

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Filimonovic, Dragan; Rutzer, Christian; Macher, Jeffrey T.; Weder, Rolf

# Working Paper Does early regional scientific leadership translate into lasting innovation advantage?

WWZ Working Paper, No. 2024/11

**Provided in Cooperation with:** Center of Business and Economics (WWZ), University of Basel

*Suggested Citation:* Filimonovic, Dragan; Rutzer, Christian; Macher, Jeffrey T.; Weder, Rolf (2024) : Does early regional scientific leadership translate into lasting innovation advantage?, WWZ Working Paper, No. 2024/11, University of Basel, Center of Business and Economics (WWZ), Basel, https://doi.org/10.5451/unibas-ep96718

This Version is available at: https://hdl.handle.net/10419/304430

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU



Faculty of Business and Economics



October 2024

# Does Early Regional Scientific Leadership Translate Into Lasting Innovation Advantage?

WWZ Working Paper 2024/11 Dragan Filimonovic, Christian Rutzer, Jeffrey T. Macher, Rolf Weder

A publication of the Center of Business and Economics (WWZ), University of Basel. © WWZ 2024 and the authors. Reproduction for other purposes than the personal use needs the permission of the author

Universität Basel Peter Merian-Weg 6 4052 Basel, Switzerland wwz.unibas.ch Corresponding Author: Dr. Christian Rutzer Tel. +41 61 207 33 00 Mail : <u>christian.rutzer@unibas.ch</u>

# Does Early Regional Scientific Leadership Translate Into Lasting Innovation Advantage?

Dragan Filimonovic<sup>\*</sup>, Christian Rutzer<sup>†</sup>, Jeffrey T. Macher<sup>‡</sup>, and Rolf Weder<sup>§</sup>

October 2, 2024

#### Abstract

We examine whether 'pioneer' regions - early leaders in generating new ideas in emerging scientific fields - develop and maintain an innovation advantage in the same fields over time. Our analysis covers 24 disruptive technologies (e.g. AI, cloud computing) in thousands of OECD regions over 20 years. The results show that pioneer regions gain a significant and growing innovation advantage over non-pioneer regions. This advantage is most pronounced in "super-cluster" regions, which are leaders in both science and related innovation. These findings highlight the importance of early scientific leadership for sustained regional innovation and suggest important policy implications.

*Keywords:* Science, Innovation, Regional Advantage, Emerging Technology *JEL*: O30, O33, R11

<sup>\*</sup>Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland; University of Basel; dragan.filimonovic@unibas.ch.

<sup>&</sup>lt;sup>†</sup>Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland; christian.rutzer@unibas.ch.

<sup>&</sup>lt;sup>‡</sup>Robert E. McDonough School of Business, Georgetown University, 37th and O Streets NW, Washington, DC 20057, USA; jeffrey.macher@georgetown.edu.

<sup>&</sup>lt;sup>§</sup>Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland; rolf.weder@unibas.ch.

We are grateful to Scott Ganz, Bradford Jensen, and Nathan Miller for providing helpful comments. We also acknowledge the constructive feedback received in seminars at the University of Basel (WWZ), Georgetown University (McDonough), the University of Maryland (Smith), and the University of Connecticut and in a session at the DRUID Conference. We are also grateful for the support by the Claudine and Hans-Heiner Zaeslin-Bustany-Foundation.

### 1 Introduction

Basic research is considered the birthplace of breakthrough ideas, whereas applied research is the workplace where these ideas are further developed and become innovations which eventually lead to new products and processes and solve real-world problems. Both research activities are key components of the innovation life-cycle process, which posits that new technologies undergo several stages before they reach their full potential.<sup>1</sup> The scientific (basic research) stage occurs when the scientific community begins to produce and cite research papers in a new field, signifying its acceptance and maturity. The innovation (applied research) stage follows as these scientific ideas and discoveries consolidate into innovations, characterized by dynamic patenting activity. The market (commercialization) stage occurs when firms use and implement these technologies in new products and processes.

The innovation life-cycle process highlights in particular how knowledge diffuses across these stages over time. However, knowledge diffusion occurs not only over time, but also across space: i.e., within and across geographical units. This spatial dimension of knowledge can lead to the formation of industry clusters.<sup>2</sup> Factors contributing to the strengthening of industry clusters include common input sharing,<sup>3</sup> high quality labor market matches,<sup>4</sup> and direct (tacit) knowledge transfers.<sup>5</sup> Many studies also document a positive relationship between a region's knowledge capital and innovation activities,<sup>6</sup> the level of wages,<sup>7</sup> and growth.<sup>8</sup> Although these studies indirectly assess the importance of intra-regional knowledge flows, they do not analyze why other regions may struggle to utilize and commercialize new knowledge for their own purposes. As emphasized by Heimeriks et al. (2019), the complexity and structure of certain types of knowledge can make it difficult for non-local actors to imitate, absorb, or use that knowledge effectively. Geographical proximity may therefore facilitate or simplify knowledge exchange.<sup>9</sup>

The existing literature reveals two key insights: (1) the innovation life cycle process highlights the importance of basic research for innovation in the early stages of emerging technologies; and (2) geographical proximity seems to significantly influence knowledge transfer in general. However, it remains unclear whether scientific proximity is particularly crucial for innovation in the early stages of emerging technologies. In this paper, we address this important gap by examining whether pioneer regions that lead in the generation of new ideas in an emerging scientific field gain and subsequently maintain an advantage in generating innovations in the same field, compared to non-pioneer regions. We thus contribute to a growing literature that embeds

 ${}^{4}$ See e.g., Wheeler (2001), Fallick et al. (2006), Strange et al. (2006).

 $^7\mathrm{See}$  e.g., Moretti (2004), Rosenthal & Strange (2008), Wang (2016), de la Roca (2017).

 $<sup>^1 \</sup>mathrm{See},$  e.g., Ernst (1997), Chanchetti et al. (2016), Bornkessel et al. (2016), Baumann et al. (2021).

<sup>&</sup>lt;sup>2</sup>For example, Arora et al. (2021) finds that high-tech innovators more readily locate in places with strong innovation activities in the same field, despite the higher incurred costs relative to other locations. <sup>3</sup>See Saxenian (1991), Porter (1998), Sturgeon (2002), Ellison et al. (2010), Kolympiris et al. (2011).

 $<sup>{}^{5}</sup>$ See e.g., Peri (2005), Alcácer & Gittelman (2006), Singh & Marx (2013), Balland & Rigby (2017), Head et al. (2019).

<sup>&</sup>lt;sup>6</sup>See e.g., Jaffe (1989), Jaffe et al. (1993), Anselin et al. (1997), Kerr (2010), Moretti (2021).

 $<sup>^8 \</sup>mathrm{See}$ e.g., Duranton (2007), Duranton & Puga (2014), Davis & Dingel (2019), Davies & Maré (2021), Mewes & Broekel (2022).

<sup>&</sup>lt;sup>9</sup>The interplay between basic and applied research at the local level is complex and not straightforward (Balland & Boschma 2022). Some scientific discoveries are not relevant to local industries (Cohen & Walsh 2002). Life sciences (e.g., chemistry, biology, pharmacology) and deep-tech (e.g., superconductors, AI) innovations exhibit stronger convergence with scientific research (Narin et al. 1995, Verbeek et al. 2002, Guan & He 2007, Glänzel & Zhou 2011, Ahmadpoor & Jones 2017, Gazni 2020). Differences exist even within industries: Sung et al. (2015) highlight variations in science-innovation linkages among top patenting firms in Computers and Communications; Glänzel & Zhou (2011) note stronger linkages in "red" biotechnology fields (e.g., genetics, cell biology) than in "white" biotechnology (e.g., applied chemistry, food science). Wang & Li (2021) emphasize geographic proximity may be important for the transfer of so-called "tacit" knowledge.

the process of endogenous innovation and creative destruction into a geographic context with insights around industry clustering, geographic concentration, hysteresis and industrial policy.<sup>10</sup> Understanding these dynamics is important for informing firm location strategies and early-stage innovation policies.

Bikard & Marx (2020) and Balland & Boschma (2022) represent the most closely related research. The former finds that geographically concentrated hubs enhance knowledge creation and exchange between local institutions and distant firms. The latter emphasizes that, at the regional level, greater scientific activity correlates with stronger innovation within the same broad technological domain. We complement and extend these research efforts in three key ways. First, we focus on specific disruptive technologies emerging in dynamically developing regions. Second, we incorporate an international perspective, examining regions in dozens of industrialized countries. Third, we examine how regional scientific capabilities in an emerging field influence innovation capabilities in the same field over time. Our approach provides insights into the persistence of regional science-innovation linkages throughout the technology-specific innovation life cycle. And we address a significant research gap regarding the role of geographic proximity to science centers in driving innovation during the early stages of emerging technologies.

We utilize and build upon Bloom et al. (2021) to identify several emergent scientific fields that have had a meaningful disruptive impact. These authors analyze the relationship between applied research and the geographic diffusion of new technologies into the labor market, whereas we analyze the relationship between basic research and the geographic diffusion of applied research. Bloom et al. (2021) thus examine the latter stages in the innovation life-cycle process – i.e., from innovation to market commercialization – while we analyze the earlier stages – i.e., from science to innovation. We measure basic research through scientific publications and applied research through patents.<sup>11</sup> Our empirical strategy uses unigram (single word sequence) and bigram (two-word sequence) keywords associated with emerging technologies to identify the universe of related scientific papers published (via the SCOPUS database) and patents granted (via the PatentsView database from the USPTO) within particular OECD country regions. For each emergent technology, we define pioneer regions as those that lead in scientific publication output when the technology reaches a specified maturity threshold and establish a unique 20-year period centered around its own scientific maturity. We then analyze the extent, magnitude, and duration of the patenting advantage that pioneering regions enjoy over non-pioneering regions due to their scientific lead over this 20-year period.

Our baseline dynamic fixed-effect model estimation results indicate that pioneer regions build a substantial patenting advantage in the early stages of the innovation life cycle – even prior to the emergent technology reaching scientific maturity – and maintain and grow this advantage over time. At the end of the observation window, pioneer regions generate approximately 50% more patents per year than non-pioneer regions. Refined estimations which exploit the heterogeneity of pioneer regions show that major and first-mover pioneer regions achieve larger, faster and more durable patenting advantages, respectively, in comparison to minor and second-mover pioneer regions. "Super-cluster" regions – viz., those with a scientific advantage in an emergent technology and an innovative advantage in related fields – achieve the largest, fastest and most durable patenting advantage, in comparison to strictly scientific-oriented or innovation-oriented regions. Alternative dependent variable estimations indicate that pioneer regions generate more highly-cited and more unique patents, suggesting that their innovation advantage pertains not only to quantity, but also to impact and novelty. Several other extensions and robustness tests around alternative time windows, pioneer region definitions, and emergent technology categories

<sup>&</sup>lt;sup>10</sup>This literature is based on the pioneering contributions of Romer (1990), Porter (1990), Krugman (1991) and Aghion & Howitt (1992), among others.

<sup>&</sup>lt;sup>11</sup>We assume basic research is effectively captured by scientific papers and applied research is effectively captured by patents. We nevertheless recognize that both innovation life cycle stages might entail activities that cannot be effectively captured by our constructs.

provide additional empirical support that region scientific advantage translates into substantial and lasting innovation advantage.

The empirical results have important firm location and innovation policy implications. For firms, the results suggest that innovation-related advantages accrue to locating in regions with strong scientific research bases, potentially due to common input sharing, R&D activity clustering, and tacit knowledge exchange. Specifically, we identify two research-intensive location options for firms: pioneer regions and super-clusters. For innovation policy, the results suggest that location matters and timing matters. We find that a basic-applied science relationship exists in emerging technologies at the regional level, operating either in a linear or a recursive fashion. We also find that regional innovation advantage in an emerging technology is created via early basic research prominence. These result underscore the importance of securing funding for initial scientific efforts in an emerging field and for strengthening the scientific capacity of particular regions. Fostering basic research in emerging technologies by supporting geographically-concentrated research clusters through targeted policies may facilitate long-term innovation advantage.

The rest of this paper is organized as follows. Section 2 presents our conceptual framework and empirical strategy. Section 3 describes the data and the identification of the key variables used in the empirical analysis. Section 4 presents the baseline and pioneer region heterogeneity empirical results. Section 5 discusses and presents various extensions and robustness tests of our analysis. Section 6 summarizes the empirical findings, discusses the potential limitations, and offers implications for firms and innovation policy.

# 2 Conceptual Framework and Empirical Strategy

To illustrate regional science and innovation interactions, consider two regions: A and B. Suppose Region A is a scientific leader in a particular emergent technology (i.e., a pioneer region), while Region B is a scientific laggard (i.e., a non-pioneer region). Further suppose that both regions have two active research groups: one in basic research; one in applied research. When the former research group makes new discoveries, their findings are typically communicated through scientific publications. The latter group focuses on developing innovations that can be patented or are kept as trade secrets.<sup>12</sup> Both groups use as the basis for their research the existing stock of available knowledge: for basic research, the body of related published scientific research; for applied research, the stock of related granted patents and trade secrets. Each group can function independently in their respective region, but also may benefit from the activities of the other group due to geographic proximity that facilitates knowledge transfers: i.e., new regional insights in basic research may trigger innovations in applied research and/or regional innovations in applied research may spark interest in basic research.

With this setup, consider the possible relationships and dynamics between the two regions. Assume that geographic distance between Region A and Region B provides a temporary intraregion information exchange advantage between the two groups. In particular, basic research breakthroughs in Region A (B) benefit applied research in Region A (B), while leaving applied research in Region B (A) largely unaffected—at least to some extent and for some time. An initial innovation advantage in Region A due to its scientific leadership in the emergent technology thus may be enduring. However, as knowledge generated in Region A (gradually) diffuses to Region B – either in basic research or in applied research – its innovation advantage may erode.

It is thus an empirical question as to whether, to what extent, and for how long pioneer regions have an innovation advantage over non-pioneer regions. We investigate this research

<sup>&</sup>lt;sup>12</sup>The former group is typically part of a university, whereas the latter group is mandated with a focused research objective and is typically part of a firm. See, e.g., Aghion et al. (2008) for a discussion of the different incentives and objectives of these groups.

question directly by employing an estimation strategy designed to assess whether regions with an initial lead in scientific publications (viz., Region A) are able to translate it into a sustained patenting advantage in comparison to regions without any such lead (viz., Region B) for several emergent technologies. Our baseline specification applies a dynamic version of the Bloom et al. (2021) model as follows:

$$Patents_{i,j,t} = \alpha + Pioneer_{i,j} + (t - t_{0,j}) + Pioneer_{i,j} \times (t - t_{0,j})$$
$$TotalPatents_{i,t} + Region_i + Technology_i + Year_t + \epsilon_{i,j,t},$$
(1)

where  $Patents_{i,j,t}$  is the log-transformed (+1) count of patents in region *i* for a specific emergent technology *j* at time *t*. *Pioneer*<sub>*i*,*j*</sub> is an indicator of whether region *i* is a pioneer in emergent technology *j*, and thus provides the average patenting advantage of pioneer regions within the same technology relative to non-pioneer regions. The variables  $(t-t_{0,j})$  are indicators of the years both prior to and since the emerging technology *j* reaches scientific maturity  $(t_{0,j})$  (defined in detail below). *TotalPatents*<sub>*i*,*t*</sub> is a count of the total number of patents of region *i* in year *t* across all technologies, and helps control for region size effects: e.g., regions with higher overall patenting activity may have inflated patent counts in specific technologies given their larger innovation base. *Region*<sub>*i*</sub>, *Technology*<sub>*j*</sub>, and *Year*<sub>*t*</sub> represent fixed effects to respectively control for additional regional, technological or temporal heterogeneity. Standard errors ( $\epsilon_{i,j,t}$ ) are clustered at the regional level. Our primary coefficients of interest are the linear combination of the average treatment effect of pioneer regions (*Pioneer*<sub>*i*,*j*</sub>) and its interactions with the temporal variables (*Pioneer*<sub>*i*,*j*</sub> × ( $t - t_{0,j}$ )). These measures together provide the overall patenting effect of pioneer regions over time.

We modify baseline specification (1) below to distinguish the effects of regional scientific and innovation capacity and to examine empirical robustness. In particular, we consider alternative pioneer region definitions (major vs. minor; first-mover vs. second-mover; super-cluster vs. scientific pioneer and innovation stronghold), alternative dependent variables (impact and novelty), emergent technology categories (physical vs. digital), as well as other measures and time windows.

