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

Carlo BOTTAI
Gaétan DE RASSENFOSSE
Emilio RAITERI

September 2025

Innovation and Intellectual Property Policy
Working Paper series no. 28

Available at: https://ideas.repec.org/p/iip/wpaper/28.html

EPFL Working Paper Series
STiP lab

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*

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 the commercialization of patented products, 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 3,070 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.

Keywords: government-funded research, invention commercialization, patents, policy evaluation, web-based evidence

*Corresponding author

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 3,070 granted patents with filing years ranging from 1984 to 2019 and assigned to 1,182 distinct companies. These patents acknowledge 2,213 different procurement contracts.

We identify traces of commercialization for 642 out of the 3,070 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.

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.

STI policy spans the full spectrum of STI activities, with measures targeting commercialization holding 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

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

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.

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, in-

[2]See `https://www.sbir.gov/`, last accessed December 4th, 2024.

volving 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 the resulting commercialized products. 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 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 contain relevant information providing evidence of 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

[3]Specifically, for an entity like 'MyBiz Corp.', we run the query `('MYBIZ 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.

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 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).

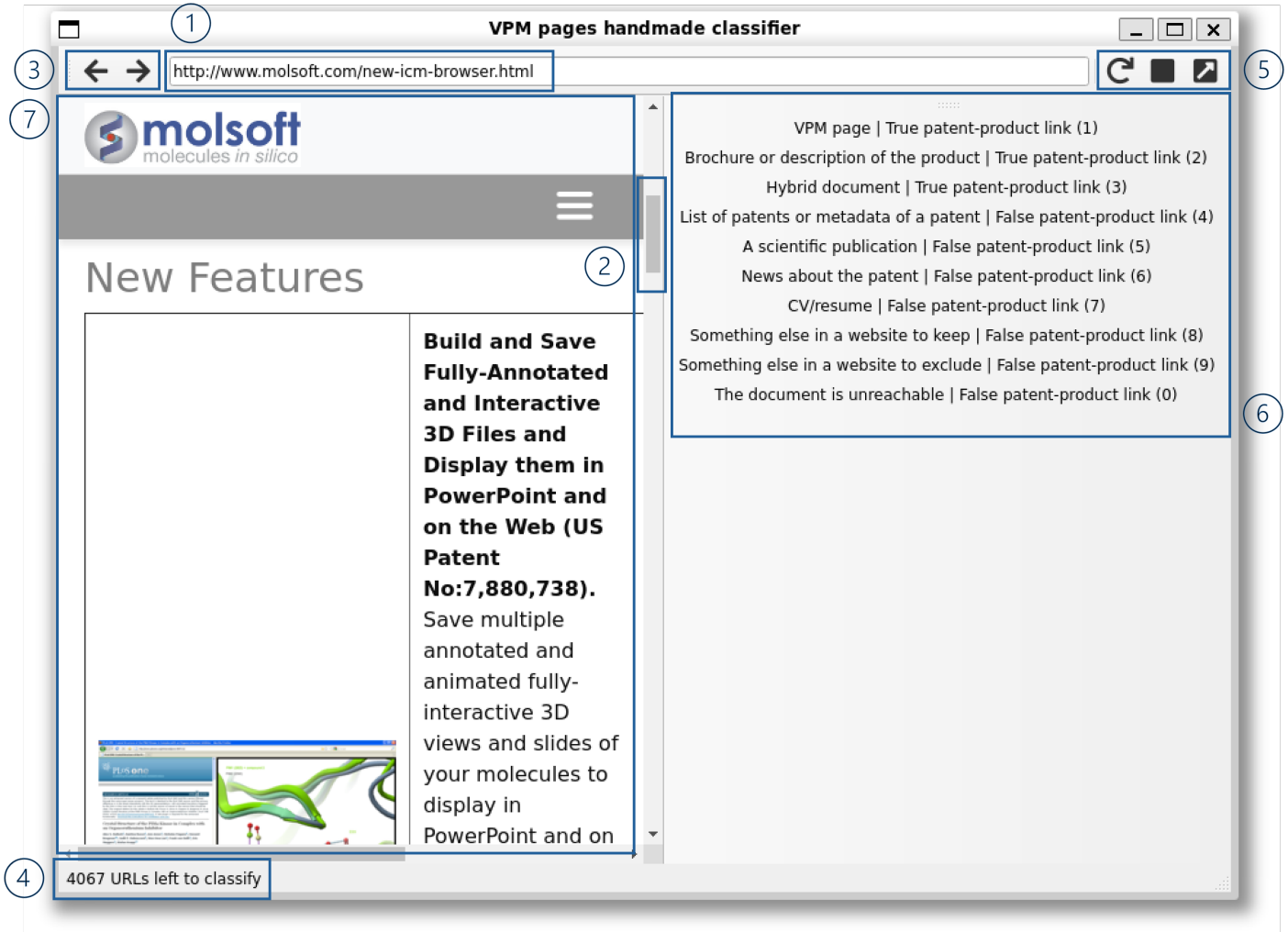
Step 3.2 concerns pages marked as 'uncertain' by the automatic classifier. We manually reviewed all of them via a browser-like interface to speed up processing. As Figure 1 illustrates, the interface displays the web page to classify, and the user can use the buttons in the right panel to classify it into several (positive or negative) commercialization outcome categories. Appendix A.2.2 describes the details of the classification process and the steps taken to ensure its accuracy.

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.

Figure 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.

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 a firm receives subsequent venture capital and has large, positive impacts on patenting and revenue." Feldman et al. (2022) examine SBIR recipients' commercial

[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`

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 3,070 patents linked to DoD SBIR awards.[8] Our approach

[8]Appendix A.2.1 provides a detailed explanation of the procedure adopted to extract the contract identifiers—the Procurement Instrument Identifiers (PIID). Data about the government in-

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 Rassenfosse 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 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. Figure 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.

