide Database Designs for Tracking Innovation Impact: COMETS, STARS and Nanobank, Annals of Economics and Statistics, Annals of Economics and Statistics No. 115-116, (December), pp. 277-311

Table 1: Largest Clusters in "Computer Science", "Biology and Chemistry" and "Semiconductors" - 2007

	Size
(A) Computer Science	
San Jose-San Francisco-Oakland, CA	.2617478
New York-Newark-Bridgeport, NY-NJ-CT-PA	.091019
Seattle-Tacoma-Olympia, WA	.0839432
Austin-Round Rock, TX	.0596916
Boston-Worcester-Manchester, MA-NH	.0478379
Los Angeles-Long Beach-Riverside, CA	.0399758
Minneapolis-St. Paul-St. Cloud, MN-WI	.0338676
Raleigh-Durham-Cary, NC	.0287269
Denver-Aurora-Boulder, CO	.0237678
San Diego-Carlsbad-San Marcos, CA	.0227396
Portland-Vancouver-Beaverton, OR-WA	.0221953
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	.0194134
Dallas-Fort Worth, TX	.0155428
Chicago-Naperville-Michigan City, IL-IN-WI	.0153614
(B) Biology and Chemistry	
New York-Newark-Bridgeport, NY-NJ-CT-PA	.113863
San Jose-San Francisco-Oakland, CA	.111812
Boston-Worcester-Manchester, MA-NH	.0697689
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	.0632221
Los Angeles-Long Beach-Riverside, CA	.0593068
San Diego-Carlsbad-San Marcos, CA	.0443517
Minneapolis-St. Paul-St. Cloud, MN-WI	.0376765
Houston-Baytown-Huntsville, TX	.0314506
Chicago-Naperville-Michigan City, IL-IN-WI	.0309371
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	.028113
Raleigh-Durham-Cary, NC	.0200899
Seattle-Tacoma-Olympia, WA	.0195764
Cincinnati-Middletown-Wilmington, OH-KY-IN	.0162388
Indianapolis-Anderson-Columbus, IN	.0156611
(C) Semiconductors	
San Jose-San Francisco-Oakland, CA	.255391
New York-Newark-Bridgeport, NY-NJ-CT-PA	.149261
Los Angeles-Long Beach-Riverside, CA	.061945
Dallas-Fort Worth, TX	.0503171
Phoenix-Mesa-Scottsdale, AZ	.0498943
Boise City-Nampa, ID	.0469345
Portland-Vancouver-Beaverton, OR-WA	.0463002
Austin-Round Rock, TX	.0399577
Burlington-South Burlington, VT	.0361522
Boston-Worcester-Manchester, MA-NH	.035518
Albany-Schenectady-Amsterdam, NY	.0215645
Minneapolis-St. Paul-St. Cloud, MN-WI	.0158562
Raleigh-Durham-Cary, NC	.0145877
San Diego-Carlsbad-San Marcos, CA	.0133192

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

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

	(1)	(2)	(3)	(4)	(5) Weighted
Panel (A)					
Rochester × 2007	-0.0658***	-0.0696***	-0.0835***	-0.0984***	-0.101***
	(0.00759)	(0.00678)	(0.00630)	(0.00640)	(0.00802)
Rochester	-0.0132	-0.0346***	-0.0296***	-	-
	(0.0105)	(0.0101)	(0.00982)		
2007	-0.190***	-0.189***	-	-	-
	(0.00759)	(0.00707)			
N	194110	194110	194110	194110	193194
Field		Y	Y	Y	Y
Field × Year			Y	Y	Y
Field × City				Y	Y
Panel (B): Within-Inventor Difference in Difference Estimates					
Rochester × 2007	-0.206***	-0.220***	-0.256***	-0.321***	-0.403***
	(0.0771)	(0.0766)	(0.0800)	(0.0744)	(0.0375)
2007	-0.206***	-0.206***	-	-	-
	(0.0190)	(0.0193)			
N	16426	16426	16426	16426	16375
Inventor	Y	Y	Y	Y	Y
Field		Y	Y	Y	Y
Field × Year			Y	Y	Y
Field × City				Y	Y

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 fields, 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 clustered by city in parenthesis.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Size	0.0514***	0.0751***	0.0870***	0.0897***	0.0668***	0.0910***	0.0514***	0.0662***
	(0.00812)	(0.0166)	(0.0186)	(0.00924)	(0.00861)	(0.00980)	(0.0116)	(0.0138)
N	932008	932008	932008	932008	932008	932008	932008	823359
Year	y	y	y	y	y	y	y	y
City	y	y	y	y	y	y	y	y
Field	y	y	y	y	y	y	y	y
Class	y	y	y	y	y	y	y	y
City × Field		y	y	y	y	y	y	y
City × Class			y	y	y	y	y	y
Field × Year				y	y	y	y	y
Class × Year					y	y	y	y
Inventor						y	y	y
City × Year							y	y
Firm								y

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.

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

	(1)	(2)	(3)	(4)	(5) Weighted
First Stage					
Rochester× 2007	-0.492***	-0.464***	-0.451***	-0.396***	-0.387***
	(0.128)	(0.113)	(0.115)	(0.0455)	(0.0460)
F stat.	14.68	16.84	15.31	75.71	70.54
Second Stage					
Cluster Size	0.134***	0.150***	0.185***	0.249***	0.260***
	(0.0380)	(0.0379)	(0.0499)	(0.0398)	(0.0459)
Rochester	Y	Y	Y	Y	Y
2007	Y	Y	Y	Y	Y
Field		Y	Y	Y	Y
Field × Year			Y	Y	Y
Field × City				Y	Y

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 years 1996 and 2007, excluding those employed in the Photography sector. Standard errors clustered by city in parenthesis.

