

NBER WORKING PAPER SERIES

PLACE-BASED PRODUCTIVITY AND COSTS IN SCIENCE

Jonathan Gruber  
Simon Johnson  
Enrico Moretti

Working Paper 30416  
<http://www.nber.org/papers/w30416>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2022

This paper was written for the volume "Entrepreneurship and Innovation Policy and the Economy", B. Jones and J. Lerner Eds., University of Chicago Press, forthcoming. We are grateful to Ben Jones, Josh Lerner and participants at the NBER conference Entrepreneurship and Innovation Policy and the Economy for their helpful comments. Lauren Rice provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w30416.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Jonathan Gruber, Simon Johnson, and Enrico Moretti. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Place-Based Productivity and Costs in Science  
Jonathan Gruber, Simon Johnson, and Enrico Moretti  
NBER Working Paper No. 30416  
September 2022  
JEL No. H0,J0,R0

### **ABSTRACT**

Cities with a larger concentration of scientists have been shown to be more productive places for additional scientists to do Research and Development. At the same time, these urban areas tend to be associated with higher costs of doing research, in terms of both wages and land. While the literature on the benefits of agglomeration economies is extensive, it offers no direct evidence of how productivity gains from agglomeration compare with higher costs of production. This paper aims to shed light on the balance between local productivity and local costs in science. Using a novel dataset, we estimate place-based costs of carrying out R&D in each US metro area and assess how these place-based costs vary with the density of scientists in each area. We then compare these costs with estimates of the corresponding productivity benefits of more scientist density from Moretti (2021). Adding more scientists to a city increases both productivity and production costs, but the rise in productivity is larger than the rise in production costs. In particular, each 10% rise in the stock of scientists is associated with a 0.11% rise in costs and a 0.67% rise in productivity. This implies that firms moving from cities with a small agglomeration of scientists to cities with a large agglomeration of scientists experience productivity gains that are 6 times larger than the increase in production costs. This finding is consistent with the increased concentration of R&D activity observed over the past 30 years. However, while the productivity estimate has only modest non-linearities, the cost estimates suggest much larger non-linearities as the concentration of scientists increases. For the most concentrated R&D cities, the difference between productivity gains and cost increases is close to zero.

Jonathan Gruber  
Department of Economics, E52-434  
MIT  
77 Massachusetts Avenue  
Cambridge, MA 02139  
and NBER  
gruberj@mit.edu

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

Simon Johnson  
MIT Sloan School of Management  
100 Main Street, E52-562  
Cambridge, MA 02142  
and NBER  
sjohnson@mit.edu

## **1. Introduction**

In recent decades the U.S. has experienced increased concentration in the location of innovation (Moretti 2012). Cities that have come to dominate the information technology and biotech sectors, primarily on the east and west coasts, have increasingly pulled away from the rest of the country, including other large urban areas. Such so-called superstar cities have become the predominant loci of innovation in the U.S., to a degree not previously experienced (Atkinson, Muro, and Whiton, 2019). For example, the top ten cities in the fields of "Computer Science," "Semiconductors," and "Biology and Chemistry" account for 70%, 79%, and 59% of all U.S. inventors in 2009, respectively (Moretti, 2021). At the same time, restrictive zoning policies in these cities keep housing prices high and limit the inflow of population. As a result, these places and the firms located there are unable to take full advantage of the implied agglomeration economies, depressing overall U.S. growth below what would otherwise have been achievable (Hsieh and Moretti, 2019).

The agglomeration of innovative activity raises important questions about the economic geography of the innovation sector. Why does private research and development (R&D) activity tend to be so geographically concentrated, despite the higher costs? Firms deciding where to locate their R&D activities presumably consider both costs and benefits offered by each location.

On the one hand, large technology clusters have been shown to increase individual and firm productivity, as working in large clusters tends to make scientists and engineers more creative and innovative – thanks to localized agglomeration economies. Marshallian spillovers stemming from human capital externalities and labor pooling have long been thought to be an important determinant of productivity and innovation, especially in the high-tech sector (Trajtenberg, and Henderson, 1993). Moretti (2021), for example, estimates the productivity advantages of large

clusters relative to small clusters and finds that scientists located in areas with a 10% larger stock of scientists in their specific research field produce 0.5-0.9% more patents per year. This effect appears to be causal, rather than driven by selection of the best scientists into the largest clusters.

On the other hand, it has become notoriously expensive to live and operate a business in the existing coastal superstar cities. Labor and real estate costs in places such as the San Francisco Bay Area, Boston and New York City are among the highest in the nation, by a considerable margin.

Thus, there appears to be a clear trade-off between productivity and production costs, with large, established high-tech clusters offering high productivity and high production costs, while smaller clusters offer lower productivity along with more affordable costs. From the point of view of an innovation-oriented firm deciding where to locate its operations – or the federal government deciding where to pursue place-based science policies – what matters is how productivity in a location compares to costs in that location. If an area with 10% higher output per scientist is 20% more expensive as a location to carry out R&D, then it is an inefficient location for a new lab, whether private or public.

This paper aims to shed light on the balance between local productivity and local costs in science. We assemble a novel dataset on cross-area costs of doing R&D, and we use this to measure place-based costs of carrying out R&D in each US metro area and to assess how place-based costs vary with the density of scientists in each area. We compare these costs with estimates of the corresponding productivity benefits of higher scientist density from Moretti (2021).