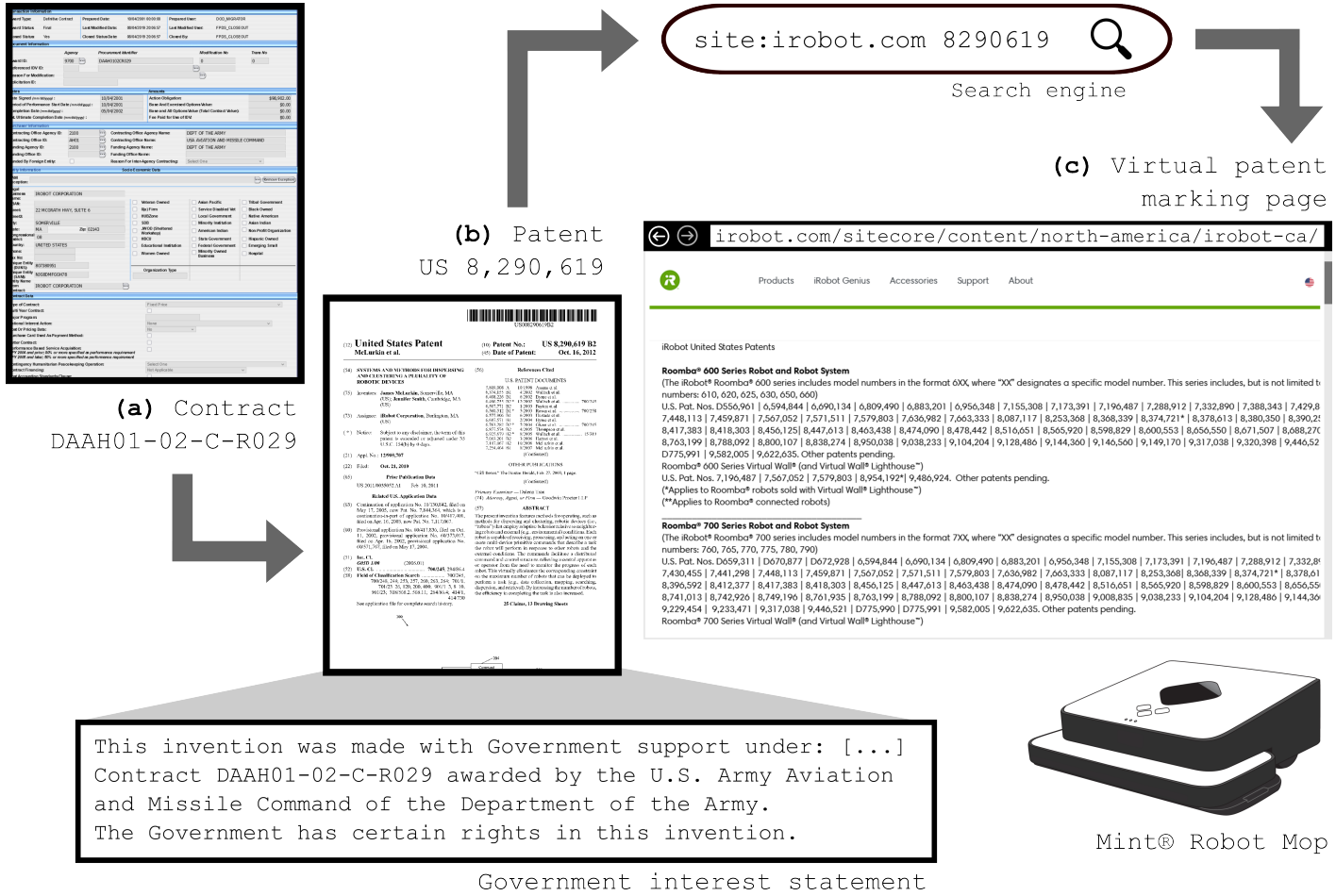
An indirect path occurs when the SBIR-funded patent is cited by a subsequent patent for which we found evidence of commercialization. Given the technical function of patent citations as signals of existing prior knowledge relevant to the new invention (Jaffe and de Rassenfosse, 2017), we also consider this second path as providing evidence of a successful commercialization event. Figure 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.

---

terest 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.

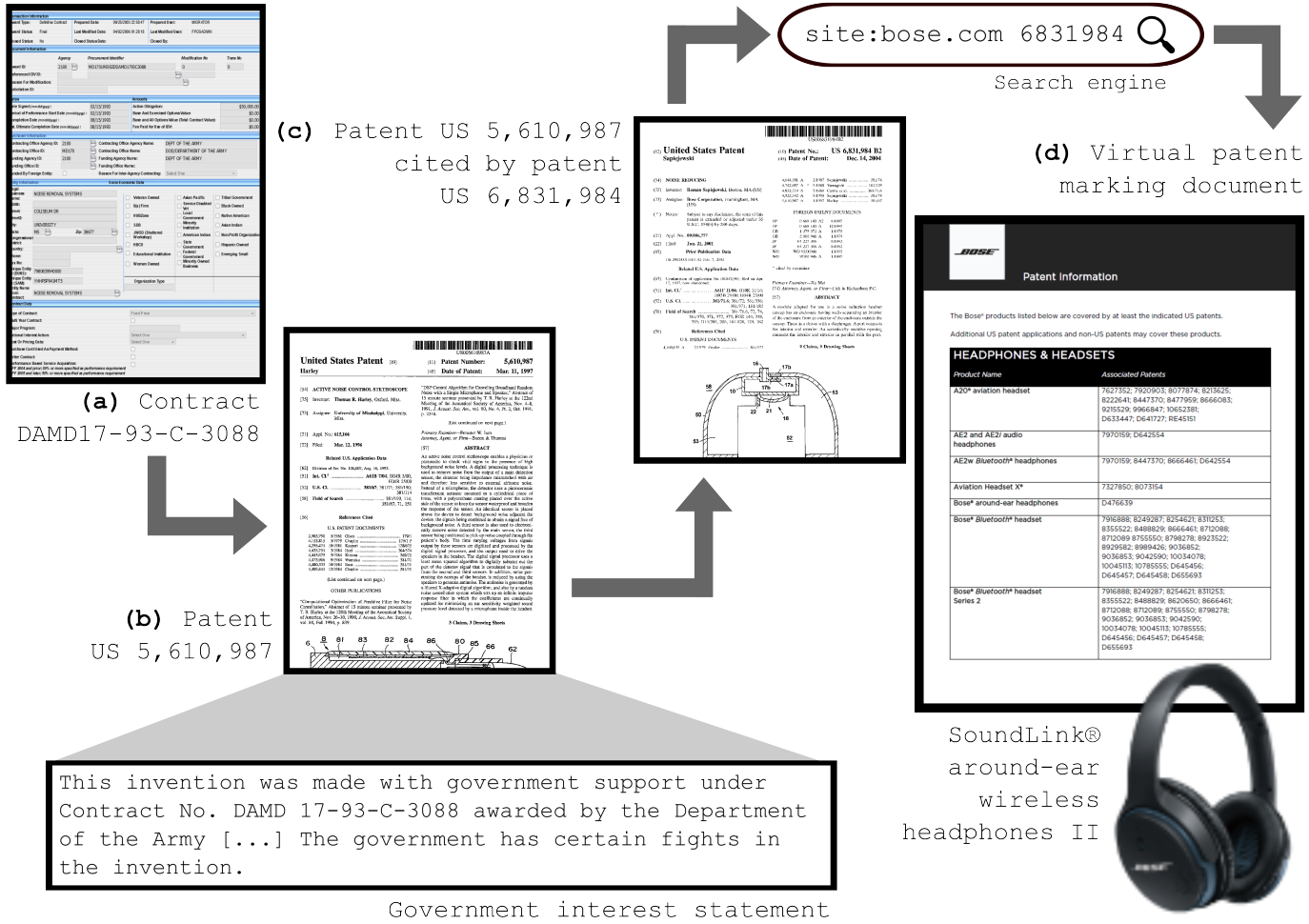
[9]More specifically, the fourth digit in the product and service code (PSC) identifies the stage of the R&D effort required by a given contract with: (1) Basic Research; (2) Applied Research and Exploratory Development; (3) Advanced Development; (4) Engineering Development; (5) Operational Systems Development;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