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

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	2SLS (7)	2SLS (8)	2SLS (9)	2SLS (10)	2SLS (11)	2SLS (12)
Panel (A): All												
Δ log Size	0.0144***	0.0148***	0.0156***	0.0169***	0.0167***	0.0162***	0.0291	0.0475**	0.0345*	0.0315**	0.0310*	0.0307**
	(0.00376)	(0.00378)	(0.00366)	(0.00387)	(0.00381)	(0.00368)	(0.0177)	(0.0200)	(0.0179)	(0.0128)	(0.0160)	(0.0147)
First Stage							0.983***	0.948***	0.965***	1.238***	1.272***	1.278***
							(0.105)	(0.127)	(0.126)	(0.189)	(0.169)	(0.168)
F stat.							87.3	56.0	59.0	42.9	56.6	57.6
N	419599	419599	419568	405095	405095	403942	419599	419599	419568	405095	405095	403942
Panel (B): Stayers – Same City and Field												
Δ log Size	0.0340***	0.0336***	0.0348***	0.0448***	0.0438***	0.0415***	0.110*	0.131*	0.0897	0.128**	0.107*	0.124*
	(0.00803)	(0.00804)	(0.00806)	(0.00935)	(0.00917)	(0.00906)	(0.0474)	(0.0536)	(0.0431)	(0.0373)	(0.0474)	(0.0433)
First Stage							0.283***	0.266***	0.273***	0.368***	0.380***	0.380***
							(0.0520)	(0.0508)	(0.0476)	(0.0521)	(0.0489)	(0.0503)
F stat.							29.7	27.4	32.8	49.8	60.2	56.9
N	176263	176263	176245	166253	166251	164454	176263	176263	176245	166253	166251	164454
Year	y	y	y	y	y	y	y	y	y	y	y	y
Field		y	y	y	y	y		y	y	y	y	y
Class			y	y	y	y			y	y	y	y
Firm				y	y	y				y	y	y
Field × Year					y	y					y	y
Class × Year						y						y

Notes: Each entry is a separate regression. Dependent variable is the change in the log number of patents in a year. The sample in Panel A includes inventors observed in consecutive years. The sample in Panel B includes inventors observed in consecutive years who do not change city or field in the sample period. The model estimated is equation 3. The instrumental variable for workers in firm j in cluster fct is defined as $IV_{jfct} = \sum_{s \neq j} D_{sfc(t-1)} \frac{\Delta N_{sfc(-c)t}}{\Delta N_{ft}}$ where $D_{jcf(t-1)}$ is an indicator equal to 1 if firm j has at least 1 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9) Same City	(10) Different City
			(A) Dependent Variable: Number of Citations							
Log Size	0.119***	0.168**	0.192**	0.202***	0.226***	0.205***	0.154***	0.156***	0.199**	0.0279
	(0.0335)	(0.0782)	(0.0937)	(0.0449)	(0.0391)	(0.0306)	(0.0422)	(0.0435)	(0.0790)	(0.0302)
			(B) Dependent Variable: Number of Citations Per Patent							
Log Size	0.0676**	0.0928	0.105	0.113***	0.160***	0.114***	0.102***	0.0894**	0.133*	-0.0375
	(0.0303)	(0.0672)	(0.0812)	(0.0435)	(0.0370)	(0.0279)	(0.0394)	(0.0400)	(0.0748)	(0.0273)
N	932008	932008	932008	932008	932008	932008	932008	823359	730280	730280
Year	y	y	y	y	y	y	y	y	y	y
City	y	y	y	y	y	y	y	y	y	y
Field	y	y	y	y	y	y	y	y	y	y
Class	y	y	y	y	y	y	y	y	y	y
City × Field		y	y	y	y	y	y	y	y	y
City × Class			y	y	y	y	y	y	y	y
Field × Year				y	y	y	y	y	y	y
Class × Year					y	y	y	y	y	y
Inventor						y	y	y	y	y
City × Year							y	y	y	y
Firm								y	y	y

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 9 is for citations by inventors located in the same city as the focal inventor. Column 10 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

Table 7: Citations By Focal Inventor

	Citations	Citations Per Patent	Share Same City	Share Same Field
	(1)	(2)	(3)	(4)
Log Size	0.158***	0.0922***	0.00771***	0.00561*
	(0.0166)	(0.0138)	(0.00269)	(0.00321)
N	810480	810480	810480	810480

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.

Table 8: Heterogeneity in Elasticity By Cluster Size and By Firm Productivity

	(1)	(2)
(A) Heterogeneity By Cluster Size		
First Quartile (Smallest)	0.0523***	0.0685***
	(0.0113)	(0.0136)
Second Quartile	0.0563***	0.0743***
	(0.0118)	(0.0140)
Third Quartile	0.0537***	0.0720***
	(0.0121)	(0.0144)
Fourth Quartile (Largest)	0.0571***	0.0776***
	(0.0137)	(0.0162)
N	932008	823359
p-value equal elasticities	0.22	0.09
Firm		y
(B) Heterogeneity By Firm Productivity		
First Quartile (Least Productive)	0.0528***	0.0617***
	(0.0122)	(0.0142)
Second Quartile	0.0514***	0.0670***
	(0.0117)	(0.0140)
Third Quartile	0.0528***	0.0737***
	(0.0116)	(0.0142)
Fourth Quartile (Most Productive)	0.0545***	0.0736***
	(0.0118)	(0.0141)
N	932008	823359
p-value equal elasticities	0.94	0.12
Firm		y

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.

Table 9: 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	-22.52%
New York-Newark-Bridgeport, NY-NJ-CT-PA	-16.91%
Seattle-Tacoma-Olympia, WA	-16.47%
Austin-Round Rock, TX	-14.56%
Boston-Worcester-Manchester, MA-NH	-13.30%
Minneapolis-St. Paul-St. Cloud, MN-WI	-11.29%
Raleigh-Durham-Cary, NC	-10.32%
San Diego-Carlsbad-San Marcos, CA	-8.92%
Portland-Vancouver-Beaverton, OR-WA	-8.78%
Pittsburgh-New Castle, PA	-2.63%
Boise City-Nampa, ID	-2.53%
Examples of Winners	
Miami-Fort Lauderdale-Miami Beach, FL	1.50%
Kansas City-Overland Park-Kansas City, MO-KS	3.18%
Buffalo-Niagara-Cattaraugus, NY	6.37%
Omaha-Council Bluffs-Fremont, NE-IA	13.23%
Des Moines-Newton-Pella, IA	14.39%
Portland-Lewiston-South Portland, ME	17.50%
Scranton-Wilkes-Barre, PA	21.22%
Toledo-Fremont, OH	23.02%
Memphis, TN-MS-AR	23.02%
Oklahoma City-Shawnee, OK	25.39%
New Orleans-Metairie-Bogalusa, LA	34.85%

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

Table 10: Effect of Cluster Size Equalization on Inventor Productivity - All Fields, 2007

	Percent Change
0-25th Percentile	27.22%
25th-50th Percentile	18.22%
50th-75th Percentile	9.29%
75th-90th Percentile	-0.50%
90th-95th Percentile	-8.09%
95th-100th Percentile	-14.56%

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

Table 11: Aggregate Effects of Cluster Size Equalization – 2007

	Constant Elasticity (1)	Heterogeneous Elasticty (2)
"Computer Science"	-13.18%	-14.54%
"Biology and Chemistry"	-9.92%	-11.23%
"Semiconductors"	-14.68%	-16.09%
"Other Engineering"	-7.62%	-8.62%
"Other Science"	-9.64%	-10.94%
All Fields	-11.07%	-12.35%

Notes: Entries are the estimated percent difference between total number of patents created in the US in the counterfactual scenario and observed number of patents created in the US. 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).