While the literature on the benefits of agglomeration economies is extensive, it offers no direct evidence of how productivity gains from agglomeration compare with higher costs of production in science. It is often assumed that, in the long-run equilibrium, localized productivity

gains from agglomeration are exactly offset by higher local costs. But actual empirical evidence on the costs of the R&D-intense “innovation sector” is scarce. Moreover, it is not obvious even in theory that localized industry-specific productivity advantages need to be exactly offset by higher costs in each industry and city, if cities contain multiple industries.<sup>1</sup>

The exact nature of the trade-off between productivity and costs matters for our understanding of the drivers of agglomeration of innovative private sector firms. But understanding this trade-off is not just an academic question: It also has important implications for the efficiency of a new set of proposed place-based initiatives designed to boost federal spending on science and innovation. For example, Gruber and Johnson (2019) and Atkinson et al. (2019) propose ambitious agendas for “place-based science,” with the aim of creating new technology hubs around the country that can complement the existing coastal superstar cities. By late 2021, the idea had been picked up in multiple legislative proposals, including: the bipartisan Endless Frontiers Act, which would commit \$10 billion over the next five years for grants to create 10-12 new technology hubs; the Innovation Centers Acceleration Act, which would provide \$80 billion over ten years for a competition for cities to become technology centers; and the Federal Institute of Technology Act, which would invest nearly \$1 trillion in public R&D over 10 years and would target a significant share of those funds to new technology centers.<sup>2</sup> Shifting additional publicly-supported R&D activity from being centered in established technology clusters towards other

---

<sup>1</sup> In equilibrium, labor and land costs are determined at the city level based on demand and supply forces in all sectors, while productivity may vary across sectors within a city as a function of sector-specific local factors, e.g., the size of that sector’s particular cluster. In the case of multiple sectors within a city, the spatial equilibrium should be such that marginal worker and the marginal firm are indifferent between cities. Different sectors may have a different ratio of productivity to costs.

<sup>2</sup> Endless Frontiers Act: <https://www.young.senate.gov/newsroom/press-releases/young-schumer-unveil-endless-frontier-act-to-bolster-us-tech-leadership-and-combat-china>. Innovation Centers Acceleration Act: <https://www.coons.senate.gov/news/press-releases/sens-coons-durbin-announce-legislation-to-expand-federal-randd-extend-tech-economy-to-more-cities-across-america>. Federal Institute of Technology Act: <https://khanna.house.gov/media/in-the-news/lawmaker-proposes-federal-institute-technology-and-new-contracting-set-aside>.

places likely creates efficiency trade-offs.<sup>3</sup> Properly evaluating the efficiency of such place-based policies requires measuring both costs and benefits.

To quantify local costs of R&D, we gather data from a variety of sources on the place-based costs of conducting scientific research. Using data from the Bureau of Economic Analysis (BEA) and the National Science Foundation (NSF) Business R&D Survey, we decompose the costs of Research and Development into components that vary by location (wages and building costs) from those that do not (machines). We use data from Glassdoor---the largest privately available source of information on wages in the R&D sector---to estimate wage costs for scientific personnel by city. For land values and rents, we use CoStar, one of the largest and most comprehensive databases of commercial real estate in the U.S., and the American Community Survey (ACS).

Combining these sources of data, we estimate the area-specific costs of carrying out R&D for a sample of 133 U.S. cities. We match this information to data on the stock of scientists by city to estimate how costs vary with the stock of scientists, and then compare these costs with Moretti's (2021) estimates of how productivity varies with the stock of scientists.

Using linear models, we find that adding more scientists to a city increases both productivity and production costs, but the rise in productivity is larger than the rise in production costs. In particular, we uncover a statistically significant but economically modest effects of cluster size on costs. Each 10% rise in the number of scientists in a city is associated with a 0.105% rise

---

<sup>3</sup> In the presence of large agglomeration externalities, federal place-based policies aimed at shifting additional R&D jobs toward new technology hubs could be costly in terms of overall innovation produced in the US. Indeed, Moretti (2021) estimates that in an extreme scenario where the quality of U.S. inventors is held constant and their geographical location is changed so that all cities have the same number of inventors in each field, the overall number and quality of patents produced in the U.S. in a year would drop significantly.

in costs, mostly due to higher wage costs. This is well below Moretti's estimate of productivity gains as the stock of scientists increases (also confirmed in our sample).

Thus, larger and more established clusters offer productivity advantages that more than offset increased costs at the margin. Our estimates imply that firms moving from cities with a small agglomeration of scientists to cities with a large agglomeration of scientists, experience productivity gains that are six times larger than the increase in production costs. This finding is consistent with the significant increase in the spatial concentration of innovative activity observed since the 1970s.

However, while Moretti's estimate of increased output has only very modest non-linearities, our estimates for costs have much larger non-linearities – meaning that costs increase a great deal when there are already many scientists in an area. Estimates from our non-linear models suggest that the relationship between productivity gains and costs increases varies significantly across areas. Using a spline regression specification, we find that while productivity gains remain larger than cost increases in cities with a sizable presence of scientists, the difference between productivity gains and cost increases is closer to zero for the most R&D intensive cities.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 presents our empirical findings. Section 4 concludes.

## 2. Data: Measuring the Area-Specific Costs of R&D

We measure area-specific R&D costs as a weighted average of the costs of the various factors needed in the R&D process, with weights reflecting the relative importance of each factor in the production function of R&D. We proceed in two steps. The first is to measure the factor weights in the R&D production function, i.e., the relative importance of various factors of production. The second is to measure the area-specific costs of the two spatially varying components of R&D costs, wages and building/land costs.