### 3 Data and Identification

This section describes the data and identification approach utilized in the empirical analyses. It first details the region as the level of aggregation. It then explains how scientific publications and patents associated with the Bloom et al. (2021) emergent technologies are identified and assigned to regions. It then describes the approach used to determine when emergent technologies reach scientific maturity, as well as the criteria used to define pioneer regions. It finally describes the final sample used in the empirical analyses.

#### 3.1 Regions

The data include all OECD countries in Europe, North America, Asia and Oceania. The geographic level of aggregation is at the region: (1) OECD TL3 regions outside of the U.S.;<sup>13</sup> and (2) Office of Management and Budget (OMB) Core-Based Statistical Areas (CBSAs) inside the U.S. (Bloom et al. 2021). The alternative level of aggregation for the U.S. is beneficial because OECD U.S. TL3 regions are generally larger in comparison.<sup>14</sup> We assign scientific publications and patents to regions based on their latitude and longitude coordinates. Appendix Table 15 lists the number of regions per country in the dataset.

 $<sup>^{13}</sup>$ See e.g., Usai (2011), Rodríguez-Pose & Wilkie (2019), Balland & Boschma (2022) as examples using OECD region classification in regional innovation analyses.

<sup>&</sup>lt;sup>14</sup>The U.S. has 939 CBSAs but only 179 TL3 regions. By comparison, Germany has 401 TL3 regions and the United Kingdom has 179. See OECD (2022) for the number of OECD regions by country.

#### 3.2 Scientific Papers

We use the Scopus Web API to identify all scientific papers related to the Bloom et al. (2021) emergent technologies.<sup>15</sup> Appendix Table 16 lists these emergent technologies along with their unigram and bigram keywords. We use these keywords to search the titles and abstracts of all articles in the Scopus database, using stemmed versions of the keywords to capture conceptually identical variations: e.g., "autonom\*" for "autonomous" and "manufactur\*" for "manufactur-ing". An article that contains at least one of the keywords is considered connected to the respective emergent technology.

We next determine the geographic locations where the research was conducted, using author affiliations. We obtain geocodes (i.e., latitude and longitude values) using the Google Maps API, which allow us to assign papers to specific regions. Papers with multi-region authors are assigned to each region accordingly. Papers with multi-affiliated authors are assigned to the first affiliation listed for each author, as this typically represents the primary institution. Our approach successfully assigns more than 95% of all emergent technology-identified scientific papers to specific regions.<sup>16</sup>

#### 3.3 Patents

We use the United States Patent and Trademark Office (USPTO) PatentsView database to identify all patents granted (since 1976) related to the Bloom et al. (2021) emergent technologies. We implement a keyword search within the title, abstract, and claims of these USPTO patents, similar to the approach used to identify the Scopus scientific papers.

We control for the potential of similar invention double-counting by leveraging Patstat database information on DOCDB patent families from the European Patent Office (EPO).<sup>17</sup> Within each patent family, we select the first USPTO patent filed as our reference patent, which allows us to assign the invention to a specific year and to a specific region using the latitude and longitude of the inventors' addresses (as provided by PatentsView).<sup>18</sup> Patents with multi-region authors are assigned to each region accordingly. Our approach successfully assigns nearly 99% of all emergent technology-identified USPTO patents to specific regions.<sup>19</sup>

<sup>&</sup>lt;sup>15</sup>The Scopus database covers publications since 1788 and contains more than 90 million records (as of March 2023). Along with Web of Science, it is one of the leading bibliographic sources of scientific information and content (Elsevier 2023, Singh et al. 2021). We include all Scopus-listed publication types: peer-reviewed scientific articles, surveys, editorials, book chapters, books, scientific reports, conference papers, and data papers. Scopus lists each paper only once, removing earlier listings (e.g., conference papers) after the final peer-reviewed article is published. We include all publication types, as research in some emerging technologies (e.g., AI) is often published in other formats (viz., conference proceedings).

<sup>&</sup>lt;sup>16</sup>Appendix Table 17 provides the number and percentage of unmatched scientific paper-geocode combinations by country.

<sup>&</sup>lt;sup>17</sup>We do not include patents from other patent offices covered by the Patstat database for two primary reasons. First, USPTO patents from PatentsView provide full text (title, abstract, and claims) for patent attribution to an emergent technology. Second, geo-coordinates are needed to assign a patent to a specific region. The Patstat database only provides the title and abstract and lacks the latitude and longitude of inventors' addresses. We use the Patstat data to account for USPTO patents that belong to the same family.

<sup>&</sup>lt;sup>18</sup>We assign patents to regions based on the listed inventor address. There are cases where the inventor address may not accurately reflect the region where the actual innovation occurred. This limitation is well documented and widely accepted in the literature. If this discrepancy affects our results, it would likely lead to an underestimation of patenting activity in pioneering regions. This is because inventors often reside outside urban areas, where research and development labs are typically located.

<sup>&</sup>lt;sup>19</sup>Appendix Table 18 provides the number and proportion of unmatched geocode-patent combinations by country.

#### 3.4 Pioneer Regions

Pioneer regions are those that lead in scientific output as an emergent technology reaches scientific maturity and are determined in two steps. First, we define scientific maturity for each emergent technology as the point at which it becomes an established scientific field – characterized by a sufficiently large body of published research. Following the Bloom et al. (2021) methodology, we calculate the cumulative number of scientific publications associated with each emergent technology over time, starting from the first publication year up to the most recent available year. We create a standardized measure for each emergent technology by dividing this cumulative number (up to a given year) by the total number of papers of that technology (over all available years). Scientific maturity is then defined as the year (designated  $t_0$ ) when this measure exceeds ten percent. Arrow (1) in Figure 1 provides a graphical representation.





Second, we identify the regions that lead in scientific publications at scientific maturity for each emergent technology. We again adopt the Bloom et al. (2021) methodology and calculate the number of scientific papers a region published in an emergent technology during the decade prior to scientific maturity: i.e.,  $t_{-10}$  to  $t_0$ , as shown by arrow (2) in Figure 1. This decade can be conceptualized as a period of basic research: i.e., when scientific ideas and breakthroughs occur in the emergent technology. We divide the number of region-specific scientific papers by the total number of scientific papers in that emergent technology in the decade up to scientific maturity. Finally, we rank-order the regions according to their respective share in the emergent technology and calculate the cumulative sum of these shares. Regions that fall within the top 50 percent of the cumulative sum are classified as pioneer regions for that emergent technology; all other regions are classified as non-pioneer regions.<sup>20</sup> We further examine pioneer region heterogeneity by size, timing and type, which we discuss in detail below.

#### 3.5 Final Sample

Several regions have no or very few patents due to myriad factors, such as low innovation capacity (common in mountainous or structurally weak regions) or the inability to assign papers or patents to regions given incomplete geographic information. To mitigate the latter concern, we only include regions where at least 50 patents can be assigned across all years and all technologies (including non-emergent technologies). There are also large delays between the filing and the granting of patents.<sup>21</sup> We therefore limit our analysis to patents filed up to and including 2015, which provides a more comprehensive tally of regional innovation activity—at least in terms of patentable innovations. In addition, we focus only on emergent technologies that have existed long enough to allow for meaningful analysis of their evolution and impact. Specifically, we include technologies that emerged on or before 2005, which provides at least a ten-year window for analysis after reaching scientific maturity. As a result, we exclude the following Bloom et al. (2021) technologies: Wireless Charging, 3D printing, Social Networking, RFID tags, and Smart Devices.

Table 1 illustrates that the post-exclusion and restriction final sample comprises 2,286 regions and 24 (out of 29) Bloom et al. (2021) emergent technologies.<sup>22</sup> The endogenously determined scientific maturity year varies substantially across the emergent technologies: Millimeter Wave, Solar Power, and Fracking are the oldest; OLED Displays, Software-Defined Radio, Electronic Gaming, and WiFi are the most recent.<sup>23</sup> The proportion of regions classified as pioneers similarly varies across the emergent technologies: Cloud Computing (2.58%), Search Engines (2.19%), and Machine Learning/AI (2.01%) are among the most widespread; Lane Departure Warning Systems (0.31%) and Fingerprint Sensors (0.22%) are among the most concentrated. Appendix Figure 5 provides a visual representation of pioneer regions (denoted by thick borders) and the diffusion of their innovation activities in U.S. and European regions.

# 4 Empirical Results

We first present our baseline estimation that distinguishes between pioneer regions and nonpioneer regions. We then examine whether heterogeneity in pioneer regions impacts their innovation advantage.

<sup>&</sup>lt;sup>20</sup>We utilize a particular threshold to denote scientific maturity in an emergent technology, which we define as the ratio of cumulative to total scientific papers published. We nevertheless recognize there are other potential methodologies to define scientific maturity: e.g., lower or higher thresholds; exceeding a specific paper growth rate or a specific paper to patent ratio; etc., and leave it to future research to explore these alternatives. As we show below, however, the core findings remain highly robust to alternative thresholds, which suggests that the specific method used to define scientific maturity is not a significant driver of our results.

<sup>&</sup>lt;sup>21</sup>The average time from filing to grant for a USPTO patent is 23.3 months. Some patents experience substantially longer delays (USPTO 2021).

<sup>&</sup>lt;sup>22</sup>Appendix Table 15 shows the number of regions per country before and after applying these exclusions and restrictions. Appendix Table 20 provides the number of unique technology-region combinations per country and the proportion classified as pioneer regions.

<sup>&</sup>lt;sup>23</sup>Figure 6 shows graphically the years of scientific maturity for each technology. Appendix Table 19 lists the total number of patents and papers for each emergent technology. Appendix Tables 21 and 22 respectively list the 30 regions with the most scientific papers and patents across all emergent technologies.

Emergent Technology	Scientific Maturity Year	Pioneer Regions	Total Regions	Pioneer Share
Autonomous Cars	1995	21	2286	0.92
Bispecific monoclonal antibody	1992	13	2286	0.57
Cloud computing	1993	59	2286	2.58
Computer vision	1995	39	2286	1.71
Drug conjugates	1999	35	2286	1.53
Electronic gaming	2004	38	2286	1.66
Fingerprint sensor	2000	5	2286	0.22
Fracking	1990	14	2286	0.61
GPS	1996	14	2286	0.61
Hybrid vehicle electric car	2000	32	2286	1.40
Lane departure warning	2003	7	2286	0.31
Lithium battery	1997	28	2286	1.22
Machine Learning AI	1994	46	2286	2.01
Millimeter wave	1987	16	2286	0.70
Mobile payment	2003	24	2286	1.05
Oled display	2004	24	2286	1.05
Online streaming	1998	27	2286	1.18
Search Engine	2001	50	2286	2.19
Software defined radio	2004	14	2286	0.61
Solar Power	1987	28	2286	1.22
Stent graft	2000	26	2286	1.14
Touch screen	2001	30	2286	1.31
Virtual Reality	2000	42	2286	1.84
Wifi	2004	28	2286	1.22

 Table 1: Emergent Technologies and Pioneer Regions

*Notes:* The table displays the year of scientific maturity, the number of pioneer regions, the total number of regions, and the share of pioneer regions after applying the following restrictions: including only patents filed up to 2015, regions with more than 50 patents in total, and technologies with an emergence year at least ten years prior to 2015, i.e., 2005. Consequently, the technologies Social Networking, Smart Devices, RFID Tags, Wireless Charging, and 3D Printing are excluded from our final dataset.

#### 4.1 Baseline Estimation

Figure 2 plots the patenting advantage estimates of pioneer regions relative to non-pioneer regions. Appendix Table 3 provides the corresponding regression estimation. The results indicate that in the earliest periods prior to scientific maturity  $(t_{-10} \text{ to } t_{-7})$ , pioneer regions do not possess any innovation advantage, in comparison to non-pioneer regions. As the emergent technology progresses, however, pioneer regions achieve and grow a substantial patenting advantage: viz., pioneers file nearly 40% more patents in an emergent technology at scientific maturity  $(t_0)$  and roughly 50% more patents a decade later  $(t_{10})$ . These results suggest that a sufficient number of basic research breakthroughs are necessary before they effectively facilitate or drive applied research and follow-on innovations.

This result is in contrast to Bloom et al. (2021), who find that any initial regional advantage in technology adoption diminishes over time and eventually disappears. In particular, these authors examine the impact of pioneer regions in applied research on the number of job advertisements (a proxy for market commercialization) and find that the advantage over non-pioneer regions disappears at around 16 years. Although their analysis does not provide annual estimates, it is reasonable to infer a gradual advantage erosion that begins shortly after technology emergence.

Two plausible arguments might explain why we observe no pioneer region patenting advantage degradation. First, the knowledge diffusion barriers in our empirical setting may be more

Figure 2: Patenting Advantage of Pioneer Regions



<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions over time and relative to non-pioneer regions. The plotted estimated effect represents the linear combination of pioneer regions and time on patenting: (1) the average treatment effect of a pioneer region; and (2) the temporal effect (expressed yearly) of a pioneer region prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

substantial than those observed in Bloom et al. (2021), suggesting that linkages between the innovation life-cycle stages are distinct and operate differently. Geographic proximity might thus play a more critical role between basic and applied research than between applied research and market commercialization. Second, our observation period since scientific maturity may be insufficient to capture when any initial advantage begins to wane. We explore this potential directly in the next section.

#### 4.2 Heterogeneous Pioneer Region Estimations

Our baseline specification distinguishes between pioneer regions and non-pioneer regions. We further categorize regions to determine whether any differences prior to scientific maturity lead to a patenting advantage in an emergent technology: first, by the size and pace of basic research activities in a specific emergent technology; and second, by the presence – in either or both – basic research in a specific emergent technology and applied research activities in aggregated fields related to the specific emergent technology.

**Does the size and pace of basic research matter?** Previous research suggests that regional scientific specialization may be largely explained by the existing stock of scientific knowledge (Boschma et al. 2014). Regional scientific focus can also shift toward more radical research activities, however, building upon existing but less related expertise (Balland et al. 2019). Regional scientific activity in newly emerging fields may consequently involve strategic decisionmaking, such as when to engage in an emerging field and to what extent.

Similar to firms, regions can achieve advantages from having a larger volume of research or being first-movers in specific fields. For example, university researchers who are early entrants in a field tend to maintain their scientific dominance in the long run (Sabatier & Chollet 2017). Firms that are first to engage in scientific research within a given field often profit more than their competitors, particularly in non-crowded emerging areas (Arora et al. 2023). Disadvantages may also occur, however, such as high fixed-cost investment inflexibilities (Lieberman & Montgomery 1988) or free-riding (Jensen 2003). We therefore evaluate the effects of scale and timing related to scientific paper production in emerging technologies for pioneer regions in the following specification:

$$Patents_{i,j,t} = \alpha + PioneerCat_{i,j} + (t - t_{0,j}) + PioneerCat_{i,j} \times (t - t_{0,j}) + TotalPatents_{i,t} + Region_i + Technology_j + Year_t + \epsilon_{i,j,t},$$
(2)

where *PioneerCat* categorizes pioneer regions by the size or pace of basic research in an emergent technology. The size analysis considers the volume of scientific publications at scientific maturity, distinguishing between regions as: major pioneers (top 30 percent of cumulative sum of scientific output in the decade before scientific maturity); minor pioneers (top 30 to 50 percent of cumulative sum in the decade before scientific maturity); and non-pioneers. The pace analysis considers the speed of scientific publications at scientific maturity. We first apply a stricter criterion for determining the scientific maturity year: viz., when the cumulative number of scientific publications surpasses five percent in an emergent technology. This analysis distinguishes between regions as: first-mover pioneers (top 50 percent of cumulative sum of scientific output under the stricter criterion); second-mover pioneers (remaining pioneer regions as in the baseline specification); and non-pioneers. In both the size and pace analyses, non-pioneer regions serve as the reference category. All other components remain identical to baseline specification (1).

Figure 3 presents the estimated patenting advantage results using specification (2): panel (a) considers pioneer region research size; panel (b) considers pioneer region research pace. Appendix Tables 4 and 5 provide the corresponding regression estimations. Panel (a) shows that in the earliest stages prior to scientific maturity  $(t_{-10} \text{ to } t_{-8})$ , major and minor pioneer regions exhibit no to little patenting advantage, in comparison to non-pioneer regions. Both pioneer region-types establish a patenting advantage shortly thereafter (i.e.,  $t_{-7}$  for majors;  $t_{-6}$  for minors) and subsequently grow this advantage over the remaining sample window. Major pioneers outperform minor pioneers in patenting advantage over the entire sample window.