Figure 1: Share of Top Ten Cities Over Time

"Computer Science"

"Biology and Chemistry"

"Semiconductors"

Figure 2: Kodak Decline

Kodak Stock Price
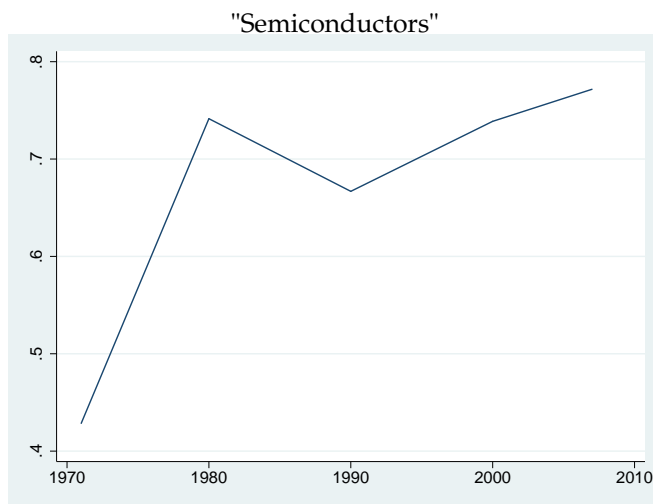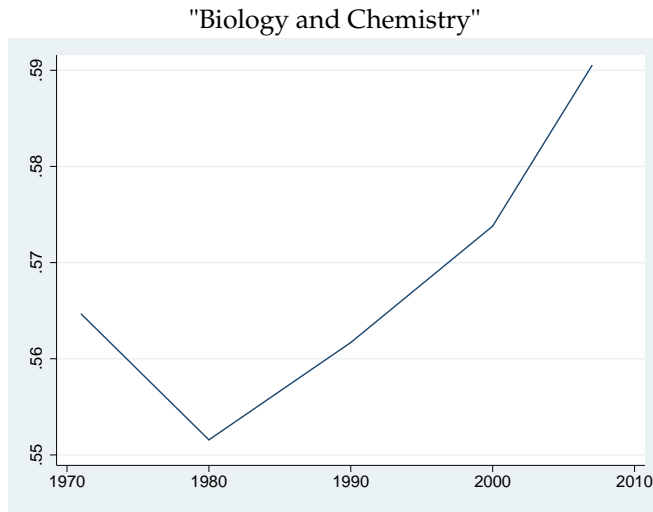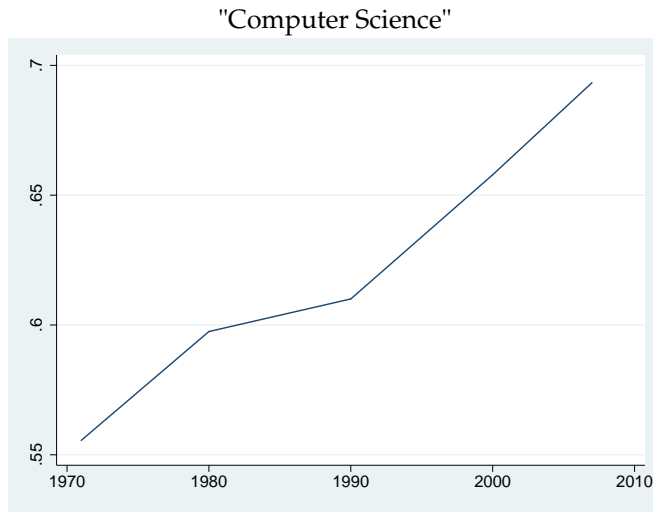
Kodak Inventors in Rochester
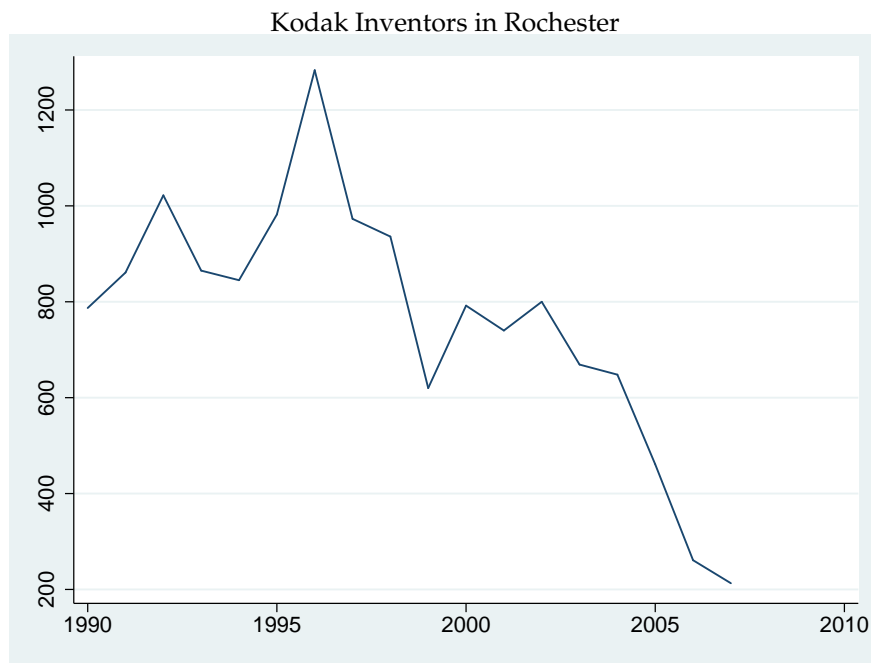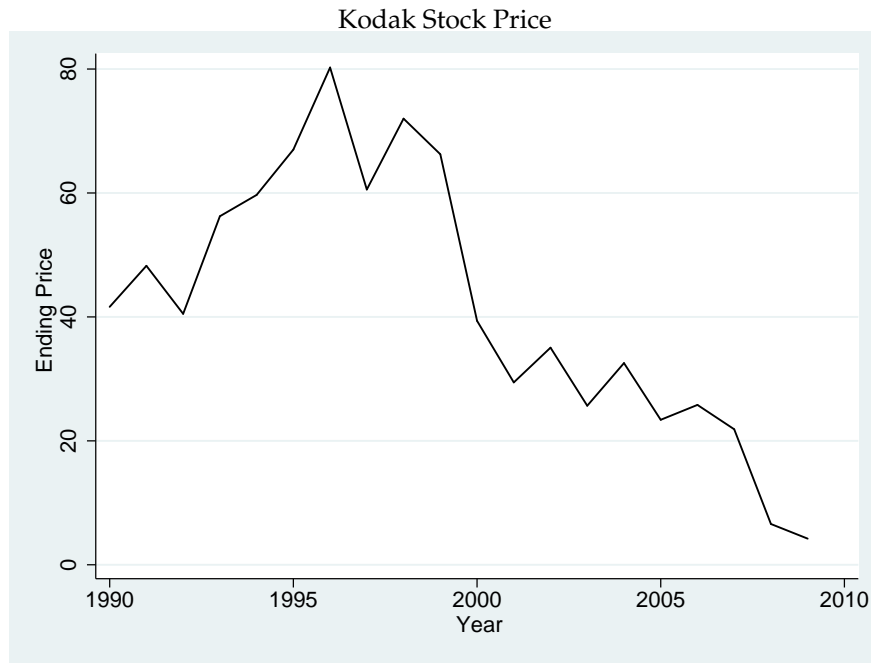
Figure 3: Average Inventor Productivity in Rochester <u>Outside Kodak</u>
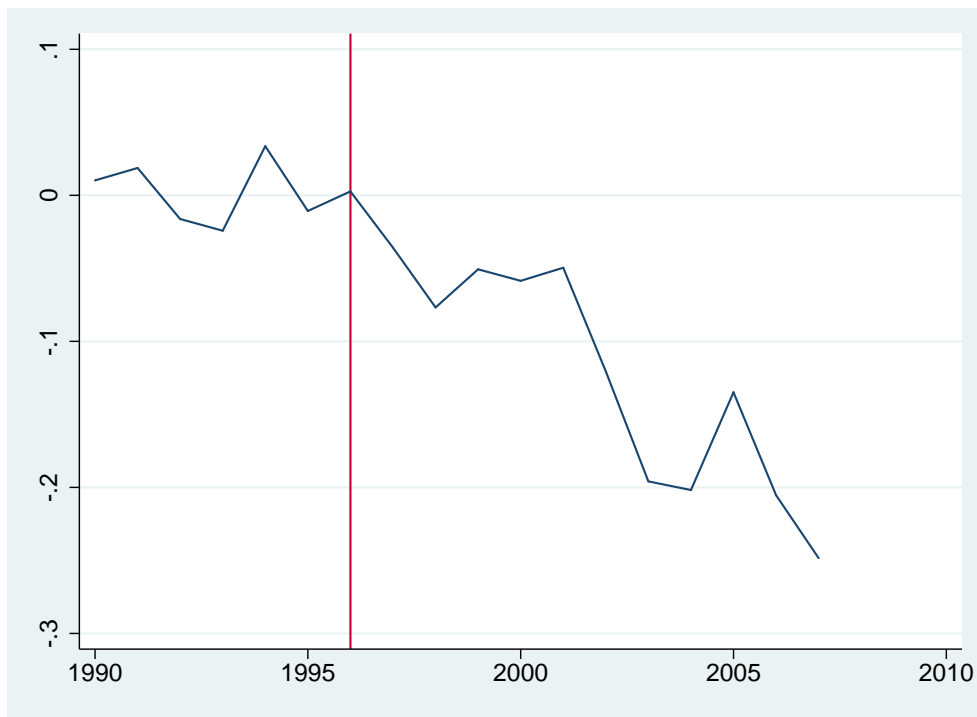
Notes: Controls include research field dummies.