**Weights.** To measure factor weights, we rely on Robbins, Belay, Donahoe, and Lee (2012), who use NSF Business R&D Survey Data to calculate the average share of expenditures in R&D activities for five basic spending categories: wages for scientists and engineers, wages for support personnel, materials and supplies, current cost depreciation, and other R&D costs. To identify the real estate share, we combine the Robbins, Belay, Donahoe, and Lee shares with data from the Input-Output tables to estimate the share of intermediate inputs, defined as materials, supplies, and other R&D costs, spent on real estate and other leasing services. Specifically, we use BEA Industry Input-Output Data for the industry “Miscellaneous Professional, Scientific, and Technical Services Industry,” which includes Scientific R&D. We assume that the cost of physical capital does not vary across cities, once installed. It appears plausible that the market for machines and other forms of physical capital is national in scope.

We find that labor accounts for 38% of the costs of R&D, office space accounts for 5.1%, and physical capital (e.g., machines and equipment of all kinds) and other non-area specific costs (e.g., raw materials for laboratories) account for the remaining 56.9%. We note that our factor

weights do not vary by location. It is possible that firms adjust their inputs based on local prices, so that shares vary geographically, but our data preclude estimates of area specific cost shares.

**Costs for Factors of Production.** To measure labor costs of R&D personnel by location, we use two sources: Glassdoor for 2020, and the American Community Survey (ACS) for 2015-2020. While the ACS is more representative, Glassdoor has information on salaries at a much finer occupational level. For example, the Glassdoor data categorizes workers as “scientists and engineers” separately from “support personnel.”

In Glassdoor we focus on 33 occupations in the Biology, Chemistry and Materials, and Computer and Information Research industries. Glassdoor annual salary averages per occupation are calculated cumulatively using all available entries, which means our data span 2010-2020. In the ACS, we focus on data from 2014-2018 for eleven occupations in the following industries: Computer and Mathematical; Architecture and Engineering; and Life, Physical, and Social Science. In some cases, certain occupations are missing information; in those cases, we average data for the available occupations. Our baseline estimates are based on Glassdoor labor costs. We also show estimates based on ACS data.

To measure area differences in real estate costs, we combine ACS data on housing prices with real estate data from CoStar to estimate the average sale price per square foot for R&D labs. The CoStar data is based on industrial and commercial land properties using public record comparable sales from 2017 to 2020. The ACS data was sampled for household values in MSAs in 2018. We use the CoStar R&D category specifications with the least amount of missing data: Class B commercial real estate, and an area of 20,000-50,000 square feet. We calculate a weighted average sale price of \$137.54 per square foot for R&D facilities. The CoStar data for R&D space

costs is available for only 253 MSAs. In the 253 cities for which we do have CoStar commercial prices, the average CoStar commercial prices for all commercial properties are highly correlated with average ACS house prices. A regression of CoStar commercial prices on ACS mean house prices yields a coefficient of 0.154 (standard error of .008) and an R-squared of 0.5966.

To construct land costs, we find the ratio of the average sale price per square foot of R&D facilities from CoStar data to the average house value from ACS data. We then scale the average house values in each city by this ratio of average R&D sale price to average house value. Land costs are therefore at the level of R&D real estate costs, shifted relatively across MSAs according to variation in house values from the ACS.

**Geographical Differences in Total Costs.** Finally, we combine wage costs, land costs, and other costs to create overall area-specific costs of R&D. We also restrict the analysis to the 133 BEA Economic Areas, since this is the level at which the stock of scientists is measured. In most cases, “economic areas” are similar to a metropolitan statistical area (MSA). For large areas like the San Francisco Bay Area, Boston, or New York City, 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, we will refer to economic areas as “cities.”

Geographical variation in total area-specific costs of R&D is mainly driven by differences in wage costs, as they account for more than one-third of total costs. Land costs do vary across locations and contribute to spatial variation in overall costs, but their share is only 5.1% – about one-seventh of the labor share.

Table I reports summary statistics of annual wages for scientists and engineers, annual wages for support personnel, land costs, and a cost index with a mean value of 100 (which use the input factor weights described above). That is, for the index, we first normalize the value in each category of costs relative to the national average, and then we take a weighted average of these normalized indices using input factor weights.

Figure I shows the roughly normal distribution of overall costs for the 133 cities in our full sample. Most cities have an overall cost that is within 10% of the average (mean) city – namely between 90 and 110 – but there is an important right tail of significantly more expensive cities.

Table II lists the ten most expensive cities, the median city (Clarksburg-Morgantown, West Virginia), and the ten least expensive cities (all “cities” are actually MSAs). The top 10 cities have systematically higher costs in all categories. San Jose-San Francisco-Oakland, CA, Honolulu, HI, and Boston-Worcester-Manchester, MA-NH, are the metro areas where the production of R&D is the most expensive. By contrast, Dayton-Springfield-Greenville, OH, Milwaukee-Racine-Waukesha, WI, and Grand Rapids-Muskegon-Holland, MI are the metro areas where the production of R&D is the least expensive.

