NBER WORKING PAPER SERIES

CITIES AND IDEAS

Mikko Packalen
Jay Bhattacharya

Working Paper 20921
http://www.nber.org/papers/w20921

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2015

We thank Lee Fleming, Bruce Weinberg, and seminar participants at Aalto University, UC-Berkeley Innovation Seminar, and Innovation in an Aging Society working group for comments. We acknowledge financial support from the National Institute on Aging grant P01-AG039347. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Cities and Ideas
Mikko Packalen and Jay Bhattacharya
NBER Working Paper No. 20921
January 2015
JEL No. O18,O31,O32,O33,R12

ABSTRACT

Faster technological progress has long been considered a key potential benefit of agglomeration. Physical proximity to others may help inventors adopt new ideas in their work by increasing awareness about which new ideas exist and by enhancing understanding of the properties and usefulness of new ideas through a vigorous debate on the ideas' merits (Marshall, 1920). We test a key empirical prediction of this theory: that inventions in large cities build on newer ideas than inventions in smaller cities. We analyze the idea inputs of nearly every US patent granted during 1836–2010. We find that a larger city size provided a considerable advantage in inventive activities during most of the 20th century but that in recent decades this advantage has eroded.

Mikko Packalen
University of Waterloo
Department of Economics
200 University Avenue West
Waterloo, ON N2L 3G1
Canada
packalen@uwaterloo.ca

Jay Bhattacharya
117 Encina Commons
CHP/PCOR
Stanford University
Stanford, CA 94305-6019
and NBER
jay@stanford.edu

1 Introduction

Physical proximity to other creative people is often a key factor in invention. It enables inventors to adopt the best new ideas faster in their own work. In larger cities, the costs of encountering new ideas are lower and inventions 'have their merits promptly discussed' (Marshall, 1920). The debate on new ideas is important because ideas are initially raw and poorly understood. Thus, every idea, if it is to develop from a germ into a substantial innovation, needs to be tried out and discussed by many people including those not responsible for originating the idea (Marshall, 1920; Usher, 1929; Kuhn, 1962).

On the other hand, modern communication technologies may have undermined or even eliminated any advantage that physical proximity to others may once have provided in terms of adopting new ideas. Marshall (1920) himself explored the implications of such 'cheapening of the means of communication,' and raised the possibility that location might play only a minor role in knowledge production as knowledge would 'depend chiefly on the aggregate volume of production in the whole civilized world.' Alternatively, there are advantages to face-to-face communication even in an internet-connected world, especially in the sharing of tacit knowledge (von Hippel, 1994).

It is also possible that agglomeration actually hinders the adoption of new ideas. For instance, suppose that the presence of large incumbent firms leads to the development of large cities. In that case, inventors in big cities will have strength in incumbent technologies and may be tilted against adopting new ideas that might lead to radical inventions and devalue existing competencies (Brezis and Krugman, 1997).

From a theoretical perspective, agglomeration may thus have a positive, a negative, or no impact at all on the adoption of new ideas as inputs to the inventive process. In this paper, our main purpose is to measure how the tendency to adopt new ideas as inputs in invention varies by the size of the city of invention.

Uncovering the sign and magnitude of the relationship between new knowledge adoption and agglomeration is important for allocating private and public research and development funds efficiently. This relationship is especially important given that the extent and share of population located in metropoli-

tan areas continues to increase (Duranton and Puga, 2013). If we find that inventors in large cities build on fresh ideas more often than inventors in smaller cities, the evidence would quantify a specific benefit to locating inventive activities in large cities. On the other hand, if we find that inventors in large cities are no more likely to try out new ideas in their work than inventors in smaller cities, the evidence would suggest that location may be largely irrelevant for inventive performance.

Our approach to examining the agglomeration–invention link is novel. The existing approaches (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Thompson, 2006; Murata et al., 2013) examine how adoption of new ideas in invention varies by distance to the origin of the idea ("localization around the origin"). Our approach is distinct as we instead examine how adoption of any new ideas in invention varies by city size. We adopt this approach for several reasons.

First, a focus on localization around the origin would limit the analysis mainly to only one type of potential benefit from agglomeration — the benefit of becoming aware of inventions developed nearby. Our approach captures also the benefit that agglomeration may have on idea adoption through its impact on inventors' ability to debate the merits of new ideas. To us it seems – and the reader can readily judge based on personal experience – that in creative work this latter channel (debating merits of a discovery) is at least as important a benefit of human interaction as the former channel (becoming aware of a discovery). Moreover, this benefit of agglomeration may continue to be present even if awareness of the existence of each new idea is practically instantaneous throughout the world, as may well be the case today.

A second advantage to our approach is that it reveals whether it is smaller or larger cities that tend to adopt new ideas faster. This represents a more direct measure of whether agglomeration benefits invention, the hypothesis put forth in Marshall (1920). Our approach retains this advantage even over an approach that focuses on localization in general rather than just localization around the origin.

A third advantage of our approach is that in a localization based approach, no matter how finely research areas are defined, any evidence for localization can reflect either genuine locational advantages or mere differences in research approaches pursued (Thompson and Fox-Kean, 2005; Henderson et al.,

2005). Our approach, by contrast, is not subject to this limitation. Even if research approaches are different in larger and smaller cities within the same research area (no matter how finely defined), evidence on which cities employ newer idea inputs in invention still speaks to the question of which cities are closer to the technological frontier.

2 Approach

Our aim is to uncover the empirical relationship between agglomeration and use of new ideas in invention. We measure invention from US patents granted during 1836–2010. Below, we explain how we identify idea inputs upon which each patent is built and how we calculate the age of those idea inputs. We also explain how we link the patent-level observations on the age of idea inputs with the size of the city where the patent originated. We measure city size by population density. Density should capture the scope for frequent interactions with others better than other city size measures, though we also perform a robustness check in which city size is measured by total population.