Panel (b) shows that in the earliest stages prior to scientific maturity  $(t_{-10} \text{ to } t_{-7})$ , firstand second-mover pioneer regions exhibit negligible patenting advantage, in comparison to nonpioneer regions. Both pioneer region-types eventually establish patenting advantages: at  $t_{-6}$  for first-movers; and at  $t_{-2}$  for minors. Both pioneer region-types subsequently grow this advantage over the remaining sample window. First-mover pioneers outperform second-mover pioneers in patenting advantage, although the latter does shrink this advantage in the later years.

In summary, the pioneer region heterogeneity estimations indicate: (1) minimal patenting advantages in the earliest pre-scientific maturity stages (which may reflect generally low emergent technology patenting levels at the beginning of the sample window); (2) sizeable and increasing patenting advantages thereafter; and (3) sensible patenting advantages from research size (viz., major pioneers outperform minor pioneers) and pace (viz., first-mover pioneers mostly outperform second-mover pioneers). At the same time, the respective patenting advantages of minor pioneers and second-mover pioneers, in comparison to non-pioneers, are substantial. These findings thus suggest that the size and pace of scientific discovery are important.

**Does a basic and/or applied research presence matter?** Our previous specification examines whether the size and pace of basic research in an emergent technology within a pioneer region facilitates applied research in that same emergent technology and pioneer region. In other words, these specifications assume a "first science, then innovation" relationship–often referred to as the "linear model". Building on Balland & Boschma (2022), we consider the possibility that innovation advantages may arise from factors beyond scientific leadership alone. In particular, we address critiques related to assumptions about the precise nature of the science-innovation





<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions over time and relative to non-pioneer regions. Panel (a) examines the research size of pioneer regions; Panel (b) examines the research pace of pioneer regions. The plotted estimated effects represent the linear combination of pioneer region type and time on patenting: (1) the average treatment effect of a pioneer region type; and (2) the temporal effect (expressed yearly) of a pioneer region type prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

relationship.<sup>24</sup> The linear model assumes that innovation begins with basic research (via scientific papers), progresses through applied research (via patents), and culminates in the market commercialization of new products and processes (Godin 2006). As the relationship between science and innovation might be multi-directional, we implement the following specification:

$$Patents_{i,j,t} = \alpha + PioneerType_{i,j} + (t - t_{0,j}) + PioneerType_{i,j} \times (t - t_{0,j}) + TotalPatents_{i,t} + Region_i + Technology_i + Year_t + \epsilon_{i,j,t},$$
(3)

where  $PioneerType_{i,j}$  categorizes regions as science pioneers, innovation strongholds, superclusters, and laggards. Science pioneers are regions that lead in scientific research output of an emergent technology at scientific maturity. These regions are identical to pioneers as in baseline specification (1), but lack any substantive corresponding innovation activity. Innovation strongholds are regions that lead in patenting activity in the broad fields that relate to specific emergent technologies, but lack any substantive corresponding scientific research output in the

<sup>&</sup>lt;sup>24</sup>As noted by Balconi et al. (2010), critics often point to the limited generalizability of the linear model and its oversimplification of the complex interactions among the various actors in the innovation process. The linear model nevertheless remains valid and useful, especially for analyses of science-intensive industries. Moreover, as policymakers and the general public seek to understand the tangible benefits of funding basic research (Poege et al. 2019), the literature continues to explore the science-innovation relationship using core linear model premises.

specific emergent technologies.<sup>25</sup> Super-clusters are regions that meet both the scientific pioneer and innovation stronghold criteria. These regions not only exhibit significant basic research (viz., scientific paper publications) in an emergent technology at its scientific maturity, but also exhibit substantial applied research (viz. patent grants) in related broader fields. Laggards are all other regions not in the aforementioned categories. All other components of this specification remain identical to baseline specification (1).

Figure 4 provides the estimated patenting advantage by region type using specification (3). Appendix Table 6 provides the corresponding regression estimate. Over the earliest periods prior to scientific maturity  $(t_{-10} \text{ to } t_{-4})$ , only super-clusters exhibit any patenting advantage relative to laggard regions. In contrast, science pioneers and innovation strongholds possess no such advantage and, in some cases, exhibit a disadvantage over this period. Moreover, over the entire sample window, super-clusters outperform all other region types: viz., their patenting advantage is nearly 100% at  $t_{+10}$ , indicating roughly twice as many new innovations on average as laggard regions. Science pioneers show a steady increase in patenting activity from  $t_{-5}$ , but never reach the same level of super-clusters. Innovation strongholds under-perform over the entire sample window – patenting at essentially the same level as laggard regions. These results thus suggest the following: first, intra-region complementarities appear to exist between scientific and technological capabilities; and second, technological capabilities alone – without a strong scientific base – is insufficient in achieving innovation success.

These results further suggest that super-clusters possess two important advantages: first, a strong basic research emphasis during the formative stages of an emergent technology and prior to its scientific maturity; and second, the technological capacity to develop these scientific ideas and insights into novel innovations. Geographical proximity of basic and applied researchers seems to foster these synergies, which may lead to the formation of knowledge bases that remain localized rather than diffusing to other regions. Regions with robust scientific capacity but lacking any initial technological capacity nevertheless do achieve a patenting advantage, relative to innovation strongholds and laggards. This finding is consistent with Balland & Boschma (2022), who observe that broader technologies are more likely to emerge in regions with a strong scientific research base in related fields. It also aligns with Gersbach et al. (2023), who note that firms are attracted to these regions because "basic research provides domestic firms with problem solvers, trained scientists, access to scientific networks, and, in general, better access to new knowledge" (p. 5).

Figure 4 also addresses a potential criticism of the baseline (Figure 2) and the heterogeneous pioneer region (Figure 3) specifications: namely, that any distinction between pioneer regions and non-pioneer regions may involve comparing research- and innovation-advantaged regions to heterogeneous groups of non-pioneer regions, which may be over-represented by underdevelopment and low R&D capacity. Our analysis notably shows, however, that science pioneers outperform innovation strongholds—a result that suggests scientific leadership not only confers a competitive advantage in patenting, but also perhaps protects regions from being overtaken or copied by other technologically-advanced regions. While scientific pioneer advantage is reduced relative to the baseline estimates, it follows a similar increasing trend over time. This consistency reinforces the

<sup>&</sup>lt;sup>25</sup>We identify innovation stronghold regions using a systematic approach: The World Intellectual Property Organization (WIPO) defines 35 aggregate technology categories for patents (Schmoch 2008). Using PatentsView data, we assign each USPTO patent to one of these categories. For each emergent technology j, we calculate the fraction of patents within each WIPO category. We then aggregate these fractions for each emergent technology j and region i over the ten years preceding its scientific maturity: t - 10to t = 0. Due to computational constraints, we consider only the four largest WIPO categories for each emergent technology. This process yields a weighted patent count within WIPO fields related to each emergent technology, which approximates each region's technological potential of being innovative in an emergent technology. Finally, we identify the regions with the highest revealed technological potential for each emergent technology – using the same methodology used to identify scientific pioneers – but use the weighted number of emergent technology patents as the defining factor.





<u>Notes</u>: The figure shows the estimated patenting advantage of region types over time and relative to laggard regions. The plotted estimated effects represent the linear combination of region type and time on patenting: (1) the average treatment effect of region type; and (2) the temporal effect (expressed yearly) of region type prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

robustness of our findings and shows the difficulty of achieving comparable innovation levels for the "average" non-pioneer region. Finally, the analysis suggests that technological leapfrogging – at least in this context – is a daunting task.

## 5 Extensions and Robustness Checks

This section presents several extensions and empirical robustness tests of our baseline and heterogeneous pioneer region estimations. It first examines alternative dependent variables via patent novelty and impact. It next examines longer time windows and distinct emergent technology categories. It finally conducts a series of alternative measurement approaches for pioneer regions and for patents.

**Innovation Impact and Novelty.** Our analyses have thus far considered the innovation advantage of pioneer regions in terms of numbers of patents. We next examine whether pioneer regions differ from non-pioneer regions in terms of patent impact and novelty. We use the number of forward citations a focal patent obtains to measure impact. We leverage the Patstat database to count the number of forward citations that the respective DOCDB patent family (which the focal USPTO patent belongs to) receives from other DOCDB patent families through 2022. We define high-impact patents as those among the ten percent most cited patents in an emergent technology and filing year.<sup>26</sup> In addition, we use data provided by Arts et al. (2021) to count the number of patents that introduce new words to measure novelty. The rationale is that patents that introduce major new concepts are often required to introduce new semantic terms to describe them.

 $<sup>^{26}</sup>$ Using a scientific field and year specific cut-off is common practice in the literature to account for different citation practices across technology areas and over time. See, e.g., Ahuja & Morris Lampert (2001), Arts & Veugelers (2015).

Appendix Figure 7 presents the alternative dependent variable results using specification (1): panel (a) uses the log-transformed (+1) number of highly-cited patents; panel (b) uses the log-transformed (+1) number of highly-novel patents (i.e., patents that introduce at least one new word).<sup>27</sup> These alternative measures have very similar effects – both to each other and with the baseline count measure. Pioneer regions thus provide patenting advantages relative to non-pioneer regions not only in number, but also in impact and novelty.

**Long-run Advantage.** We next examine alternative time windows. Our baseline specification requires ten years both prior to and since scientific maturity for meaningful analysis of the patenting evolution and impact of pioneer regions in an emergent technology. A 10-year post-scientific maturity window might not be sufficient, however, to determine the long-run patenting advantage effects. We examine this potential by restricting the sample to those emergent technologies that exist by 2015 for at least 15 years post-scientific maturity and re-estimate specification (1).

Appendix Figure 8 illustrates that the patenting advantage effect not only follows a similar trend as the baseline specification, but also continues to increase beyond the tenth post-scientific maturity year.<sup>28</sup> These results further reinforce the argument of a persistent patenting advantage of pioneer regions due to the difficulties that non-pioneer regions face in acquiring or absorbing highly-complex knowledge and ideas without a strong science foundation in the relevant scientific field. The accumulation of advanced scientific know-how thus appears a critical determinant in not only creating an initial innovation advantage, but also in maintaining that advantage over time. Pioneer region patenting advantage thus appears persistent in the medium- to long-run after an emergent technology reaches scientific maturity.

**Heterogeneous Technologies.** We next consider whether pioneer region patenting advantage is affected by the type of emergent technology. We split the emergent technologies into digital and physical categories,<sup>29</sup> for the following reasons. First, as knowledge can be categorized as codified (transferable) or tacit, new ideas and innovations in the digital domain may be more easily transferred or copied than those in the physical domain. Second, R&D in the physical domain may require more sophisticated and expensive investment (e.g., infrastructure, equipment) than is required in the digital domain. Finally, intellectual property rights protection may be more challenging for digital innovations (e.g., code) than for physical innovations. Appendix Table 16 provides the list of all emergent technologies by their digital or physical aspects.

Appendix Figure 9 provides the emergent technology category estimation results:<sup>30</sup> panel (a) considers digital emergent technologies; and panel (b) considers physical emergent technologies. Pioneer region patenting advantage increases consistently over the entire sample window in both categories but in different ways: for physical technologies, it results from the outset but increases gradually; for digital technologies, it lags from the outset but increases rapidly (especially prior to scientific maturity). These findings thus suggest that our baseline estimates are not substantially influenced by this emergent technology categorization. Regions with scientific leads in either a digital or physical emergent technology achieve largely comparable innovation advantages.

**Continuous Pioneer Region Measure.** We next examine the empirical robustness of our baseline results to alternative pioneer regions definitions. In particular, we replace the indicator

<sup>&</sup>lt;sup>27</sup>Appendix Table 7 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>28</sup>Appendix Table 8 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>29</sup>We classify the following as digital emergent technologies: "Machine Learning AI", "Computer vision", "GPS", "Online streaming", "Virtual Reality", "Search Engine", "Cloud computing" and "Electronic gaming". All others are classified as physical emergent technologies.

<sup>&</sup>lt;sup>30</sup>Appendix Table 9 provides the corresponding regression estimations.

variable that distinguishes pioneer and non-pioneer regions with a continuous measure of the number of scientific papers. In particular we use specification (1) and replace  $Pioneer_{i,j}$  with  $Papers_{i,j}$ , which measures the log-transformed (+1) number of scientific papers of region *i* in emergent technology *j* published until the year of scientific maturity (t = 0). This approach provides a rough estimate of the marginal product of basic research: i.e., the additional patented innovations a region generates when it achieves an extra unit (i.e., scientific paper) of basic research output.

Appendix Figure 10 provides the continuous treatment estimation results,<sup>31</sup> which indicate that the marginal product of basic research is roughly 0.06% at scientific maturity and subsequently grows over the remaining sample window to nearly 0.11%. Each additional 10% increase in scientific output in a given emergent technology increases patenting activity in a pioneer region by almost 1.1% by the end of the observed period (t = 10). These results further indicate that the marginal benefit of scientific research increases over time—similar to our baseline specification results.

**Distinct Continents.** We next examine the empirical robustness of our baseline results to specific geographic locations. In particular, different geographic regions might exhibit variations in factors associated with knowledge transfer, such as culture, political systems, sociodemographics, quality of education, and quality of transport and business infrastructure. These factors might impact the pace and effectiveness of knowledge flows between basic and applied research. We therefore divide our baseline sample into three continent sub-samples: North America (USA and Canada); Europe; and Asia. We then estimate baseline specification (1) separately for each geographic location.

Appendix Figure 11 provides the continent-specific estimation results:<sup>32</sup> panel (a) considers North America; panel (b) considers Asia; and panel (c) considers Europe. Pioneer regions achieve an innovation advantage regardless of geographic location, but inter-continent variation is exhibited: viz., North America achieves the largest patenting advantage; Europe achieves the smallest patenting advantage.<sup>33</sup> Asian pioneer regions interestingly start at a similar patenting advantage as their European counterparts but increase it more rapidly over the sample window.

Several factors might explain the more limited European pioneer region patenting advantage. One explanation that we directly examine is the size distribution of pioneer regions across continents. Appendix Table 2 indicates minor pioneer regions are predominantly in Europe (accounting for nearly 76% of all pioneer regions), whereas major and minor pioneer regions are more evenly distributed in North America and Asia. This result may also suggests that our estimates might be influenced to some extent by a U.S. emergent technology bias, as they are identified by Bloom et al. (2021) using U.S.-centric data. A different set of emergent technologies—perhaps identified using global data—might show more balanced inter-continent effects.

**Distance to Pioneers.** We next examine the empirical robustness of our baseline results to region geographic distance. As Section 2 suggests, an important determinant of knowledge transfer is proximity. We therefore examine whether innovation advantage varies for pioneer regions by their geographic distance to non-pioneer regions. In particular, we examine whether any scientific advantage "spills over" to more proximate non-pioneer neighbours leading to an innovation advantage relative to more distant non-pioneer regions.

We first determine neighbour regions for each pioneer region-emergent technology pair. We classify non-pioneer regions as neighbours if they are within 50, 100, or 200 km of a pioneer region.

<sup>&</sup>lt;sup>31</sup>Appendix Table 10 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>32</sup>Appendix Table 11 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>33</sup>As Appendix Figure 14 shows, the results remain robust when the analysis is limited to triadic patents. The use of triadic patents thus helps mitigate potential upward bias in favor of U.S. domiciled inventors due to reliance on USPTO patent data.

All other non-pioneer regions are classified as non-neighbors. Distance for each region is measured using mean longitudes and latitudes determined by averaging the geographical coordinates of all patent inventors within the region. This intra-region proximity measure thus takes into account distances across locations with varying innovation activity. We modify  $Pioneer_{i,j}$  in specification (1) into three categories: (1) pioneer regions; (2) non-pioneer neighbours; and (3) non-pioneer non-neighbours. We re-estimate the specification keeping all other components the same. The coefficients of interest are the total patenting advantage effect of pioneers [ $Pioneer_{i,j}(Pioneer) + Pioneer_{i,j}(Pioneer) \times (t-t_0)$ ] and the total patenting advantage effect of proximate neighbours [ $Pioneer_{i,j}(Neighbours) + Pioneer_{i,j}(Neighbours) \times (t-t_0)$ ], relative to distant neighbours.