Costs in the metro area at the top of the distribution --San Jose-San Francisco-Oakland, CA – are 38% higher than costs in the metro area at the bottom --Grand Rapids-Muskegon-Holland, MI. While these spatial differences are large, they are somewhat smaller in magnitude than the differences in the consumer cost of living indexes for high-income households and significantly smaller in magnitude than the differences in the consumer cost of living indexes for low-income households. In particular, Diamond and Moretti (2021) estimates that the overall cost of living in the commuting zone that is the most expensive for high-income households is 49% higher than the overall cost of living in the commuting zone that is the least expensive for high-

income households. The corresponding difference for low-income households is 99%. The range of prices that low-income families are exposed to is much wider than the range of prices that high-income families are exposed to, because low-income households put a higher weight on housing expenditure, which is the item in the consumption basket whose price varies the most across cities.

### **3. Comparing Costs and Productivity in R&D**

Empirically determining the productivity advantage of Silicon Valley-style clusters is difficult, since location is endogenous. 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.

In a recent paper, Moretti (2021) uses longitudinal data on inventors to identify the productivity benefits for inventors who locate in Silicon Valley-style clusters. He defines a cluster as city  $\times$  research field and estimates how inventors' productivity---defined as number of patents produced in a year---varies with the size of the relevant cluster, measured by the number of other inventors in the same city and field, excluding the focal inventor.

He first studies the experience of inventors in Rochester, New York, where the high-tech cluster declined due to the demise of its main employer, Kodak. Kodak was the market leader in films for cameras and one of the most prolific patenters in the United States. But due to the diffusion of digital photography and the decline of physical film, Kodak employment collapsed after 1996. Essentially, demand for Kodak's main product evaporated due to a global technology shock. By 2007, the number of Kodak inventors in Rochester had declined by 84 percent. Moretti shows that Kodak's decline had a profound effect on the broader Rochester high-tech cluster. Measured by the number of inventors in all fields, its size declined by 49.2 percent relative to other cities, dragged down by Kodak's downsizing. The shock was large and

arguably exogenous, as it was caused by the advent of digital photography and not factors specific to Rochester's local economy. The experience of Rochester therefore offers an interesting case study for testing the hypothesis that high-tech clusters' size affects inventors' productivity. Moretti focuses on non-Kodak inventors outside the photography sector. He finds 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. The *within inventor* estimates indicate that the log productivity of non-Kodak inventors in Rochester declined by 0.206 (0.077) relative to other cities. This is consistent with the existence of important productivity spillovers in the high-tech sector stemming from geographical agglomeration.

Next, Moretti uses data for all U.S. clusters and presents estimates based on 109,846 inventors observed between 1971 and 2007, located in 895 clusters (179 cities  $\times$  5 research fields). He regresses the patents held by a particular scientist on the field and city-specific stock of scientists. His approach uses moves to identify the effects of each scientist.

In his richest specification, he finds that a scientist in a city-field with 10% more scientists produces 0.67 more patents. This indicates that a 10% increase in cluster size is associated with a 0.67% increase the number of patents produced by a scientist in a year.

To get a sense of the magnitude implied, consider an inventor in computer science who moves from the median cluster to the cluster at the 75th percentile of the size distribution. Moretti's estimate suggests that the scientist 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 to the 75th percentile cluster would be associated with a productivity gain of 8.4%, holding constant the inventor and the firm.

Of course, that scientist (or the firm that employs her) would also face a higher cost of carrying out research in areas that have the most scientific expertise. To see how R&D costs vary across cities as a function of the size of the local R&D sector, we regress our estimate of costs on the number of scientists who are active in the relevant metropolitan area. The stock of active scientists in a metropolitan area is from Moretti (2021) and is measured as the ratio of the number of inventors who file for a patent in in any research field in a year over the number of all active inventor in the US in that year. These estimates are based on data on the universe of U.S. patents filed between 1971 and 2007 from the COMETS patent database.

Ideally, we would measure field and city-specific costs, but our cost data is not ideal in this respect. While we can measure wage costs for a broad set of fields by cities, we cannot create field-specific land costs. However, there is no specific reason to expect that the cost of land or office space varies significantly within a city for different research fields. Thus, in our baseline estimates, we consider only city-specific costs. In an extension, we present additional analyses for city and field-specific labor costs which allow us to include city fixed effects.

The results of our baseline regressions are shown in Table III. Columns (1)-(3) show regressions of the log of total area costs on the stock of scientists, for different specifications. Columns (4)-(6) show the regressions for equivalent specifications using productivity data from Moretti (2021); the dependent variable is the log number of patents produced in a given year by an inventor. The level of observation is an inventor-year pair.

We begin in column 1 with a log-linear regression of log area-specific costs on the log of the stock of scientists. We find an elasticity of 0.0105, indicating that each 10% rise in the stock of scientists is associated with a 0.105% rise in total area costs. Column 4 reports the estimate for productivity using the same log-linear specification. A comparison of column 1 and 4 suggests that the increase in cost associated with a larger cluster size is about one-sixth of the Moretti

estimate for the increase in productivity. This specification suggests that the higher productivity enjoyed by scientists in larger innovation clusters is only partially offset by the higher costs of carrying out research in those larger clusters.

Taken at face value, this specification indicates that by moving to bigger innovation clusters, firms will experience productivity gains that are significantly larger than the increase in production costs – a result that is broadly consistent with the increase in concentration of innovative activity that we have seen over the last 30 years. For example, Moretti (2021) reports that the share of inventors in Computer Science, Semiconductors, Biology, and Chemistry in the top 10 largest clusters is larger today than it was in the 1970s.