2.1 Measuring Idea Inputs and the Use of Recent Ideas

We capture idea inputs in invention from patents' textual content. First, we index all words and 2- and 3-word sequences in each patent. We refer to these words and word sequences as *concepts*. For each concept we then determine the year it first appeared in the patent database. We refer to that year as the *cohort* year of a concept. The age of concept is defined as the number of years elapsed since the cohort year of the concept.[1]

This textual approach organically captures ideas that are well-known to have been important inputs to invention (e.g. *microprocessor*, *polymerase chain reaction*), as we have shown in our prior work (Packalen and Bhattacharya, 2012). Accordingly, we interchangeably refer to the concepts mentioned

[1]Because of optical character recognition (OCR) errors, some concepts appear in the data before they actually appear in patents. To address this issue, we exclude early mentions of any concept that is mentioned less than 5 times in the 10 years that follow the concept's first mention. For such concepts the cohort is re-assigned as the first year such that (1) the concept was mentioned that year *and* (2) the concept was mentioned at least 5 times in the subsequent 10 years.

in each patent as idea inputs to the invention. We manually cull through the initial concept lists to exclude concepts that likely do not represent an idea input. The lists of popular idea inputs that are uncovered by our approach, as well as the list of manually excluded concepts, are included in Packalen and Bhattacharya (2015b). Throughout the 20th century, these lists are dominated by ideas that any reader can recognize as having been important building blocks in technological innovation.[2]

For each patent, we calculate a summary measure *Age of Idea Inputs*, defined as the age of the newest idea input mentioned in it. This measure captures the vintage of idea inputs for each invention. A lower value for the measure indicates that the invention built on newer ideas; the converse indicates a reliance on well-established ideas.

Because the hypotheses about the agglomeration–invention link concern adoption of the newest ideas, we employ as the dependent variable an indicator measure that captures whether idea inputs for a patent are among the top 5% newest based on the *Age of Idea Inputs* measure. This dummy variable, *Top 5% by Age of Idea Inputs*, is constructed based on comparisons of patents granted in the same year in the same 3-digit technology class.[3] In robustness checks, we define a similar variables using a narrower comparison group and a different percentile cutoff.

In our main analyses, we consider only mentions of the 100 most popular idea inputs in each cohort, with popularity defined as the total number of patents that mention the idea. These analyses reveal whether there are differences by degree of agglomeration in terms of using the best new knowledge. In robustness checks, we extend the analysis to the top 10,000 ideas in each cohort, revealing whether any differences extend also to the use of new ideas more generally.

The closest related literature (Jaffe et al. 1993; Thompson and Fox-Kean, 2005; Thompson, 2006; Murata et al., 2013) has relied on citations to capture idea inputs. Instead, we rely on the textual content of patents. We do this for several reasons: (1) citations can give rise to spurious geographic patterns (Thompson, 2006), (2) at least half of all citations are unrelated with any concept of information flow

[2]The approach, of course, does not capture idea inputs that are typically expressed as combinations of non-consecutive words (e.g. *silicon* and *transistor*, *ulcer* and *bacteria*). Analysis of such ideas is beyond the scope of the present paper.

[3]We previously employed such a variable in Packalen and Bhattacharya (2015a).

(Jaffe et al., 2000), and (3) citations capture only a very narrow set of idea inputs (Rosenberg, 1982).

Spurious geographic patterns in citations arise when knowledge about ideas and associated patents travels by word-of-mouth. The more important part of this knowledge – the idea – tends to travel further than the information about who thought of the idea. A geographic difference in tendency to cite a given patent may then arise not just due to differences in the tendency to apply the idea but also because of differences in the tendency to know to whom to attribute the idea. While people do of course sometimes refer to the same idea by different names, it seems to us that there is much less room for similar artificial geographic patterns to arise when idea inputs are captured from text.

The other limitations of using citations to capture idea inputs arise because citations have no requirement to list even all patented technologies that the invention built upon. Rather, the purpose of citations is to delineate the boundaries of the patent (Jaffe et al., 1993). As a result, many research inputs (e.g. *database*, *microprocessor*) either cannot be identified from citations at all or can be identified from text much more reliably. That the majority of patent citations do not reflect any form of knowledge flow (Jaffe et al., 2000) is another consequence of the delineation aspect of patent citations. Unfortunately, it is unknown which citations reflect inputs and which do not. When even researchers who have relied on citations extensively consider the cup only 'half-full' when it comes to using citations as measures of knowledge inputs (Jaffe et al., 2000), it seems worthwhile to pursue alternative approaches that might turn out to be either better than citations or at least capture some of what is missing from the metaphorical cup.

2.2 Estimation of the Link between Agglomeration and Use of New Ideas

We estimate the empirical relationship between the degree of agglomeration and the tendency to use new ideas in invention. As discussed above, our preferred measure of the recency of idea inputs is the constructed patent-level dummy variable *Top 5% by Age of Idea Inputs*, and our preferred measure of city size is population density. We use the linear probability model and the corresponding conditional logit model to estimate the relationship between these variables. In the appendix, we provide also a

non-parametric analysis. The linear probability model that we estimate is

$$\textit{Top 5\% by Age of Idea Inputs}_{\text{city,year,i}} = \beta \times \log(\textit{Population Density}_{\text{city,year}}) + \alpha_{\text{year,tech\_class}} + \epsilon_{\text{city,year,i}}. \quad (1)$$

The subscript *i* differentiates patents granted in the same *city* in the same *year*. The triplet *city*,*year*,*i* thus uniquely identifies each patent and links each patent to the city and year in which the patent was granted. We estimate this specification separately for each decade. A separate fixed effect $\alpha_{\text{year,tech\_class}}$ is included for each year and technology class pair. Each patent is thus compared only to patents granted in the same technology class in the same year.

An upward-sloping relationship is the predicted empirical link when agglomeration increases use of new ideas. A flat relationship is the predicted empirical link when advances in communication technologies have rendered distance irrelevant. A downward-sloping relationship is the predicted empirical link when incumbents shy away from experimenting with new ideas. Because we estimate specification (1) separately for each decade, we uncover how the agglomeration–use of new ideas link has evolved over time. We estimate specification (1) also using an instrumental variables approach in which we instrument current population density with measures of population density from 50 years prior.[4]

An increasing number of patents list multiple inventors (e.g. Jones, 2010). In our baseline approach we assign invention location based on the first inventor, but we also examine the agglomeration–use of new ideas link in specifications that control for lone vs. team inventor status.

3 Data

We analyze US patents granted during 1836–2010. For 1976–2010, patent documents are available as ASCII data, whereas for 1836–1975 patents are available as scanned images. To obtain the older patents' textual content, we applied optical character recognition ("OCR") algorithms to the images of

[4]In our specifications, we do not include city dummies on the basis that most of the cross-sectional variation in city size is driven by events that occurred long time ago and are exogenous to incentives to adopt new ideas by inventors. Thus adding city dummies would remove a source of "good variation" that is the historical differences in city size, and would leave only "bad variation" that is current changes in population density which could be due to success of recent inventions.

the 4+ million patents granted between 1836 and 1975.[5] The patent texts reveal the ideas that each patent built upon as well as the age of those ideas (please see section 2.1 above and Packalen and Bhattacharya (2015b) for further details). For 1836–1974, we extract the number of inventors in each patent and the location of each inventor. For inventors located in the U.S. we extract information on either the city or county in order to match each patent to a Primary Metropolitan Statistical Area ("PMSA"). For inventors located in other countries we extract only the country information. For 1975–2010 we use the location data of Lai et al. (2013).