Figure 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.

Figure 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.

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 3,070 SBIR-funded patents (of which 2,304 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 6,647 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 9,411 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 3,131 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 semi-automatic 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 3,070 granted patents with filing years ranging from 1984 to 2019 and assigned to 1,182 distinct companies. These patents acknowledge 2,213 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 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 1,088 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, 1,582 patents (51.5%) acknowledge at least one Phase I contract and 1,252 patents acknowledge Phase I contracts exclusively. A total of 1,818 patents

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

(59.2%) acknowledge one or more Phase II contracts.[11]

Figure 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 was 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 Figure 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.

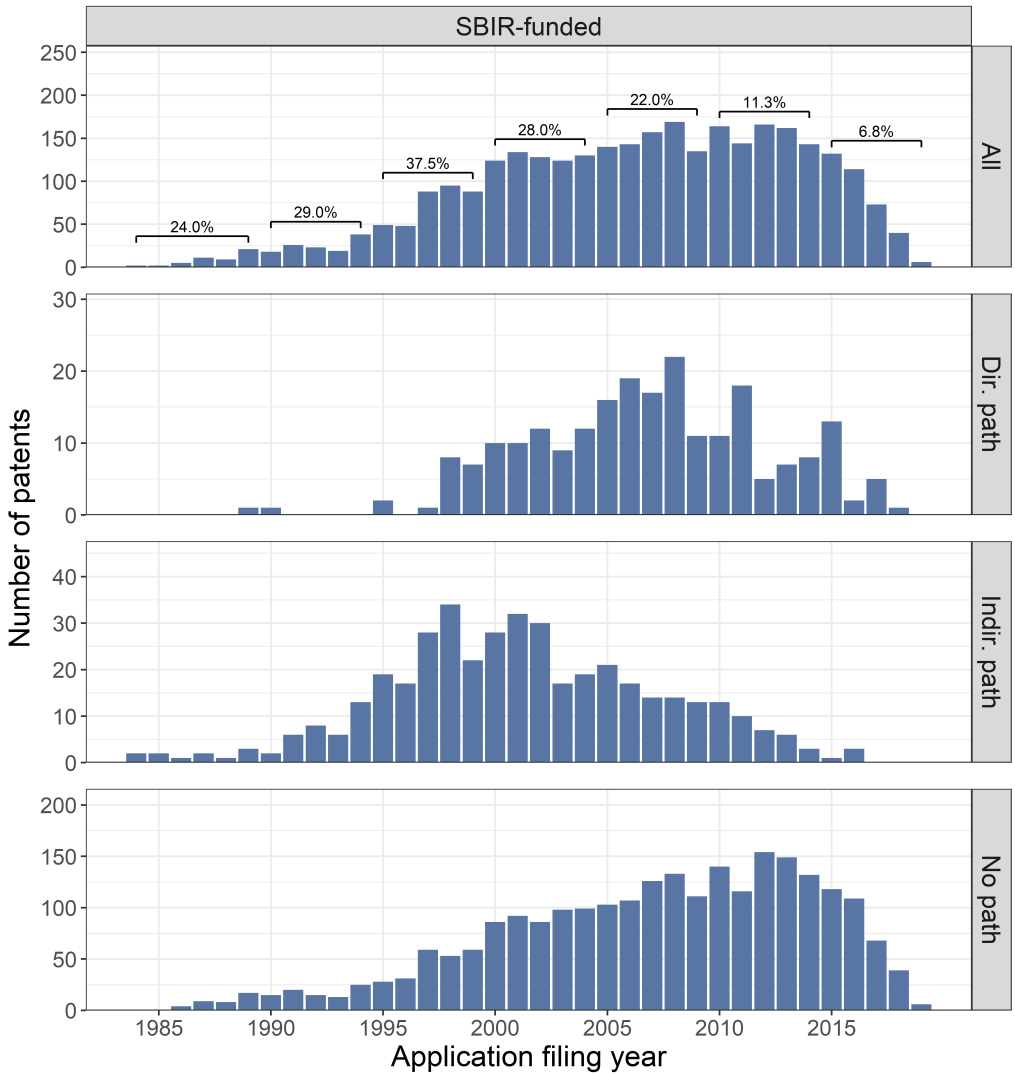
Lastly, turning to the spatial distribution of the data, the top panel of Figure 6 illustrates that SBIR-funded patents are unevenly concentrated in a few metropolitan areas (MSAs) around the United States. This observation is consistent with the geography of innovation literature (Feldman and Kogler, 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