Figure II shows a graph of the relationship between cluster size and costs. The upward slope in the data is apparent, but what is more striking is the apparent non-linearity for the cities with the largest numbers of scientists (i.e., the so-called superstar cities). To illustrate this, we show fitted regression lines for a log-linear regression, a regression that is quadratic in logs, and a three-piece spline with cutoffs of  $\log(\text{stock})$  of -8 and -5. The latter cutoffs were chosen to model the trends in the Lowess smoother and the Local-Linear Kernel Regression. The value of -8 is at the 13<sup>th</sup> percentile of the  $\log(\text{stock})$  distribution, while -5 is at the 76<sup>th</sup> percentile. Visually, these non-linear models fit the data much better, reflecting the fact that wages and real estate costs---and therefore total costs---are much higher in a handful of cities than even in other relatively dynamic (and high productivity) cities.

The second and third columns of Table III present regression results from these non-linear specifications. The quadratic log specification (column 2) suggests a rapidly rising cost for the places with the largest stock of scientists, for example implying that for a place at the 80<sup>th</sup> percentile

in the scientist stock distribution, adding 10% more scientists raises costs by 0.46%, about four times the estimate from the linear specification.

The spline specification is even more interesting. It shows that for places with relatively few scientists, adding 10% more scientists raises costs by 0.37%. For places in the typical range of our sample, the relationship between cost and stock of scientists is negative and marginally significant, albeit quite small. For places with the most scientists, in contrast, the effect is large, with each additional 10% of scientists raising costs by 0.53%.

For an apples-to-apples comparison, columns 5 and 6 of the Table replicate the Moretti productivity regressions for the same non-linear specifications, using his richest model of covariates. As noted earlier, the Moretti linear coefficient is much larger (about six times) than the associated linear cost coefficient. But the Moretti productivity relationship is more linear (in a log-log model) than what appears on the cost side. In the quadratic specification in column 5, the quadratic term is statistically significant. But the implied curvature (i.e., increasing returns to more scientists) is limited. The estimated coefficients imply that adding 10% more scientists in a city at the 80<sup>th</sup> percentile of the distribution (i.e., Kansas City), raises productivity by 0.74%, not too far from the linear specification. The spline specification in column 6 suggests that for scientists in the largest clusters, each 10% rise in the stock of scientists raises productivity by 0.77%. The productivity increase is still above the corresponding cost increase in column 3, which is 0.52%, but the difference between the productivity and cost increases is now quantitatively smaller and not statistically significant (using the 95 percent confidence interval).

Overall, our non-linear specifications suggest that the relationship between productivity gains and costs increases varies significantly across areas. Productivity gains from adding scientists are much larger than cost increases in cities that have a limited presence of scientists. At

the same time, productivity gains remain larger than costs increases but not by very much in cities with a sizable presence of scientists.

As noted above, for labor costs, we can go further and analyze field-specific estimates. We do so in Table IV, re-estimating our models at the city\*field level for the three fields for which we are able to measure labor costs: biology; chemistry and materials; and computer and information research. We create these three fields by assigning the 38 Glassdoor occupations to each category, which allowed us to fully populate each of our 133 cities.

Our linear regression in Table IV gives a weaker result than for the overall costs in Table III, suggesting that the linear impact of scientist stock on costs in Table II is driven by land (real estate) and other physical plant costs. When we move to non-linear specifications, we see a more extreme version of our earlier finding, with a strong positive effect on labor costs from adding scientists at the top and bottom of the distribution, and no relationship in the middle.

Table V assesses the robustness of our results to the source of labor cost data. We show our base results using Glassdoor data (columns 1 and 2), as well as the estimates that instead use data from the ACS, which provides large samples but less precise worker classifications (columns 3 through 6). For the ACS, we show the results using both unconditional values (columns 3 and 4) and values conditional on worker age, sex, marital status, race, and education level (columns 5 and 6). Using the ACS data reduces the sample size slightly (from 133 to 98), but we still find results that are consistent with Glassdoor data.

#### 4. Conclusions

Scientists located in areas with a larger stock of related scientists tend to be more productive (Moretti, 2021). This is likely to be an important factor in explaining the geographical concentration of innovative activity in the U.S. and in other developed countries (Kerr, 2020).

Our findings suggest that, on average across U.S. cities, the productivity gains stemming from agglomeration exceed the higher research costs that characterize larger clusters. However, the ratio of productivity gains and costs increases varies significantly across areas. For areas that currently have a small cluster, the productivity gains of adding an additional scientist are much larger than the corresponding cost increases. For the largest innovation clusters (i.e., those with most scientists), the productivity gains of adding an additional scientist are still larger than the corresponding cost increases, but not by much.

A natural question is whether there is a case for place-based government technology policy, even in the absence of costs that offset productivity differentials. A first issue to consider in this respect is equity. In a world of imperfect mobility and imperfect information on underlying needs, place has a role in redistribution (e.g. Gaubert et. al., 2021). Redistributing by place allows a tool for targeting needy individuals that are missed by other redistributive systems. This is particularly important given the findings of Bell et al. (2019) on the intergenerational correlation of patenting.

A second argument in favor of these policies is robustness to geographic shocks – particularly in a nation as large as the United States. Catastrophes, man-made or natural, would have an outsized effect on the U.S. if they happen to occur in the very few, most agglomerated locations. A broader portfolio of technology centers provides a form of insurance against geographically focused shocks.