The time series on the extent of patenting shows the familiar steady increase in patenting over the past century (please see the top panel of Figure A1 in the appendix). Due to OCR errors the location extraction analysis does not yield any location information for some patents. However, for the vast majority of years we match over 85% of patents to a location (please see the bottom panel of Figure A1 in the appendix).[6]

During the early years of the sample the measured idea input ages start low and increase quickly (please see Figure A2 in the appendix). This is unsurprising given how we determine whether an idea is new – we measure how recently the concept first appeared in patents – and given the small number of patents granted during the early years of the sample. We address this issue by ignoring mentions of concepts that belong to pre–1870 cohorts and by limiting the analysis of the agglomeration–use of new ideas link to cover the period from the decade that Marshall started writing his economics textbook to the present era (that is, years 1880 through 2005). We exclude years 2006–2010 from the regression analyses because for the most recent idea cohorts it is not yet known which ideas will stand the test of time and represent the cohorts' "best ideas".

Population density is measured by individuals per square mile (please see Figure A3 in the appendix for summary statistics). We perform the analysis at the PMSA level for two reasons. First, by design

[5]For 1920–1975 there does exist an alternate OCR transfer of patent texts, which we used in Packalen and Bhattacharya (2012). Here and in Packalen and Bhattacharya (2015b), we rely only on data from the new OCR transfer.

[6]For 1926–1933 the match rate is considerably lower due to changes in the reporting of location in the patent text and subsequent difficulty for deciphering the inventor information using OCR. Importantly, these lower match rates only affect our estimates for 1920s and 1930s and do not affect the estimates for the other decades.

PMSAs form somewhat more meaningful borders for opportunities to interact than city or county limits. Second, conducting either a county- or city-level analyses would introduce additional barriers that stem from the fact that whereas for patents granted before 1920s county information can be extracted more reliably than city information, the reverse is true for patents granted after 1920s. By conducting the analysis at the PMSA-level we can map patents to PMSAs based on information on either the county or the city of invention.[7]

4 Results

Figure 1 shows decade-specific estimates obtained using the conditional logit equivalent of specification (1). The corresponding numerical values are reported in Table A1 in the appendix. The dependent variable is the dummy variable *Top 5% by Age of Newest Idea Input*. The main regressor is the logarithm of population density. For ease of interpretation, the estimates shown in Figure 1 depict how much more often inventions in large cities built on new ideas relative to inventions in average sized cities. Formally, each estimate of β is expressed as the percentage change in the dependent variable that is implied by the estimate and a two standard deviation increase in the regressor. This, of course, roughly corresponds to a comparison of cities in the 95th and 50th percentiles of the city size distribution.

The results in Figure 1 indicate that agglomeration did indeed have a considerable positive impact on the adoption of new ideas as inputs to the inventive process throughout most of the 20th century. Our results confirm that Marshall both accurately described the importance of agglomeration in the innovative process and that he was prescient in predicting the importance of this link in the decades following his writings.

The results in Figure 1 also indicate that the advantage of larger cities may now be declining. This is

[7] In defining which counties, cities, and towns form each PMSA we follow the 1999 definitions of the *Federal Office of Management and Budget* ("OMB"), with the exception of PMSAs that include areas in Massachusetts. The OMB definitions of PMSAs for Massachusetts map different towns within a county to different PMSAs. Because county information is more reliably extracted for patents granted before 1920s, we redefine PMSAs that include areas in Massachusetts so that also every county in Massachusetts is mapped to just one PMSA. For each PMSA we construct an annual population density time series based on US Census data on county-level population and area size (see Minnesota Population Center, 2011).

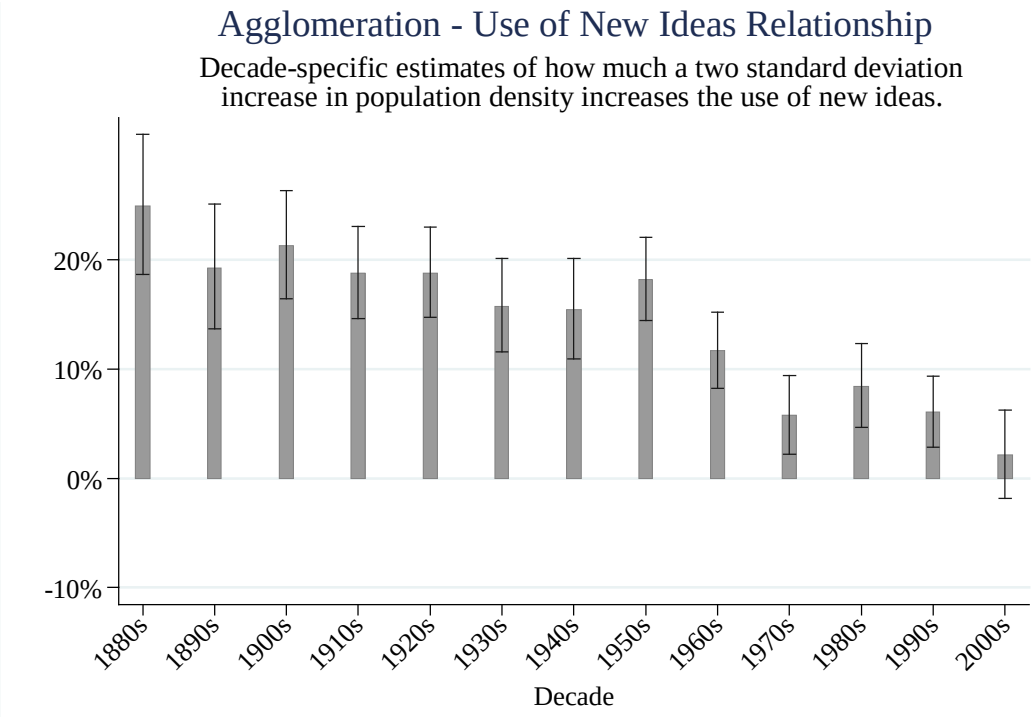

Figure 1: Estimates of the agglomeration–age of idea inputs relationship by decade. The outcome variable is the patent-level top 5% status by the age of the newest idea input in a patent (*Top 5% by Age of Idea Inputs*). The main regressor is logarithm of the population density of the city of the first inventor in the patent. Estimates reflect a comparison of cities in the 95th and 50th percentiles of the city size distribution. A positive estimate implies that inventors in larger cities use newer idea inputs than inventors in smaller cities. Capped lines indicate 95% confidence intervals.

suggested by a comparison of the estimates for the most recent decades (1980s through 2000s) against the results for decades following Marshall's writing. Results from a formal test of the hypothesis that the agglomeration–use of new ideas link has eroded in recent decades are shown in Table 1.