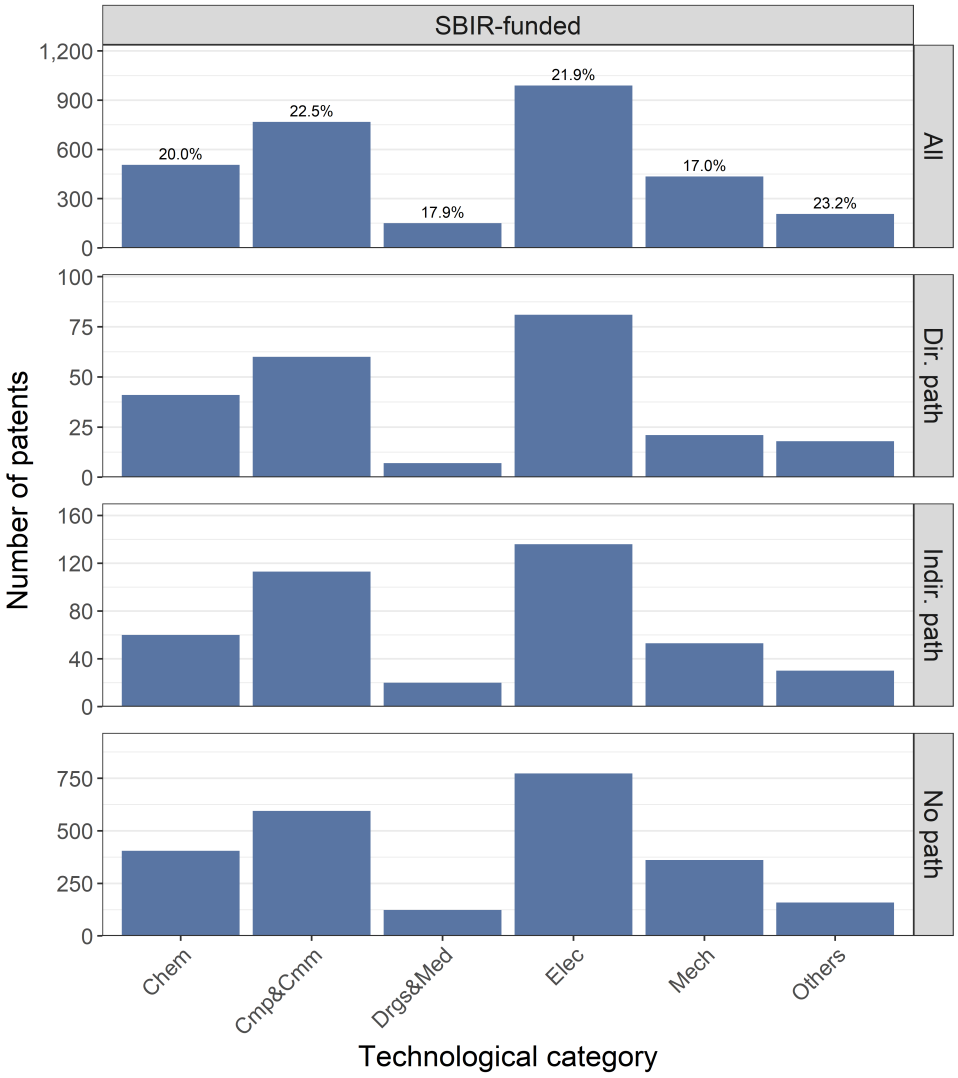
[11]For patents linked only to Phase I contracts, we also determine if the project never reached Phase II or if a Phase II contract exists, but the patent simply did not mention it (see Appendix A.2.1 for further explanation). Accounting for Phase I contracts later extended to a Phase II contract not acknowledged in the patent document, we find that 2,374 patents (82.0%) are connected to Phase II funding.

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



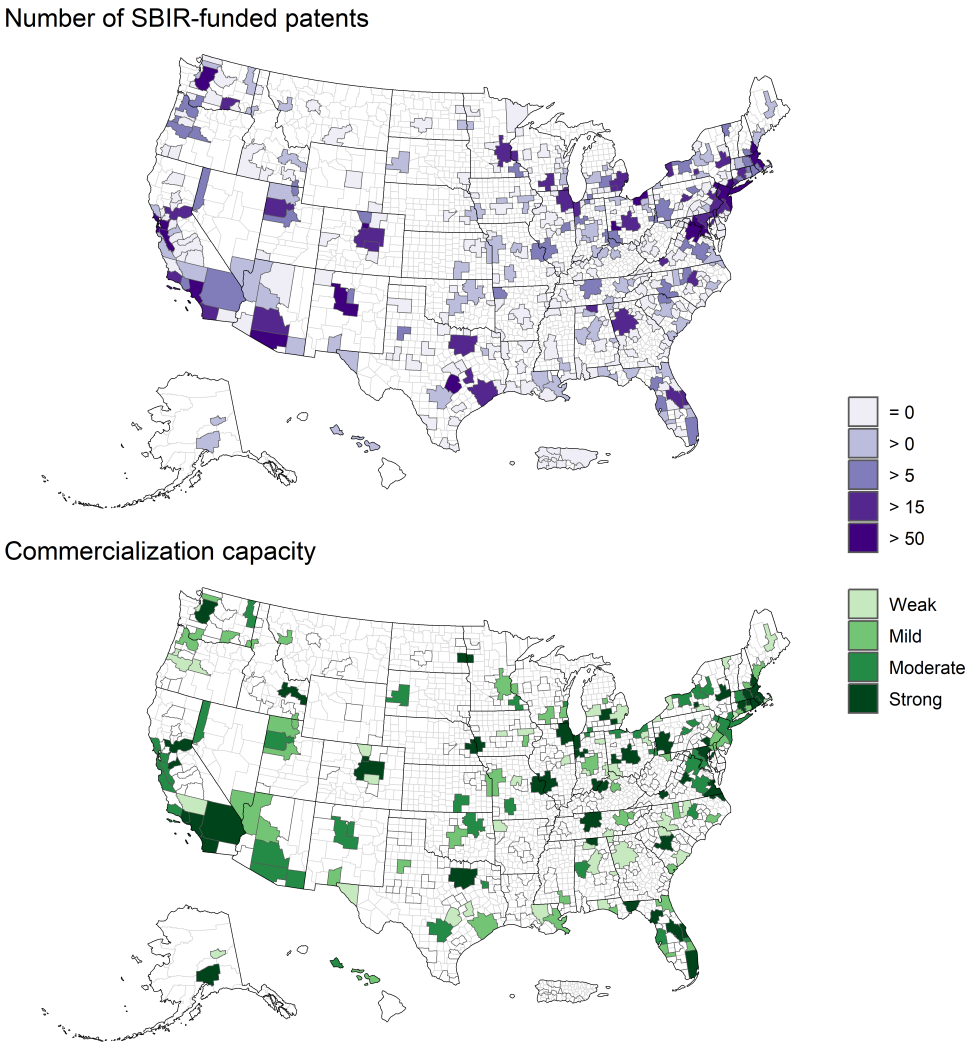
Notes. The figure distinguishes between patents with a direct commercialization path, patents with an indirect commercialization path, and patents for which we did not find any commercialization trace. The percentages reported in the top panel represent the fraction of patents protecting products over 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.

