



Short communication

A new approach to measuring invention commercialization: An application to the SBIR program

Carlo Bottai ^{a,b}, Gaétan de Rassenfosse ^c, Emilio Raiteri ^b ^{*}

^a Department of Economics, Management and Statistics, University of Milano–Bicocca, P.zza dell'Ateneo Nuovo 1, Milano, 20126, Italy

^b Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, P.O. Box 513, Eindhoven, 5600 MB, The Netherlands

^c College of Management of Technology, École polytechnique fédérale de Lausanne, ODY 201.1, Station 5, Lausanne, 1015, Switzerland



ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.16779955>, <https://github.com/n3ssuno/iris-scraper/releases/tag/v0.9-rc>, <https://github.com/n3ssuno/iris-classifier/releases/tag/v0.9-rc>

Keywords:

Government-funded research
Invention commercialization
Patents
Policy evaluation
Web-based evidence

ABSTRACT

Measuring the commercialization of patented inventions remains a key challenge in innovation studies. This paper introduces a novel, web-based method for tracking the commercialization of patented inventions. The method leverages targeted web searches to identify online traces of commercialization, offering a scalable alternative to surveys and case studies. We apply this method to patents arising from the U.S. Department of Defense's Small Business Innovation Research program, linking 3070 patents to procurement contracts and assessing their commercialization outcomes. The results indicate that 21.5% of these patents show signs of commercialization, with variations across R&D stages and contract phases. The method provides a systematic way to identify market adoption of patented technologies and can be extended to other contexts where identifying commercialized patents is relevant.

1. Introduction

Science, Technology, and Innovation (STI) policy aims to foster R&D investments and stimulate inventive activities. A key objective is to drive the development of novel products and services that, when commercialized, deliver significant benefits to consumers. However, while commercialization is a central goal of STI policy, measuring commercialization outcomes remains challenging, and research in this area often relies on survey data or isolated success stories. For example, Ruttan (2006) describes how the U.S. Department of Defense (DoD) was instrumental in the launch of the commercial Internet and GPS technology. Mazzucato (2013, 2021) stresses that popular consumer products, such as the iPhone or the iPad, and services like Siri benefited strongly from public support. A key obstacle to systematic evaluation is data scarcity, which hampers the ability to track the full journey from invention to market launch. This paper addresses this challenge by introducing a novel, web-based method for measuring invention commercialization.

Building on a growing body of literature that exploits web-based data to measure innovation (Gök et al., 2015; Kinne and Axenbeck, 2020; Rammer and Es-Sadki, 2023), our method involves performing targeted searches on the web to identify traces of invention commercialization. Specifically, we leverage recent changes in U.S. patent

law (de Rassenfosse, 2018) and innovators' publicizing their commercialization success online to identify patent-protected products and services.

We apply our method to the study of the DoD's Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs, two related public funding programs that seek to encourage U.S. small businesses to engage in federal R&D projects with commercialization potential. The DoD accounts for the majority of the SBIR funding, whose total budget for 2025 reaches about \$4 billion. The focus on SBIR funding allows us to link with reasonable certainty the procurement contracts to the associated patents using government interest statements in patent documents. The final dataset consists of 3070 granted patents with filing years ranging from 1984 to 2019 and assigned to 1182 distinct companies. These patents acknowledge 2213 different procurement contracts.

We identify traces of commercialization for 642 out of the 3070 patents in our sample. The commercialization rate is higher for Applied or Development R&D contracts compared to Basic R&D contracts, and for Phase II contracts compared to Phase I contracts. These findings, consistent with expectations, suggest that our method effectively captures meaningful signals of commercialization.

* Corresponding author.

E-mail address: e.raiteri@tue.nl (E. Raiteri).

In a further analysis, we compare these commercialization outcomes to a benchmark set of privately funded patents with otherwise similar observable characteristics to the SBIR-funded patents. The results indicate that SBIR-funded patents are 17 percent more likely to be commercialized than the benchmark patents. However, we refrain from interpreting this finding as evidence of a causal effect of the SBIR program on commercialization.

The rest of the paper is organized as follows. Section 2 provides background information on the importance of commercialization in STI policy, and the difficulty in measuring it. Section 3 describes the growing use of web-based data in innovation studies, explains previous attempts to measure commercialization, and introduces our method. Section 4 illustrates the method with an application to patents acknowledging funding from the DoD SBIR/STTR programs. Section 5 reports the results of an exploratory analysis of the collected data, and Section 6 concludes.

2. Background

2.1. Commercialization is a central goal of STI policy

Endogenous growth theory has long emphasized STI's central role in driving economic growth (Romer, 1990; Grossman and Helpman, 1993; Aghion and Howitt, 1998). STI enhances productivity by enabling more efficient production processes, allowing economies to produce more output with the same or even fewer inputs.

Realizing these productivity gains requires the creation, implementation, and diffusion of new knowledge. While knowledge creation often dominates the collective imagination — think of the scientist in the lab performing experiments — the implementation and diffusion of newly-created knowledge are equally critical. Implementation ensures that this knowledge is put to practical use, enabling productivity gains, while diffusion broadens its impact, maximizing those gains across sectors and regions.

The implementation phase, in particular, is central to the very concept of innovation, commonly defined as the application or practical use of an invention to create economic value. Put differently, innovation occurs when inventions reach the market, making invention commercialization a pivotal milestone in STI activities.

While STI policy spans all the phases of STI activities, measures targeting commercialization hold a prominent place in the policy toolbox. For example, in the European Union, the European Innovation Council (EIC) was established with the specific mission of supporting the commercialization of high-risk, high-impact technologies. It has a budget of €1.4 billion for 2025 (approximately \$1.5 billion).¹ In the United States, the SBIR program plays a similar role, offering competitive grants to small businesses to support the commercialization of innovative technologies arising from federally funded R&D. Its budget for 2025 is projected at approximately \$4 billion.²

2.2. Commercialization is the poor relation of STI policy studies

Studies evaluating STI policies have focused predominantly on estimating the so-called input additionality effect—whether specific policy tools such as R&D subsidies, R&D tax credits, or innovation procurement contracts increase firms' private investment in R&D (García-Quevedo, 2004; Dimos and Pugh, 2016). In contrast, studies examining output additionality — the extent to which subsidies lead to the introduction of new products, processes, or services — are less numerous, largely due to data restrictions. Research in this area generally employs three primary approaches to measure innovation output.

¹ See https://eic.ec.europa.eu/eic-2025-work-programme_en, last accessed December 4th, 2024.

² See <https://www.sbir.gov/>, last accessed December 4th, 2024.

One approach builds on the foundational work of Pakes and Griliches (1980) using patents as indicators of successful R&D projects—arguably an intermediate measure of innovation (Bronzini and Piselli, 2016; Guo et al., 2016; Czarnitzki and Hussinger, 2018). Another approach relies on self-reported survey data, using variables such as the number of new products introduced by a firm or revenues generated from innovative products (Hussinger, 2008; Guo et al., 2016; Radicic and Pugh, 2017; Prencipe et al., 2024). Finally, a third approach infers the effect of STI policies using production functions à la Griliches (1979). Works in the stream include, e.g., Karhunen and Huovari (2015), Cin et al. (2017), and Li et al. (2022).

These approaches have considerably enriched our understanding of STI policy; however, they suffer from limitations that hamper further progress. For example, evaluations based on patent counts or survey data about newly introduced products rarely establish a direct link between a patent or a new product and the specific government support instrument, which is particularly problematic for firms that benefit from several support instruments simultaneously. Furthermore, studies relying on patent data face the well-documented issue of patent value skewness. While a small subset of patents may be highly valuable, the majority are of little economic significance (Scherer and Harhoff, 2000). Although (imperfect) methods exist to account for patent value (Higham et al., 2021), patents are filed and maintained for a host of reasons, some of which bear little relevance for innovation measurement purposes (e.g., patents for defensive or strategic reasons). Interpreting patenting activity as indicative of commercialization efforts might be somewhat misleading. The commercialization of products based on granted patents is a highly uncertain process, involving expensive development and testing with unpredictable results. Many patents fail to lead to commercialized products, while a single product may fall under the scope of numerous patents, or a single patent may relate to multiple commercialized products.

To address this challenge, some studies have focused on contexts where the link between patents and products is straightforward. For example, in the U.S. pharmaceutical industry, the FDA's Orange Book provides a resource for linking medical drugs to the patents protecting them. Azoulay et al. (2019) leverage these data to assess the impact of public support for scientific research provided by the National Institutes of Health. Other studies have used in-depth surveys and interviews with company managers to identify connections between patented inventions and new products (Svensson, 2007; Braunerhjelm and Svensson, 2024).

To the best of our knowledge, no existing study has developed a method capable of systematically tracking, at scale, the direct relationship between innovation policy instruments, the inventions they generate, and information about commercialization. Our web-based approach represents an important step toward bridging this methodological gap in the literature.

3. Web-based assessment of commercialization

3.1. Web-based data to measure innovation

Our approach fits within a wide and growing literature that has established that corporate websites are a useful and reliable information source for economic studies. Companies today employ their websites as digital shopfronts to showcase their products, convey operational information, and establish their corporate identity. Since corporate websites reflect an organization's economic activities, are public, regularly updated, and intentionally created by the businesses themselves, they have become a precious information source for social science researchers (Domènech et al., 2012; Edelman, 2012; Kinne and Axenbeck, 2020; Arora et al., 2021; Rammer and Es-Sadki, 2023). This source is particularly valuable in the context of micro, small, and medium-sized enterprises, for which the availability of conventional sources, such as balance sheets and survey data, is limited.

Creative ways of using web-based data to measure innovation are constantly emerging. For instance, Arora et al. (2013) have used corporate websites to examine SMEs' commercialization of emerging technologies. Gök et al. (2015) demonstrated that web-extracted data from UK SMEs yielded supplementary insights beyond traditional sources, such as patents and scientific publications, about firms' innovation activity. Among other things, the authors highlight that, while patents and publications are superior at capturing early phases of the R&D activity, web mining offers better insights about the downstream or customer-oriented part of the innovation process.

Libaers et al. (2016) employed corporate websites' text to develop a business-model taxonomy for small, innovation-driven firms focused on technology commercialization. Additionally, several scholarly works established the relevance of the textual content of corporate websites to identify product innovators (Daas and van der Doef, 2020; Kinne and Lenz, 2021; Axenbeck and Breithaupt, 2021; Ashouri et al., 2022). In a similar vein, Guzman and Li (2023) employed textual content from corporate websites of over 12,000 U.S. startups to assess their strategic differentiation relative to incumbent competitors.

3.2. Virtual patent marks as a valuable source of information

Besides the textual content of corporate websites, some specific pages or documents hosted on these websites contain valuable information on commercialized products. In particular, "virtual patent marking" (VPM) web pages offer a detailed mapping between a firm's products and the patents protecting them. VPM is the online equivalent of "physical marking," which involves printing or engraving the relevant patent numbers on a product to notify the public of its patent protection. VPM was enabled in the United States by the Leahy-Smith America Invents Act (AIA), signed into law on September 16, 2011. The AIA amended 35 U.S.C. §287(a), the so-called "marking" statute in U.S. patent law, allowing firms to post the marking information online.