For the analysis in Table 1, we first divided the data into four time periods: 1880s-1910s (the time of Marshall's writing), 1920s-1960s, 1970s-1980s, and 1990s-2000s. We then interacted the main regressor in specification (1) with dummy variables representing each time period (as opposed to estimating the model separately for each decade or time period). In this table, column 1 shows the results for the conditional logit specification, column 2 shows the result for the linear probability model, column 3 shows the result for the linear probability model when city size is instrumented with values from 50 years prior, and column 4 shows the result for the alternative measure of city size (total population). Across the columns the estimates in Table 1 show a decline in the agglomeration-use of new ideas relationship from the period following Marshall's writing to the most recent decades. This finding is also supported by the formal tests, which are shown on the last two rows of the table.

Analyses reported in Table A2 in the appendix explore the robustness of this pattern to alternative specifications. Columns 1 and 2 explore the robustness to using narrower comparison groups.[8] Column 3 explores the robustness to constructing the dependent variable based on the top 20% status by recency of idea inputs rather than the top 5% status. Column 4 extends the analysis from the mentions of the top 100 concepts to mentions of the top 10,000 concepts in each cohort. Columns 5 and 6 limit the analysis to lone- and team-authored patents, respectively. In each case the results reported in Table A2 continue to follow the qualitative pattern found above, though the flattening of the agglomeration–use of new ideas relationship is not statistically significant when the sample is limited to team-authored patents (column 6 of Table A2). Non-parametric estimates reported in Figure A4 in the appendix provide further

[8]In column (1) the comparison group for each patent is patents granted in the same patent subclass in the same year. This departure from the specification used in Table 1 decreases sample size considerably because many comparison groups then have only one patent. It also changes the interpretation of the estimates because many of the remaining comparison groups have very few observations, which implies that being in the top 5% by the recency of idea inputs is no longer as selective a characteristic (even in comparison groups with very few patents, the patent with the most recent idea input is assigned the top 5% status by the recency of the age of idea inputs). To provide a comparison to results obtained using broader vs. narrower comparison groups, in column 2 we use the same dependent variable as in column 1 but include a fixed effect for each patent class and year pair rather than each patent subclass and year pair.

Table 1: Estimates of agglomeration–use of new ideas relationship by time period. A separate estimate of β for each time period is obtained by interacting the logarithm of the city size variable (population density or total population) with time period dummies. Each city size variable is normalized so that the estimates of the β coefficients represent changes associated with a 2 standard deviation increase in city size.

Dependent variable: *Top 5% by Age of Idea Inputs*.

	(1)	(2)	(3)	(4)
Specification:	Conditional Logit	Linear	Linear	Conditional Logit
Estimation:	ML	LS	IV	LS
Measure of City Size:	Population Density	Population Density	Population Density	Total Population
	Odds ratios:	Coefficients:	Coefficients:	Odds ratios:
β for 1880s-1910s	1.20***	.008***	.008***	1.20***
	(.013)	(.001)	(.001)	(.012)
β for 1920s-1960s	1.15***	.007***	.006***	1.09***
	(.009)	(.000)	(.000)	(.008)
β for 1970s-1980s	1.08***	.003***	-.001	1.00
	(.014)	(.001)	(.001)	(.011)
β for 1990s-2000s	1.04***	.002***	-.005***	1.02
	(.013)	(.001)	(.001)	(.012)
Fixed Effects	Year-Tech-Class Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs
Observations	3752553	3752553	3570511	3752553
Number of Fixed Effects	46264	46264	43144	46264
Mean of Dep. Var.		.044	.046	
		(.000)	(.000)	
Test of $\beta_{1970s-1980s} = \beta_{1920s-1960s}$	p= .000	p=.000	p=.000	p=.000
Test of $\beta_{1990s-2000s} = \beta_{1920s-1960s}$	p= .000	p=.000	p=.000	p=.000

A separate fixed effect is included for each technology class and year pair.

Standard errors in parentheses; clustered by year and technology class pair. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

corroborating evidence.

To gauge the importance of agglomeration in the use of new ideas relative to other factors, we conduct two additional sets of analyses. In one set of analyses we examine the effect of collaboration on new idea adoption: we regress the outcome variable *Top 5% by Age of Newest Idea Input* on a dummy variable measuring whether a patent lists multiple inventors or a lone inventor. For this analysis we focus on patents with a US first inventor. In the other set of analyses we examine the effect of being located in the US on new idea adoption: we regress the preferred outcome variable on a dummy variable measuring whether the first inventor was located in the US or in a foreign country. In both sets of additional analyses, we again include also a separate fixed effect for each technology class and year pair.

The results from these additional analyses are depicted in Figures 2 and 3. The corresponding numerical values are reported in Tables A2 and A3 in the appendix. For ease of interpretation, in Figures 2 and 3 each estimate of the coefficient of interest is expressed as the percentage change in the dependent variable that is implied by the estimate and a unit increase in the regressor.

In Figure 2, a positive estimate implies that patents by a team of inventors employ newer ideas than patents by a lone inventor. We find that teams of inventors are much more likely to apply fresh knowledge than lone inventors. An intriguing avenue for future work is examining to which extent this result arises because each inventor in a team brings in their own knowledge of existing ideas to the team and to which extent the result arises because inventors working in teams can solve the mysteries of new ideas faster than lone inventors through a vigorous debate on the new ideas' merits.

In Figure 3, a positive estimate implies that patents which first inventor lives in the U.S. employ newer ideas than patents which first inventor lives elsewhere. The results show that patents developed by inventors living in the U.S. are more likely to use newer ideas than patents developed by inventors living elsewhere. This result provides direct evidence on the extent and evolution of US leadership in technological invention. Taken together our results suggest that in the late 20th century agglomeration has become less important to invention in both absolute terms and relative to other factors – like collaboration – that predict the use of newer ideas.