Figure 4: Average Log Number of Patents Per Inventor Per Year and Log Cluster Size — All Years and Fields
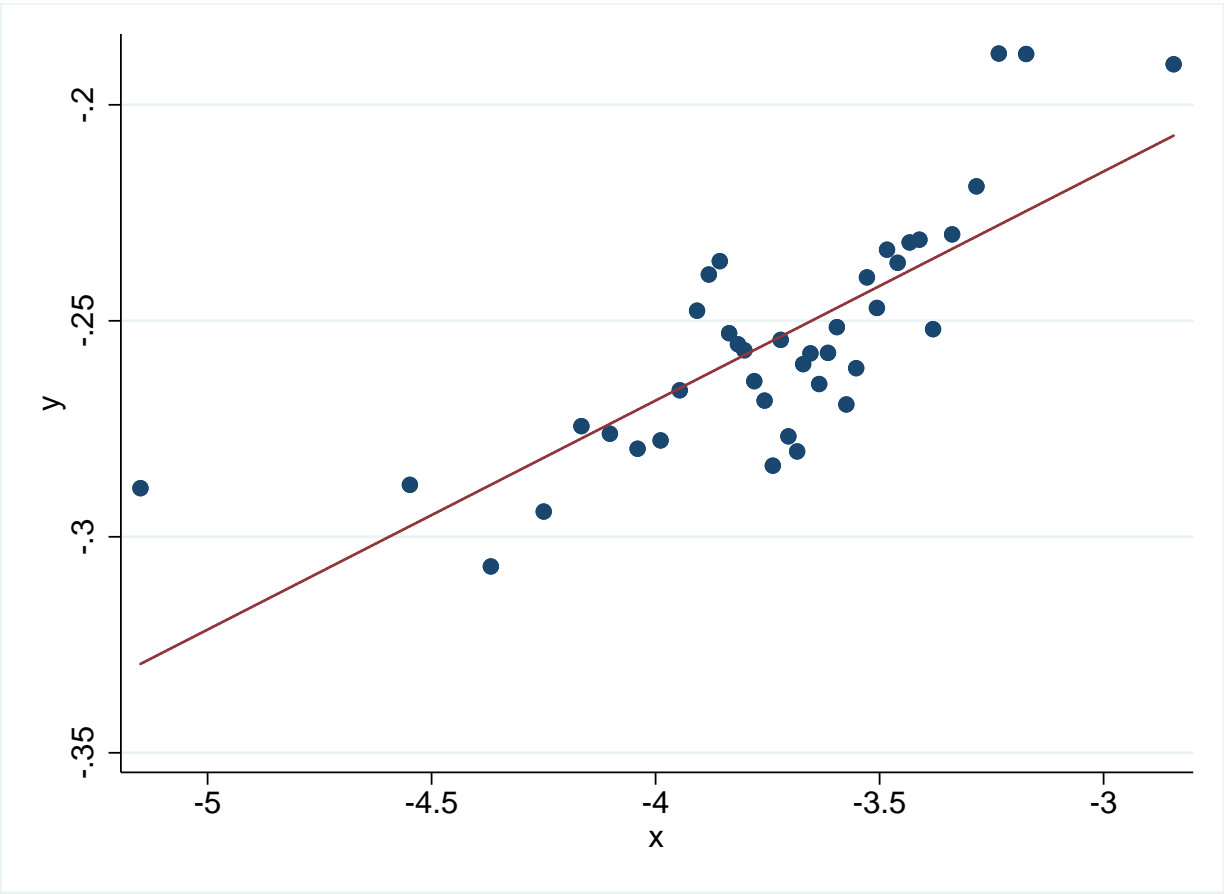
Notes: The slope is 0.053 (0.008). Controls include dummies for year, field and city.