A recent project has exploited this legislative change to build a database of patent-product pairs. The IProduct initiative performs large-scale crawls of the web in search of VPM web pages. It then extracts and harmonizes the patent-product correspondences, which is made available via the [iproduct.io](https://www.iproduct.io) platform (de Rassenfosse, 2018). Note that, in the process, the crawler also captures other classes of pages that provide patent-product links, such as press releases, product catalogues, and product description sheets. Such data have already been used in academic studies, for instance in de Rassenfosse and Zhou (2020) and Devarakonda et al. (2024). The present study is an extension of the IProduct project.

3.3. Application to STI policy instruments

Building on previous works, we start from the observation that patents constitute a key milestone in the commercialization of new technology-based products. We then search for traces of patent commercialization using dedicated web searches.

Our starting point is the list of all patented inventions arising from the innovation policy instrument under review. In the United States, this can be achieved by leveraging the Bayh-Dole Act of 1980 and its integration into the U.S. Federal Acquisition Regulation (FAR), as detailed by de Rassenfosse et al. (2019). The Bayh-Dole Act mandates that private entities acknowledge federal funding and government rights in the written specification of any U.S. patent application for inventions arising from federally-funded research. Additionally, the FAR requires patent applicants to disclose the specific government agency and the associated contract or grant number in the patent document.

Next, we identify web pages that provide evidence of patent commercialization. These web pages are not necessarily VPM pages. They could also relate, for instance, to product data sheets or company promotional material. We refer to web pages containing relevant commercialization information simply as "relevant pages." Note that this

study does not exploit information on the actual products—we are simply interested in finding traces of patent commercialization. Therefore, unlike the IProduct project, our approach does not integrate information on the product-patent relationship. Furthermore, we leverage information on the assignees (and awardees of the acknowledged federal contracts) along with the patent numbers to perform targeted searches of the web. Our process involves three steps.

In Step 1, we systematically collect a list of potential companies' website URLs. To do so, we search for the legal names of assignees on Google Search, and extract domain names from each search result, retaining the ten most relevant domains (excluding duplicates).³ Since we will focus on patents arising from government contracts, we can also search for the awardee names in addition to the patent assignee names. Considering both assignees and awardees increases the chance of identifying pertinent web domains. Indeed, assignees may differ from awardees due to mergers, acquisitions, or individual patent transfers.⁴ Note that we do not assess the pertinency of the domains collected at this stage, as any false positives will be filtered out during the subsequent steps.⁵

In Step 2, we search specifically for each patent numbers on each of the web domains retrieved in Step 1, utilizing queries like (`site:mybiz.com`) AND (8502792 OR 8898242). This process leads to the retrieval of multiple web pages from the assignee's web domain(s) containing a string of characters that matches one of the patent numbers of interest. The string of characters may correspond, say, to a phone number or a patent. If it is a patent, it may not link to a product (e.g., a notice of patent issuance).

Step 3 filters out irrelevant web pages. We classify each of the pages containing a patent number as a positive or negative case. We combine a classifier developed for the study (Step 3.1) and a semi-automatic approach (Step 3.2). The classifier in Step 3.1 uses predefined rules and regular expressions to classify pages as positive or negative cases based on content patterns. The group of negatives includes, for example, PDF files of patent documents as released by the USPTO and PDF files of legal forms required by the U.S. Securities and Exchange Commission. The group of positives includes pages explicitly mentioning *virtual patent marking* or referring to the associated legislation; pages where the text also includes a trademark (™) or registered trademark (®) symbol in the proximity of the patent numbers being analyzed; pages where the text reports expressions frequently associated with patent-protection of a product, such as *covered by* or *employs our patent*, near one of the relevant patent numbers. Pages that cannot be assigned a positive or negative outcome automatically are labeled as "uncertain" and will be reviewed manually in Step 3.2. While this classifier is fairly simple, a cross-validation with manually classified pages shows an overall accuracy rate of approximately 90 percent (i.e., the proportion of correct predictions among total predictions).

Step 3.2 concerns pages marked as "uncertain" by the automatic classifier. We manually reviewed all of them via a browser-like interface to speed up processing. As Fig. 1 illustrates, the interface displays the web page to classify, and the user can use the buttons in the right panel to classify it into several (positive or negative) commercialization outcome categories. Appendix A.22 describes the details of the classification process and the steps taken to ensure its accuracy.

³ Specifically, for an entity like 'MyBiz Corp.', we run the query (`'MY-BIZ CORP'` OR `'MYBIZ CORPORATION'`) `-site:gov -site:edu -site:mil -site:int -site:bloomberg.com`.

⁴ However, our method is likely to miss patents for which the ownership transfer was not recorded at the USPTO. In such cases, complementing the approach with data from a large-scale, untargeted crawl, as used in IProduct, can help mitigate this issue.

⁵ Since we will search for each patent number within the domain in Step 2 and assess the relevance of each page retrieved in Step 3, domains not associated with patent numbers or pages deemed as non-relevant will naturally be filtered out.

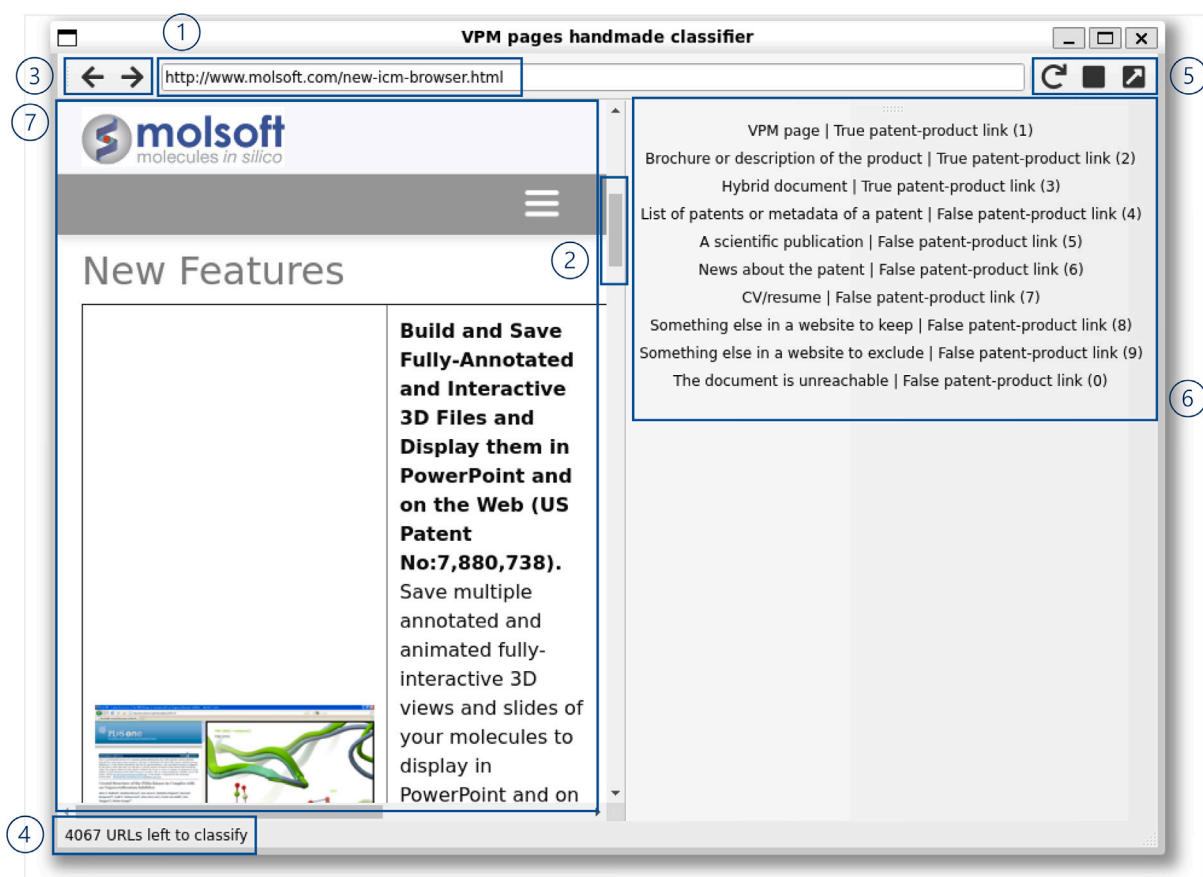


Fig. 1. Main features of the manual classifier interface.

Notes. (1) The URL address of the web page to classify. (2) A scrollbar to navigate within the page. (3) Navigation buttons. (4) A counter of the number of web pages left to classify. (5) A button to open the web page in the user operating system's main browser. (6) Classification panel. (7) Panel that displays the web page to classify.

At the end of the process, we have linked a specific policy instrument to the resulting patents and the associated commercialization traces. These data can then be used for investigating the impact of innovation policies on market outcomes. In the next section, we demonstrate the practical application of this approach using the DoD's SBIR program as a case study.

4. Application to the DoD SBIR and STTR programs

Before delving into the details of the dataset construction in Section 4.3, we briefly present the SBIR and STTR programs and take stock of the literature assessing their impact on commercialization.

4.1. The SBIR and STTR programs

The SBIR program was introduced by the Small Business Innovation Development Act of 1982, whose objectives include the increase of private sector commercialization of innovations derived from federal R&D. Its explicit goals are to (i) stimulate technological innovation, (ii) use small business to meet federal R&D needs, (iii) foster and encourage participation in innovation and entrepreneurship by women and socially or economically disadvantaged persons, and (iv) increase private-sector commercialization of innovations derived from federal R&D funding.⁶ The STTR came a decade later, in 1992. The two programs share a similar structure and purpose, primarily distinguished by the collaboration requirement; the SBIR program allows optional

research partnerships, while the STTR mandates them. Given their close alignment, we consider the two programs a joint funding scheme for the purpose of this paper. As such, from now on, we will use the term "SBIR" to refer to both, unless stated otherwise.

The U.S. Small Business Administration (SBA) coordinates the programs that involve eleven participating agencies. The expected contribution of these federal agencies amounts to \$4 billion of SBIR funding for the year 2025. The SBIR program has two main phases. Phase I funds initial research to establish the technical merit, feasibility, and commercial potential of an R&D project. Successful Phase I participants may proceed to Phase II, where they receive more significant funding to pursue the research started in Phase I. In our study period, Phase I awards generally amount to \$50,000–150,000 for six months or one year, whereas Phase II awards may reach \$1 million and last for two years. The DoD accounts for the majority of SBIR funding, contributing over 60 percent of the total annual budget.⁷

4.2. Commercialization and the SBIR program

Policymakers and scholars alike have devoted considerable effort to assessing the "impact" of the SBIR program in terms of commercialization. A handful of academic studies exploit sales and patent applications as proxies for commercialization, including Audretsch et al. (2002), Link and Scott (2010), Dutta et al. (2022). Howell (2017) uses data on grant applications to the U.S. Department of Energy's SBIR program and shows that a Phase I award "approximately doubles the probability that

⁶ For further details about the program, see the Small Business Act (15 U.S.C. § 638), as well as <https://www.sbir.gov/about>.