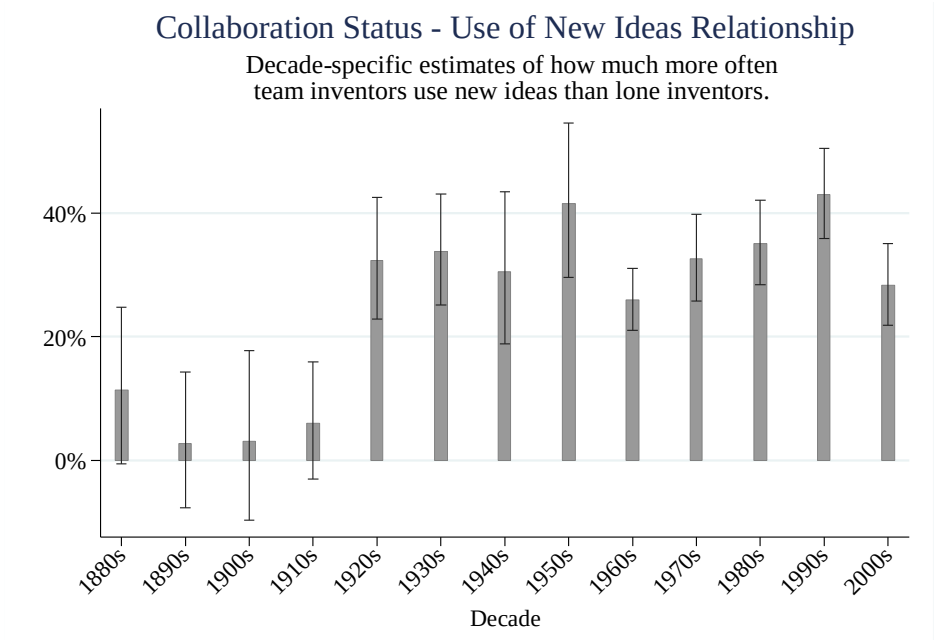

Figure 2: Estimates of collaboration status–use of new idea inputs relationship by decade. The outcome variable is the top 5% status by recency of the newest idea input in a patent (*Top 5% by Age of Idea Inputs*). A positive estimate implies that inventor teams use newer idea inputs than lone inventors. Capped lines indicate 95% confidence intervals.

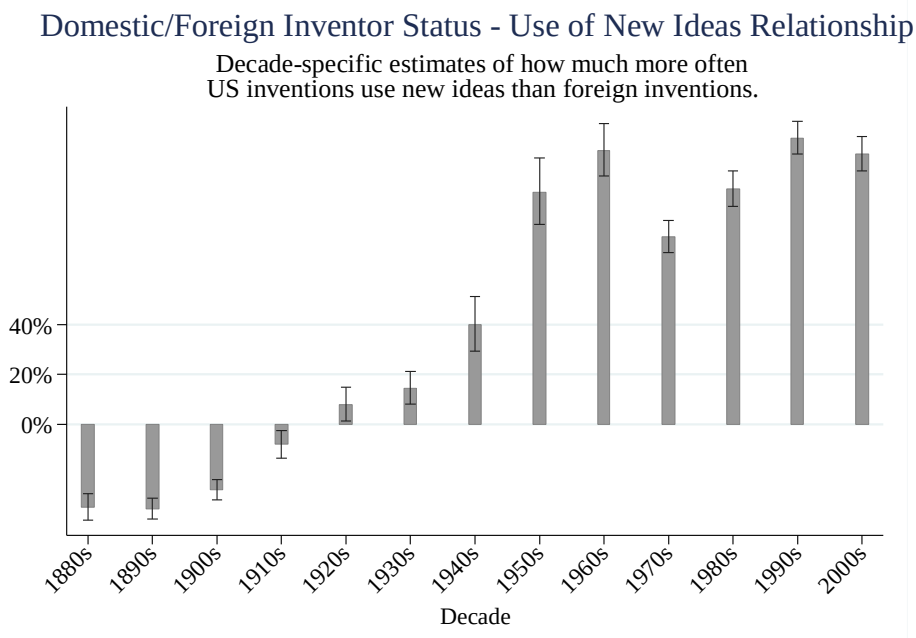

Figure 3: Estimates of domestic inventor status–use of new idea inputs relationship by decade. The outcome variable is the top 5% status by recency of the newest idea input in a patent (*Top 5% by Age of Idea Inputs*). A positive estimate implies inventors located in the U.S. use newer idea inputs than inventors located elsewhere. Capped lines indicate 95% confidence intervals.

5 Conclusion

Our empirical findings indicate that during the 20th century inventions in large US cities built on recent advances much more often than comparable inventions in smaller US cities. The findings also indicate that during the most recent decades this advantage of large cities has waned. The advantage of locating R&D resources in large cities over locating the same resources in smaller cities thus seems to be much smaller now than it has been in the past.

It is interesting to speculate why the convergence occurred. One possibility is that the spread of the internet and other communication technologies has made new ideas available more cheaply to all, regardless of where they were developed. Another possibility is that these new technologies have expanded the sizes of the communities that examine and debate the merits of each a new idea from a local to a wider scale. At the same time, our results on US leadership relative to other countries suggest that the communities have *not* yet become truly global as Marshall (1920) speculated might one day happen.

A particularly intriguing direction for future work is distinguishing between skill driven differences and physical proximity driven differences. Neither our application nor applications of the localization based approaches (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Thompson, 2006; Murata et al., 2013) have distinguished between benefits of agglomeration that may arise from knowledge differences due to physical proximity to others and benefits of agglomeration that may arise from knowledge differences due to skill-differences. In particular, more able inventors may prefer to live in larger cities and be better equipped to solve the mysteries of new ideas and thus adopt them quicker as inputs in their own work. While the reduced-form application provided here is an important first step, and while efficient allocation of R&D resources does not necessarily require knowing the mechanism by which larger cities may have an advantage, distinguishing between the skill and physical proximity mechanisms seems a valuable direction for future work.

References

Brezis, E. S. and P. R. Krugman, 1997, "Technology and the Life Cycle of Cities," *Journal of Economic Growth*, 2, pp. 369-83.

Duranton, G. and D. Puga, 2013, "Growth of Cities," Mimeo.

Hall, B. H., Jaffe, A. B. and M. Trajtenberg, 2001, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper No. 8498.

Henderson, R., Jaffe, A. B. and M. Trajtenberg, 2005, "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment," *American Economic Review*, 95(1), pp. 461-4.

von Hippel, E., 1994, "Sticky Information" and the Locus of Problem Solving: Implications for Innovation," *Management Science*, 40(4), pp. 429-39.