Figure 5: Dynamic Response Following a Change in Cluster Size
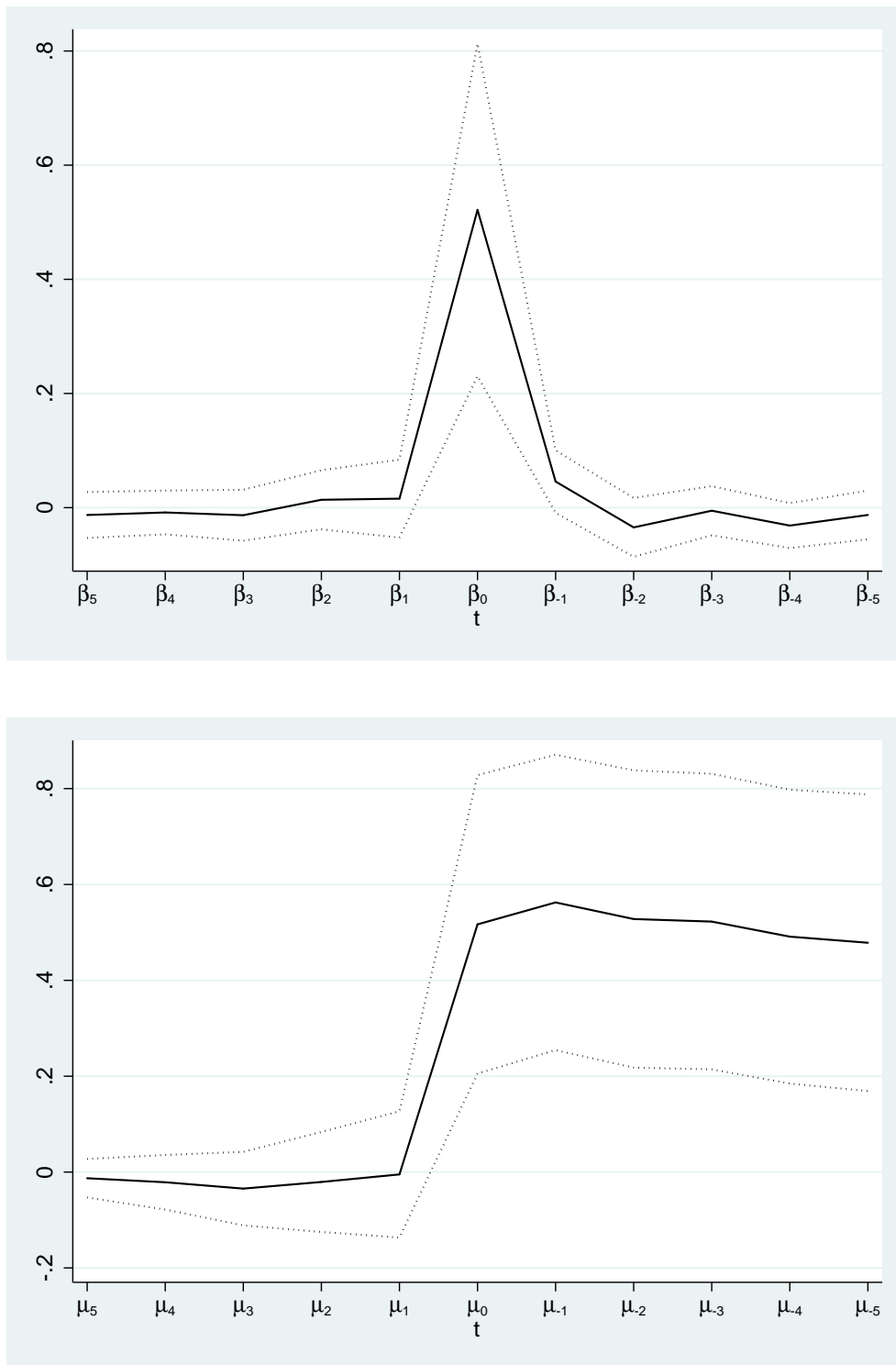
Notes: This figure is based on equation 2 in the text. In the top panel I plot the estimated β coefficients in equation 2 on the lag and lead terms. For example, β_5 is the coefficient on the fifth lead term. In the bottom panel, I plot the cumulative response, where the μ's are defined as: $\mu_n = \beta_5 + \beta_4 + ... + \beta_n$ for $n = -5$ through 5. Only inventors who are 11 consecutive years in the data are included. N=21787. Standard errors are clustered by city $\times$ research field.