⁷ <https://www.sbir.gov/participating-agencies>

a firm receives subsequent venture capital and has large, positive impacts on patenting and revenue". Feldman et al. (2022) examine SBIR recipients' commercial activities using a variety of metrics, including manual web searches and find that the top ten highly awarded SBIR firms engage in significant commercial activity.

Since 2000, the National Academies have undertaken a quadrennial assessment of each agency's SBIR program, using case studies and survey data. The DoD reports assert the program's positive effect on commercialization. According to these assessments, nearly half of Phase II projects are associated with sales from products developed with SBIR funds (National Research Council, 2009a,b, 2014).

A few contributions highlight some potential limitations of the SBIR evaluations conducted so far. A Government Accountability Office report emphasizes that studies carried out by military departments mainly focus on selected success stories (Mak, 2014). A recent study by the National Academies of Sciences, Engineering, and Medicine (2020) stresses how extant evaluations do not always capture product market introductions. The DoD considers SBIR-funded projects as having a successful transition to commercialization if supported firms report any positive revenues from a product or service developed in the performance of the project. However, these revenues may originate from non-SBIR contracts awarded by the DoD itself.

All in all, the program is considered to be largely successful, being extensively studied and emulated internationally (National Academies of Sciences, Engineering, and Medicine, 2020). With its explicit focus on commercialization, substantial scale, and ongoing calls to enhance the tracking of commercialization outcomes, the DoD SBIR program serves as an ideal setting to test our method.

4.3. Data construction

As explained, our method starts with the identification of patents connected to a specific policy instrument. It then involves locating and validating web pages that provide evidence of commercialization for these patents.

We rely on publicly available federal award data from the Defense Contract Action Data System (DCADS) and the USAspending.gov databases to identify patents related to the DoD SBIR program. These resources provide comprehensive information on U.S. federal contracts, grants, and other financial awards. We retrieved data on DoD contracts from 1983 to 2018, focusing specifically on SBIR Phase I and Phase II awards. Using patent records from the USPTO's PatentsView database, we linked these awards to relevant patents by extracting contract identifiers from government interest statements included in patent documents. This process resulted in the identification of 3070 patents linked to DoD SBIR awards.⁸ Our approach to link SBIR contracts to patents is similar to the one adopted for the construction of the 3PFL dataset, but with a key difference (de Rassenfossé et al., 2019). While 3PFL covers all federal agencies and various contract types, we focus exclusively on DoD SBIR contracts. This more targeted scope allows us to extend the time coverage back to the program's inception and adopt tailored methods to collect richer contract- and recipient-level details.

For each of the SBIR contracts linked to a patent, we augment the base data with contract-level information from the Federal Procurement Database System (FPDS). We specifically retain key details about the contract start and end dates; the awarding sub-agency and office; the recipient's name and DUNS number; the total dollar amount awarded; the product or service code (PSC); and the SBIR Phase. The

⁸ Appendix A.21 provides a detailed explanation of the procedure adopted to extract the contract identifiers—the Procurement Instrument Identifiers (PIID). Data about the government interest statements in patents are from PatentsView (Jones and Madhavan, 2020). Data about the awards comes from the DCADS for the years 1984–2001 and from USAspending.gov for the years 2001–2018.

PSC allows us to identify the stages of R&D efforts for which a contract is awarded, from basic research to more advanced development activities. We use this information to classify the contracts as *basic*, *applied*, or *development* research.⁹

To capture the commercialization outcome of a contract in a more comprehensive manner, we consider two paths leading to a product. A direct path occurs when a patent acknowledging SBIR support protects a product as identified on a relevant page belonging to the patent assignee. Fig. 2 illustrates this case with an autonomous home floor mopper. The website of the company commercializing the product lists the patents protecting it. One of these patented inventions was first developed in the performance of an SBIR contract awarded by the Army Aviation and Missile Command.

An indirect path occurs when the SBIR-funded patent is cited by a subsequent patent for which we found evidence of commercialization. Given the technical function of patent citations as signals of existing prior knowledge relevant to the new invention (Jaffe and de Rassenfossé, 2017), we also consider this second path as providing evidence of a successful commercialization event. Fig. 3 reports the example of a set of noise-canceling headphones. One of the key patents protecting the noise-canceling technology embedded in these headphones builds on a patented invention developed with the support of an Army SBIR contract awarded in 1993.

Relying on these two paths implies that we must look for traces of commercialization not only for the SBIR-funded patents but also for the patents that cite them. Thus, the set of patents for which we will search the web includes the original sample of 3070 SBIR-funded patents (of which 2304 received at least one citation from another patent) and a sample of 40,020 patents citing a SBIR-funded patent.

We follow the three-step process described in Section 3.3. The first step involves identifying and recovering the websites associated with patent assignees (and contract awardees). The patents in our working sample are associated with 6647 distinct entities. Searching for them on Google Search leads to 11,731 web domains.¹⁰ After removing information aggregator websites and obvious incorrect attributions, the sample was reduced to 9411 unique domains.

We then scan this list of web domains iteratively to identify web pages mentioning the patent numbers. Specifically, we performed Google searches using Puppeteer, a JavaScript library that simulates a browser to automate tasks on web pages to find patent numbers on the web domains associated with a specific assignee. This process retrieves 3131 web pages containing a string of characters compatible with at least one of the patent numbers of interest.

In the final step, we use the classification approach described above to identify the web pages that provide evidence of commercialization. We find that 44.3 percent of the web pages collected by the scraper are relevant.

4.4. Data overview

The final dataset consists of 3070 granted patents with filing years ranging from 1984 to 2019 and assigned to 1182 distinct companies. These patents acknowledge 2213 different procurement contracts, with 14.6 percent of the patents reporting the support of multiple awards. We find a direct path of commercialization for about eight percent of

⁹ More specifically, the fourth digit in the product and service code (PSC) identifies the stage of the R&D effort required by a given contract with: (1) Basic Research; (2) Applied Research and Exploratory Development; (3) Advanced Development; (4) Engineering Development; (5) Operational Systems Development; (6) Management and Support; (7) Commercialization. Contracts classified as development research include those at levels (3)–(5).

¹⁰ To increase the precision of results, we searched for the assignee's website also on [Bloomberg.com](https://www.bloomberg.com) and the official SBIR program's website (<https://www.sbir.gov>).

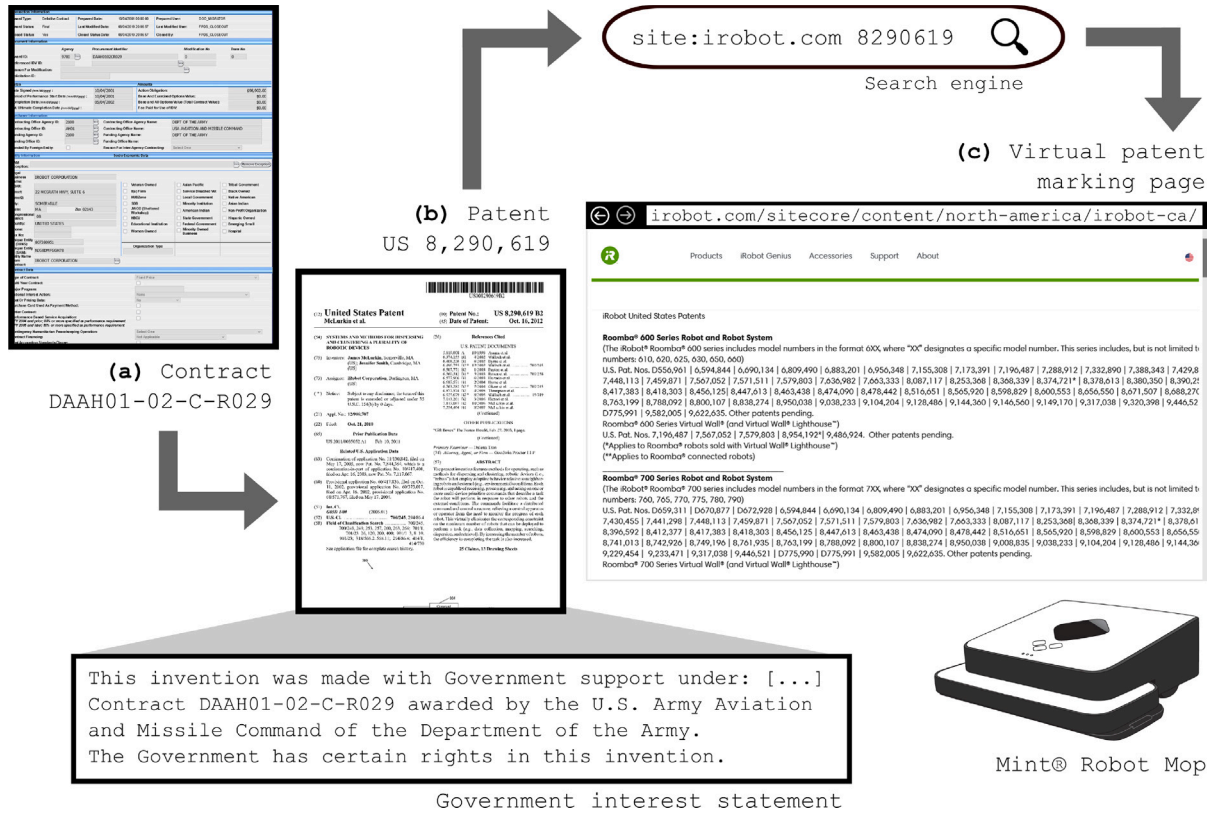


Fig. 2. Illustrative example of a direct path. Notes. (a) In 2001, the U.S. Army signed contract No. DAAH01-02-C-R029 with iRobot, Corp.. (b) The company applied for a patent, granted in 2012 as US-8290619-B2, acknowledging the government's support for this invention. (c) As disclosed by iRobot on its website, this patent protects the company's Mint® Robot Mop, Mint Plus® Robot Mop, and Braava® Robot Mop products.

the patents and an indirect path for about 17 percent of them. Accounting for the fact that some patents are linked to a product through both direct and indirect paths, we find evidence of commercialization for about 21 percent of the patents.

A total of 1088 patents acknowledge at least one basic research contract, 990 an applied research contract, and 489 a development contract. Regarding the phase of the SBIR contract, 1582 patents (51.5%) acknowledge at least one Phase I contract and 1252 patents acknowledge Phase I contracts exclusively. A total of 1818 patents (59.2%) acknowledge one or more Phase II contracts.¹¹

Fig. 4 illustrates that most patents acknowledging support from the DoD SBIR program concern recent years, with the median patent being applied to the USPTO in 2007. In particular, the chart shows a significant increase in patenting activity by DoD-SBIR recipients from 1997 onwards. This pattern aligns with the growth in overall patenting activity (Danguy et al., 2014) over that period but also reflects the fact that the compliance rate in reporting mandated by the Bay-Dole Act was generally lower in the earlier years of the time window (Rai and Sampat, 2012).

The commercialization of DoD-SBIR-funded technologies appears particularly strong during 1990–2004, with 28 to 37.5 percent of funded patents linked to a product, compared to about 6.8 to 24 percent in 1986–1989 and 2005–2019. This trend reflects the limited web presence in earlier years and the time lag for newer patents to reach

commercialization, especially through indirect paths, for more recent years.