Jaffe, A. B., Trajtenberg, M. and R. Henderson, 1993, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108(3), pp. 577-98.

Jaffe, A. B., Trajtenberg, M. and M. S. Fogarty, 2000, "The Meaning of Patent Citations: A Report on the NBER/Case-Western Survey of Patentees," NBER Working Paper No. 7631.

Jones, B. F., 2010, "Age and Great Invention," *Review of Economics and Statistics*, 92(1), pp. 1-14.

Kuhn, T. S., 1962, *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lai, R., D'Amour, A., Yu, A., Sun, Y. and L. Fleming, 2013, "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)," Mimeo.

Marshall, A., 1920, *Principles of Economics*, 8th ed., London: Macmillan and Co.

Murata, Y., Nakajima, R., Okamoto, R. and R. Tamura, 2013, "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," Mimeo.

Minnesota Population Center, 2011, National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota. www.nhgis.org.

Packalen, M. and J. Bhattacharya, 2012, "Words in Patents: Research Inputs and the Value of Innovativeness in Invention," NBER Working Paper No. 18494.

Packalen, M. and J. Bhattacharya, 2015a, "Age and the Trying Out of New Ideas," Mimeo.

Packalen, M. and J. Bhattacharya, 2015b, "New Ideas in Invention," Mimeo.

Rosenberg, N., 1982, *Inside the Black Box: Technology and Economics*. Cambridge University Press.

Thompson, P. and M. Fox-Kean, 2005, "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review*, 95(1), pp. 450-60.

Thompson, P., 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations," *Review of Economics and Statistics*, 88(2), pp. 383-8.

Usher, A. P., 1929, *A History of Mechanical Inventions*. New York: McGraw-Hill.

Appendix: Additional Figures and Tables

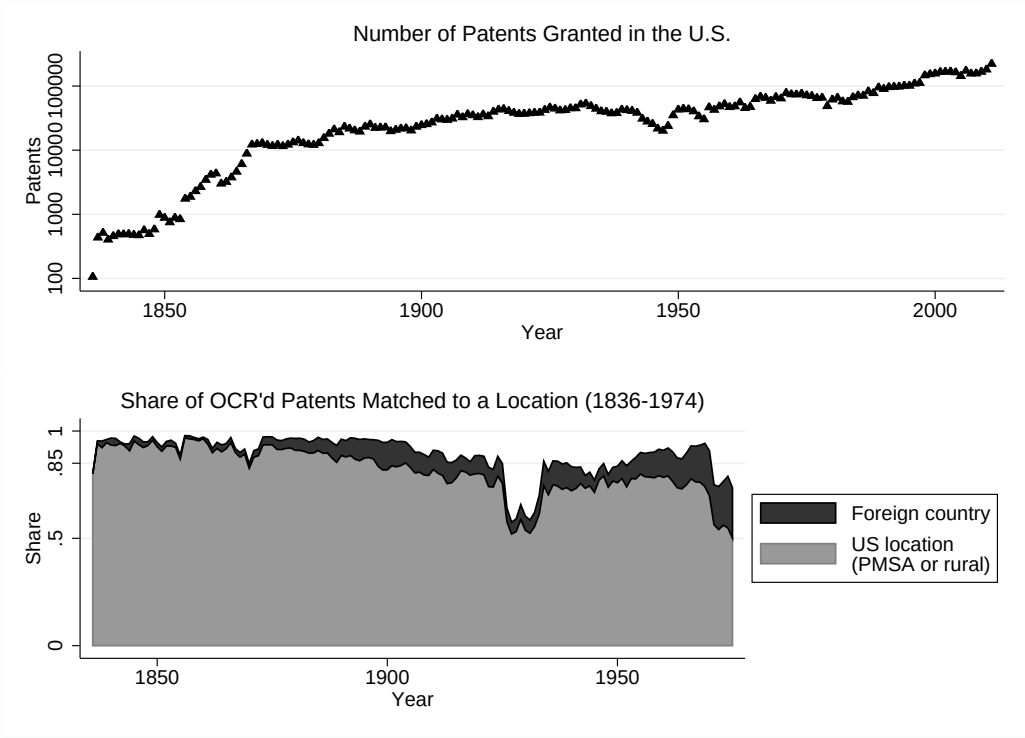

Figure A1: Number of US patents granted (top panel) and match rate of OCR'd patents to a PMSA or rural US location or to a foreign country (bottom panel). In the top panel, the vertical axis depicts the logarithm of the number of patents.

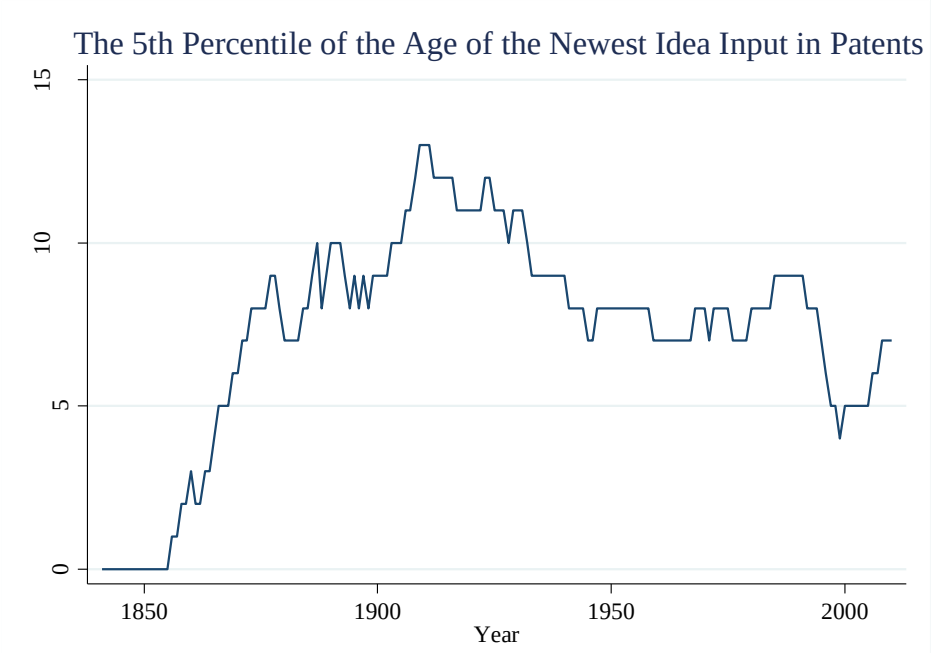

Figure A2: The 5th percentile of the age of the newest idea input in patents, calculated based on the top 100 idea inputs in each idea cohort.