Figure 6: Dynamic Response Following a Change in Cluster Size – Movers
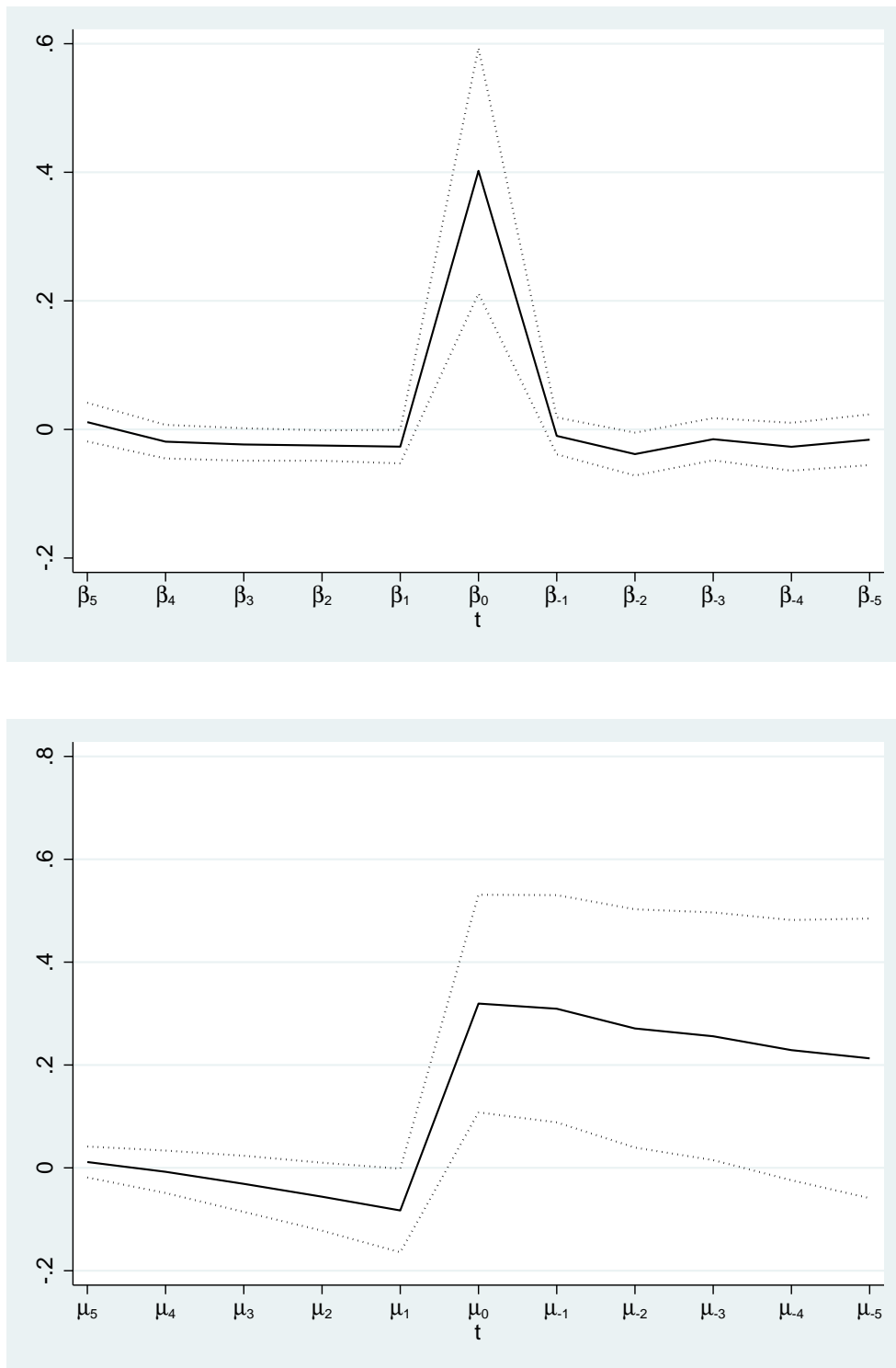
Notes: In the top panel I plot the estimated β coefficients on the lag and lead terms. In the bottom panel, I plot the cumulative response, where the μ's are defined as: $\mu_n = \beta_5 + \beta_4 + ... + \beta_n$ for $n = -5$ through 5. Only inventors who are 11 consecutive years in the data and move across cities are included. Standard errors are clustered by city $\times$ research field.

Online Appendix

Appendix Table A.1: Largest Patent Assignees by Total Number of Patents 1971-2007

	Number of Patents
INT BUSINESS MACHINES	155708
GEN ELECT	69019
MICROSOFT	43550
INTEL	42087
EASTMAN KODAK	41553
MOTOROLA	40977
XEROX	35042
MICRON TECHNOLOGY	32010
TEXAS INSTRUMENTS	27886
E I DU PONT DE NEMOURS	25291
HEWLETT-PACKARD DEV LP	25022
AT&T	24891
ADV MICRO DEVICES	21280
DOW CHEM	19866
GEN MOTORS	19756
US OF AMER AS REPRESENTED BY SECR. OF NAVY	19698
LUCENT TECHNOLOGIES	18863
EXXON RES ENGN	18562
HONEYWELL INT	18195
PROCTER GAMBLE	17746
REGENTS OF UNIV OF CALIFORNIA	16800
APPL MATERIALS	16441
HEWLETT-PACKARD	15980
SUN MICROSYSTEMS	15357
BOEING	14698

Appendix Table A.2: Some Examples of Firm-Specific Productivity Spillovers – 2007

Firm	City	Estimated Productivity Spillover
(A) Computer Science		
MICROSOFT	Seattle-Tacoma-Olympia, WA	.0806
INT BUSINESS MACHINES	Minneapolis-St. Paul-St. Cloud, MN-WI	.0415
CISCO TECHNOLOGY	San Jose-San Francisco-Oakland, CA	.0020
DELL PROD	Austin-Round Rock, TX	.0061
TEXAS INSTRUMENTS	Dallas-Fort Worth, TX	.0184
CATERPILLAR	Peoria-Canton, IL	.1047
MOTOROLA	Chicago-Naperville-Michigan City, IL-IN-WI	.0088
HEWLETT-PACKARD	San Jose-San Francisco-Oakland, CA	.0007
Average Firm	**Average City**	.0032
(B) Biology and Chemistry		
E I DU PONT DE NEMOURS	Philadelphia-Camden-Vineland, PA-NJ-DE-MD	.0117
BRISTOL-MYERS SQUIBB	New York-Newark-Bridgeport, NY-NJ-CT-PA	.0054
PROCTER GAMBLE	Cincinnati-Middletown-Wilmington, OH-KY-IN	.0347
AMGEN	Los Angeles-Long Beach-Riverside, CA	.0065
CHEVRON RES	San Jose-San Francisco-Oakland, CA	.0014
3M	Duluth, MN-WI	.0899
PFIZER	Hartford-West Hartford-Willimantic, CT	.0124
EXXON RES ENGN	Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	.0043
Average Firm	**Average City**	.0024

Notes: Entries reflect the impact that a specific firm is estimated to have on the productivity of scientists in other firms in the same cluster in 2007. For a given firm j, field f and city c, entries are obtained as $\hat{\alpha}\Delta S_{-jfct}$ where $\hat{\alpha} = 0.066$ is the estimated elasticity in Table 3, column 8 and $\Delta \ln S_{-jfct}$ is the difference in log cluster size with and without a given firm. In particular, $\Delta \ln S_{-jfct} = [\ln(N_{fct}/N_{ft}) - \ln(N_{jfct}/N_{ft})]$, where N_{fct} in 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 $t = 2007$.