As Fig. 5 illustrates, the DoD-SBIR-funded patents are concentrated in a few technological fields, reflecting the DoD's R&D needs. A total of 32.2 percent of the patents relate to electrical and electronic technologies, 25 percent to the domain of computers and communications, 16.5 percent to chemical, and 14.2 percent to mechanical fields. The proportion of commercialized patents is surprisingly similar across the technological categories (from 17.9% to 23.2%), suggesting little technology-specific effects.

Lastly, turning to the spatial distribution of the data, the top panel of Fig. 6 illustrates that SBIR-funded patents are unevenly concentrated in a few metropolitan areas (MSAs) around the United States. This observation is consistent with the geography of innovation literature (Feldman and Kogler, 2010). The bottom panel of the figure depicts the commercialization rate of SBIR-funded patents. Looking at the two maps combined suggests no correlation between an MSA's share of patents and its commercialization rate (Pearson's correlation coefficient of -0.007).

5. Exploratory analysis

This section compares the commercialization odds of SBIR-funded patents with those of comparable, non-SBIR-funded patents. This exploratory analysis highlights the relevance of the collected data, revealing aspects of the SBIR program that correlate with shifts in commercialization probability.

5.1. Building a set of comparable patents

We construct a set of benchmark patents with similar characteristics to the SBIR-funded patents in the sample. For each SBIR-funded

¹¹ For patents linked only to Phase I contracts, we also determine if the project never reached Phase II or if a Phase II contract exists, but the patent simply did not mention it (see Appendix A.21 for further explanation). Accounting for Phase I contracts later extended to a Phase II contract not acknowledged in the patent document, we find that 2374 patents (82.0%) are connected to Phase II funding.

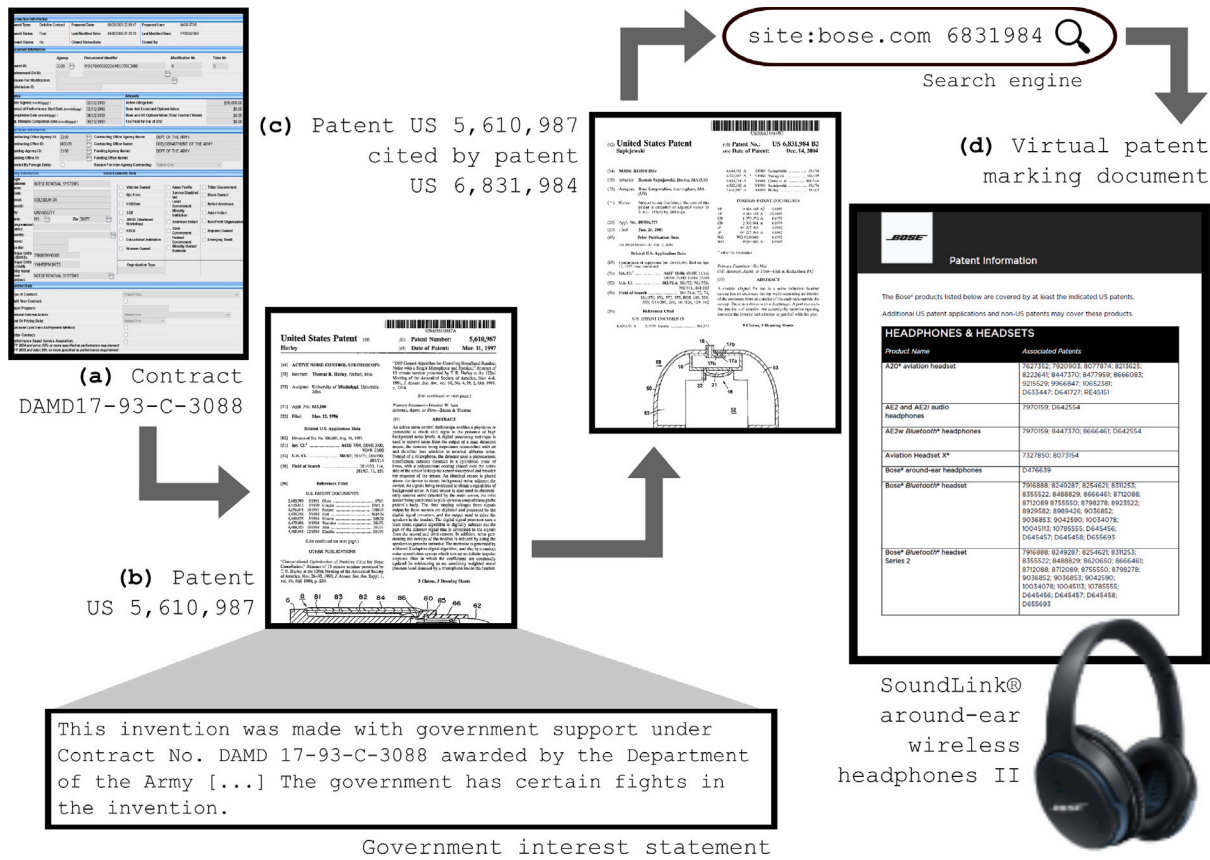


Fig. 3. Illustrative example of an indirect path. Notes. (a) Contract No. DAMD17-93-C-3088, signed between the U.S. Army and Noise Removal Systems in 1993. (b) This contract is acknowledged in patent US-5610987-B2, granted by the USPTO in 1997. (c) This patent is cited as relevant prior-art by patent US-6831984-B2, granted to Bose, Corp. in 2004. (d) Bose uses its website to notify the public that this last patent is protecting its SoundLink® around-ear wireless headphones II and its A20® aviation headset.

patent, we selected up to three benchmark patents from a pool of patents assigned to a private company classified as a small entity by the USPTO and applied for between 1984 and 2019. Each of the selected benchmark patents shares the main USPC technological class and the filing year of its respective SBIR-funded patent (exact matching). The matching procedure starts from 3070 SBIR-funded patents and 4828 benchmark patents. However, for seventeen SBIR-funded patents, we did not find any benchmark candidates and had to drop them. Moreover, 157 SBIR-funded patents and four benchmark candidates do not have any assignee’s organization—but are assigned either to the inventor or to the non-inventor applicant. Dropping the 202 benchmark candidates for these 157 SBIR-funded patents, we end up with a final sample of 2896 SBIR-funded patents, assigned to 1060 distinct companies, and 4622 benchmark patents, assigned to 3892 distinct companies, to be used in our comparison exercise.

The next step involves looking for commercialization traces of the patents in the benchmark set (and their associated 58,881 citing patents). Tables 1 and 2 report the results of this web search. Table 1 provides an overview of the page types. We find 3144 relevant pages mentioning the benchmark patents (or the patents that cite these patents), and we are able to classify automatically 25.4 percent of them. Among the remaining 2345 pages that we had to classify manually, 19.7 percent are VPM pages (without any clear mention of the body of the patent marking legislation, otherwise included among the “automatically classified” pages), 43.5 “brochures,” which we define as HTML or PDF documents describing the characteristics of a company’s products, and the remaining 11.4 percent of pages are hybrid documents, such as press releases. These numbers are comparable with those for the web pages relevant to the SBIR-funded patents, for which we

Table 1
Overview of relevant web pages.

	SBIR-funded		Benchmark patents	
	N	Percent.	N	Percent.
Automatically classified	579	25.7%	799	25.4%
Manually classified	1676	74.3%	2345	74.6%
VPM page	544	(24.1%)	620	(19.7%)
Brochure	928	(41.2%)	1368	(43.5%)
Hybrid document	204	(9.1%)	357	(11.4%)
Relevant pages	2255	100%	3144	100%

Notes. Brochures include any HTML or PDF document describing the characteristics of a company’s products. Proportions in parenthesis sum up to 100 percent within the column and reflect the proportion of manually-classified page types.

Table 2
Number of patents in the sample by commercialization path.

	SBIR-funded		Benchmark patents		Diff. in prop.
	N	Percent.	N	Percent.	
Patents with a direct path	225	7.8%	277	6.0%	1.8***
Patents with an indirect path	498	17.2%	696	15.1%	2.1**
Patents with any path	623	21.5%	856	18.5%	3.0***
Patents	2896		4622		

Notes. The same patent can simultaneously have both direct and indirect paths. The tests for differences in proportions in the last column report the estimates in percentages and the associated p -value from χ^2 tests. Significance levels are indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

detected, proportionally, slightly more VPM pages and fewer hybrid documents.

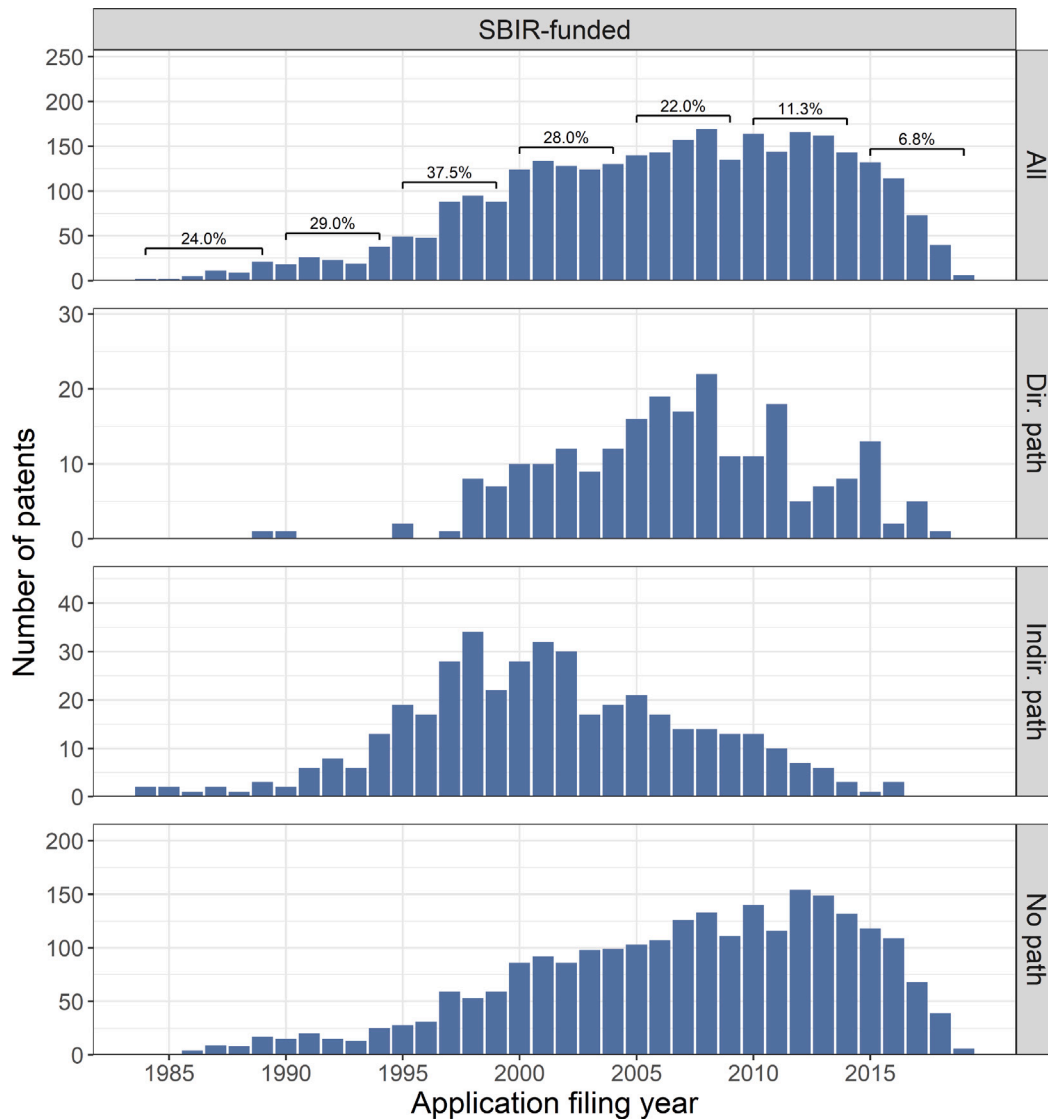


Fig. 4. Distribution of SBIR-funded patents by patent application year.