Appendix Figure 12 provides the geographic proximity estimation results:<sup>34</sup> panel (a) panel (c) provide respectively increasing geographic distance thresholds. For pioneer regions, the results indicate a patenting advantage strongly similar to the baseline specification. For non-pioneer regions, the results indicate a (small) patenting advantage based upon proximity. Non-pioneer neighbors gain an early patenting advantage (at around  $t_{-4}$ ), and achieve roughly 3-5% more patents at scientific maturity and roughly 5-6% more patents at the end of the sample window, in comparison to non-pioneer non-neighbours. Varying geographic distance thresholds demonstrates little change in these results, suggesting that effective scientific-innovation linkages requires "geographic proximity".

**Triadic Patents.** We finally examine the empirical robustness of our baseline results using triadic patents as the patent data source. Triadic patents are filed in three major patent offices: the USPTO, EPO, and the Japan Patent Office (JPO). These patents are often considered to be of greater value because applicants are willing to incur the additional cost and effort to protect these innovations across multiple jurisdictions. The use of triadic patents might thus reduce the influence of any single jurisdiction's patenting practices, legal standards, or economic conditions. Given that our baseline sample is USPTO patents, the use of triadic patents also helps estimate the presence and extent of home advantage bias in our analysis – viz., given domestic innovators tend to file proportionally higher numbers of patents in their home country than abroad (Dernis & Khan 2004). Criscuolo (2006) notes, in particular, that triadic patents do not exhibit a home advantage bias favoring any specific country – unlike their USPTO and EPO counterparts. We estimate specification (1) with the dependent variable redefined as the log (+1) of the number of triadic patents.

Appendix Figure 13 provides the triadic patent estimation results,<sup>35</sup> which are strongly similar to the baseline USPTO estimation results albeit somewhat smaller in magnitude. We attribute this reduction to the generally lower number of triadic patents. Nevertheless, the results provide support that our findings are not driven by different patenting practices related to home bias or by any institutional differences. Appendix Figure 14 further supports this argument in an estimation of patenting advantage for triadic patents disaggregated by continent.<sup>36</sup> Overall, the use of triadic patents yields strongly similar results to those in the baseline estimation, helping confirm the robustness of our findings.

# 6 Discussion and Concluding Remarks

This paper examines whether, to what extent, and for how long regions that pioneer in producing science (basic research) in an emergent technology have an advantage in producing innovations (applied research) in the same technology, in comparison to non-pioneer regions. The empirical strategy examines 24 recent emergent technologies that have had a disruptive impact (Bloom

 $<sup>^{34}\</sup>mathrm{Appendix}$  Table 12 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>35</sup>Appendix Table 13 provides the corresponding regression estimations.

<sup>&</sup>lt;sup>36</sup>Appendix Table 14 provides the corresponding regression estimations.

et al. 2021). We use these technologies and their keywords to identify the universe of related scientific papers published (via the SCOPUS database) and patents granted (via the USPTO PatentsView database). We define pioneer regions as those that lead in scientific output at the time of scientific maturity of an emergent technology.

Our approach complements Bloom et al. (2021) by extending their analysis of innovation diffusion in the labor market to the earlier stages of the innovation life-cycle. These authors examine the demand for skills associated with emergent technologies in job advertisements (a proxy for technological adoption and commercialization), and observe a growing geographic dispersion over time. We focus on the same set of emergent technologies, and observe limited geographic dispersion in patenting advantage over time. Our combined analyses thus offer a more comprehensive view that maps the trajectory of emergent technologies from science and basic research, through innovation and applied research, and to commercialization and the practical application of new technologies. Our results imply that the role of geographic clustering for emergent technologies is more important in the early stages of the innovation life-cycle. For firms that develop innovations in new technologies, geographic proximity to scientific expertise in the relevant field appears critical. For other areas – e.g., where a firm uses but does not develop emergent technologies– proximity to scientific centers appears less critical as technologies tend to diffuse across geographic distances, making adaptation and implementation easier regardless of location (Bloom et al. 2021).

Our empirical approach also extends Balland & Boschma (2022) who suggest that the dynamic interplay between science and innovation – as well as regions beyond Europe – are worth studying. We address these research gaps directly by analyzing the interplay between science and innovation in dozens of emerging and disruptive technologies – each over a 20-year period centered around their scientific maturity and across thousands of OECD country regions. We find that pioneer regions achieve a significant and lasting patenting advantage that begins early and grows consistently over time. At the end of the 20-year period, pioneer regions have nearly 50% more granted patents than their non-pioneer counterparts. Moreover, our empirical analysis of the effects of different continents show that pioneer regions in North America gain a greater innovation advantage over time compared to their Asian and European counterparts.<sup>37</sup> This finding supports the notion that the U.S. is better at translating new scientific knowledge into innovation than other countries. And this finding is consistent with prior research: e.g., Conti & Gaule (2011) highlight the superior performance of U.S. universities in technology licensing; Fisch et al. (2015) show higher patenting propensities among U.S. universities compared to their European counterparts.

Our results point to the need to strengthen the scientific capacity of regions early in the technology life-cycle. The funding of scientific ideas in emerging technologies in areas related to a region's innovation potential appears paramount: i.e., both location and timing matter. Our empirical results indicate that super-cluster regions with relatively strong scientific expertise in an emerging technology and robust innovation capacity in the broader fields to which the technology belongs gain the largest and most durable innovation advantage in that technology. A smart innovation policy thus suggests funding basic research in emerging fields where a region has an established innovation advantage. An example in the field of lifesciences: Route 128 (Boston) in the U.S. or Basel in Switzerland. We recognize the associated difficulties in implementing this, as funding is required early in the process and targeted to specific areas. Such an approach would nevertheless help ensure that when breakthroughs do occur, these regions are well-positioned to evolve into super-clusters in the emerging technology, as local firms are better equipped to absorb and apply scientific discoveries. Our results also suggest that regions with strong innovation capabilities in the broader fields to which the emerging technology belongs, but weak

<sup>&</sup>lt;sup>37</sup>This finding is obtained using both USPTO patent data and triadic patent data, reinforcing the empirical robustness of our results. We acknowledge, however, the potential to replicate and extend these findings using other patent datasets.

scientific capabilities in that technology, generate only slightly more innovation over time than regions lacking in either science or innovation. This result implies that free riding on the scientific competences of others – even though scientific publications are widely available – is a difficult and under-performing endeavor. Innovation policy should instead strengthen the scientific capacity of such regions - especially in areas aligned with innovation capacity.

Some limitations of our analysis are worth noting. First, our empirical results may be influenced by our reliance on Bloom et al. (2021) as the source of emerging technologies, which are predominantly U.S.-centric. Technologies such as Cloud Computing, Machine Learning, AI systems, and Search Engines may disproportionately benefit North American regions where digital technology ecosystems are more established. This may subsequently bias the results toward greater innovation advantages in pioneer regions located there. Future research might replicate our findings using alternative or complementary sets of emerging technologies.

Second, our empirical results indicate a strong correlation between science and innovation in emerging technologies, but do not establish a direct causal relationship. It is possible that both processes occur simultaneously or that one precedes the other; it is also possible that the interplay may vary across emerging technologies. In certain instances, scientific advancements may occur prior to technological innovations, thereby establishing the foundation for novel products or processes. In other instances, the requirements of innovation may prompt scientific investigation, establishing a feedback loop in which the two processes evolve concurrently. We address this to some extent in our empirical analysis by examining four distinct region types. In addition, so long as there is no strong reverse causality from innovation to science, the exact nature of the relationship does not affect our policy implications of strengthening regional scientific capacity.

Finally, we do not examine the potential channels through which a strong science base in emerging technologies enables a region to become a leader in these technologies. We leave the identification of the most important channels of knowledge transfer to future research. Such analyses would further enhance the understanding of how regional scientific expertise and innovation advantage in emerging technologies are related.

# References

- Aghion, P., Dewatripont, M. & Stein, J. C. (2008), 'Academic freedom, private-sector focus, and the process of innovation', The RAND Journal of Economics 39(3), 617–635.
- Aghion, P. & Howitt, P. (1992), 'A model of growth through creative destruction', <u>Econometrica</u> 60(2), 323–351.
- Ahmadpoor, M. & Jones, B. F. (2017), 'The dual frontier: Patented inventions and prior scientific advance', Science 357(6351), 583–587.
- Ahuja, G. & Morris Lampert, C. (2001), 'Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions', <u>Strategic management</u> journal 22(6-7), 521–543.
- Alcácer, J. & Gittelman, M. (2006), 'Patent citations as a measure of knowledge flows: The influence of examiner citations', The Review of Economics and Statistics 88(4), 774–779.
- Anselin, L., Varga, A. & Acs, Z. (1997), 'Local geographic spillovers between university research and high technology innovations', Journal of Urban Economics 42(3), 422–448.
- Arora, A., Belenzon, S. & Dionisi, B. (2023), 'First-mover advantage and the private value of public science', Research Policy 52(9), 104867.
- Arora, A., Belenzon, S. & Sheer, L. (2021), 'Knowledge spillovers and corporate investment in scientific research', American Economic Review 111(3), 871–898.
- Arts, S., Hou, J. & Gomez, J. C. (2021), 'Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures', <u>Research</u> <u>Policy</u> 50(2).
- Arts, S. & Veugelers, R. (2015), 'Technology familiarity, recombinant novelty, and breakthrough invention', Industrial and Corporate Change **24**(6), 1215–1246.
- Balconi, M., Brusoni, S. & Orsenigo, L. (2010), 'In defence of the linear model: An essay', Research Policy **39**(1), 1–13.
- Balland, P.-A. & Boschma, R. (2022), 'Do scientific capabilities in specific domains matter for technological diversification in european regions?', Research Policy **51**.
- Balland, P. A., Boschma, R., Crespo, J. & Rigby, D. L. (2019), 'Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification', <u>Regional</u> Studies 53(9), 1252–1268.
- Balland, P. A. & Rigby, D. (2017), 'The Geography of Complex Knowledge', <u>Economic</u> <u>Geography</u> **93**(1), 1–23.
- Baumann, M., Domnik, T., Haase, M., Wulf, C., Emmerich, P., Rösch, C., Zapp, P., Naegler, T. & Weil, M. (2021), 'Comparative patent analysis for the identification of global research trends for the case of battery storage, hydrogen and bioenergy', <u>Technological Forecasting and</u> <u>Social Change</u> 165.
- Bikard, M. & Marx, M. (2020), 'Bridging academia and industry: How geographic hubs connect university science and corporate technology', Management Science **66**(8), 3425–3443.
- Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J. & Tahoun, A. (2021), 'The diffusion of disruptive technologies', NBER Working Paper 28999.

- Bornkessel, S., Bröring, S. & Omta, S. O. (2016), 'Crossing industrial boundaries at the pharmanutrition interface in probiotics: A life cycle perspective', PharmaNutrition 4(1), 29–37.
- Boschma, R., Heimeriks, G. & Balland, P.-A. (2014), 'Scientific knowledge dynamics and relatedness in biotech cities', Research Policy **43**(1), 107–114.
- Chanchetti, L. F., Oviedo Diaz, S. M., Milanez, D. H., Leiva, D. R., de Faria, L. I. L. & Ishikawa, T. T. (2016), 'Technological forecasting of hydrogen storage materials using patent indicators', International Journal of Hydrogen Energy 41(41), 18301–18310.
- Cohen, Wes M., R. R. N. & Walsh, J. P. (2002), 'Links and impacts: the influence of public research on industrial rd', Management Science **38**, 1–23.
- Conti, A. & Gaule, P. (2011), 'Is the us outperforming europe in university technology licensing? a new perspective on the european paradox', Research Policy **40**(1), 123–135.
- Criscuolo, P. (2006), 'The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO', Scientometrics **66**(1), 23–41.
- Davies, B. & Maré, D. C. (2021), 'Relatedness, complexity and local growth', <u>Regional Studies</u> 55(3), 479–494.
- Davis, D. R. & Dingel, J. I. (2019), 'A spatial knowledge economy', <u>American Economic Review</u> 109(1), 153–170.
- de la Roca, Jorge, P. D. (2017), 'Learning by working in big cities', <u>The Review of Economic</u> Studies **84**(1), 106–142.
- Dernis, H. & Khan, M. (2004), 'Triadic Patent Families Methodology', <u>OECD Science</u>, Technology and Industry Working Papers **2004**/**02**.
- Duranton, G. (2007), 'Urban evolutions: The fast, the slow, and the still', <u>American Economic</u> Review **97**(1), 197–221.
- Duranton, G. & Puga, D. (2014), The growth of cities, in 'Handbook of Economic Growth', Vol. 2B, Elsevier, pp. 781–853.
- Ellison, G., Glaeser, E. L. & Kerr, W. R. (2010), 'What causes industry agglomeration? evidence from coagglomeration patterns', American Economic Review **100**(3), 1195–1213.
- Elsevier (2023), 'The scopus content coverage guide: A complete overview of the content coverage in scopus and corresponding policies'.
- Ernst, H. (1997), 'The Use of Patent Data for Technological Forecasting: The Diffusion of CNC-Technology in the Machine Tool Industry', Small Business Economics **9**(4), 361–381.
- Fallick, B., Fleischman, C. & Rebitzer, J. (2006), 'Job-hopping in silicon valley: Some evidence concerning the microfoundations of a high-technology cluster', <u>The Review of Economics and</u> Statistics 88(3), 472–481.
- Fisch, C. O., Hassel, T. M., Sandner, P. G. & Block, J. H. (2015), 'University patenting: a comparison of 300 leading universities worldwide', <u>The Journal of Technology Transfer</u> 40, 318– 345.
- Gazni, A. (2020), 'The growing number of patent citations to scientific papers: Changes in the world, nations, and fields', Technology in Society **62**(November 2019), 101276.

- Gersbach, H., Schetter, U. & Schmassmann, S. (2023), 'From local to global: A theory of public basic research in a globalized world', European Economic Review **160**, 104530.
- Glänzel, W. & Zhou, P. (2011), 'Publication activity, citation impact and bi-directional links between publications and patents in biotechnology', Scientometrics 86(2), 505–525.
- Godin, B. (2006), 'The Linear Model of Innovation', <u>Science</u>, <u>Technology</u>, <u>Human Values</u> **31**(6), 639–667.
- Guan, J. & He, Y. (2007), 'Patent-bibliometric analysis on the Chinese science Technology linkages', Scientometrics 72(3), 403–425.
- Head, K., Amber Li, Y. & Minondo, A. (2019), 'Geography, ties, and knowledge flows: Evidence from citations in mathematics', The Review of Economics and Statistics **101**(4), 713–727.
- Heimeriks, G., Li, D., Lamers, W., Meijer, I. & Yegros, A. (2019), 'Scientific knowledge production in european regions: patterns of growth, diversity and complexity', <u>European Planning</u> <u>Studies</u> 27(11), 2123–2143.
- Jaffe, A. (1989), 'Real effects of academic research', <u>The American Economic Review</u> **79**(5), 957–970.
- Jaffe, A., Trajtenberg, M. & Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', <u>The Quarterly Journal of Economics</u> **108**(3), 577– 598.
- Jensen, R. (2003), 'Innovative leadership: First-mover advantages in new product adoption', Economic Theory **21**(1), 97 – 116.
- Kerr, W. R. (2010), 'Breakthrough inventions and migrating clusters of innovation', <u>Journal of</u> Urban Economics **67**(1), 46–60.
- Kolympiris, C., Kalaitzandonakes, N. & Miller, D. (2011), 'Spatial collocation and venture capital in the US biotechnology industry', Research Policy 40(9), 1188–1199.
- Krugman, P. (1991), 'Increasing returns and economic geography', <u>Journal of political economy</u> **99**(3), 483–499.
- Lieberman, M. B. & Montgomery, D. B. (1988), 'First-Mover Advantages', <u>Strategic Management</u> Journal 9, 41–58.
- Mewes, L. & Broekel, T. (2022), 'Technological complexity and economic growth of regions', Research Policy 51(8), 104156.
- Moretti, E. (2004), 'Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data', Journal of Econometrics **121**(1-2), 175–212.
- Moretti, E. (2021), 'The effect of high-tech clusters on the productivity of top inventors', American Economic Review **111**(10), 3328–3375.
- Narin, F., Hamilton, K. S. & Olivastro, D. (1995), 'Linkage between agency-supported research and patented industrial technology', Research Evaluation 5(3), 183–187.
- OECD (2022), 'Regpat database'. https://stats.oecd.org/wbos/fileview2.aspx?IDFile=cebce94d-9474-4ffc-b72a-d731fbdb75b9.