A third, and by far the most speculative argument, relates to politics. Gruber and Johnson (2019) point out that one of the reasons for the weak public support in the U.S. for public investment in R&D is the geographical concentration of such investment. In a nation where voting is related to population and geography, not income or productivity, investments that concentrate their benefits in small geographic (even if densely populated) places may suffer from a lack of political support. Gruber and Johnson (2019) argue that even if there is some efficiency loss from redistributing R&D, the rate of return to more R&D is high enough that more geographic dispersion may lead to more overall efficiency by raising the level of support for public science spending.

## References

- Atkinson, Robert D., Mark Muro, and Jacob Whiton. 2019. "The Case for Growth Centers: How to Spread Tech Innovation Across America." Brookings, December.  
[https://www.brookings.edu/wp-content/uploads/2019/12/Full-Report-Growth-Centers\\_PDF\\_BrookingsMetro-BassCenter-ITIF.pdf](https://www.brookings.edu/wp-content/uploads/2019/12/Full-Report-Growth-Centers_PDF_BrookingsMetro-BassCenter-ITIF.pdf)
- Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *The Quarterly Journal of Economics*, Volume 134, Issue 2, May, Pages 647–713,
- Bureau of Economic Analysis. 2019. "Industry Economic Accounts Data: Input-Output, The Use of Commodities by Industries." U.S. Department of Commerce.  
[https://apps.bea.gov/iTable/iTable.cfm?reqid=52&step=102&isuri=1&table\\_list=4&aggregation=sum](https://apps.bea.gov/iTable/iTable.cfm?reqid=52&step=102&isuri=1&table_list=4&aggregation=sum)
- Diamond, Rebecca and Enrico Moretti. 2022 "Where is Standard of Living the Highest? Local Prices and the Geography of Consumption." Unpublished working paper, Stanford and Berkeley, July. <https://eml.berkeley.edu/~moretti/consumption.pdf>
- Gaubert, Cecile, Patrick Kline, and Danny Yagan (2021). "Place based redistribution," NBER working paper # 28337, January 2021.
- Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics*, 11 (2): 1-39.
- Gruber, Jonathan and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. Public Affairs, New York, April.
- Kerr, William and Frederic Robert-Nicoud. 2020. "Tech Clusters." NBER Working Paper 27421.
- Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt, Boston.
- Moretti, Enrico. 2021. "The Effect of High-Tech Clusters on the Productivity of Top Inventors." *American Economic Review*, 2021.
- Robbins, Carol, Olympia Belay, Matthew Donahoe, and Jennifer Lee. 2012. "Industry-level Output Price Indexes for R&D: An Input-cost Approach with R&D Productivity Adjustment." [BEA Working Papers 0090](#), Bureau of Economic Analysis.

**Table I: Summary Statistics of City Costs**

<u>Cost Variable</u>	Mean	Std. dev.	Median	N
Annual Wages for Scientists/Engineers	61,839.34	8,792.77	62,677.19	133
Annual Wages for Support Personnel	32,170.34	3,368.66	32,012.29	133
Land Costs (Sale Price/sqft)	52.407	22.817	45.912	133
Overall Cost Index	100	5.870	99.882	133

**Table II: Costs of Ten Least Expensive Cities, Median City, and Ten Most Expensive Cities**

	Overall Cost Index	Annual Wages for Scientists and Engineers	Annual Wages for Support Personnel	Land Costs (Sale Price/sqft)
	(1)	(2)	(3)	(4)
<b><u>Ten Most Expensive Cities</u></b>				
San Jose-San Francisco-Oakland, CA	127.27	96,786.74	39,810.49	151.59
Honolulu, HI	119.13	77,912.53	36,557.43	162.44
Boston-Worcester-Manchester, MA-NH	116.12	71,873.22	55,434.76	86.59
San Diego-Carlsbad-San Marcos, CA	115.56	74,253.94	35,605.35	145.20
Anchorage, AK	113.04	85,182.56	37,201.02	65.61
New York-Newark-Bridgeport, NY-NJ-CT-PA	110.81	74,256.59	36,957.06	91.28
Portland-Vancouver-Beaverton, OR-WA	110.68	75,790.26	34,563.31	92.22
Bend-Prineville, OR	109.00	70,822.55	34,394.40	97.34
Sacramento-Arden-Arcade-Truckee, CA-NV	108.68	72,759.07	36,529.79	77.48
Washington-Baltimore-N Virginia, DC-MD-VA-WV	108.40	69,137.43	36,233.01	91.58
<b><u>Median City</u></b>				
Clarksburg-Morgantown, WV	99.88	63,419.17	30,750.27	49.53
<b><u>Ten Least Expensive Cities</u></b>				
Joplin, MO	86.85	40,015.67	26,275.52	34.46
Evansville, IN-KY	89.05	45,680.54	25,866.85	33.89
Raleigh-Durham-Cary, NC	89.94	44,649.64	28,295.99	38.42
Greensboro-Winston-Salem-High Point, NC	90.44	42,384.03	32,198.56	38.69
Huntsville-Decatur, AL	90.50	44,463.75	30,742.14	35.68
Kennewick-Richland-Pasco, WA	91.14	44,646.15	29,288.61	47.01
Harrisburg-Carlisle-Lebanon, PA	91.26	46,155.57	27,903.62	46.87
Dayton-Springfield-Greenville, OH	91.67	47,296.16	31,950.61	30.83
Milwaukee-Racine-Waukesha, WI	92.12	44,240.18	32,714.46	45.91
Grand Rapids-Muskegon-Holland, MI	92.49	47,767.18	31,424.06	39.24