Figure 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.

Figure 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 $\text{CC}_c = (\text{CP}_c/\text{FP}_c)/(\sum_{i=1}^{C} \text{CP}_i / \sum_{i=1}^{C} \text{FP}_i)$, 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.

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 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 3,070 SBIR-funded patents and 4,828 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 2,896 SBIR-funded patents, assigned to 1,060 distinct companies, and 4,622 benchmark patents, assigned to 3,892 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 3,144 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 2,345 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 detected, proportionally, slightly more VPM pages and fewer hybrid documents.

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 1: Overview of relevant web pages.

	SBIR-funded		Benchmark patents	
	N	Percent.	N	Percent.
Automatically classified	579	25.7%	799	25.4%
Manually classified	1,676	74.3%	2,345	74.6%
VPM page	544	(24.1%)	620	(19.7%)
Brochure	928	(41.2%)	1,368	(43.5%)
Hybrid document	204	(9.1%)	357	(11.4%)
Relevant pages	2,255	100%	3,144	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	2,896		4,622		

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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 are extended in fewer countries, with the difference being statistically significant at the 1 percent probability threshold. The mean difference in the number of forward citations does not appear to be statistically significantly different from zero.

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	`claims`	3.07	2.24	2.87	2.11	0.195***
Backward citations	`bwd_cit`	20.0	33.2	23.7	51.2	-3.74***
NPL citations	`npl_cit`	12.5	33.3	11.4	50.2	1.12*
Forward cit. (3 years)	`fwd_cit`	2.04	5.55	1.95	5.43	0.0889
Geographic family	`geo_fam`	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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$.

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:

$$\Pi_i = \beta_0 + \beta_1 \cdot \text{SBIR}_i + \mathbf{X_i} \cdot \beta + \gamma_i + \delta_i + \varepsilon_i \tag{1}$$

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 equation (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

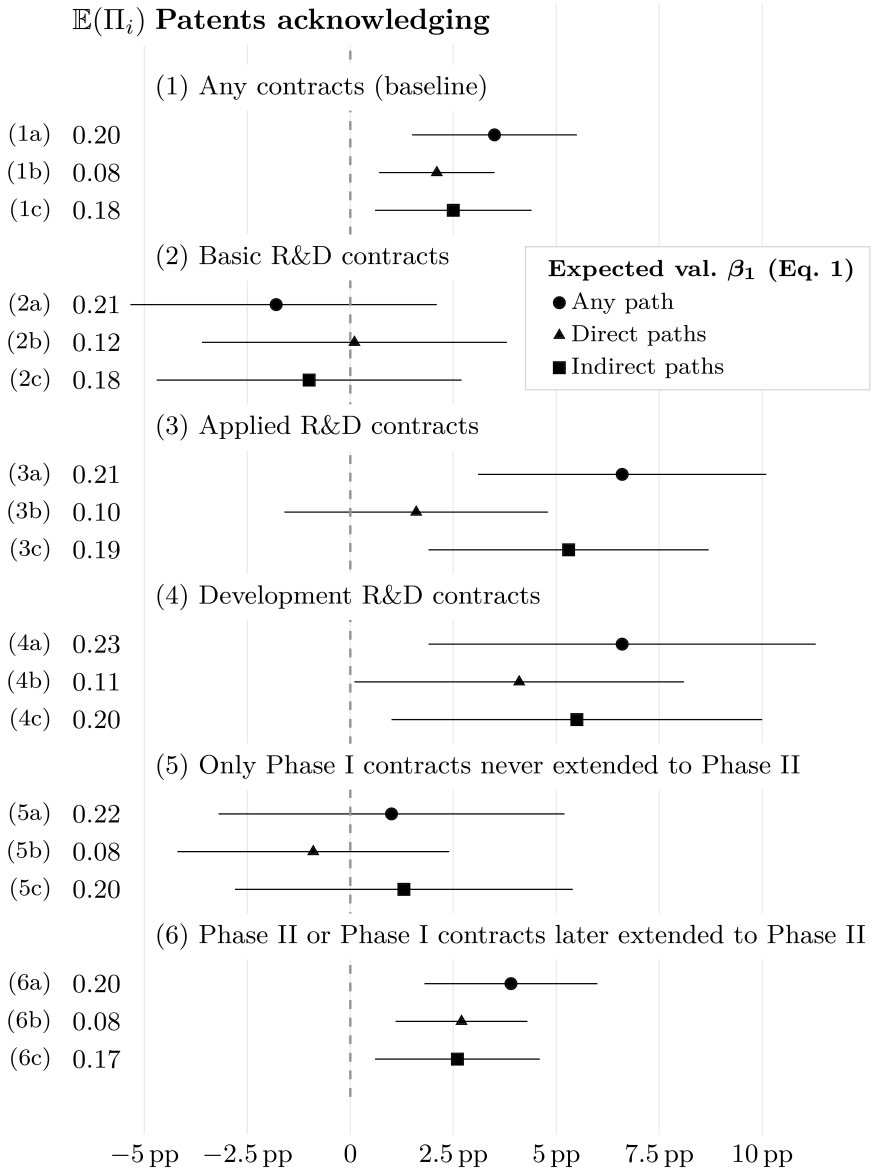
Figure 7 provides a visual overview of the regression coefficient β_1. It reports the point estimate and the 95-percent confidence interval of the SBIR predictor for 18 LPM regressions. (Appendix C reports the regression tables as well as the results for Probit regression models.) The top part of the 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 applied