Notes. The figure distinguishes between patents with a direct commercialization path, patents with an indirect commercialization path, and patents for which we did not find any commercialization trace. The percentages reported in the top panel represent the fraction of patents protecting products over the total number of SBIR-funded patents in each five-year group. Note that a patent linked both directly and indirectly to a relevant page is counted only as a direct path.

Turning now to commercialization events, in Table 2, we find evidence of direct commercialization for about six percent of the benchmark patents and indirect commercialization for 15.1 percent of them. All in all, 18.5 percent of the benchmark patents appear on relevant web pages, either directly or indirectly. These proportions are statistically significantly lower than for SBIR-funded patents, as reported in the last column.

Table 3 presents descriptive statistics for the groups of SBIR-funded and benchmark patents, focusing on the following five dimensions: the number of independent claims in the patent (*claims*); the number of citations made to other patents (*bwd_cit*) and the non-patent literature (*npl_cit*); the number of citations received by the patent in the first three years after its application date (*fwd_cit*); and its geographic family size, namely, the number of countries in which patent protection is sought (*geo_fam*).

Overall, the table suggests that SBIR-funded patents have more independent claims than benchmark patents, make fewer citations to prior patent literature but rely more on the non-patent literature, and

Table 3
Summary statistics of key patent-level covariates.

Variable	Acronym	SBIR-funded		Benchmark patents		Diff. in means
		Mean	Std dev.	Mean	Std dev.	
Independent claims	<i>claims</i>	3.07	2.24	2.87	2.11	0.195***
Backward citations	<i>bwd_cit</i>	20.0	33.2	23.7	51.2	-3.74***
NPL citations	<i>npl_cit</i>	12.5	33.3	11.4	50.2	1.12*
Forward cit. (3 years)	<i>fwd_cit</i>	2.04	5.55	1.95	5.43	0.0889
Geographic family	<i>geo_fam</i>	1.91	2.08	2.28	2.51	-0.365***

Notes. The tests for differences in means between the two groups report the estimated mean difference and the *p*-value of paired *t*-tests: * *p* < 0.10, ** *p* < 0.05, *** *p* < 0.010.

are extended in fewer countries, with the difference being statistically significant at the 1 percent probability threshold. The mean difference

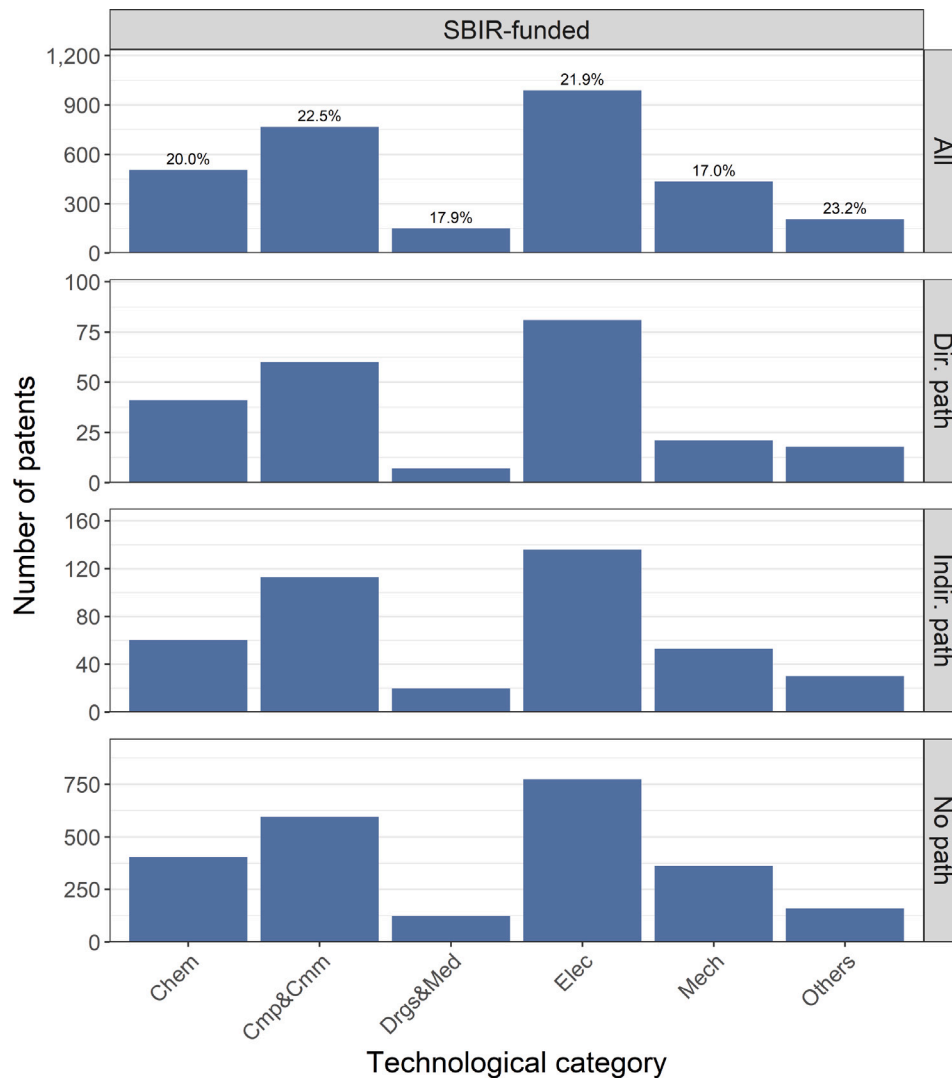


Fig. 5. Distribution of SBIR-funded patents by NBER technological category.

Notes. ‘Chem’: Chemical; ‘Cmp&Cmm’: Computers & Communications; ‘Drgs&Med’: Drugs & Medical; ‘Elec’: Electrical & Electronic; ‘Mech’: Mechanical. The percentage reported represents the fraction of product-protecting patents over the total number of SBIR-funded patents in each technological category. A patent linked both directly and indirectly to a relevant page is counted only as a direct path. For this figure, we also dropped thirteen observations with an unknown USPC class (0.4 percent) and mapped the following USPC patent classes, not included in Hall et al. (2001), into NBER classes as follows: 364 and 506 into Chem; 398 into Mech; 371, 703, 715, 717, 718, 719, 725, and 726 into Cmp&Cmm; 716 into Elec; and 850 into Drgs&Med. The patents classified into one of these USPC classes account for 4.2 percent of all patents.

in the number of forward citations does not appear to be statistically significantly different from zero.

5.2. Regression model

We estimate the following linear probability model (LPM) to assess differences in the commercialization probability between SBIR-funded and benchmark patents:

$$\Pi_i = \beta_0 + \beta_1 \cdot \text{SBIR}_i + \mathbf{X}_i \cdot \beta + \gamma_i + \delta_i + \varepsilon_i \quad (1)$$

The outcome variable Π_i takes the value 1 if patent i is commercialized, and 0 otherwise. We construct three different versions of Π_i , based on the commercialization path: direct, indirect, or any of the two paths. The variable of interest is SBIR_i , which takes the value 1 if patent i acknowledges funding from the DoD SBIR program, and 0 otherwise. The vector \mathbf{X}_i includes patent-level control variables that might correlate with the commercialization outcome, as listed in Table 3. Lastly, the model includes fixed effects for the patent i 's priority year, γ_i , and

USPC patent class, δ_i , to control for time- and technology-dependent factors.¹²

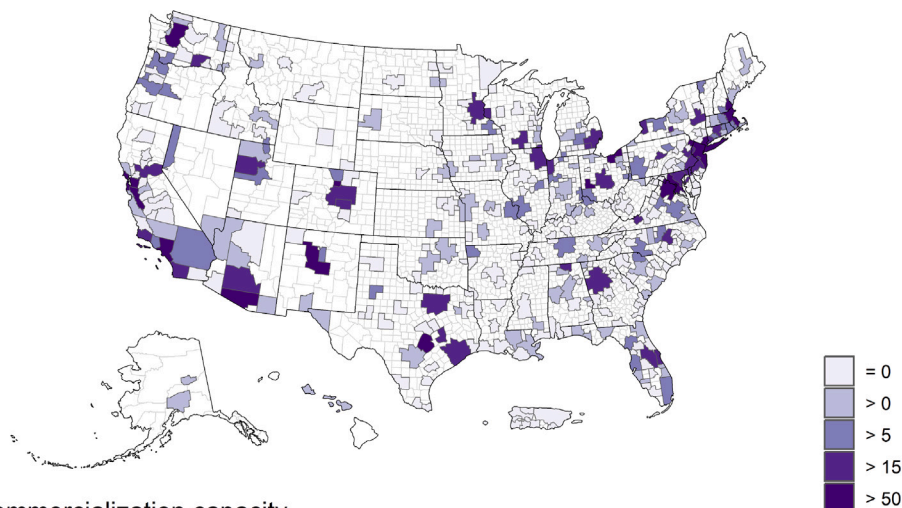
In addition to the baseline regression model specified in Eq. (1), we exploit the contract-level information to analyze whether specific characteristics of an SBIR contract disproportionately correlate with the probability of commercialization of the inventions arising from that contract. In particular, we focus on the stage of the R&D work procured by DoD (basic, applied, or development research stage) and on the phase of the contract (Phase I or Phase II).

5.3. Econometric results

Fig. 7 provides a visual overview of the regression coefficient β_1 . It reports the point estimate and the 95-percent confidence interval of the SBIR predictor for 18 LPM regressions. (Appendix Section C reports

¹² The observations have been weighted so that the weights assigned to the benchmark patents with a given USPC patent class and application year sum to the number of the SBIR-funded patents they are linked with. Appendix Section C reports additional details.