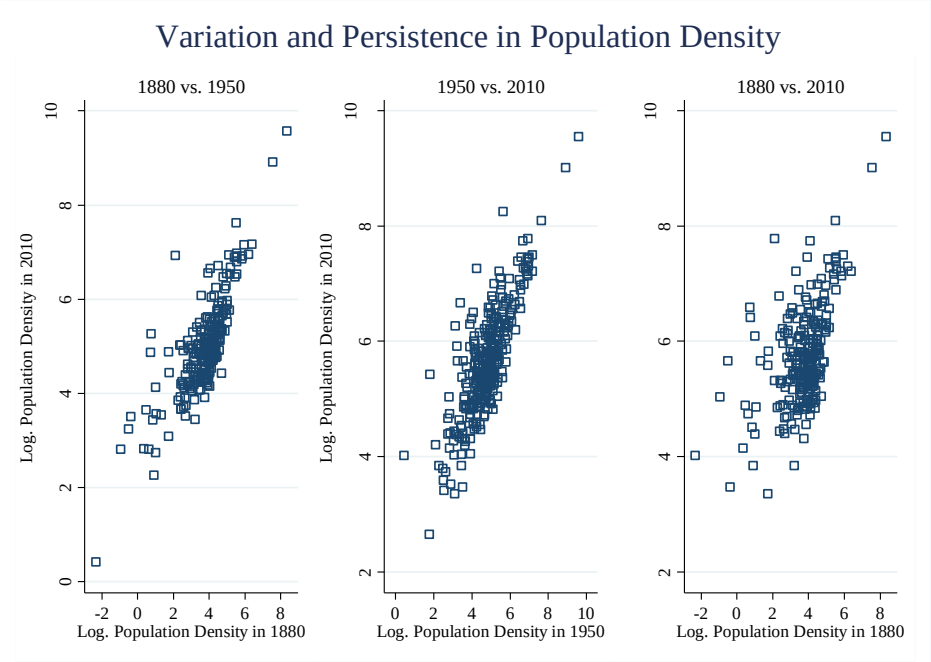

Figure A3: Population density in 1880 vs. 2010, 1950 vs. 2010, and 1880 vs. 2010. This figure provides a summary of how the distribution of population density has evolved over the last 130 years. Population density is measured by individuals per square mile. The figure depicts observations on the logarithm of population density for 1880 vs. 1950 (left panel), for 1950 vs. 2010 (center panel), and for 1880 and 2010 (right panel). It is striking that the most population dense cities remain near or at the top of the distribution. For instance, the two densest PMSAs remain New York and New Jersey throughout these periods. The persistence and the considerable within-year variation in population density enable us to identify the link between population density and use of new knowledge in invention.

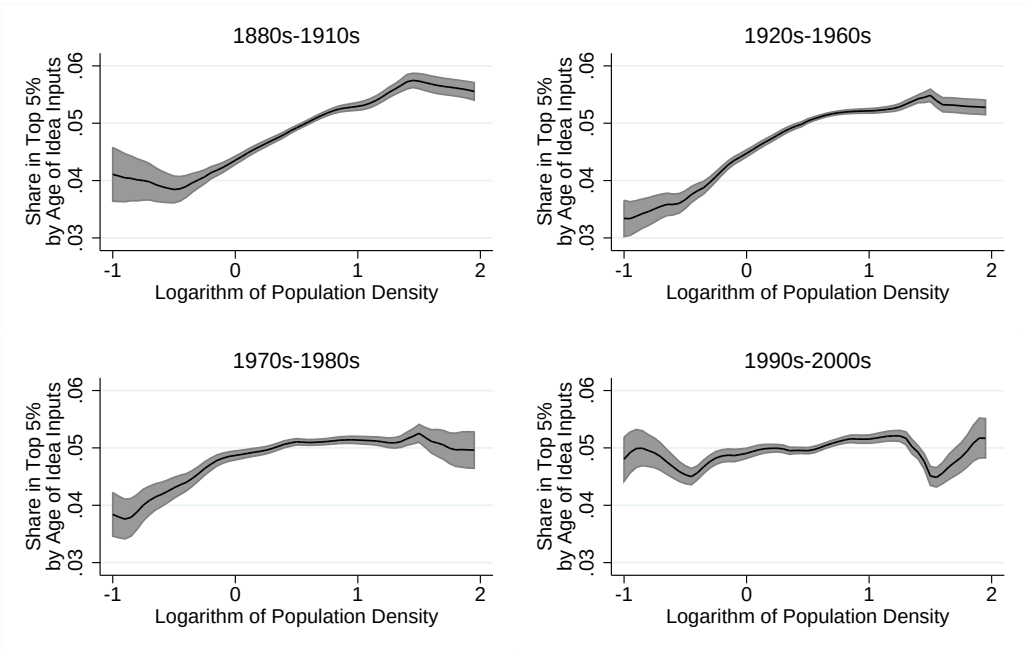

Figure A4: Non-parametric estimates of the agglomeration–age of idea inputs relationship by time period. The outcome variable is the patent-level top 5% status by the age of the newest idea input in a patent (*Top 5% by Age of Idea Inputs*). Observations on the explanatory variable, Logarithm of Population Density, are normalized for each year so that for cities this measure has mean of zero and standard deviation of one (the normalization is at the city-level). We calculate the weighted mean of the outcome variable at different values of the explanatory variable using the Epanechnikov kernel and a bandwidth of 0.5. Analysis is limited to those values of the explanatory variable that are between -1 and 2 because the sparsity of observations outside of this region renders the associated estimates much less informative. Shaded area indicates 95% confidence interval.

Table A1: Estimates of agglomeration–use of new ideas relationship by decade. The city size variable is normalized so the reported odds ratios represent changes associated with a 2 standard deviation increase in city size.

Dependent variable: *Top 5% by Age of Idea Inputs*.

Panel A. Model: Conditional Logit.