- Peri, G. (2005), 'Determinants of Knowledge Flows and Their Effect on Innovation', <u>The Review</u> of Economics and Statistics 87(2), 308–322.
- Poege, F., Harhoff, D., Gaessler, F. & Baruffaldi, S. (2019), 'Science quality and the value of inventions', Science Advances 5(12), 1–8.
- Porter, M. E. (1990), 'The competitive advantage of nations', Harv. Bus. Rev. 68(2), 73–93.
- Porter, M. E. (1998), 'Clusters and the new economics of competition', <u>Harvard Business Review</u> November-December, 77–90.
- Rodríguez-Pose, A. & Wilkie, C. (2019), 'Innovating in less developed regions: What drives patenting in the lagging regions of europe and north america', <u>Growth and Change</u> **50**(1), 4–37.
- Romer, P. (1990), 'Endogenous technological change', <u>Journal of Political Economy</u> **98**(5), 71–102.
- Rosenthal, S. S. & Strange, W. C. (2008), 'The attenuation of human capital spillovers', <u>Journal</u> of Urban Economics **64**(2), 373–389.
- Sabatier, M. & Chollet, B. (2017), 'Is there a first mover advantage in science? Pioneering behavior and scientific production in nanotechnology', Research Policy 46(2), 522–533.
- Saxenian, A. (1991), 'The origins and dynamics of production networks in silicon valley', <u>Research</u> Policy **1991**(20), 423–437.
- Schmoch, U. (2008), Concept of a technology classification for country comparisons, Technical report, WIPO.
- Singh, J. & Marx, M. (2013), 'Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity', Management Science 59(9), 2056–2078.
- Singh, V. K., Singh, P., Karmakar, M., Leta, J. & Mayr, P. (2021), 'The journal coverage of web of science, scopus and dimensions: A comparative analysis', Scientometrics 126, 5113–5142.
- Strange, W., Hejazi, W. & Tang, J. (2006), 'The uncertain city: Competitive instability, skills, innovation and the strategy of agglomeration', Journal of Urban Economics **59**(3), 331–351.
- Sturgeon, T. (2002), 'Modular production networks: a new american model of industrial organization', Industrial and Corporate Change 11(3), 451–496.
- Sung, H. Y., Wang, C. C., Huang, M. H. & Chen, D. Z. (2015), 'Measuring science-based science linkage and non-science-based linkage of patents through non-patent references', <u>Journal of</u> Informetrics 9(3), 488–498.
- Usai, S. (2011), 'The geography of inventive activity in oecd regions', <u>Regional Studies</u> **45**(6), 711–731.
- USPTO (2021), Performance and accountability report, Technical report, USPTO.
- Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E. & Deleus, F. (2002), 'Linking science to technology: Using bibliographic references in patents to build linkage schemes', Scientometrics 54(3), 399–420.
- Wang, L. & Li, Z. (2021), 'Knowledge flows from public science to industrial technologies', <u>The</u> Journal of Technology Transfer 46(4), 1232–1255.

- Wang, Z. (2016), 'Wage growth, ability sorting, and location choice at labor-force entry: New evidence from u.s. census data', Journal of Urban Economics **96**, 112–120.
- Wheeler, C. (2001), 'Search, sorting, and urban agglomeration', <u>Journal of Labor Economics</u> **19**(4), 879–899.

# 7 Appendix

Continent	Type	Total Pioneers	Pioneers	Share $(\%)$
	01	per Region-Tech	per Type	( )
Asia/Oceania	major	112	41	36.61
Asia/Oceania	minor	112	71	63.39
Europe	major	178	43	24.16
Europe	minor	178	135	75.84
North America	$\operatorname{major}$	370	137	37.03
North America	minor	370	233	62.97

 Table 2: Pioneer Region Type Distribution by Continent

Notes: The table presents pioneer regions across continents by type, showing the total number of pioneers per region-technology combination, the number of pioneers per type, and their respective shares.

Coefficient	Estimate	SE	Lower CI	Upper CI
Pioneer	-0.050	0.021	-0.092	-0.008
Pioneer + Pioneer* $(t - t0) = -9$	-0.016	0.021	-0.058	0.026
Pioneer + Pioneer*(t - t0) = -8	-0.007	0.022	-0.050	0.036
Pioneer + Pioneer*(t - t0) = -7	0.035	0.022	-0.008	0.079
Pioneer + Pioneer*(t - t0) = -6	0.097	0.032	0.034	0.160
Pioneer + Pioneer*(t - t0) = -5	0.169	0.035	0.101	0.238
Pioneer + Pioneer*(t - t0) = -4	0.205	0.038	0.131	0.278
Pioneer + Pioneer $(t - t0) = -3$	0.248	0.036	0.177	0.320
Pioneer + Pioneer $(t - t0) = -2$	0.317	0.042	0.234	0.399
Pioneer + Pioneer*(t - t0) = -1	0.314	0.037	0.242	0.386
Pioneer + Pioneer*(t - t0)=0	0.338	0.042	0.256	0.419
Pioneer + Pioneer*(t - t0)=1	0.349	0.045	0.260	0.438
Pioneer + Pioneer*(t - t0)=2	0.368	0.045	0.280	0.457
Pioneer + Pioneer*(t - t0)=3	0.372	0.050	0.275	0.469
Pioneer + Pioneer*(t - t0)=4	0.387	0.049	0.291	0.483
Pioneer + Pioneer*(t - t0)=5	0.415	0.049	0.319	0.511
Pioneer + Pioneer*(t - t0)=6	0.438	0.050	0.341	0.536
Pioneer + Pioneer*(t - t0) = 7	0.461	0.057	0.349	0.572
Pioneer + Pioneer*(t - t0) = 8	0.486	0.050	0.388	0.584
Pioneer + Pioneer*(t - t0)=9	0.531	0.058	0.418	0.644
Pioneer + Pioneer*(t - t0)=10	0.505	0.058	0.392	0.618
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				1153656

 Table 3: Patenting Advantage of Pioneer Regions

*Notes:* The dependent variable is the log of number of patents. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	Estimate	SE	Lower CI	Upper CI
Major Pioneer	0.068	0.062	-0.054	0.191
Major Pioneer + Major Pioneer*(t - $t0$ )=-9	0.086	0.055	-0.022	0.195
Major Pioneer + Major Pioneer $(t - t0) = -8$	0.095	0.064	-0.030	0.220
Major Pioneer + Major Pioneer $(t - t0) = -7$	0.190	0.073	0.048	0.332
Major Pioneer + Major Pioneer $(t - t0) = -6$	0.275	0.072	0.133	0.417
Major Pioneer + Major Pioneer $(t - t0) = -5$	0.356	0.082	0.196	0.517
Major Pioneer + Major Pioneer $(t - t0) = -4$	0.417	0.086	0.249	0.585
Major Pioneer + Major Pioneer $(t - t0) = -3$	0.511	0.082	0.350	0.671
Major Pioneer + Major Pioneer $(t - t0) = -2$	0.587	0.094	0.403	0.770
Major Pioneer + Major Pioneer $(t - t0) = -1$	0.555	0.087	0.383	0.726
Major Pioneer + Major Pioneer $(t - t0) = 0$	0.585	0.093	0.402	0.768
Major Pioneer + Major Pioneer $(t - t0) = 1$	0.565	0.094	0.381	0.750
Major Pioneer + Major Pioneer $(t - t0) = 2$	0.611	0.087	0.441	0.781
Major Pioneer + Major Pioneer $(t - t0) = 3$	0.598	0.095	0.412	0.784
Major Pioneer + Major Pioneer $(t - t0) = 4$	0.636	0.100	0.439	0.832
Major Pioneer + Major Pioneer $(t - t0) = 5$	0.665	0.097	0.474	0.856
Major Pioneer + Major Pioneer $(t - t0) = 6$	0.654	0.104	0.451	0.858
Major Pioneer + Major Pioneer $(t - t0) = 7$	0.703	0.119	0.469	0.937
Major Pioneer + Major Pioneer $(t - t0) = 8$	0.721	0.098	0.530	0.913
Major Pioneer + Major Pioneer $(t - t0) = 9$	0.783	0.112	0.563	1.003
Major Pioneer + Major Pioneer $(t - t0) = 10$	0.769	0.107	0.560	0.978
Minor Pioneer	-0.069	0.023	-0.113	-0.025
Minor Pioneer + Minor Pioneer $(t - t0) = -9$	-0.027	0.022	-0.070	0.016
Minor Pioneer + Minor Pioneer $(t - t0) = -8$	-0.017	0.027	-0.070	0.036
Minor Pioneer + Minor Pioneer $(t - t0) = -7$	-0.001	0.026	-0.051	0.049
Minor Pioneer + Minor Pioneer $(t - t0) = -6$	0.050	0.033	-0.015	0.115
Minor Pioneer + Minor Pioneer $(t - t0) = -5$	0.119	0.038	0.044	0.193
Minor Pioneer + Minor Pioneer $(t - t0) = -4$	0.143	0.040	0.064	0.222
Minor Pioneer + Minor Pioneer $(t - t0) = -3$	0.162	0.036	0.092	0.232
Minor Pioneer + Minor Pioneer $(t - t0) = -2$	0.227	0.044	0.141	0.312
Minor Pioneer + Minor Pioneer $(t - t0) = -1$	0.240	0.038	0.165	0.315
Minor Pioneer + Minor Pioneer*(t - $t0$ )=0	0.261	0.043	0.177	0.345
Minor Pioneer + Minor Pioneer*(t - $t0$ )=1	0.288	0.047	0.196	0.380
Minor Pioneer + Minor Pioneer*(t - $t0$ )=2	0.295	0.049	0.198	0.391
Minor Pioneer + Minor Pioneer*(t - $t0$ )=3	0.307	0.051	0.206	0.408
Minor Pioneer + Minor Pioneer*(t - $t0$ )=4	0.311	0.046	0.220	0.402
Minor Pioneer + Minor Pioneer*(t - $t0$ )=5	0.339	0.049	0.243	0.435
Minor Pioneer + Minor Pioneer*(t - $t0$ )=6	0.380	0.050	0.283	0.477
Minor Pioneer + Minor Pioneer*(t - $t0$ )=7	0.390	0.055	0.283	0.497
Minor Pioneer + Minor Pioneer*(t - $t0$ )=8	0.418	0.050	0.319	0.517
Minor Pioneer + Minor Pioneer*(t - $t0$ )=9	0.457	0.053	0.353	0.562
$\frac{\text{Minor Pioneer} + \text{Minor Pioneer}*(t - t0) = 10}{10}$	0.426	0.057	0.314	0.538
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				1153656

#### Table 4: Patenting Advantage of Pioneer Regions by Size

*Notes:* The dependant variable is the log of number of patents. Coefficients for each period represent linear combinations of the effect of being a minor or major research pioneer hub (Pioneer minor or Pioneer major) and the effect in each period (Pioneer minor\*(t-t0) or Pioneer major\*(t-t0)) since the scientific maturity, relative to being a non-pioneer region. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	Estimate	SE	Lower CI	Upper CI
First Pioneer	-0.011	0.029	-0.068	0.046
First Pioneer + First Pioneer $(t - t0) = -9$	0.004	0.028	-0.051	0.059
First Pioneer + First Pioneer $(t - t0) = -8$	0.042	0.031	-0.017	0.102
First Pioneer + First Pioneer $(t - t0) = -7$	0.067	0.032	0.004	0.131
First Pioneer + First Pioneer $(t - t0) = -6$	0.145	0.044	0.058	0.232
First Pioneer + First Pioneer $(t - t0) = -5$	0.235	0.047	0.144	0.326
First Pioneer + First Pioneer $(t - t0) = -4$	0.273	0.047	0.180	0.365
First Pioneer + First Pioneer $(t - t0) = -3$	0.336	0.049	0.240	0.433
First Pioneer + First Pioneer $(t - t0) = -2$	0.404	0.056	0.295	0.513
First Pioneer + First Pioneer $(t - t0) = -1$	0.375	0.049	0.280	0.471
First Pioneer + First Pioneer $(t - t0) = 0$	0.402	0.052	0.301	0.504
First Pioneer + First Pioneer $(t - t0) = 1$	0.414	0.054	0.309	0.520
First Pioneer + First Pioneer $(t - t0) = 2$	0.427	0.056	0.317	0.537
First Pioneer + First Pioneer $(t - t0) = 3$	0.436	0.061	0.317	0.554
First Pioneer + First Pioneer $(t - t0) = 4$	0.450	0.061	0.330	0.570
First Pioneer + First Pioneer $(t - t0) = 5$	0.494	0.061	0.374	0.613
First Pioneer + First Pioneer $(t - t0) = 6$	0.510	0.062	0.389	0.631
First Pioneer + First Pioneer $(t - t0) = 7$	0.534	0.068	0.399	0.668
First Pioneer + First Pioneer $(t - t0) = 8$	0.551	0.059	0.435	0.666
First Pioneer + First Pioneer $(t - t0) = 9$	0.611	0.069	0.476	0.745
First Pioneer + First Pioneer $(t - t0) = 10$	0.571	0.068	0.437	0.705
Second Pioneer	-0.098	0.023	-0.143	-0.053
Second Pioneer + Second Pioneer $(t - t0) = -9$	-0.034	0.029	-0.092	0.024
Second Pioneer + Second Pioneer $(t - t0) = -8$	-0.056	0.031	-0.117	0.005
Second Pioneer + Second Pioneer $(t - t0) = -7$	-0.018	0.035	-0.086	0.050
Second Pioneer + Second Pioneer $(t - t0) = -6$	0.005	0.039	-0.072	0.082
Second Pioneer + Second Pioneer $(t - t0) = -5$	0.031	0.045	-0.058	0.119
Second Pioneer + Second Pioneer $(t - t0) = -4$	0.054	0.047	-0.038	0.146
Second Pioneer + Second Pioneer $(t - t0) = -3$	0.075	0.043	-0.010	0.160
Second Pioneer + Second Pioneer $(t - t0) = -2$	0.113	0.056	0.003	0.224
Second Pioneer + Second Pioneer $(t - t0) = -1$	0.173	0.054	0.067	0.280
Second Pioneer + Second Pioneer $(t - t0) = 0$	0.188	0.054	0.082	0.294
Second Pioneer + Second Pioneer $(t - t0) = 1$	0.195	0.054	0.090	0.301
Second Pioneer + Second Pioneer $(t - t0) = 2$	0.226	0.056	0.117	0.335
Second Pioneer + Second Pioneer $(t - t0) = 3$	0.219	0.062	0.098	0.341
Second Pioneer + Second Pioneer $(t - t0) = 4$	0.236	0.059	0.121	0.351
Second Pioneer + Second Pioneer $(t - t0) = 5$	0.242	0.056	0.133	0.351
Second Pioneer + Second Pioneer $(t - t0) = 6$	0.273	0.058	0.158	0.387
Second Pioneer + Second Pioneer $(t - t0) = 7$	0.302	0.065	0.175	0.430
Second Pioneer + Second Pioneer $(t - t0) = 8$	0.338	0.062	0.216	0.459
Second Pioneer + Second Pioneer $(t - t0) = 9$	0.360	0.069	0.225	0.496
Second Pioneer + Second Pioneer*(t - $t0$ )=10	0.357	0.067	0.226	0.487
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				1153656