**Table III: Regressions of Log Total Area Costs on Log Stock of Scientists**

	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dependent Variable</u>	<b>Log Total Area Costs</b>			<b>Log Inventor Productivity</b>		
Log (Stock)	0.0105 (0.0038)	0.0654 (0.0251)		0.0676 (0.0139)	0.2876 (0.0346)	
Log (Stock) <sup>2</sup>		0.0046 (0.0020)			0.0236 (0.0031)	
1 <sup>st</sup> spline			0.0373 (0.0156)			0.0636 (0.0132)
2 <sup>nd</sup> spline			-0.0095 (0.0051)			0.0721 (0.0142)
3 <sup>rd</sup> spline			0.0525 (0.0125)			0.0772 (0.0148)
N	133	133	133	823,375	823,375	823,375

Note: Columns 1-3 show regressions from our data of the log of total area costs on the stock of scientists from Moretti (2021). Columns (4)-(6) show regressions using data from Moretti (2021). Columns (1) and (4) show linear specification; (2) and (5) show quadratic specification; (3) and (6) show three-piece spline. Standard errors in parentheses.

**Table IV: Regressions of Log Area Labor Costs on Log Stock of Scientists with Field Fixed Effects**

	(1)	(2)	(3)
<u>Dependent Variable</u>	<u>Log Area Labor Costs</u>		
Log (Stock)	0.0030 (0.0037)	0.0576 (0.0245)	
Log (Stock) <sup>2</sup>		0.0045 (0.0020)	
1 <sup>st</sup> Line			0.0401 (0.0153)
2 <sup>nd</sup> Line			-0.0187 (0.0064)
3 <sup>rd</sup> Line			0.0460 (0.0131)
N	399	399	399

Note: Regressions from our data on the log of labor costs by city and field on the stock of scientists in that city and field from Moretti (2021). Regressions include field fixed effects. C=). Column (1) shows linear specification; (2) shows quadratic specification; (3) shows three-piece spline. Standard errors in parentheses.

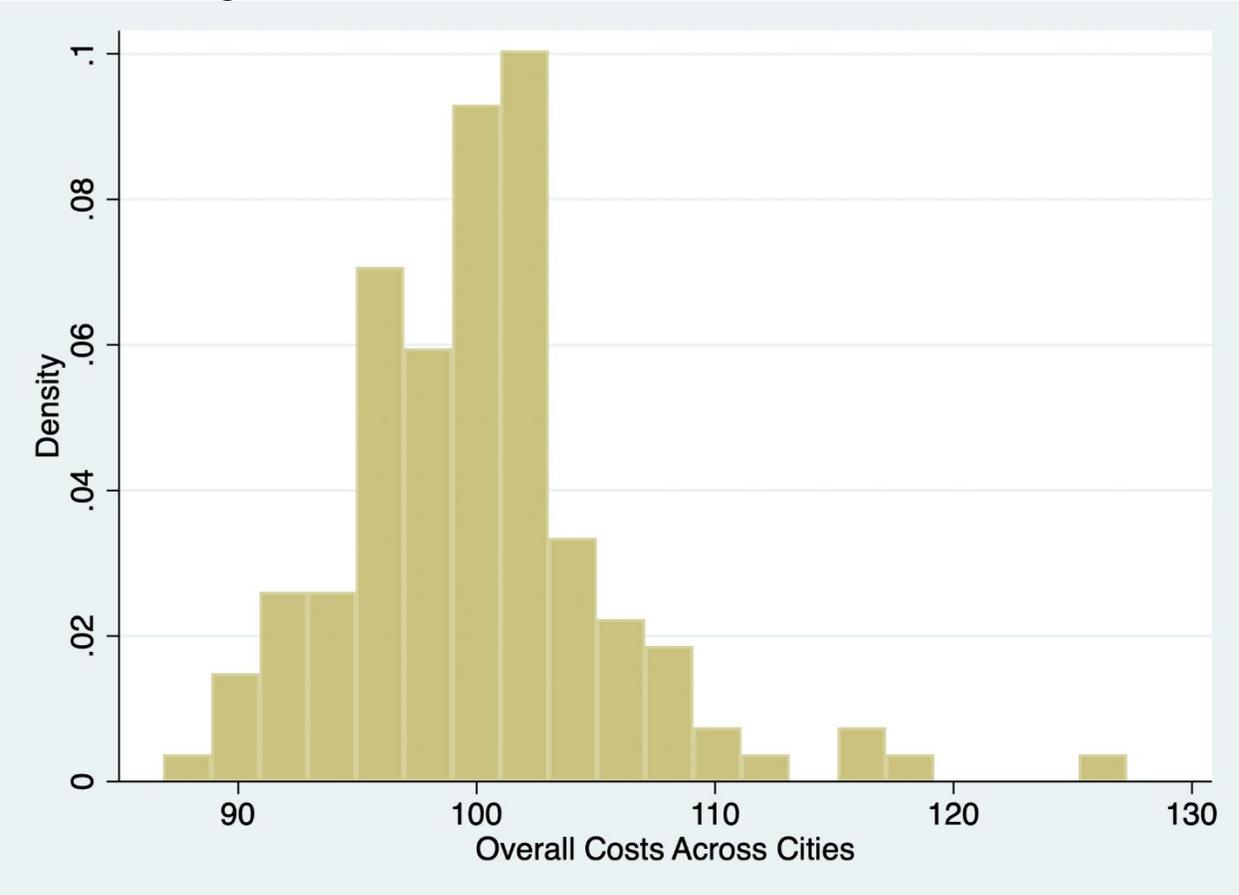
**Table V: Robustness Table for Varying Wage Data Sources  
for Regressions of Total Area Costs on Log Stock of Scientists**