Appendix Table A.3: Field-Specific Elasticities

	Comp. Science	Biol. and Chem.	Semiconductors	Other Engineer.	Other Science
	(1)	(2)	(3)	(4)	(5)
Log Size	0.194***	0.140***	0.258***	0.105***	0.0766***
	(0.0400)	(0.0199)	(0.0650)	(0.0213)	(0.0289)
N	77263	198803	38319	427985	80989

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. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, and Firm effects. City $\times$ Year are not included. Standard errors are clustered by city $\times$ research field.

Appendix Table A.4: Models in Differences: Effect of Changes in Cluster Size on Changes in Inventor Productivity – OLS and IV Estimates — Single Location Firms

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
Panel (A): All						
Δ log Size	0.00121	0.00167	0.00245	0.0770**	0.0698**	0.0505
	(0.00667)	(0.00677)	(0.00670)	(0.0444)	(0.0408)	(0.0410)
First Stage				1.230***	1.263***	1.300***
				(0.171)	(0.175)	(0.161)
F stat.				51.8	52.0	64.9
N	53610	53608	51182	53610	53608	51182
Panel (B): Stayers – Same City and Field						
Δ log Size	0.00751	0.00693	0.00758	0.153**	0.145**	0.105
	(0.00851)	(0.00863)	(0.00890)	(0.0725)	(0.0763)	(0.114)
First Stage				0.868***	0.897***	0.915***
				(0.129)	(0.141)	(0.124)
F stat.				45.4	40.3	54.3
N	37767	37762	35124	37767	37762	35124
Year	y	y	y	y	y	y
Field	y	y	y	y	y	y
Class	y	y	y	y	y	y
Field $\times$ Year		y	y		y	y
Class $\times$ Year			y			y

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 instrumental variable for workers 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 $D_{jcf(t-1)}$ is an indicator equal to 1 if firm j has at least 1 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.

Appendix Table A.5: Cross-Field Spillovers

	All	Biology and Chem.	Computer Sc.	Other Eng.	Other Sci.	Semicond.
	(1)	(2)	(3)	(4)	(5)	(6)
Own Field	0.0621***	0.167***	0.237***	0.123***	0.0980**	0.207***
	(0.0206)	(0.0371)	(0.0399)	(0.0317)	(0.0480)	(0.0645)
Mean of Other Fields	-0.01297					
	(0.04628)					
Biology and Chem			-0.0592	-0.0182	0.0370	-0.0729
			(0.0569)	(0.0202)	(0.0487)	(0.0861)
Computer Sc.		0.0329**		0.0139	-0.00710	0.162***
		(0.0147)		(0.00890)	(0.0283)	(0.0594)
Other Eng.		-0.0102	0.0377		-0.0101	-0.185
		(0.0391)	(0.103)		(0.0734)	(0.206)
Other Sc.		-0.0348	-0.00925	-0.00993		0.192***
		(0.0250)	(0.0608)	(0.0129)		(0.0548)
Semicond.		0.00551	-0.0376	0.00822	0.00433	
		(0.00667)	(0.0302)	(0.00527)	(0.0172)	
N	822291	174840	74831	367503	74138	38043

Notes: Each column is a regression. Column 1 includes all inventors. The sample in columns 2 to 6 includes inventors in the field specified at the top. Entries in a given row show the coefficient on the field-specific cluster size. For example, the entry in row 3, column 3 show the effect of the size of the Biology and Chemistry cluster on the productivity of Computer Scientists. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, and Firm effects.

Appendix Table A.6: Interpolation

		Interpolation	
	Baseline Sample	1 Year	2 Years
	(1)	(2)	(3)
(A): Inverse Hyperbolic Sine			
log Size	0.0533***	0.0651***	0.0670***
	(0.0086)	(0.0082)	(0.0081)
(B): log(patents+1)			
log Size	0.0401***	0.0496***	0.0513***
	(0.0064)	(0.0062)	(0.0061)
N	823359	860753	873311

Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, City $\times$ Year and Firm effects. Standard errors are clustered by city $\times$ research field.

Appendix Table A.7: Alternative Definitions of Star Inventors

	All	Top 25%	Top 10% (Baseline)	Top 5%	Top 1%	Top 0.5%
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Years in Sample	2.14	4.58	7.01	8.94	13.20	13.60
log Size	0.0189	0.0380***	0.0662***	0.104***	0.247***	0.269***
	(0.0117)	(0.0115)	(0.0138)	(0.0179)	(0.0406)	(0.0419)
N	2355529	1344991	823359	526344	155210	131805

Notes: Dependent variable is log number of patents in a year. Baseline is the entry in Col 8 of Table 3. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, City $\times$ Year and Firm effects. Standard errors are clustered by city$\times$research field.

Appendix Table A.8: Alternative Units of Time

	1 Month	2 Months	3 Months	6 Months	1 Year (Baseline)	2 Years	3 Years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) All							
Log Size	-0.0249***	-0.0114	0.000793	0.0302***	0.0662***	0.130***	0.164***
	(0.00852)	(0.00933)	(0.00985)	(0.0107)	(0.0138)	(0.0147)	(0.0192)
N	1321679	1234595	1165614	1013440	823359	610122	500797
(B) Top 1% Inventors							
Log Size	0.0197	0.0490**	0.0743***	0.138***	0.247***	0.352***	0.410***
	(0.0165)	(0.0203)	(0.0221)	(0.0276)	(0.0406)	(0.0502)	(0.0652)
N	344992	304424	274620	216426	155210	100738	77521

Notes: Dependent variable is log number of patents in unit of time. Baseline in Panel A is the entry in Col 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.

Appendix Table A.9: Alternative Measures of Cluster Size

|	(1)	(2)	(3)	(4)	(5)	(6)
(a) Baseline	0.0662***					
	(0.0138)					
(b) 1 Year Ago		0.0799***				
		(0.0103)				
(c) 2 Years Ago			0.0531***			
			(0.0106)			
(d) Mean Over 3 Years				0.100***		
				(0.0150)		
(e) Min Over 3 Years					0.0637***	
					(0.0114)	
(f) Max Over 3 Years						0.105***
						(0.0150)
N	823359	821644	820155	823572	823572	823572

Notes: Dependent variable is log number of patents in a year. Baseline is the entry in Col 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.