#### Table 5: Patenting Advantage of Pioneer Regions by Pace

*Notes:* The dependant variable is the log of number of patents. Coefficients for each period represent linear combinations of the effect of being a first or second research pioneer moover (First or Second Pioneer) and the effect in each period (First pioneer\*(t-t0) or Second pioneer\*(t-t0)) since the Scientific maturity, relative to being a non-pioneer region. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	Estimate	SE	Lower CI	Upper CI
Super-cluster	0.152	0.068	0.019	0.286
Super-cluster + Super-cluster* $(t - t0) = -9$	0.202	0.065	0.074	0.330
Super-cluster + Super-cluster* $(t - t0) = -8$	0.226	0.066	0.097	0.354
Super-cluster + Super-cluster* $(t - t0) = -7$	0.317	0.067	0.187	0.448
Super-cluster + Super-cluster* $(t - t0) = -5$	0.415	0.078	0.202	0.508
Super-cluster + Super-cluster* $(t - t0) = -3$	0.549	0.077	0.398	0.099
Super-cluster + Super-cluster* $(t - t0) = -3$	0.650	0.078	0.498	0.801
Super-cluster + Super-cluster* $(t - t0) = -2$	0.779	0.077	0.627	0.931
Super-cluster + Super-cluster $(t - t0) = -1$	0.701	0.075	0.553	0.848
Super-cluster + Super-cluster* $(t - t0)=0$	0.757	0.085	0.590	0.923
Super-cluster + Super-cluster*(t - t0)=1	0.775	0.088	0.602	0.949
Super-cluster + Super-cluster*(t - t0)=2	0.817	0.073	0.673	0.960
Super-cluster + Super-cluster* $(t - t0) = 3$	0.845	0.087	0.674	1.016
Super-cluster + Super-cluster* $(t - t0) = 4$	0.847	0.091	0.669	1.026
Super-cluster + Super-cluster* $(t - t0) = 5$	0.839	0.095	0.653	1.026
Super-cluster + Super-cluster* $(t - t0) = 0$	0.912	0.088	0.740	1.085
Super-cluster + Super-cluster* $(t - t0) = 8$	0.940	0.105	0.741	1 150
Super-cluster + Super-cluster $(t - t0) = 9$	1.038	0.110	0.822	1.253
Super-cluster + Super-cluster* $(t - t0) = 10$	0.991	0.112	0.771	1.210
Science Pioneer	-0.087	0.019	-0.125	-0.049
Science Pioneer + Science Pioneer* $(t - t0) = -9$	-0.059	0.019	-0.095	-0.023
Science Pioneer + Science Pioneer $(t - t0) = -8$	-0.056	0.018	-0.092	-0.020
Science Pioneer + Science Pioneer $(t - t0) = -7$	-0.038	0.021	-0.078	0.002
Science Pioneer + Science Pioneer* $(t - t0) = -6$	0.006	0.025	-0.043	0.055
Science Pioneer + Science Pioneer* $(t - t0) = -5$	0.046	0.027	-0.007	0.100
Science Pioneer + Science Pioneer* $(t - t0) = -4$	0.079	0.030	0.019	0.138
Science Pioneer + Science Pioneer* $(t - t0) = -3$	0.121	0.020	0.071	0.220
Science Pioneer + Science Pioneer $(t - t0) = 1$	0.203	0.032	0.140	0.266
Science Pioneer + Science Pioneer $(t - t0) = 0$	0.212	0.037	0.139	0.285
Science Pioneer + Science Pioneer*(t - t0)=1	0.222	0.041	0.142	0.302
Science Pioneer + Science Pioneer $(t - t0) = 2$	0.231	0.040	0.152	0.310
Science Pioneer + Science Pioneer* $(t - t0)=3$	0.224	0.039	0.147	0.301
Science Pioneer + Science Pioneer $(t - t0) = 4$	0.248	0.040	0.169	0.327
Science Pioneer + Science Pioneer* $(t - t0) = 5$	0.299	0.040	0.221	0.377
Science Pioneer + Science Pioneer* $(t - t0) = 0$	0.296	0.043	0.211	0.381
Science Pioneer + Science Pioneer* $(t - t0) = i$	0.310	0.047	0.224	0.407
Science Pioneer + Science Pioneer* $(t - t0) = 8$	0.348	0.044	0.201	0.434
Science Pioneer + Science Pioneer* $(t - t0)=10$	0.370	0.049	0.273	0.466
Innovation Stronghold	-0.178	0.036	-0.249	-0.107
Innovation Stronghold + Innovation Stronghold*(t - t0)=-9	-0.155	0.040	-0.234	-0.076
Innovation Stronghold + Innovation Stronghold*(t - t0)=-8	-0.161	0.039	-0.237	-0.084
Innovation Stronghold + Innovation Stronghold* $(t - t0) = -7$	-0.118	0.042	-0.201	-0.035
Innovation Stronghold + Innovation Stronghold* $(t - t0) = -6$	-0.101	0.045	-0.190	-0.013
Innovation Stronghold + Innovation Stronghold $*(t - t0) = -5$	-0.096	0.048	-0.191	-0.001
Innovation Stronghold + Innovation Stronghold* $(t - t0) = -4$	-0.024	0.050	-0.122	0.073
Innovation Stronghold + Innovation Stronghold* $(t - t0) = -3$	-0.025	0.048	-0.119	0.068
Innovation Stronghold + Innovation Stronghold* $(t - t0) = -2$	0.0002	0.051	-0.100	0.100
Innovation Stronghold + Innovation Stronghold* $(t - t0) = 0$	-0.003	0.030	-0.093	0.100
Innovation Stronghold + Innovation Stronghold* $(t - t0) = 1$	-0.026	0.043	-0.111	0.059
Innovation Stronghold + Innovation Stronghold* $(t - t0)=2$	-0.001	0.046	-0.092	0.089
Innovation Stronghold + Innovation Stronghold* $(t - t0) = 3$	0.005	0.049	-0.090	0.101
Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=4	0.048	0.046	-0.041	0.137
Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=5	0.027	0.049	-0.070	0.123
Innovation Stronghold + Innovation Stronghold $*(t - t0) = 6$	0.056	0.047	-0.036	0.148
Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=7	0.075	0.048	-0.019	0.168
Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=8	0.102	0.051	0.002	0.203
Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=9 Innovation Stronghold + Innovation Stronghold*(t - $t0$ )=10	0.117	0.057	0.005	0.228
$\frac{1}{2} = \frac{1}{2} = \frac{1}$	0.101	0.001	0.000	0.211 V
Region fixed effect				Yes
Tech fixed effect				1 es Ves
Observations				1153656

 Table 6: Patenting Advantage by Region Type

Notes: The dependant variable is the log of number of patents. Coefficients for each period represent linear combinations of the effect of a region being super-cluster, scientific stronghold or innovation leader and the effect in each period (Type\*(t-t0)) since the scientific maturity, relative to being a non-pioneer region. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.



<u>Notes</u>: The map illustrates the cumulative number of Bloom technology patents by region in the USA (a) and Europe (b) from year t=0 (the year of scientific maturity) to year t=10 (the tenth year after reaching scientific maturity). Regions with thick borders highlight areas that are pioneers in research for at least one of the Bloom technologies.



Figure 6: Cumulative Papers and Patents Across Emergent Technologies

Notes: The figure shows cumulative number of papers (blue dots) and patents (red dots) across emergent technologies over time. The black vertical line denotes the year of scientific maturity.



Figure 7: Patenting Advantage in Innovation Quality

<u>Notes</u>: The figure shows the estimated patenting advantage in innovation quality of pioneer regions over time and relative to non-pioneer regions: Panel (a) examines patent citations; Panel (b) examines patent novelty. The plotted estimated effects represent the linear combination of pioneer region and time on these measures: (1) the average treatment effect of a pioneer region type; and (2) the temporal effect (expressed yearly) of a pioneer region type prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year ( $t_0$ ) of scientific maturity.





<u>Notes</u>: The figure shows the estimated patenting advantage for a sample of technologies that exist at least 15 periods (years) post-scientific maturity. The plotted estimated effects represent the linear combination of pioneer region and time on this measure: (1) the average treatment effect of a pioneer region type; and (2) the temporal effect (expressed yearly) of a pioneer region type prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

Figure 9: Patenting Advantage of Pioneer Regions by Technology Type



<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions by emergent technology type over time and relative to non-pioneer regions: Panel (a) examines digital emergent technologies; Panel (b) examines physical emergent technologies. The plotted estimated effects represent the linear combination of pioneer region and time on these sub-samples: (1) the average treatment effect of a pioneer region type; and (2) the temporal effect (expressed yearly) of a pioneer region type prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

Figure 10: Patenting Advantage of Pioneer Regions Using Continuous Measure



Notes: The figure shows the estimated patenting advantage of pioneer regions using a continuous measure for pioneer regions (number of scientific papers) over time and relative to non-pioneer regions. The plotted estimated effects represent the linear combination of this pioneer region measure and time on this measure: (1) the number of papers of region *i* in given Bloom-technology *j* (*Papers*<sub>*i*,*j*</sub>); and (2) the temporal effect (expressed yearly) of this measure prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year ( $t_0$ ) of scientific maturity.

Figure 11: Patenting Advantage of Pioneer Regions By Continent



<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions in different geographic regions over time and relative to non-pioneer regions: Panel (a) examines North America; Panel (b) examines Europe; Panel (c) examines Asia. The plotted estimated effects represent the linear combination of pioneer region and time on these sub-samples: (1) the average treatment effect of a pioneer region; and (2) the temporal effect (expressed yearly) of a pioneer region prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.





<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions and neighbour regions over time and relative to non-pioneer regions: Panels (a) - (c) examine different distance thresholds. The plotted estimated effects represent the linear combination of pioneer region and time on these thresholds: (1) the average treatment effect of a pioneer region; and (2) the temporal effect (expressed yearly) of a pioneer region prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

Figure 13: Patenting Advantage of Pioneer Regions Using Triadic Patents



<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions using triadic patents over time and relative to non-pioneer regions. The plotted estimated effects represent the linear combination of pioneer region and time on this measure: (1) the average treatment effect of a pioneer region; and (2) the temporal effect (expressed yearly) of a pioneer region prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year ( $t_0$ ) of scientific maturity.



Figure 14: Patenting Advantage of Pioneer Regions Using Triadic Patents By Continent

<u>Notes</u>: The figure shows the estimated patenting advantage of pioneer regions using triadic patents in different geographic regions over time and relative to non-pioneer regions. Panel (a) examines North America; Panel (b) examines Europe; Panel (c) examines Asia. The plotted estimated effects represent the linear combination of pioneer region and time on these sub-samples: (1) the average treatment effect of a pioneer region; and (2) the temporal effect (expressed yearly) of a pioneer region prior to and since scientific maturity. Confidence intervals are at 95 percent levels and are calculated using cluster-robust standard errors. The black dashed line denotes the year  $(t_0)$  of scientific maturity.

Coefficient	Top patents	Novel patents
Pioneer	-0.012 (0.01)	-0.11 (0.02)
Pioneer + Pioneer*(t - t0) = -9	0.007(0.013)	-0.061(0.02)
Pioneer + Pioneer*(t - t0) = -8	-0.001 (0.013)	-0.04 (0.02)
Pioneer + Pioneer*(t - t0) = -7	0.002(0.011)	0.042(0.022)
Pioneer + Pioneer*(t - t0) = -6	0.016(0.012)	0.102(0.032)
Pioneer + Pioneer*(t - t0) = -5	0.037(0.014)	0.178(0.035)
Pioneer + Pioneer*(t - t0) = -4	0.052(0.013)	0.213(0.038)
Pioneer + Pioneer*(t - t0) = -3	0.057(0.015)	0.259(0.037)
Pioneer + Pioneer*(t - t0) = -2	0.081(0.018)	0.327(0.043)
Pioneer + Pioneer*(t - t0) = -1	0.094(0.015)	0.325(0.037)
Pioneer + Pioneer*(t - t0)=0	0.103(0.021)	0.345(0.042)
Pioneer + Pioneer*(t - t0) = 1	0.109(0.019)	0.351(0.045)
Pioneer + Pioneer*(t - t0) = 2	0.103(0.02)	0.376(0.045)
Pioneer + Pioneer*(t - t0) = 3	0.113(0.023)	0.381(0.049)
Pioneer + Pioneer*(t - t0) = 4	0.122(0.022)	0.394(0.049)
Pioneer + Pioneer*(t - t0) = 5	0.128(0.024)	0.42(0.048)
Pioneer + Pioneer*(t - t0) = 6	0.137(0.022)	0.444(0.049)
Pioneer + Pioneer*(t - t0) = 7	0.153(0.03)	0.46(0.056)
Pioneer + Pioneer*(t - t0) = 8	0.155(0.023)	0.474(0.049)
Pioneer + Pioneer*(t - t0)=9	0.166(0.028)	0.507(0.056)
Pioneer + Pioneer*(t - t0) = 10	0.181(0.028)	0.445(0.056)
Region fixed effect	Yes	Yes
Year fixed effect	Yes	Yes
Tech fixed effect	Yes	Yes
Observations	1153656	1153656

 Table 7: Patenting Advantage of Pioneer Regions in Innovation Quality

=

Notes: The dependent variable is either the log (+1) of the number of top patents or the log of the number of patents (+1) introducing new semantic terms. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. Standard errors clustered at the regional level are shown in brackets.

Coefficient	Estimate	SE	Lower CI	Upper CI
Pioneer	-0.078	0.023	-0.124	-0.032
Pioneer + Pioneer*(t - t0) = -9	-0.045	0.024	-0.092	0.002
Pioneer + Pioneer $(t - t0) = -8$	-0.020	0.024	-0.067	0.027
Pioneer + Pioneer*(t - t0) = -7	0.016	0.026	-0.034	0.065
Pioneer + Pioneer $(t - t0) = -6$	0.078	0.033	0.013	0.142
Pioneer + Pioneer $(t - t0) = -5$	0.134	0.036	0.063	0.204
Pioneer + Pioneer*(t - t0) = -4	0.151	0.037	0.078	0.223
Pioneer + Pioneer*(t - t0) = -3	0.180	0.035	0.112	0.248
Pioneer + Pioneer*(t - t0) = -2	0.235	0.040	0.156	0.314
Pioneer + Pioneer*(t - t0) = -1	0.188	0.038	0.114	0.262
Pioneer + Pioneer*(t - t0)=0	0.228	0.038	0.153	0.303
Pioneer + Pioneer*(t - t0) = 1	0.244	0.041	0.164	0.324
Pioneer + Pioneer*(t - t0)=2	0.255	0.041	0.174	0.335
Pioneer + Pioneer*(t - t0)=3	0.235	0.045	0.147	0.322
Pioneer + Pioneer*(t - t0)=4	0.267	0.046	0.177	0.358
Pioneer + Pioneer*(t - t0)=5	0.274	0.045	0.187	0.362
Pioneer + Pioneer*(t - t0)=6	0.318	0.046	0.227	0.409
Pioneer + Pioneer*(t - t0) = 7	0.332	0.049	0.236	0.429
Pioneer + Pioneer*(t - t0) = 8	0.358	0.045	0.269	0.446
Pioneer + Pioneer*(t - t0)=9	0.412	0.052	0.310	0.515
Pioneer + Pioneer*(t - t0)=10	0.406	0.052	0.304	0.508
Pioneer + Pioneer*(t - t0)=11	0.467	0.057	0.356	0.579
Pioneer + Pioneer*(t - t0) = 12	0.524	0.057	0.412	0.636
Pioneer + Pioneer*(t - t0) = 13	0.539	0.064	0.414	0.664
Pioneer + Pioneer*(t - t0) = 14	0.555	0.064	0.429	0.681
Pioneer + Pioneer*(t - $t0$ )=15	0.525	0.061	0.405	0.646
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				952224

 Table 8: Patenting Advantage of Pioneer Regions in Longer Time Windows

*Notes:* The dependent variable is the log of number of patents. This sample includes only those technologies that exist at least 15 years since the scientific maturity. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	Digital tech	Physical tech
Pioneer	-0.124(0.034)	$0.096\ (0.036)$
Pioneer + Pioneer*(t - t0) = -9	-0.087 (0.032)	0.127(0.038)
Pioneer + Pioneer*(t - t0) = -8	-0.079(0.031)	0.135(0.04)
Pioneer + Pioneer*(t - t0) = -7	-0.038 (0.03)	0.18(0.045)
Pioneer + Pioneer $(t - t0) = -6$	0.03(0.04)	0.232(0.045)
Pioneer + Pioneer $(t - t0) = -5$	0.132(0.042)	0.251(0.049)
Pioneer + Pioneer $(t - t0) = -4$	0.19(0.048)	0.247(0.054)
Pioneer + Pioneer*(t - t0) = -3	0.228(0.048)	0.299(0.048)
Pioneer + Pioneer*(t - t0) = -2	0.293(0.055)	0.369(0.056)
Pioneer + Pioneer*(t - t0) = -1	0.319(0.051)	0.316(0.054)
Pioneer + Pioneer*(t - t0)=0	0.305(0.056)	0.401(0.06)
Pioneer + Pioneer*(t - t0) = 1	0.338(0.058)	0.375(0.062)
Pioneer + Pioneer (t - t0) = 2	0.382(0.062)	0.353(0.06)
Pioneer + Pioneer*(t - t0) = 3	0.389(0.066)	0.35(0.062)
Pioneer + Pioneer*(t - t0) = 4	0.375(0.065)	0.412(0.062)
Pioneer + Pioneer*(t - t0) = 5	0.439(0.064)	0.379(0.057)
Pioneer + Pioneer (t - t0) = 6	0.433(0.066)	0.448(0.064)
Pioneer + Pioneer*(t - t0) = 7	0.441(0.074)	0.493(0.071)
Pioneer + Pioneer (t - t0) = 8	0.437(0.067)	0.566(0.073)
Pioneer + Pioneer*(t - t0) = 9	0.488(0.075)	0.602(0.075)
Pioneer + Pioneer*(t - t0) = 10	0.482(0.075)	0.539(0.081)
Region fixed effect	Yes	Yes
Year fixed effect	Yes	Yes
Tech fixed effect	Yes	Yes
Observations	672966	480690

Table 9: Patenting Advantage of Pioneer Regions by Technology Type

*Notes:* The dependent variable is the log of the number of patents. The table presents the estimated patenting advantage for two samples: digital and physical technologies. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. Standard errors clustered at the regional level are shown in brackets.