Number of SBIR-funded patents



Commercialization capacity

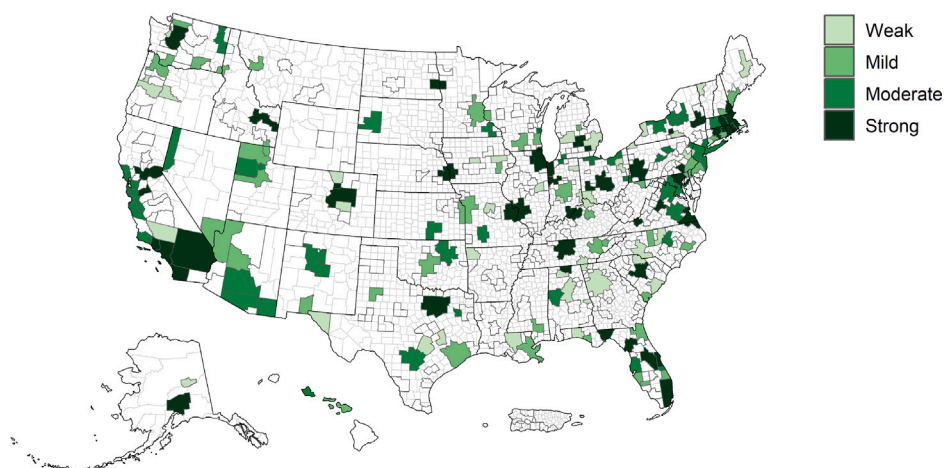


Fig. 6. Spatial distribution of our data by U.S. Metropolitan Statistical Area.

Notes. Spatial distribution of SBIR-funded patents (*top panel*) and commercialization capacity (*bottom panel*). The commercialization capacity measures an MSA's ability to commercialize SBIR-funded patents. It is defined as $CC_c = (CP_c / FP_c) / (\sum_{i=1}^C CP_i / \sum_{i=1}^C FP_i)$, where CP_c is the number of patent-to-product paths and FP_c is the number of SBIR-funded patents in MSA c . In the maps, non-metropolitan counties and micropolitan areas are colored in white, and each patent has been assigned to an MSA considering the area where the majority of its inventors reside (we choose at random if two or more MSAs were equally likely). Less than 1.5 percent of the SBIR-funded patents in our data are not attached to an MSA.

the regression tables as well as the results for Probit regression models.) The top part of the figure depicts the results of the baseline regression model for the three outcome variables. As regression results (1a)–(1c) show, an invention created with the support of a DoD-SBIR contract exhibits a higher likelihood of commercialization than a benchmark invention. The effect appears to be sizable: SBIR support is associated with a 17-percent increase in the probability of a commercial product introduction (any path).¹³ We find a similar effect if we consider only direct (1b) or indirect paths (1c).

Remember that we observe an indirect path when a patented invention connected to a product cites one of the focal patents. The channel through which such an association arises remains subject to speculation. However, a careful look at the data reveals that for about 40 percent of the patents that are linked to a product indirectly, the connecting citation is a self-citation, *i.e.*, it comes from a patent

¹³ The point estimate reported in result (1a) suggests a 3.52 percentage point increase in the likelihood of commercialization. The average patent in our regression sample has a probability of 20.48 percent to be linked, directly or indirectly, to a commercial product, leading to a 17.2 percent increase in the probability of commercialization.

applied for by the same assignee as the focal patent. Accordingly, we run the baseline model for the indirect path on two distinct sets of focal patents: patents that did receive at least one self-citation from a subsequent patent and patents that did not receive any self-citation. Interestingly, the association between SBIR support and commercialization disappears — and even turns negative — when we consider patents with no ensuing self-citations. By contrast, the results are in line with the baseline model (1c) when we consider patents with self-citations exclusively, with a 3.8-percentage-point higher probability of commercialization for SBIR-supported patents (see Table C.11 in the Appendix for an in-depth reporting of this analysis). This finding suggests that the long-term, indirect association with commercialization is observed only if the company that received SBIR support is actively involved with further technological developments—and, therefore, if the indirect path is closely connected to the SBIR funding. This finding is consistent with an “input additionality” effect, in line with existing literature documenting the presence of spillovers generated by the Department of Energy SBIR program (Myers and Lanahan, 2022).

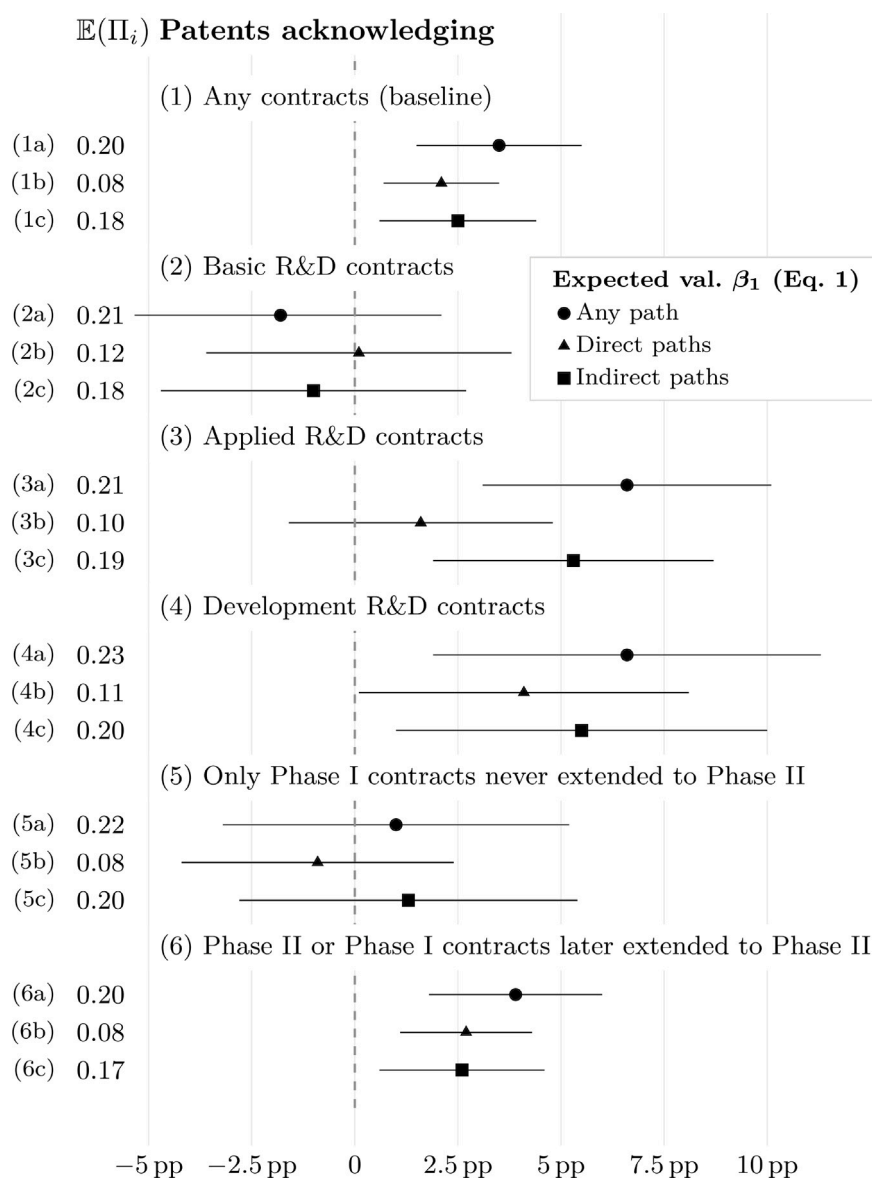


Fig. 7. SBIR program characteristics and the commercialization likelihood of patents.

Notes. Point estimates of the coefficient β_1 with corresponding 95-percent confidence intervals. The econometric method is a linear probability model. Figures on the left report the average value of the dependent variable for each model. Some patents have been zero-weighted in some models except (1a)–(1c). Zero-weighting occurs for patents that are not associated with the contract characteristic under consideration. Moreover, since a patent can acknowledge more than one contract, the three R&D stages or the two SBIR phases are not mutually exclusive.

Role of contract characteristics

The baseline results suggest a strong and positive association between SBIR funding and commercialization outcomes. To better understand the nature of this relationship, we evaluate the role of specific contract characteristics. We start by considering the stage of the R&D work characterizing the award. To do so, we split the sample of SBIR-funded patents into three subsamples — basic, applied, or development R&D — based on the features of the contract connected to each invention. We then couple each of the patents in these subsamples with its respective benchmark patents and run the baseline model on each subsample separately.

Fig. 7 reports the summary results of these regressions for the three outcome variables. Turning to patents connected to basic R&D contracts, the effect of SBIR support on direct or indirect commercialization outcomes appears to be null (models (2a)–(2c)). Regarding applied R&D contracts, SBIR support correlates with an increased commercialization likelihood, model (3a). This result seems to be driven by the indirect paths, model (3c), where SBIR-supported inventions

have a 5.3-percent higher likelihood of being indirectly connected to a product. In contrast, we observe no effect associated with direct paths, see model (3b). This result has some intuitive appeal, for applied R&D contracts are presumably still too far from commercialization, and the underlying inventions require follow-on development by the firm. Looking at patents connected to development R&D contracts, the data show a strong positive association for both direct and indirect paths to commercial products (models (4a)–(4c)). Overall, the results of this split sample analysis suggest that SBIR funding correlates more strongly with commercialization events for more downstream R&D stages.

Another key characteristic of SBIR contracts is their *phase*.¹⁴ As discussed above, Phase I projects can receive Phase II funding based on the results achieved in Phase I. The second phase allows the recipient to develop further the ideas and technologies generated during the initial

¹⁴ See <https://www.sbir.gov/about/policies> (last accessed February 21, 2024) for a thorough discussion.

phase. Therefore, by design, Phase II projects are closer to commercialization. In addition, the bulk of the funding that successful applicants receive arrives in Phase II, where the award size is an order of magnitude larger than in Phase I. If the SBIR program is, indeed, effective at spurring commercialization, we would expect a stronger commercialization likelihood for Phase II projects. The results of models (5a)–(5c) and (6a)–(6c) in Fig. 7 contrast the impact of the two phases and confirm this intuition. Focusing on Phase I projects that never reached Phase II in models (5a)–(5c), the difference between the SBIR-funded and the benchmark group is never statistically significantly different from zero. By contrast, the commercialization likelihood is markedly higher for patents linked to projects that obtained Phase II funding. Phase I projects are awarded to assess both the capacity of an SME to perform R&D and the quality of an innovative idea; therefore, the likelihood for an invention generated by a Phase I project to reach the commercialization stage is not particularly higher than for a “comparable” but privately-funded invention. The effect we observe for Phase II projects may stem from the DoD agencies’ accumulated informational capital, enabling superior selection of projects with high commercialization potential. Alternatively, it may reflect the impact of increased financial resources allocated at the Phase II stage.