[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 C reports additional details.

[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.

Figure 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.

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).

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.

Figure 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 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 Figure 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) 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 match-

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

ing 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.

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

[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	
Direct path	(1)	(2)	(3)	(4)
Phase II	0.042	-0.040	0.041	-0.041
	(0.031)	(0.053)	(0.035)	(0.051)
Post 2000	0.054*	-0.054	0.060**	-0.048
	(0.030)	(0.066)	(0.028)	(0.067)
Phase II × Post 2000		0.128*		0.128*
		(0.069)		(0.071)
log(`claims`)	0.006	0.005	0.004	0.004
	(0.021)	(0.020)	(0.019)	(0.019)
log(`bwd_cit`)	0.005	0.006	0.004	0.005
	(0.014)	(0.014)	(0.013)	(0.013)
log(`npl_cit`)	0.019	0.020*	0.018*	0.019*
	(0.012)	(0.012)	(0.011)	(0.011)
log(`geo_fam`)	0.007	0.008	0.004	0.005
	(0.023)	(0.023)	(0.020)	(0.020)
log(`fwd_cit`)	0.038**	0.036**	0.035**	0.033**
	(0.016)	(0.016)	(0.014)	(0.014)
Constant	0.313	0.387		
	(0.324)	(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: $\mathbb{E}(\Pi_i) = 0.17$. and $\mathbb{E}(\Pi_i|\text{Award pre 2000}) = 0.14$. 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.

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 protection (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. 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 pro-

tection, 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.

Data accessibility

The data are available at `https://doi.org/10.5281/zenodo.16779954`, under a Creative Commons Attribution 4.0 International license.

The web scraping component of the project relies on website 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`.

Acknowledgements

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.

Declarations of interest

The authors declare no competing interests.

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. Journal of Official Statistics 37, 1–30. doi:doi: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, 1189–1207. doi: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 in Brief 42, 108246. doi:10.1016/j.dib.2022.108246.

Audretsch, D.B., Link, A.N., Scott, J.T., 2002. Public/private technology partnerships: Evaluating SBIR-supported research. Research Policy 31, 145–158. doi: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, 1–23. doi: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. The Review of Economic Studies 86, 117–152. doi:10.1093/restud/rdy034.

Braunerhjelm, P., Svensson, R., 2024. Inventions, commercialization strategies, and knowledge spillovers in SMEs. Small Business Economics 63, 275–297. doi:10.1007/s11187-023-00812-z.

Bronzini, R., Piselli, P., 2016. The impact of R&D subsidies on firm innovation. Research Policy 45, 442–457. doi: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 Business Economics 48, 345–360. doi: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). NBER Working Paper doi:10.3386/w7552.

Czarnitzki, D., Hussinger, K., 2018. Input and output additionality of R&D subsidies. Applied Economics 50, 1324–1341. doi:10.1080/00036846.2017.1361010.

Daas, P.J.H., van der Doef, S., 2020. Detecting innovative companies via their website. Statistical Journal of the IAOS 36, 1239–1251. doi: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. Industrial and Corporate Change 23, 535–572. doi:10.1093/icc/dtt042.

de Rassenfosse, G., 2018. Notice failure revisited: Evidence on the use of virtual patent marking. Working Paper 24288. National Bureau of Economic Research. doi:10.3386/w24288.

de Rassenfosse, G., Jaffe, A.B., Raiteri, E., 2019. The procurement of innovation by the U.S. government. PLOS ONE 14, e0218927. doi:10.1371/journal.pone.0218927.

de Rassenfosse, G., Pellegrino, G., Raiteri, E., 2024. Do patents enable disclosure? Evidence from the invention secrecy act. International Journal of Industrial Organization 92, 103044. doi: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. Strategic Management Journal doi:10.1002/smj.3682.

Dimos, C., Pugh, G., 2016. The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature. Research Policy 45, 797–815. doi:10.1016/j.respol.2016.01.002.

Domènech, J., de la Ossa, B., Pont, A., Gil, J.A., Martinez, 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, IEEE Computer Society, Washington (DC, USA). pp. 573–578. doi: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. Academy of Management Discoveries 8, 103–138. doi:10.5465/amd.2019.0078.

Edelman, B., 2012. Using Internet data for economic research. Journal of Economic Perspectives 26, 189–206. doi: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. Research Policy 51, 104539. doi: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. North-Holland. volume 1 of *Handbook of the Economics of Innovation*. chapter 8, pp. 381–410. doi: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, 87–102. doi:10.1111/j.0023-5962.2004.00244.x.

Gök, A., Waterworth, A., Shapira, P., 2015. Use of web mining in studying innovation. Scientometrics 102, 653–671. doi:10.1007/s11192-014-1434-0.

Graham, S.J.H., Marco, A.C., Miller, R., 2015. The USPTO Patent Examination Research Dataset: A Window on the Process of Patent Examination. USPTO Economic Working Paper 2015-4. USPTO. doi:10.2139/ssrn.2848549.

Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. The Bell Journal of Economics 10, 92–116. doi:10.2307/3003321.

Gross, D.P., 2023. The hidden costs of securing innovation: the manifold impacts of compulsory invention secrecy. Management Science 69, 2318–2338. doi: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. Research Policy 45, 1129–1144. doi:10.1016/j.respol.2016.03.002.