Coefficient	Estimate	SE	Lower CI	Upper CI
Papers	0.005	0.002	0.0002	0.009
Papers + Papers*(t - t0) = -9	0.008	0.002	0.003	0.013
Papers + Papers*(t - t0) = -8	0.012	0.003	0.007	0.018
Papers + Papers*(t - t0) = -7	0.017	0.003	0.011	0.023
Papers + Papers*(t - t0) = -6	0.025	0.004	0.018	0.032
Papers + Papers*(t - t0) = -5	0.034	0.004	0.026	0.043
Papers + Papers*(t - t0) = -4	0.043	0.005	0.034	0.053
Papers + Papers*(t - t0) = -3	0.048	0.005	0.038	0.058
Papers + Papers*(t - t0) = -2	0.056	0.006	0.044	0.067
Papers + Papers*(t - t0) = -1	0.059	0.005	0.048	0.069
Papers + Papers*(t - t0)=0	0.065	0.006	0.054	0.076
Papers + Papers*(t - t0) = 1	0.065	0.006	0.053	0.076
Papers + Papers*(t - t0)=2	0.071	0.006	0.059	0.082
Papers + Papers*(t - t0)=3	0.075	0.006	0.063	0.087
Papers + Papers*(t - t0)=4	0.083	0.006	0.070	0.095
Papers + Papers*(t - t0) = 5	0.083	0.006	0.070	0.095
Papers + Papers*(t - t0)=6	0.087	0.007	0.074	0.100
Papers + Papers*(t - t0) = 7	0.092	0.007	0.079	0.105
Papers + Papers*(t - t0) = 8	0.096	0.007	0.084	0.109
Papers + Papers*(t - t0)=9	0.101	0.007	0.087	0.115
Papers + Papers*(t - t0) = 10	0.106	0.007	0.093	0.120
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				1153656

Table 10: Patenting Advantage of Pioneer Regions Using Continuous Measure

*Notes:* The dependant variable is the number of patents in logs. Coefficients for each period represent linear combinations of the average effect of the log number of papers (Papers) and the effect of having a certain log number of papers in each period (Papers\*(t-t0)) since the emergence year. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	North America	Europe	Asia
Pioneer	-0.025(0.039)	-0.024(0.013)	0.032(0.047)
Pioneer + Pioneer*(t - t0) = -9	0.019(0.035)	0.006(0.032)	0.032(0.052)
Pioneer + Pioneer*(t - t0) = -8	0.034(0.037)	-0.011 (0.017)	0.075(0.059)
Pioneer + Pioneer*(t - t0) = -7	0.096(0.041)	0.009(0.02)	0.063(0.051)
Pioneer + Pioneer*(t - t0) = -6	0.179(0.052)	$0.011 \ (0.029)$	0.133(0.067)
Pioneer + Pioneer*(t - t0) = -5	0.294(0.052)	0.001(0.022)	0.194(0.081)
Pioneer + Pioneer*(t - t0) = -4	$0.325 \ (0.056)$	0.002(0.029)	0.272(0.087)
Pioneer + Pioneer*(t - t0) = -3	0.387 (0.054)	0.022(0.026)	0.289(0.078)
Pioneer + Pioneer*(t - t0) = -2	0.473(0.06)	0.028(0.024)	0.374(0.082)
Pioneer + Pioneer*(t - t0) = -1	0.469(0.051)	0.087(0.031)	0.259(0.082)
Pioneer + Pioneer*(t - t0) = 0	0.511(0.061)	0.085(0.036)	0.275(0.071)
Pioneer + Pioneer*(t - t0) = 1	0.533(0.067)	0.062(0.033)	0.273(0.073)
Pioneer + Pioneer*(t - t0) = 2	0.548(0.064)	0.086(0.042)	0.305(0.085)
Pioneer + Pioneer*(t - t0) = 3	0.575(0.071)	0.071(0.038)	0.239(0.084)
Pioneer + Pioneer*(t - t0) = 4	0.602(0.07)	$0.056\ (0.034)$	0.252(0.088)
Pioneer + Pioneer*(t - t0) = 5	0.623(0.071)	0.116(0.038)	$0.255\ (0.076)$
Pioneer + Pioneer*(t - t0) = 6	$0.651 \ (0.069)$	0.105(0.037)	0.302(0.083)
Pioneer + Pioneer*(t - t0) = 7	0.682(0.083)	0.152(0.045)	0.221(0.078)
Pioneer + Pioneer*(t - t0) = 8	0.687(0.07)	0.184(0.049)	0.3 (0.096)
Pioneer + Pioneer*(t - t0)=9	0.76(0.082)	$0.185\ (0.053)$	$0.294\ (0.097)$
Pioneer + Pioneer*(t - t0)=10	0.724(0.08)	$0.15 \ (0.058)$	0.298(0.107)
Region fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Tech fixed effect	Yes	Yes	Yes
Observations	527184	559440	35280

 Table 11: Patenting Advantage of Pioneer Regions by Continent

*Notes:* The dependent variable is the log of the number of patents. The table presents the estimated patenting advantage for three samples: North America, Europe and Asia. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. Standard errors clustered at the regional level are shown in brackets.

Coefficient	$50 \mathrm{km}$	$100 \mathrm{km}$	200 km
Neighbour	-0.017 (0.006)	-0.008(0.003)	0(0.002)
Neighbour + Neighbour*(t - $t0$ )=-9	-0.005 (0.005)	-0.004 (0.003)	0.001 ( $0.002$ )
Neighbour + Neighbour $(t - t0) = -8$	-0.01 (0.006)	-0.005(0.003)	0.001(0.002)
Neighbour + Neighbour $(t - t0) = -7$	0.001(0.007)	0.001(0.003)	0.004(0.002)
Neighbour + Neighbour $(t - t0) = -6$	0.004(0.006)	0.002(0.003)	0.004(0.002)
Neighbour + Neighbour $(t - t0) = -5$	0.007 (0.006)	0.004(0.003)	0.004(0.002)
Neighbour + Neighbour $(t - t0) = -4$	0.018(0.007)	0.013(0.003)	0.01(0.002)
Neighbour + Neighbour $(t - t0) = -3$	0.024(0.008)	0.016(0.004)	$0.011 \ (0.002)$
Neighbour + Neighbour $(t - t0) = -2$	0.02(0.007)	0.013(0.003)	$0.01 \ (0.002)$
Neighbour + Neighbour $(t - t0) = -1$	0.028(0.008)	0.02(0.004)	0.013(0.002)
Neighbour + Neighbour* $(t - t0)=0$	0.019(0.007)	0.016(0.004)	0.011(0.002)
Neighbour + Neighbour $(t - t0) = 1$	0.023(0.007)	0.016(0.004)	0.015(0.002)
Neighbour + Neighbour $(t - t0) = 2$	0.027(0.007)	0.023(0.004)	0.016(0.002)
Neighbour + Neighbour* $(t - t0) = 3$	0.021(0.008)	0.021(0.005)	0.014(0.003)
Neighbour + Neighbour $(t - t0) = 4$	0.038(0.009)	0.028(0.005)	0.018(0.003)
Neighbour + Neighbour $(t - t0) = 5$	0.025(0.008)	0.023(0.005)	0.016(0.003)
Neighbour + Neighbour $(t - t0) = 6$	0.032(0.009)	0.033(0.005)	0.019(0.003)
Neighbour + Neighbour $(t - t0) = 7$	0.031(0.01)	0.028(0.006)	0.017(0.003)
Neighbour + Neighbour $(t - t0) = 8$	0.04(0.01)	0.03(0.006)	0.02(0.003)
Neighbour + Neighbour $(t - t0) = 9$	0.039(0.012)	0.035(0.006)	0.023(0.004)
Neighbour + Neighbour $(t - t0) = 10$	0.05(0.012)	0.038(0.006)	0.024(0.004)
Pioneer	-0.051(0.021)	-0.051 (0.021)	-0.049 (0.021)
Pioneer + Pioneer*(t - t0) = -9	-0.016 (0.021)	-0.016 (0.021)	-0.015 (0.021)
Pioneer + Pioneer $(t - t0) = -8$	-0.008 (0.022)	-0.007 (0.022)	-0.006 (0.022)
Pioneer + Pioneer*(t - t0) = -7	0.035(0.022)	0.036(0.022)	0.037(0.022)
Pioneer + Pioneer*(t - t0) = -6	0.097(0.032)	0.098(0.032)	0.099(0.032)
Pioneer + Pioneer*(t - t0) = -5	0.17(0.035)	0.17(0.035)	0.172(0.035)
Pioneer + Pioneer*(t - t0) = -4	0.206(0.038)	0.207(0.038)	0.209(0.038)
Pioneer + Pioneer*(t - t0) = -3	0.25(0.036)	0.251(0.036)	0.253(0.037)
Pioneer + Pioneer $(t - t0) = -2$	0.318(0.042)	0.319(0.042)	0.321(0.042)
Pioneer + Pioneer $(t - t0) = -1$	0.316(0.037)	0.317(0.037)	0.319(0.037)
Pioneer + Pioneer $(t - t0) = 0$	0.339(0.042)	0.34(0.042)	0.342(0.042)
Pioneer + Pioneer $(t - t0) = 1$	0.351(0.045)	0.352(0.045)	0.354(0.045)
Pioneer + Pioneer* $(t - t0) = 2$	0.37(0.045)	0.372(0.045)	0.374(0.045)
Pioneer + Pioneer $(t - t0) = 3$	0.374(0.05)	0.376(0.05)	0.377(0.05)
Pioneer + Pioneer* $(t - t0) = 4$	0.389(0.049)	0.391(0.049)	0.393(0.049)
Pioneer + Pioneer* $(t - t0) = 5$	0.417(0.049)	0.419(0.049)	0.421(0.049)
Pioneer + Pioneer* $(t - t0) = 6$	0.441(0.05)	0.443(0.05)	0.445(0.05)
Pioneer + Pioneer* $(t - t0) = 7$	0.463 (0.057)	$0.465\ (0.057)$	0.467 (0.057)
Pioneer + Pioneer* $(t - t0) = 8$	0.489(0.05)	0.491 (0.05)	0.493(0.05)
Pioneer + Pioneer* $(t - t0) = 9$	0.534(0.058)	0.537 (0.058)	0.539(0.058)
Pioneer + Pioneer* $(t - t0) = 10$	0.508 (0.058)	$0.511 \ (0.058)$	0.513(0.058)
Region fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Tech fixed effect	Yes	Yes	Yes
Observations	1153656	1153656	1153656

Table 12: Patenting Advantage of Pioneer and Neighbour Regions by Distance

*Notes:* The dependent variable is the log of the number of patents. The table presents the estimated patenting advantage for three samples where neighbours are either 50km, 100km or 200km away from pioneers. Coefficients for each period represent linear combinations of the effect of being a research pioneer (Pioneer) or a neighbour region (Neighbour) and the effect in each period (Pioneer\*(t-t0) or Neighbour\*(t-t0)) since the scientific maturity year. Standard errors clustered at the regional level are shown in brackets.

Coefficient	Estimate	SE	Lower CI	Upper CI
Pioneer	-0.051	0.015	-0.079	-0.023
Pioneer + Pioneer* $(t - t0) = -9$	-0.034	0.015	-0.064	-0.004
Pioneer + Pioneer*(t - t0) = -8	-0.024	0.010	-0.044	-0.003
Pioneer + Pioneer*(t - t0) = -7	-0.017	0.013	-0.042	0.008
Pioneer + Pioneer*(t - t0) = -6	0.008	0.013	-0.017	0.033
Pioneer + Pioneer*(t - $t0$ )=-5	0.048	0.018	0.012	0.084
Pioneer + Pioneer $(t - t0) = -4$	0.051	0.017	0.018	0.085
Pioneer + Pioneer*(t - $t0$ )=-3	0.086	0.019	0.048	0.125
Pioneer + Pioneer*(t - t0) = -2	0.130	0.021	0.090	0.171
Pioneer + Pioneer*(t - t0) = -1	0.129	0.020	0.090	0.169
Pioneer + Pioneer*(t - t0)=0	0.125	0.022	0.083	0.168
Pioneer + Pioneer*(t - t0)=1	0.147	0.025	0.099	0.195
Pioneer + Pioneer*(t - t0)=2	0.161	0.025	0.111	0.210
Pioneer + Pioneer*(t - t0)=3	0.167	0.029	0.110	0.223
Pioneer + Pioneer*(t - t0) = 4	0.176	0.030	0.117	0.235
Pioneer + Pioneer*(t - t0)=5	0.194	0.030	0.135	0.253
Pioneer + Pioneer*(t - t0)=6	0.179	0.027	0.127	0.231
Pioneer + Pioneer*(t - t0) = 7	0.208	0.031	0.147	0.269
Pioneer + Pioneer*(t - t0) = 8	0.217	0.029	0.161	0.273
Pioneer + Pioneer*(t - t0)=9	0.241	0.033	0.176	0.305
Pioneer + Pioneer*(t - t0)=10	0.227	0.032	0.164	0.290
Region fixed effect				Yes
Year fixed effect				Yes
Tech fixed effect				Yes
Observations				764568

 Table 13: Patenting Advantage of Pioneer Regions Using Triadic Patents

*Notes:* The dependent variable is the log of number of (triadic only) patents. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. The lower and upper bounds represent the 95% confidence intervals, calculated using standard errors clustered at the regional level.

Coefficient	North America	Europe	Asia
Pioneer	-0.031(0.02)	-0.027(0.01)	0.004(0.031)
Pioneer + Pioneer* $(t - t0) = -9$	-0.005 (0.017)	0 (0.023)	-0.051 (0.048)
Pioneer + Pioneer $(t - t0) = -8$	-0.001 (0.016)	-0.019 (0.012)	0.016(0.03)
Pioneer + Pioneer*(t - t0) = -7	-0.011 (0.018)	0 (0.015)	0.038(0.032)
Pioneer + Pioneer $(t - t0) = -6$	0.011(0.017)	0.022(0.025)	0.064(0.036)
Pioneer + Pioneer $(t - t0) = -5$	0.079(0.027)	-0.019 (0.016)	0.11(0.049)
Pioneer + Pioneer $(t - t0) = -4$	0.086(0.027)	-0.004 (0.021)	0.085(0.044)
Pioneer + Pioneer $(t - t0) = -3$	0.132(0.031)	0.002(0.013)	0.108(0.048)
Pioneer + Pioneer $(t - t0) = -2$	0.187(0.03)	0.007(0.017)	0.17(0.044)
Pioneer + Pioneer*(t - t0) = -1	0.161(0.03)	0.048(0.022)	0.18(0.042)
Pioneer + Pioneer*(t - t0) = 0	0.164(0.035)	0.041(0.021)	0.149(0.039)
Pioneer + Pioneer*(t - t0) = 1	0.187(0.038)	0.04(0.025)	0.192(0.038)
Pioneer + Pioneer*(t - t0) = 2	0.197(0.035)	0.057(0.029)	0.224(0.072)
Pioneer + Pioneer*(t - t0) = 3	0.238(0.042)	0.043(0.026)	0.149(0.056)
Pioneer + Pioneer*(t - t0) = 4	0.252(0.044)	0.035(0.023)	0.151(0.05)
Pioneer + Pioneer*(t - t0) = 5	0.271(0.043)	$0.056\ (0.024)$	0.182(0.072)
Pioneer + Pioneer*(t - t0) = 6	0.22(0.041)	0.093(0.029)	0.231(0.055)
Pioneer + Pioneer*(t - t0) = 7	0.261(0.048)	0.12(0.034)	0.19(0.055)
Pioneer + Pioneer*(t - t0) = 8	0.244(0.041)	0.147(0.041)	0.268(0.073)
Pioneer + Pioneer*(t - t0)=9	0.323(0.045)	$0.095 \ (0.026)$	0.222(0.081)
Pioneer + Pioneer*(t - t0) = 10	0.272(0.043)	$0.117 \ (0.036)$	$0.286\ (0.095)$
Region fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Tech fixed effect	Yes	Yes	Yes
Observations	231336	475776	35280

Table 14: Patenting Advantage of Pioneer Regions Using Triadic Patents by Continent

*Notes:* The dependent variable is the log of the number of (triadic only) patents. The table presents the estimated patenting advantage for three samples: North America, Europe and Asia. Coefficients for each period represent linear combinations of the effect of being a research pioneer hub (Pioneer) and the effect in each period (Pioneer\*(t-t0)) since the scientific maturity year. Standard errors clustered at the regional level are shown in brackets.