Appendix Table A.10: Cluster Quality and Teams

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log Size	0.124***	0.133**	0.114***	0.133**	0.118***	0.238***	0.0815***	0.0870***
	(0.0466)	(0.0556)	(0.0376)	(0.0573)	(0.0118)	(0.0448)	(0.0130)	(0.0138)
log Size X 1(Team size $\geq$ Median)							0.0124***	
							(0.00234)	
log Size X (Solo Inventor)								-0.0194***
								(0.00405)
N	823521	823179	823490	823419	823359	823359	822884	823359

Notes: Each entry is a separate regression. Dependent variable is log number of patents in a year. In column 1, 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 3. 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. In column 4, cluster size is measured as the number of inventors with a lifetime patent citation count above 5. Column 5 controls for a quadratic in team size. In column 6, cluster size is defined excluding all members of the focal inventor's team. In column 7, cluster size is interacted with a dummy equal to 1 if the focal inventor team has size above median. In column 8, cluster size is interacted with a dummy equal to 1 if the focal inventor share of patents where he is the only inventor is above .9. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, City $\times$ Year and Firm effects. Standard errors are clustered by city $\times$ research field.

Appendix Figure A.1: Dynamic Response Following a Change in Cluster Size – Pharmaceutical Only
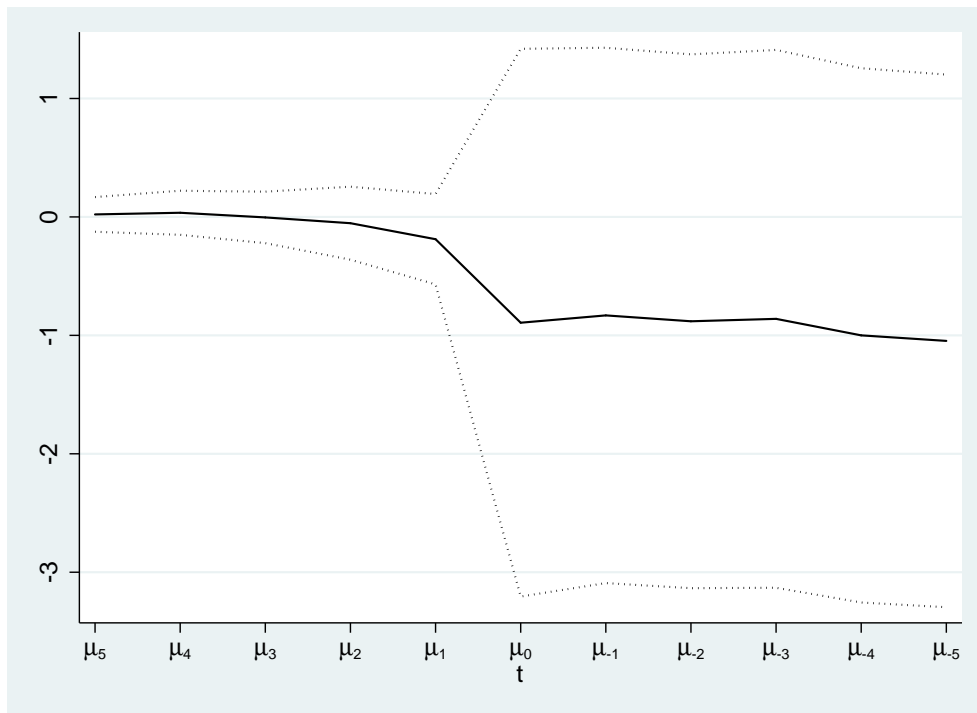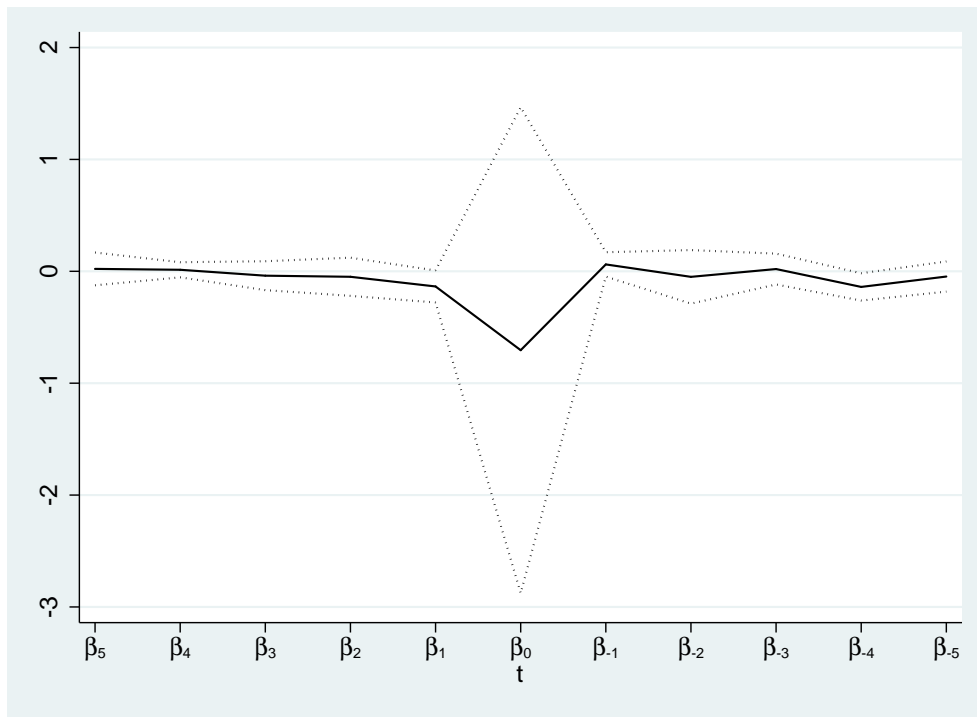
Notes: This figure is based on equation 2 in the text. In the top panel I plot the estimated β coefficients in equation 2 on the lag and lead terms. For example, β_5 is the coefficient on the fifth lead term. In the bottom panel, I plot the cumulative response, where the μ's are defined as: $\mu_n = \beta_5 + \beta_4 + ... + \beta_n$ for $n = -5$ through 5. Standard errors are clustered by city $\times$ research field.