Guzman, J., Li, A., 2023. Measuring founding strategy. Management Science 69, 101–118. doi: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. doi:10.3386/w8498.

Higham, K., De Rassenfosse, G., Jaffe, A.B., 2021. Patent quality: Towards a systematic framework for analysis and measurement. Research Policy 50, 104215. doi:10.1016/j.respol.2021.104215.

Howell, S.T., 2017. Financing innovation: Evidence from R&D grants. American Economic Review 107, 1136–1164. doi: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. Journal of Applied Econometrics 23, 729–747. doi:10.1002/jae.1016.

Jaffe, A.B., de Rassenfosse, G., 2017. Patent citation data in social science research: Overview and best practices. Journal of the Association for Information Science and Technology 68, 1360–1374. doi: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 Business Economics 45, 805–823. doi: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, 2011–2041. doi:10.1007/s11192-020-03726-9.

Kinne, J., Lenz, D., 2021. Predicting innovative firms using web mining and deep learning. PLOS ONE 16, 1–18. doi:10.1371/journal.pone.0249071.

Lemley, M.A., Shapiro, C., 2005. Probabilistic patents. Journal of Economic Perspectives 19, 75–98. doi:10.1257/0895330054048650.

Li, M., Jin, M., Kumbhakar, S.C., 2022. Do subsidies increase firm productivity? Evidence from Chinese manufacturing enterprises. European Journal of Operational Research 303, 388–400. doi:10.1016/j.ejor.2022.02.029.

Libaers, D., Hicks, D., Porter, A.L., 2016. A taxonomy of small firm technology commercialization. Industrial and Corporate Change 25, 371–405. doi:10.1093/icc/dtq039.

Link, A.N., Scott, J.T., 2010. Government as entrepreneur: Evaluating the commercialization success of SBIR projects. Research Policy 39, 589–601. doi: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 Technology Law Journal 20, 1521.

Myers, K.R., Lanahan, L., 2022. Estimating spillovers from publicly funded R&D: Evidence from the US Department of Energy. American Economic Review 112, 2393–2423. doi: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. doi:10.17226/25674.

National Research Council, 2009a. An Assessment of the SBIR Program at the Department of Defense. The National Academies Press, Washington, DC. doi:10.17226/11963.

National Research Council, 2009b. Revisiting the Department of Defense SBIR Fast Track Initiative. The National Academies Press, Washington, DC. doi:10.17226/12600.

National Research Council, 2014. SBIR at the Department of Defense. The National Academies Press, Washington, DC. doi:10.17226/18821.

Pakes, A., Griliches, Z., 1980. Patents and R&D at the firm level: A first report. Economics Letters 5, 377–381. doi: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. International Journal of Finance & Economics 29, 2278–2299. doi:10.1002/ijfe.2766.

Radicic, D., Pugh, G., 2017. R&D programmes, policy mix, and the 'European paradox': Evidence from European SMEs. Science and Public Policy 44, 497–512. doi:10.1093/scipol/scw077.

Rai, A.K., Sampat, B.N., 2012. Accountability in patenting of federally funded research. Nature Biotechnology 30, 953–956. doi:10.1038/nbt.2382.

Rammer, C., Es-Sadki, N., 2023. Using big data for generating firm-level innovation indicators - a literature review. Technological Forecasting and Social Change 197, 122874. doi:10.1016/j.techfore.2023.122874.

Romer, P.M., 1990. Endogenous technological change. Journal of Political Economy 98, S71–S102. doi:10.1086/261725.

Ruttan, V.W., 2006. Is War Necessary for Economic Growth? Military Procurement and Technology Development. Oxford University Press, Oxford, UK. doi:10.1093/0195188047.001.0001.

Scherer, F.M., Harhoff, D., 2000. Technology policy for a world of skew-distributed outcomes. Research Policy 29, 559–566. doi:10.1016/S0048-7333(99)00089-X.

Sharp, G.S., 2003. A layman's guide to intellectual property in Defense contracts. Public Contract Law Journal 33, 99–137.

Sichelman, T., 2009. Commercializing patents. Stanford Law Review 62, 341–411.

Squicciarini, M., Dernis, H., Criscuolo, C., 2013. Measuring Patent Quality: Indicators of Technological and Economic Value. OECD Science, Technology and Industry Working Papers 2013/03. OECD. doi:10.1787/5k4522wkw1r8-en.

Svensson, R., 2007. Commercialization of patents and external financing during the R&D phase. Research Policy 36, 1052–1069. doi:10.1016/j.respol.2007.04.004.

U.S. Congress, 2000. Small business reauthorization act of 2000. HR 5667. Pub. L. 106-554, Appendix I.

Appendix A. The Database

Appendix A.1. Introduction

This appendix describes the creation and structure of the database used for the analysis in the main text and better details the novel, web-based approach for tracking patented inventions commercialization proposed by the article. The ultimate goal of this database is to provide novel data allowing to better understand the extent to which R&D funding from the U.S. DoD's SBIR/STTR program translates into products for the final consumer. The database links U.S. federal contracts, awarded by the U.S. Department of Defense in the context of the Small Business Innovation Research (SBIR) and the Small Business Technology Transfer (STTR) programs, to patents granted by the U.S. Patent and Trademark Office (USPTO), and provides evidence of commercialization for SBIR/STTR-related patents. Specifically, we consider that a patent is commercialized if a web page exists on the patent assignee's corporate website that mentions that the patent protects one or several products.

The construction of these data builds on administrative data on federal procurement and patented inventions and on a novel, web-based approach to recover information on patent coverage of commercial products. Information about the SBIR contracts comes from the Defense Contract Action Data System (DCADS), for the years 1984–2001, and from USAspending.gov, for the years 2001–2018. Patent-related information is provided by PatentsView, although specific pieces of information are recovered from the USPTO's Patent Examination Research Dataset (PatEx) and the European Patent Office's PATSTAT database (v. 2020a).