	(1)	(2)	(3)	(4)	(5)	(6)
<u>Wage Source</u>	<b>Glassdoor Wage Data</b>		<b>Unconditional ACS Wage Data</b>		<b>Conditional* ACS Wage Data</b>	
Log (Stock)	0.0105 (0.0038)	0.0654 (0.0251)	0.0195 (0.0052)	0.1003 (0.0243)	0.0194 (0.0052)	0.0999 (0.0244)
Log (Stock) <sup>2</sup>		0.0046 (0.0020)		0.0068 (0.0020)		0.0068 (0.0020)
N	133	133	98	98	98	98

\*Conditional on age, sex, marital status, race, and education level

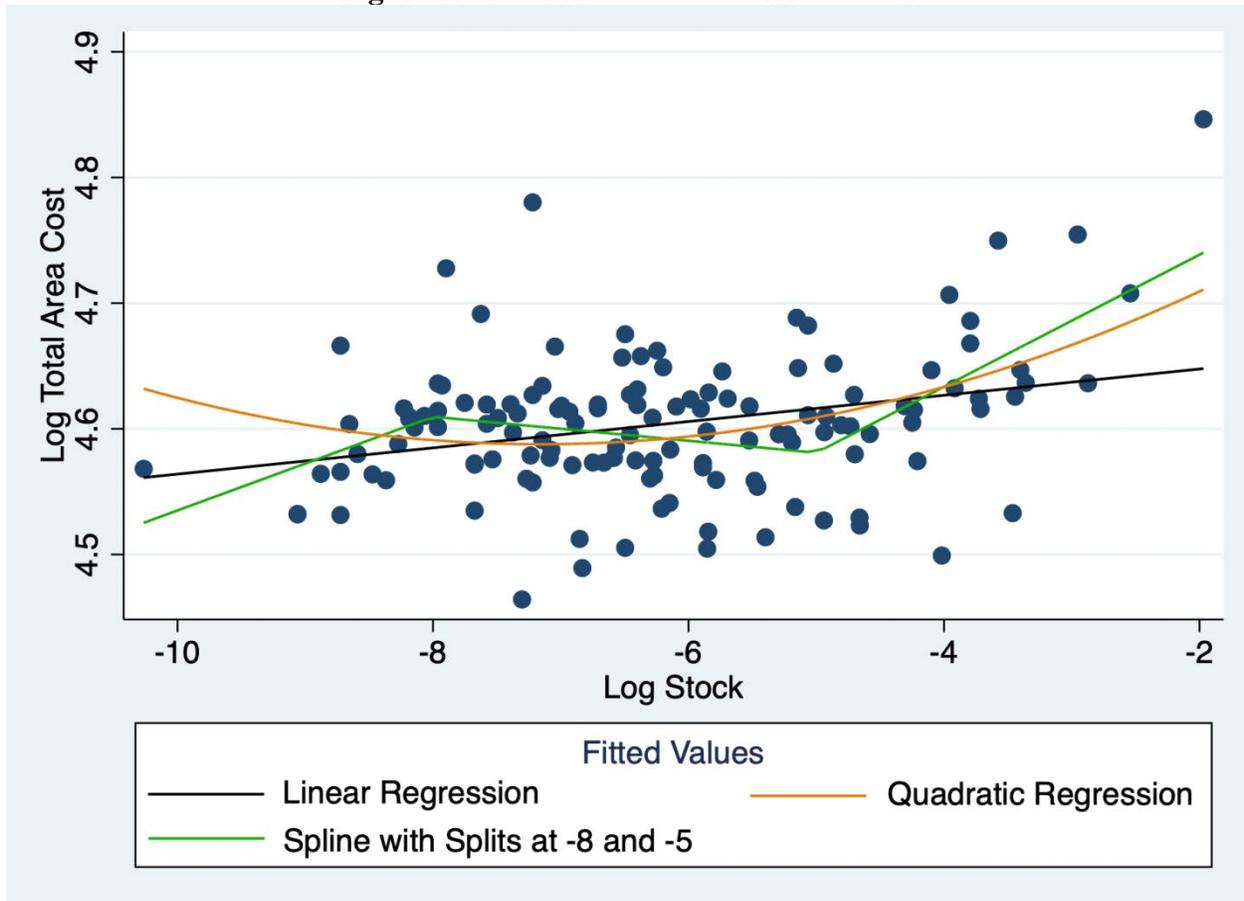
Note: Columns (1) and (2) show regressions from our data with Glassdoor wage data of the log of total area costs on the stock of scientists from Moretti (2021). Columns (3) and (4) instead use unconditional ACS data for similar occupations and industries. Columns (5) and (6) use ACS data for similar occupations and industries conditional on observable characteristics. Columns (1), (3), and (5) show linear specification; (2), (4), and (6) show quadratic specification. Standard errors in parentheses.

Figure I: Distribution of Overall Costs



Note: Histogram of the distribution of the overall costs constructed from our data.

Figure II: Total Area Costs vs. Area Stock



Note: X-axis is log stock of scientists and Y axis is log total area costs. Fitted lines from linear, quadratic and spline regressions included.