	(1) 1880s	(2) 1890s	(3) 1900s	(4) 1910s	(5) 1920s	(6) 1930s	(7) 1940s	(8) 1950s	(9) 1960s	(10) 1970s	(11) 1980s	(12) 1990s	(13) 2000s
	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:
Log Population Density	1.246***	1.198***	1.210***	1.168***	1.197***	1.151***	1.157***	1.162***	1.102***	1.063***	1.090***	1.059***	1.020
	(.033)	(.029)	(.025)	(.020)	(.021)	(.022)	(.023)	(.019)	(.018)	(.019)	(.020)	(.016)	(.021)
Fixed Effects	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs
Observations	140956	155975	202302	248598	239627	249381	203993	290532	372017	324904	350282	535103	438883
Number of Fixed Effects	3112	3215	3365	3493	3634	3727	3719	3869	3992	3632	4028	4067	2411

Standard errors in parentheses; clustered by year and technology class pair. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel B. Model: Linear Probability.

	(1) 1880s	(2) 1890s	(3) 1900s	(4) 1910s	(5) 1920s	(6) 1930s	(7) 1940s	(8) 1950s	(9) 1960s	(10) 1970s	(11) 1980s	(12) 1990s	(13) 2000s
	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:	Coefficient:
Log Population Density	.011***	.009***	.009***	.008***	.008***	.007***	.007***	.007***	.005***	.003***	.004***	.003***	.001
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Fixed Effects	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs
Observations	140956	155975	202302	248598	239627	249381	203993	290532	372017	324904	350282	535103	438883
Number of Fixed Effects	3112	3215	3365	3493	3634	3727	3719	3869	3992	3632	4028	4067	2411
Mean of Dep. Var.	.052	.048	.049	.048	.048	.048	.047	.047	.047	.047	.047	.048	.048

Standard errors in parentheses; clustered by year and technology class pair. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Estimates of agglomeration–use of new ideas relationship by time period: robustness checks A separate estimate of β for each time period is obtained by interacting the logarithm of the city size variable with time period dummies. Each city size variable is normalized so that the estimates of the β coefficients represent changes associated with a 2 standard deviation increase in city size.

Dependent variable: *Top 5% by Age of Idea Inputs*, except column 4.

Model: Conditional Logit.

Measure of City Size: *Log Population Density*.

Explanations for columns:
(1) Comparison group for each patent is other patents granted in the same technology subclass in the same year, and a separate fixed effect is included for each technology subclass and year pair.
(2) Same dependent variable and sample as in column 1 but now a separate fixed effect is included for each technology class and year pair.
(3) Age of idea inputs is calculated based on mentions of the top 10,000 idea inputs in each cohort.
(4) Dependent variable is top 20% status by recency of newest idea input.
(5) Only lone-authored patents are included in the sample.
(6) Only team-authored patents are included in the sample.

.

	(1)	(2)	(3)	(4)	(5)	(6)
	Odds ratios:	Odds ratios:	Odds ratios:	Odds ratios:	Odds ratios:	Odds ratios:
β for 1880s-1910s	1.031***	1.006	1.166***	1.141***	1.196***	1.165***
	(.008)	(.006)	(.012)	(.007)	(.013)	(.030)
β for 1920s-1960s	1.061***	1.046***	1.125***	1.125***	1.173***	1.093***
	(.006)	(.005)	(.009)	(.005)	(.010)	(.017)
β for 1970s-1980s	1.030***	1.027***	1.055***	1.075***	1.087***	1.069***
	(.008)	(.007)	(.013)	(.008)	(.016)	(.021)
β for 1990s-2000s	.998	.987*	1.062***	1.072***	1.027	1.063***
	(.007)	(.006)	(.012)	(.009)	(.016)	(.019)
Fixed Effects	Year-Tech-Subclass Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs	Year-Tech-Class Pairs
Observations	2059486	2059486	3752553	3383627	2655184	1112879
Number of Fixed Effects	687900	42150	46264	45773	45701	35043
Test of $\beta_{1970s-1980s} = \beta_{1920s-1960s}$	p=.002	p=.038	p=.000	p=.000	p=.000	p=.364
Test of $\beta_{1990s-2000s} = \beta_{1920s-1960s}$	p=.000	p=.000	p=.000	p=.000	p=.000	p=.226

Standard errors in parentheses; clustered by year and technology class pair. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Estimates of collaboration–use of new ideas relationship by decade.

Dependent variable: *Top 5% by Age of Idea Inputs*.

The main regressor, *I(Team-authored)*, is a dummy variable that is 1 for team-authored patents and 0 for lone-authored patents.

Model: Conditional Logit.

	(1) 1880s	(2) 1890s	(3) 1900s	(4) 1910s	(5) 1920s	(6) 1930s	(7) 1940s	(8) 1950s	(9) 1960s	(10) 1970s	(11) 1980s	(12) 1990s	(13) 2000s
	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:
I(Team-authored)	1.114	1.027	1.031	1.060	1.323***	1.338***	1.305***	1.415***	1.259***	1.326***	1.351***	1.430***	1.283***
	(.065)	(.056)	(.070)	(.048)	(.050)	(.046)	(.063)	(.064)	(.026)	(.036)	(.035)	(.037)	(.034)
Fixed Effects	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples	Year-Tech Class-PMSA-Triples
Observations	31214	32604	46693	62974	61117	68504	47406	68503	91856	72115	73288	157437	290718
Number of Fixed Effects	4287	4493	5751	7202	6960	7333	5948	8492	11001	9383	9828	14507	17002

Standard errors in parentheses; clustered by PMSA. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Estimates of domestic/foreign invention status–use of new ideas relationship by decade.

Dependent variable: *Top 5% by Age of Idea Inputs*.

The main regressor, *I(Domestic)*, is a dummy variable that is 1 for patents which first inventor is in the U.S. and 0 for patents which first inventor is in a foreign country.

Model: Conditional Logit.

	(1) 1880s	(2) 1890s	(3) 1900s	(4) 1910s	(5) 1920s	(6) 1930s	(7) 1940s	(8) 1950s	(9) 1960s	(10) 1970s	(11) 1980s	(12) 1990s	(13) 2000s
	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:	Odds Ratio:
I(Domestic)	.669***	.664***	.739***	.919**	1.080*	1.144***	1.398***	1.928***	2.095***	1.749***	1.941***	2.144***	2.081***
	(.027)	(.021)	(.021)	(.028)	(.034)	(.033)	(.055)	(.068)	(.054)	(.033)	(.036)	(.034)	(.035)
Fixed Effects	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs	Year-Tech Class-Pairs
Observations	194873	221073	304594	381084	414831	442771	307536	425901	567802	690372	705904	1037028	1285516
Number of Fixed Effects	3244	3385	3566	3653	3736	3820	3795	3919	4047	4106	4108	4103	4004

Standard errors in parentheses; clustered by year and technology class pair. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$