Exploiting changes in the SBIR program

To shed more light on the link between the program stage and commercialization, we exploit a policy change in the design of SBIR that puts a greater focus on commercialization. With the Small Business Reauthorization Act of 2000 (§110), the U.S. Congress (2000-12-21) demanded the Small Business Administration “to provide for the requirement of a succinct commercialization plan with each application for a Phase II award that is moving toward commercialization.” Furthermore, and specifically for the DoD, the Act also introduced the Phase II Enhancement policy — also known as Phase II Plus — to encourage further the transition of SBIR research into DoD acquisition programs as well as the private sector (National Research Council, 2009b). Under this policy, a Phase II recipient can receive additional SBIR funds matching private or public financing the company obtains from non-SBIR sources. Both these changes affected the implementation of Phase II, but not Phase I, projects and provided additional emphasis on the commercialization goals of the program. These adjustments likely had a limited impact on the technical merit of the projects selected for Phase II. We exploit the latter fact to provide tentative evidence on whether the results we observe stem from a pure selection effect (*i.e.*, DoD agencies simply selecting the projects with the highest commercialization potential) or from the support (including financial) and the explicit push towards commercialization offered by the program.

We adopt a difference-in-differences (DiD) approach and focus on SBIR-funded patents awarded in the years immediately before and after this policy change (1996–2005). More specifically, we assess whether Phase II-related patents connected to SBIR awards signed after the year 2000 have a higher chance of directly linking to a commercial product than Phase II patents connected to pre-2000 contracts, using Phase I-related patents as the benchmark group. If the results were entirely driven by selection — *i.e.*, the agencies select the most promising projects in terms of commercialization outcomes *ex-ante* — we should not observe any effect of the policy change on the commercialization likelihood.

Table 4 reports the results of the DiD analysis. As the table shows, our main variable of interest, the interaction term Phase II \times Post 2000, is positive and significant at the 10 percent level.¹⁵ In other words, it seems that the additional push towards commercialization introduced in the year 2000 correlates with a higher commercialization propensity of the average Phase II-related patent.

¹⁵ Note that these results are obtained using a smaller sample limited to SBIR-funded patents, which may help explain the weak statistical significance of the results.

Table 4
Results of the policy-change regression.

Dep. var.:	OLS		Probit	
	(1)	(2)	(3)	(4)
<i>Direct path</i>				
Phase II	0.042 (0.031)	−0.040 (0.053)	0.041 (0.035)	−0.041 (0.051)
Post 2000	0.054* (0.030)	−0.054 (0.066)	0.060** (0.028)	−0.048 (0.067)
Phase II \times Post 2000		0.128* (0.069)		0.128* (0.071)
log(claims)	0.006 (0.021)	0.005 (0.020)	0.004 (0.019)	0.004 (0.019)
log(bwd_cit)	0.005 (0.014)	0.006 (0.014)	0.004 (0.013)	0.005 (0.013)
log(np1_cit)	0.019 (0.012)	0.020* (0.012)	0.018* (0.011)	0.019* (0.011)
log(geo_fam)	0.007 (0.023)	0.008 (0.023)	0.004 (0.020)	0.005 (0.020)
log(fwd_cit)	0.038** (0.016)	0.036** (0.016)	0.035** (0.014)	0.033** (0.014)
Constant	0.313 (0.324)	0.387 (0.319)		
Observations	809	809	809	809
R ²	0.134	0.138		
Pseudo R ²			0.137	0.140

Notes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$. Average value of the dependent variable: $\mathbb{E}(I_i) = 0.17$, and $\mathbb{E}(I_i | \text{Award pre 2000}) = 0.14$. Only SBIR-funded patents, funded by contracts signed in 1996–2005, included. Phase II contracts also include Phase I ones later extended to the second phase of the SBIR program. For the extended contracts, we considered the extending contract date. All the models include fixed effects for the patent’s USPC patent class to control for technology-dependent factors.

Overall, using our web-based method, we find that SBIR-funded patented inventions are significantly more likely to transition into commercial products than a set of benchmark inventions developed by the private sector without government support. These results, consistent with prior literature linking the SBIR program to successful commercialization outcomes, highlight the relevance and practicality of our method.

6. Concluding remarks

This paper introduces a novel method for quantifying the likelihood of commercialization for patented inventions by systematically searching the web for traces of market presence. We demonstrate the method’s applicability using the DoD’s SBIR program as a case study, leveraging readily available data on SBIR-funded patents from participating agencies.

To assess the relevance of the method, we run several regression models and find that the data respond in line with expectations. In particular, we find that the commercialization rate is higher for applied or development R&D contracts than for basic R&D contracts, and for Phase II than for Phase I contracts. These results suggest that the method captures a meaningful signal of commercialization.

We also compare the commercialization probability of SBIR-funded patents with that of privately funded but otherwise broadly similar patents. We find that SBIR-funded inventions are 17 percent more likely to be commercialized, with an overall commercialization rate of 21.5 percent. While we are cautious not to interpret these results causally, they add to the body of evidence suggesting that the SBIR program is effective at stimulating the commercialization of federally funded scientific discoveries. Its overall commercialization rate seems relatively high in light of common wisdom that the majority of U.S. patents are “worthless” (Moore, 2005; Lemley and Shapiro, 2005; Sichelman, 2009). Another notable finding relates to the importance of the indirect path to commercialization, suggesting that the social benefits of the DoD’s SBIR program may extend well beyond the supported inventions, adding to the findings by Myers and Lanahan (2022).

Despite its potential impact, the method has some limitations. First and foremost, it applies only to patented inventions. Since not all inventions are patentable, and not all patentable inventions are patented, it is best suited for fields where patents are a common means of appropriation (Cohen et al., 2000). Second, for efficiency reasons, we have prioritized a targeted web crawl over a broader search. As a result, the method may miss patents for which ownership is not accurately recorded in USPTO data. This limitation can be mitigated by incorporating data on the ownership structures of patent assignees (thereby also accounting for M&A activity) and supplementing the targeted approach with large-scale web crawls as offered by the IPProduct project. Finally, scholars applying our method should carefully consider the specific features of the policy instrument under study. Commercialization pathways, patenting incentives, and disclosure practices may vary across agencies, funding instruments, and the starting point used to track patents. Our starting point is patents with government interest statements related to federal contracts, but alternatives exist, such as patents linked to publicly funded research papers, patent numbers obtained from surveys, or patents filed by universities and public research organizations. Each approach has its own nuances, making contextual awareness essential.

In our context, studying the commercialization of DoD-funded technologies solely through the lens of patented inventions provides an incomplete picture, as not all commercialized technologies are patented. Some may be unsuitable for patent protection, particularly niche inventions within the DoD technology transition pipeline. Additionally, certain research contracts may result in classified technologies subject to secrecy orders (Gross, 2023; de Rassenfosse et al., 2024). Moreover, when the DoD is the sole buyer, SBIR recipients may have less incentive to report commercialization activity online. We encourage readers to keep these considerations in mind when interpreting our results and call for more analyses of these important questions.

Overall, we believe that our method adds a valuable tool to researchers' toolboxes, complementing other commercialization signals, such as follow-on contracts and venture capital funding, or data obtained from surveys. We hope this method will encourage broader adoption of web-based techniques for measuring commercialization in innovation research.

CRediT authorship contribution statement

Carlo Bottai: Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Gaétan de Rassenfosse:** Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Emilio Raiteri:** Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the EuroTech Universities Alliance for sponsoring this work. C.B. was supported by the European Union's Marie Skłodowska-Curie program for the project *Insights on the "Real Impact" of Science* (H2020 MSCA-COFUND-2016 Action, Grant Agreement No 754462). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are grateful to Scott Stern and three anonymous reviewers for their valuable feedback.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.respol.2025.105302>.

Data availability

The data are available at <https://doi.org/10.5281/zenodo.16779955> under a Creative Commons Attribution 4.0 International license. The web scraping component of the project relies on websites' content, which frequently changes over time. Therefore, reproducing our exact results may not always be possible. Nonetheless, we provide the original scraping and classification code at the following repositories: <https://github.com/n3ssuno/iris-scraper/releases/tag/v0.9-rc> and <https://github.com/n3ssuno/iris-classifier/releases/tag/v0.9-rc>.

References

- Aghion, P., Howitt, P., 1998. *Endogenous Growth Theory*. MIT Press.
- Arora, S.K., Kelley, S., Madhavan, S., 2021. Building a sample frame of SMEs using patent, search engine, and website data. *J. Off. Stat.* 37 (1), 1–30. <http://dx.doi.org/10.2478/jos-2021-0001>.
- Arora, S.K., Youtie, J., Shapira, P., Gao, L., Ma, T., 2013. Entry strategies in an emerging technology: A pilot web-based study of graphene firms. *Scientometrics* 95 (3), 1189–1207. <http://dx.doi.org/10.1007/s11192-013-0950-7>.
- Ashouri, S., Suominen, A., Hajikhani, A., Pukelis, L., Schubert, T., Türkeli, S., Van Beers, C., Cunningham, S., 2022. Indicators on firm level innovation activities from web scraped data. *Data Brief* 42, 108246. <http://dx.doi.org/10.1016/j.dib.2022.108246>.
- Audretsch, D.B., Link, A.N., Scott, J.T., 2002. Public/private technology partnerships: Evaluating SBIR-supported research. *Res. Policy* 31 (1), 145–158. [http://dx.doi.org/10.1016/s0048-7333\(00\)00158-x](http://dx.doi.org/10.1016/s0048-7333(00)00158-x).
- Axenbeck, J., Breithaupt, P., 2021. Innovation indicators based on firm websites—Which website characteristics predict firm-level innovation activity? *PLOS ONE* 16 (4), 1–23. <http://dx.doi.org/10.1371/journal.pone.0249583>.
- Azoulay, P., Graff Zivin, J.S., Li, D., Sampat, B.N., 2019. Public R&D investments and private-sector patenting: evidence from NIH funding rules. *Rev. Econ. Stud.* 86 (1), 117–152. <http://dx.doi.org/10.1093/restud/rdy034>.
- Braunerhjelm, P., Svensson, R., 2024. Inventions, commercialization strategies, and knowledge spillovers in SMEs. *Small Bus. Econ.* 63 (1), 275–297. <http://dx.doi.org/10.1007/s11187-023-00812-z>.
- Bronzini, R., Piselli, P., 2016. The impact of R&D subsidies on firm innovation. *Res. Policy* 45 (2), 442–457. <http://dx.doi.org/10.1016/j.respol.2015.10.008>.
- Cin, B.C., Kim, Y.J., Vonortas, N.S., 2017. The impact of public R&D subsidy on small firm productivity: evidence from Korean SMEs. *Small Bus. Econ.* 48, 345–360. <http://dx.doi.org/10.1007/s11187-016-9786-x>.
- Cohen, W.M., Nelson, R., Walsh, J., 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). <http://dx.doi.org/10.3386/w7552>, NBER Working Paper.
- Czarnitzki, D., Hussinger, K., 2018. Input and output additionality of R&D subsidies. *Appl. Econ.* 50 (12), 1324–1341. <http://dx.doi.org/10.1080/00036846.2017.1361010>.
- Daas, P.J.H., van der Doef, S., 2020. Detecting innovative companies via their website. *Stat. J. the IAOS* 36 (4), 1239–1251. <http://dx.doi.org/10.3233/SJI-200627>.
- Danguy, J., de Rassenfosse, G., van Pottelsberghe de la Potterie, B., 2014. On the origins of the worldwide surge in patenting: an industry perspective on the R&D–patent relationship. *Ind. Corp. Chang.* 23 (2), 535–572. <http://dx.doi.org/10.1093/icc/dtt042>.
- de Rassenfosse, G., 2018. Notice failure revisited: Evidence on the use of virtual patent marking. Working Paper Series 24288, National Bureau of Economic Research, <http://dx.doi.org/10.3386/w24288>.
- de Rassenfosse, G., Jaffe, A.B., Raiteri, E., 2019. The procurement of innovation by the U.S. government. *PLOS ONE* 14 (8), e0218927. <http://dx.doi.org/10.1371/journal.pone.0218927>.
- de Rassenfosse, G., Pellegrino, G., Raiteri, E., 2024. Do patents enable disclosure? Evidence from the invention secrecy act. *Int. J. Ind. Organ.* 92, 103044. <http://dx.doi.org/10.1016/j.ijindorg.2023.103044>.
- de Rassenfosse, G., Zhou, L., 2020. Patents and supra-competitive prices: Evidence from consumer products. Available At SSRN 3756359.
- Devarakonda, S.V., Goossen, M.C., Mulotte, L., 2024. The allocation of resource control within the corporate structure: Evidence from post-acquisition patent reassignments. *Strat. Manag. J.* <http://dx.doi.org/10.1002/smj.3682>.
- Dimos, C., Pugh, G., 2016. The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature. *Res. Policy* 45 (4), 797–815. <http://dx.doi.org/10.1016/j.respol.2016.01.002>.