Country	Regions before subset	Regions of final dataset	Aggregation level
Australia	49	49	OECD TL3
Austria	35	34	OECD TL3
Belgium	44	44	OECD TL3
Canada	284	185	OECD TL3
Croatia	21	1	OECD TL3
Czechia	14	14	OECD TL3
Denmark	11	11	OECD TL3
Estonia	5	2	OECD TL3
Finland	19	19	OECD TL3
France	101	98	OECD TL3
Germany	388	384	OECD TL3
Greece	46	10	OECD TL3
Hungary	20	17	OECD TL3
Iceland	8	2	OECD TL3
Ireland	8	8	OECD TL3
Israel	6	6	OECD TL3
Italy	110	99	OECD TL3
Japan	47	47	OECD TL3
Latvia	6	2	OECD TL3
Lithuania	10	2	OECD TL3
Luxembourg	1	1	OECD TL3
Netherlands	40	40	OECD TL3
New Zealand	14	11	OECD TL3
Norway	18	17	OECD TL3
Poland	73	24	OECD TL3
Portugal	24	7	OECD TL3
Slovakia	8	7	OECD TL3
Slovenia	12	7	OECD TL3
South Korea	17	17	OECD TL3
Spain	56	40	OECD TL3
Sweden	21	21	OECD TL3
Switzerland	26	26	OECD TL3
United Kingdom	177	173	OECD TL3
United States	932	861	Core-based Statistical Areas

 Table 15: Number of Regions by Country

*Notes:* The table shows the number of regions to which patents are assigned in our raw data, as well as after filtering to include only regions with more than 50 patents. The 'Aggregation Level' column indicates the level of regional aggregation for each country, using the gadm database to assign patents and papers based on geocodes.

Emergent Technology	Key Words
3D Printing	3D Print, Additive Manufacturing
Autonomous Cars	Automated Driving, Autonomous Car, Autonomous Driving Autonomous Vehicle, Autonomous Vehicle, Driverless Car Driverless Truck, Robot Car, Self-Driving Car
Bispecific Monoclonal Antibody	Bispecific
Cloud Computing	Cloud Applic, Cloud Computing, Cloud Deployment, Cloud Deployment, Cloud Environment, Cloud Hosting, Cloud Infrastructure, Cloud Management, Cloud Offering, Cloud Platform, Cloud Provider, Cloud Provider, Cloud Security, Cloud Service, Cloud Solution, Cloud Storage, Community Cloud, Distributed Cloud, Enterprise Application, Enterprise Class, Enterprise Cloud, Enterprise Network, Hybrid Cloud, IAAS, PAAS, Personal Cloud, Private Cloud, Public Cloud, SAAS
Computer Vision	Computer Vision, Facial Recognition, Gesture Recognition, Image Recognition Motion Estimation, Object Recognition, Pose Estimation, Sensor Fusion, Visual Servoing
Drug Conjugates	Antibody Drug, Drug Conjugate, Kinase Inhibitor
Electronic Gaming	Electronic Gaming, Game Content, Gaming Products, Social Game, Video Gaming
Fingerprint Sensor	Fingerprint Scanner, Fingerprint Sensor
Fracking	Fraccing, Fracking, Hydraulic Fracturing, Hydrofracking, Hydrofracturing
GPS	Global Positioning, GPS System, Navigation Device
Hybrid Vehicle Electric Car	Electric Buses, Electric Car, Electric Motorcycle, Electric Vehicle, Hybrid Electric, Hybrid Vehicle, Plugin Hybrid, Vehicle Charging
Lane Departure Warning	Departure Warning, Lane Departure
Lithium Battery	Ion Batter, Lithium Batter, Lithium Ion, Lithium Polymer
Machine Learning AI	AI Technology, Artificial Intelligence, Deep Learning, Language Processing, Learning Algorithm, Machine Intelligence, Machine Learning, Reinforcement Learning, Supervised Learning, Unsupervised Learning
Millimeter Wave	Millimeter Wave
Mobile Payment	Mobile Commerce, Mobile Money, Mobile Payment, Mobile Wallet
OLED Display	OLED

# Table 16: Emergent Technologies and Associated Key Words

Online Streaming	Interactive TV, Internet Radio, Live Stream, Live Video, Mobile Video, Music Streaming, Online Streaming, Streaming Content, Streaming Media, Streaming Servic, Streaming Video, Video Conferenc, Video Ondemand, Video Streaming
RFID Tags	Keyless Entry, RFID
Search Engine	Search Engine
Smart Devices	Android Device, Android Phone, Android Smartphone, iPhone, Media Device, Mobile Device, Phones Tablet, Portable Media, Smart Band, Smart Car, Smart Car, Smart Device, Smart Phone, Smart Refrigerator, Smart Speaker, Smart Tablet, Smart TV, Smart Watch, Smartphone Tablet, Smartphones Tablet, Tablet Computer, Tablet PCs, Wearable Device
Social Networking	Networking Site, Social Channel, Social Media, Social Network, Social Platform
Software Defined Radio	Defined Radio
Solar Power	Crystalline Silicon, Photovoltaic, Rooftop Solar, Silicon Solar, Solar Application, Solar Cell, Solar Energy, Solar Grade, Solar Module, Solar Module, Solar Panel, Solar Power, Solar PV, Solar Thermal, Solar Wafer
Stent Graft	Stent Graft
Touch Screen	Touch Screen, Touch Sensor, Touchscreen
Virtual Reality	Augmented Reality, Extended Reality, Mixed Reality, Virtual Reality
Wifi	Broadband Connectivity, Wifi, Wireless Network
Wireless Charging	Inductive Charging, Wireless Charging

 $\it Notes:$  This table presents Bloom technology fields along with the key terms used to identify each technology.

Country	No regional match	Total	Share not matched (%)
Austria	437	16761	2.61
Australia	1600	54506	2.94
Belgium	390	19318	2.02
Canada	5199	76120	6.83
Switzerland	656	27861	2.35
Czechia	534	10883	4.91
Germany	3373	120700	2.79
Denmark	367	10851	3.38
Estonia	28	1247	2.25
Spain	1634	58404	2.80
Finland	263	18193	1.45
France	2847	81010	3.51
United Kingdom	2942	123372	2.38
Greece	677	19191	3.53
Croatia	105	2824	3.72
Hungary	165	6566	2.51
Ireland	401	9250	4.34
Israel	1514	15814	9.57
Iceland	23	579	3.97
Italy	3228	70075	4.61
Japan	5027	128244	3.92
South Korea	2281	71329	3.20
Lithuania	25	1627	1.54
Luxembourg	81	1348	6.01
Latvia	13	848	1.53
Netherlands	602	35497	1.70
Norway	232	8497	2.73
New Zealand	130	7031	1.85
Poland	357	18957	1.88
Portugal	669	14939	4.48
Sweden	340	21574	1.58
Slovenia	137	4434	3.09
Slovakia	48	3013	1.59
United States	37991	546548	6.95
All	79129	1607411	4.92

 Table 17: Unassigned Region Institution-Paper Combinations

*Notes:* The table shows, for all countries considered, the paper-institution combinations of papers published up to and including 2015 that (i) could not be matched to regions, (ii) the total number, and (iii) the proportion of unmatched paper-institution combinations.

Country	No regional match	Total	Share not matched $(\%)$
Austria	0	78163	0.00
Australia	91	110646	0.08
Belgium	88	107582	0.08
Canada	10794	424966	2.54
Switzerland	2	220714	0.00
Czechia	2	11793	0.02
Germany	217	1475092	0.01
Denmark	0	71624	0.00
Estonia	0	1374	0.00
Spain	21	42707	0.05
Finland	2	86163	0.00
France	777	541042	0.14
United Kingdom	9022	449616	2.01
Greece	20	4870	0.41
Croatia	2	1291	0.15
Hungary	2	11352	0.02
Ireland	8	32277	0.02
Israel	997	211603	0.47
Iceland	0	1946	0.00
Italy	144	200835	0.07
Japan	944	3062423	0.03
South Korea	622	805106	0.08
Lithuania	0	779	0.00
Luxembourg	0	5867	0.00
Latvia	0	650	0.00
Netherlands	186	188995	0.10
Norway	332	31701	1.05
New Zealand	19	15469	0.12
Poland	0	9758	0.00
Portugal	15	4886	0.31
Sweden	123	176717	0.07
Slovenia	0	2963	0.00
Slovakia	0	1752	0.00
United States	191369	11238927	1.70
All	215799	19631649	1.10

 Table 18: Unassigned Region Geocode-Patent Combinations

*Notes:* The table shows for all countries considered, the geocode-patent combinations of all USPTO patents published up to and including 2015 that (i) could not be matched to regions, (ii) the total number, and (iii) the proportion of unmatched combinations.

Technology field	Number papers	Number patents
Autonomous Cars	9307	2612
Bispecific monoclonal antibody	4938	2499
Cloud computing	80182	11776
Computer vision	178816	13637
Drug conjugates	186986	7972
Electronic gaming	15697	8422
Fingerprint sensor	504	1987
Fracking	13477	1985
GPS	53947	25975
Hybrid vehicle electric car	36099	15896
Lane departure warning	643	487
Lithium battery	62372	14471
Machine Learning AI	624995	28692
Millimeter wave	53766	3684
Mobile payment	2080	863
Oled display	16661	9049
Online streaming	39318	12637
Search Engine	43803	9604
Software defined radio	4842	773
Solar Power	238978	45806
Stent graft	9653	1926
Touch screen	8274	30593
Virtual Reality	101593	5930
Wifi	88184	38464

 Table 19:
 Emergent Technology Paper and Patent Final Sample

*Notes:* The table provides the total number of papers and patents in each technology field of all regions included in the final dataset till up to and including the year 2015.

Country	Region-Tech	#Pioneers Size	Share Size (%)	#Pioneers Time	Share Time (%)
Australia	1176	12	1.02	12	1.02
Austria	816	6	0.74	6	0.74
Belgium	1056	2	0.19	2	0.19
Canada	4440	33	0.74	37	0.83
Chile	72	0	0.00	0	0.00
Croatia	24	0	0.00	0	0.00
Czechia	336	0	0.00	0	0.00
Denmark	264	1	0.38	3	1.14
Estonia	48	0	0.00	0	0.00
Finland	456	7	1.54	8	1.75
France	2352	23	0.98	24	1.02
Germany	9216	46	0.50	55	0.60
Greece	240	3	1.25	3	1.25
Hungary	408	0	0.00	1	0.25
Iceland	48	0	0.00	0	0.00
Ireland	192	0	0.00	0	0.00
Israel	144	5	3.47	5	3.47
Italy	2376	18	0.76	20	0.84
Japan	1128	78	6.91	82	7.27
Latvia	48	0	0.00	0	0.00
Lithuania	48	0	0.00	0	0.00
Luxembourg	24	0	0.00	0	0.00
Netherlands	960	11	1.15	13	1.35
New Zealand	264	0	0.00	0	0.00
Norway	408	2	0.49	2	0.49
Poland	576	0	0.00	1	0.17
Portugal	168	0	0.00	0	0.00
Slovakia	168	0	0.00	0	0.00
Slovenia	168	0	0.00	0	0.00
South Korea	408	17	4.17	18	4.41
Spain	960	4	0.42	4	0.42
Sweden	504	6	1.19	6	1.19
Switzerland	624	13	2.08	15	2.40
United Kingdom	4152	36	0.87	39	0.94
United States	20664	337	1.63	373	1.81

Table 20: Region-Technology Combinations and Number of Pioneer Regions by Country

*Notes:* The table displays the emergence year, total number of regions, and pioneer regions after applying the following restrictions: including only patents filed up to 2015, regions with more than 50 patents in total, and technologies with an emergence year at least ten years prior to 2015, i.e., 2005. Consequently, the technologies Social Networking, Smart Devices, RFID Tags, Wireless Charging, and 3D Printing are excluded from our final dataset.

Country	Region	Number papers	Number patents
United States	New York-Newark-Jersey City, NY-NJ-PA	66981	11341
Japan	Tokyo	54795	6551
United States	Boston-Cambridge-Newton, MA-NH	36708	7439
United States	Washington-Arlington-Alexandria, DC-VA-MD-WV	35655	4397
South Korea	Seoul	32985	6928
United States	Los Angeles-Long Beach-Anaheim, CA	32040	9157
United Kingdom	Westminster	26123	1037
France	Paris	25952	707
United States	San Jose-Sunnyvale-Santa Clara, CA	24757	22380
United States	San Francisco-Oakland-Berkeley, CA	20120	18367
United States	Chicago-Naperville-Elgin, IL-IN-WI	19106	4474
Australia	Sydney	16572	643
Spain	Madrid	15891	176
United States	Pittsburgh, PA	15405	1193
Italy	Rome	15313	184
United States	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	14563	3007
South Korea	Daejeon	14348	1912
Japan	Osaka	14063	2572
United States	Seattle-Tacoma-Bellevue, WA	13871	9033
United States	Houston-The Woodlands-Sugar Land, TX	12939	2034
Australia	Melbourne	12622	326
United States	San Diego-Chula Vista-Carlsbad, CA	12344	7212
Spain	Barcelona	12306	234
Canada	Communautue-Urbaine-de-Montrueal, QC	11734	656
Germany	Muenchen, Kreisfreie Stadt	11655	813
United States	Dallas-Fort Worth-Arlington, TX	11438	3951
Canada	Toronto metropolitan municipality, ON	11075	897
United States	Atlanta-Sandy Springs-Alpharetta, GA	10901	2808
Germany	Berlin	10874	473
United States	Austin-Round Rock-Georgetown, TX	10545	3030

 Table 21: Top 30 Regions with the Most Emerging Technology Papers

*Notes:* The table shows the total number of papers and patents of all 24 considered Bloom technologies combined up to and including 2015 for the 30 regions with the highest number of papers in these technologies.

Country	Region	Number patents	Number papers
United States	San Jose-Sunnyvale-Santa Clara, CA	22380	24757
United States	San Francisco-Oakland-Berkeley, CA	18367	20120
United States	New York-Newark-Jersey City, NY-NJ-PA	11341	66981
United States	Los Angeles-Long Beach-Anaheim, CA	9157	32040
United States	Seattle-Tacoma-Bellevue, WA	9033	13871
South Korea	Gyeonggi-do	8749	10258
United States	Boston-Cambridge-Newton, MA-NH	7439	36708
United States	San Diego-Chula Vista-Carlsbad, CA	7212	12344
South Korea	Seoul	6928	32985
Japan	Tokyo	6551	54795
Japan	Kanagawa	5749	5368
United States	Chicago-Naperville-Elgin, IL-IN-WI	4474	19106
United States	Washington-Arlington-Alexandria, DC-VA-MD-WV	4397	35655
United States	Detroit-Warren-Dearborn, MI	4175	7140
United States	Dallas-Fort Worth-Arlington, TX	3951	11438
United States	Austin-Round Rock-Georgetown, TX	3030	10545
United States	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	3007	14563
United States	Portland-Vancouver-Hillsboro, OR-WA	2817	3148
United States	Atlanta-Sandy Springs-Alpharetta, GA	2808	10901
Japan	Osaka	2572	14063
Japan	Aichi	2571	10227
United States	Raleigh-Cary, NC	2303	4365
Israel	Central District	2080	1979
United States	Minneapolis-St. Paul-Bloomington, MN-WI	2072	5838
United States	Phoenix-Mesa-Chandler, AZ	2040	6773
United States	Houston-The Woodlands-Sugar Land, TX	2034	12939
United States	Denver-Aurora-Lakewood, CO	1961	8791
United States	Miami-Fort Lauderdale-Pompano Beach, FL	1939	4139
South Korea	Daejeon	1912	14348
United States	Ann Arbor, MI	1593	7855

 Table 22: Top 30 Regions with the Most Emerging Technology Patents

*Notes:* The table shows the total number of papers and patents of all 24 considered Bloom technologies combined up to and including 2015 for the 30 regions with the highest number of patents in these technologies.