- Domènech, J., de la Ossa, B., Pont, A., Gil, J.A., Martínez, M., Rubio, A., 2012. An intelligent system for retrieving economic information from corporate websites. In: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, vol. 1, IEEE Computer Society, Washington (DC, USA), pp. 573–578. <http://dx.doi.org/10.1109/WI-IAT.2012.92>.
- Dutta, S., Folta, T.B., Rodrigues, J., 2022. Do governments fund the best entrepreneurial ventures? The case of the small business innovation research program. *Acad. Manag. Discov.* 8 (1), 103–138. <http://dx.doi.org/10.5465/amd.2019.0078>.
- Edelman, B., 2012. Using Internet data for economic research. *J. Econ. Perspect.* 26 (2), 189–206. <http://dx.doi.org/10.1257/jep.26.2.189>.
- Feldman, M.P., Johnson, E.E., Bellefleur, R., Dowden, S., Talukder, E., 2022. Evaluating the tail of the distribution: the economic contributions of frequently awarded government R&D recipients. *Res. Policy* 51 (7), 104539. <http://dx.doi.org/10.1016/j.respol.2022.104539>.
- Feldman, M.P., Kogler, D.F., 2010. Stylized facts in the geography of innovation. In: Hall, B.H., Rosenberg, N. (Eds.), *Handbook of the Economics of Innovation*, vol. 1, North-Holland, pp. 381–410. [http://dx.doi.org/10.1016/S0169-7218\(10\)01008-7](http://dx.doi.org/10.1016/S0169-7218(10)01008-7).
- García-Quevedo, J., 2004. Do public subsidies complement business R&D? A meta-analysis of the econometric evidence. *Kyklos* 57 (1), 87–102. <http://dx.doi.org/10.1111/j.0023-5962.2004.00244.x>.
- Gök, A., Waterworth, A., Shapira, P., 2015. Use of web mining in studying innovation. *Scientometrics* 102 (1), 653–671. <http://dx.doi.org/10.1007/s11192-014-1434-0>.
- Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. *Bell J. Econ.* 10 (1), 92–116. <http://dx.doi.org/10.2307/3003321>.
- Gross, D.P., 2023. The hidden costs of securing innovation: the manifold impacts of compulsory invention secrecy. *Manag. Sci.* 69 (4), 2318–2338. <http://dx.doi.org/10.1287/mnsc.2022.4457>.
- Grossman, G.M., Helpman, E., 1993. *Innovation and Growth in the Global Economy*. MIT Press.
- Guo, D., Guo, Y., Jiang, K., 2016. Government-subsidized R&D and firm innovation: Evidence from China. *Res. Policy* 45 (6), 1129–1144. <http://dx.doi.org/10.1016/j.respol.2016.03.002>.
- Guzman, J., Li, A., 2023. Measuring founding strategy. *Manag. Sci.* 69 (1), 101–118. <http://dx.doi.org/10.1287/mnsc.2022.4369>.
- Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. Working Paper 8498, National Bureau of Economic Research. <http://dx.doi.org/10.3386/w8498>.
- Higham, K., De Rassenfosse, G., Jaffe, A.B., 2021. Patent quality: Towards a systematic framework for analysis and measurement. *Res. Policy* 50 (4), 104215. <http://dx.doi.org/10.1016/j.respol.2021.104215>.
- Howell, S.T., 2017. Financing innovation: Evidence from R&D grants. *Am. Econ. Rev.* 107 (4), 1136–1164. <http://dx.doi.org/10.1257/aer.20150808>.
- Hussinger, K., 2008. R&D and subsidies at the firm level: An application of parametric and semiparametric two-step selection models. *J. Appl. Econometrics* 23 (6), 729–747. <http://dx.doi.org/10.1002/jae.1016>.
- Jaffe, A.B., de Rassenfosse, G., 2017. Patent citation data in social science research: Overview and best practices. *J. Assoc. Inf. Sci. Technol.* 68 (6), 1360–1374. <http://dx.doi.org/10.1002/asi.23731>.
- Jones, C., Madhavan, S., 2020. PatentsView: Government Interest Extraction and Processing – Version 2.0. Mimeo, American Institutes for Research, URL <https://patentsview.org/government-interest>.
- Karhunen, H., Huovari, J., 2015. R&D subsidies and productivity in SMEs. *Small Bus. Econ.* 45, 805–823. <http://dx.doi.org/10.1007/s11187-015-9658-9>.
- Kinne, J., Axenbeck, J., 2020. Web mining for innovation ecosystem mapping: A framework and a large-scale pilot study. *Scientometrics* 125 (3), 2011–2041. <http://dx.doi.org/10.1007/s11192-020-03726-9>.
- Kinne, J., Lenz, D., 2021. Predicting innovative firms using web mining and deep learning. *PLOS ONE* 16 (4), 1–18. <http://dx.doi.org/10.1371/journal.pone.0249071>.
- Lemley, M.A., Shapiro, C., 2005. Probabilistic patents. *J. Econ. Perspect.* 19 (2), 75–98. <http://dx.doi.org/10.1257/0895330054048650>.
- Li, M., Jin, M., Kumbhakar, S.C., 2022. Do subsidies increase firm productivity? Evidence from Chinese manufacturing enterprises. *European J. Oper. Res.* 303 (1), 388–400. <http://dx.doi.org/10.1016/j.ejor.2022.02.029>.
- Libaers, D., Hicks, D., Porter, A.L., 2016. A taxonomy of small firm technology commercialization. *Ind. Corp. Chang.* 25 (3), 371–405. <http://dx.doi.org/10.1093/icc/dtq039>.
- Link, A.N., Scott, J.T., 2010. Government as entrepreneur: Evaluating the commercialization success of SBIR projects. *Res. Policy* 39 (5), 589–601. <http://dx.doi.org/10.1016/j.respol.2010.02.006>.
- Mak, M.A., 2014. Small Business Innovation Research: DOD's Program Has Developed Some Technologies that Support Military Users, but Lacks Comprehensive Data on Transition Outcomes. Testimony Before the House Committee on Small Business GAO-14-748T, United States Government Accountability Office.
- Mazzucato, M., 2013. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. Anthem Press, London, UK.
- Mazzucato, M., 2021. *Mission Economy: A Moonshot Guide to Changing Capitalism*. Allen Lane, London, UK.
- Moore, K.A., 2005. Worthless patents. *Berkeley Technol. Law J.* 20, 1521.
- Myers, K.R., Lanahan, L., 2022. Estimating spillovers from publicly funded R&D: Evidence from the US department of energy. *Am. Econ. Rev.* 112 (7), 2393–2423. <http://dx.doi.org/10.1257/aer.20210678>.
- National Academies of Sciences, Engineering, and Medicine, 2020. Review of the SBIR and STTR Programs at the Department of Energy. The National Academies Press, Washington, DC, <http://dx.doi.org/10.17226/25674>.
- National Research Council, 2009a. An Assessment of the SBIR Program at the Department of Defense. The National Academies Press, Washington, DC, <http://dx.doi.org/10.17226/11963>.
- National Research Council, 2009b. Revisiting the Department of Defense SBIR Fast Track Initiative. The National Academies Press, Washington, DC, <http://dx.doi.org/10.17226/12600>.
- National Research Council, 2014. SBIR at the Department of Defense. The National Academies Press, Washington, DC, <http://dx.doi.org/10.17226/18821>.
- Pakes, A., Griliches, Z., 1980. Patents and R&D at the firm level: A first report. *Econom. Lett.* 5 (4), 377–381. [http://dx.doi.org/10.1016/0165-1765\(80\)90136-6](http://dx.doi.org/10.1016/0165-1765(80)90136-6).
- Prencipe, A., D'Amico, L., Boffa, D., Corsi, C., 2024. The effect of output additionality of public funding support on firm innovation. Evidence from firms of different sizes. *Int. J. Financ. Econ.* 29 (2), 2278–2299. <http://dx.doi.org/10.1002/ijfe.2766>.
- Radicic, D., Pugh, G., 2017. R&D programmes, policy mix, and the 'European paradox': Evidence from European SMEs. *Sci. Public Policy* 44 (4), 497–512. <http://dx.doi.org/10.1093/scipol/scw077>.
- Rai, A.K., Sampat, B.N., 2012. Accountability in patenting of federally funded research. *Nature Biotechnol.* 30 (10), 953–956. <http://dx.doi.org/10.1038/nbt.2382>.
- Rammer, C., Es-Sadki, N., 2023. Using big data for generating firm-level innovation indicators - a literature review. *Technol. Forecast. Soc. Change* 197, 122874. <http://dx.doi.org/10.1016/j.techfore.2023.122874>.
- Romer, P.M., 1990. Endogenous technological change. *J. Political Econ.* 98 (5, Part 2), S71–S102. <http://dx.doi.org/10.1086/261725>.
- Ruttan, V.W., 2006. *Is War Necessary for Economic Growth? Military Procurement and Technology Development*. Oxford University Press, Oxford, UK, <http://dx.doi.org/10.1093/0195188047.001.0001>.
- Scherer, F.M., Harhoff, D., 2000. Technology policy for a world of skew-distributed outcomes. *Res. Policy* 29 (4–5), 559–566. [http://dx.doi.org/10.1016/S0048-7333\(99\)00089-X](http://dx.doi.org/10.1016/S0048-7333(99)00089-X).
- Sichelman, T., 2009. Commercializing patents. *Stanf. Law Rev.* 62 (2), 341–411.
- Svensson, R., 2007. Commercialization of patents and external financing during the R&D phase. *Res. Policy* 36 (7), 1052–1069. <http://dx.doi.org/10.1016/j.respol.2007.04.004>.
- U.S. Congress, 2000-12-21. Small business reauthorization act of 2000. HR 5667. Pub. L. 106-554, Appendix I.