The database provides information about three different objects: DoD's SBIR and STTR awards; USPTO patents; and web pages. It is composed of five *main tables*, and five *associative tables* that link the different pieces together. Each associative table has a many-to-many relationship between the two element kinds that it links. Figure A.8 displays the logical model of the database. Appendix A.3 describes the content of each table in detail.

Two of the associative tables, `award_to_patent` and `patent_to_webpage`, provide the most relevant pieces of information for our work. The `award_to_patent` table links an award identifier to a patent document, whereas the `patent_to_webpage` table links each patent document to a web page providing evidence of a connection between the patent and a commercial product. Appendix A.2 provides a detailed description of the process and the method we used to populate these tables.

The main tables report detailed information about awards, patents, and web pages. More specifically:

The `award` table reports information on procurement contracts—as recovered from

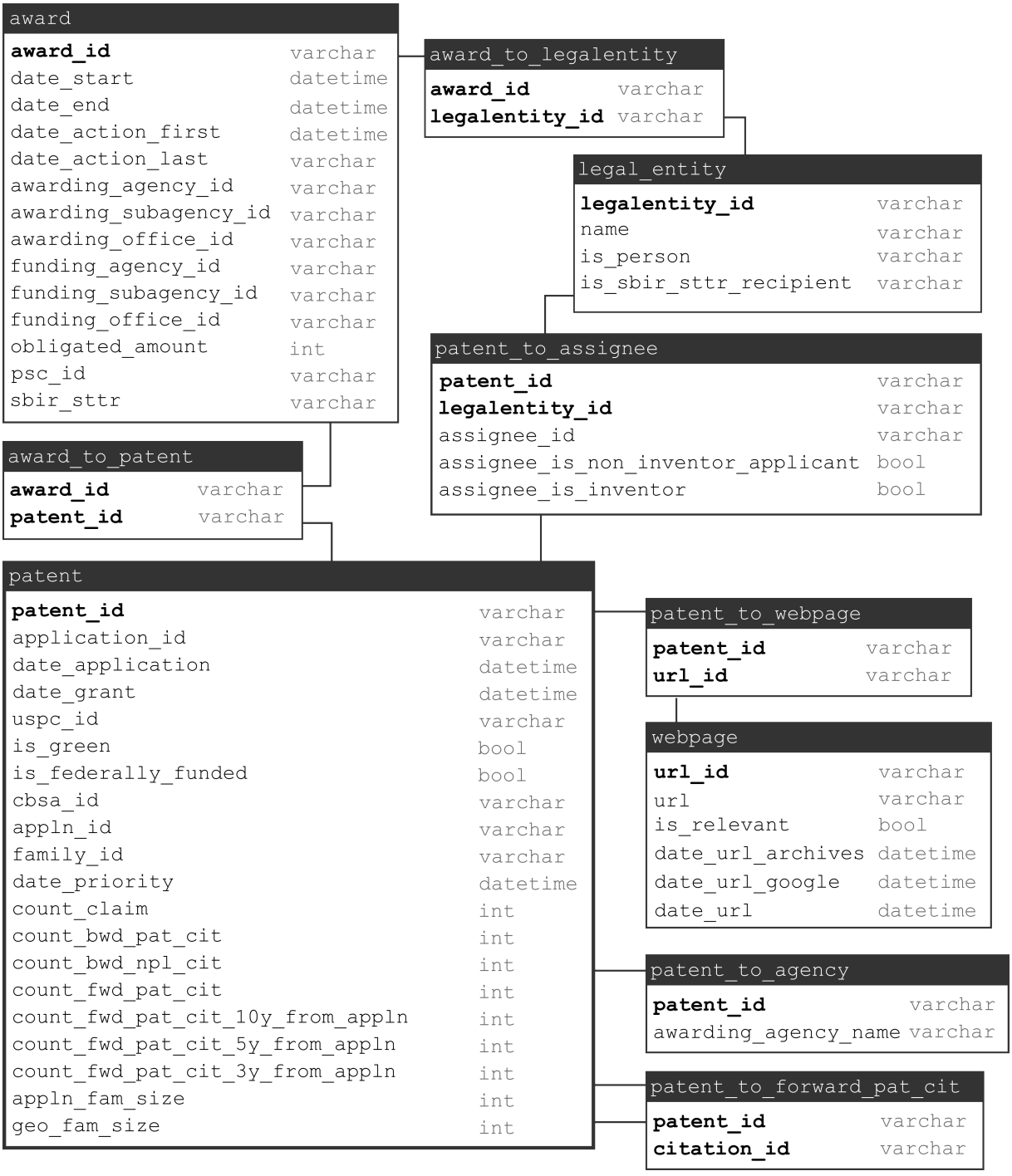


Figure A.8: Logical model of the database.

DCADS and USAspending.gov—such as the funding agency, information about the starting and ending dates, the SBIR/STTR Phase, the type of purchased good or service, the total obligated amount of the contract, the recipient name and DUNS number. This table can be linked to the `legal_entity` table to gather additional information.

The `patent` table reports information on patents such as assignment and grant date, and assignee identifier, which we recover from PatentsView, PatEx, and PATSTAT.

The `webpage` table reports information about the web pages linked to one or more patents in the `patent` table.

Appendix A.2. Construction of the database

The main objective is to link SBIR-funded awards to evidence of commercialization. To do so, we limit the scope of our work to the DoD's SBIR/STTR awards that generated at least one patented invention, and we consider two potential *paths* from an invention to a commercial product, as illustrated by Fig. 1 in the main text. First, a *direct path* that connects an award to a product through a patent that acknowledges public funding and is listed on the website of the patent owner in connection to a product. Second, we also consider an *indirect path*, where the patent acknowledging public funding receives a citation from another patent which is then listed on a web page as covering a commercial product.

Therefore, the construction of the database relies on two necessary conditions: (i) the availability of consistent information about SBIR/STTR awards, patented inventions, and patent-applicants' web pages; and (ii) the possibility of unambiguously connecting data from the different information sources.

To meet the first condition, we collected SBIR/STTR awards information from DCADS and USAspending.gov, patent information from PatentsView, and we built corporate web pages information based on search-engines results and fine-tuned scripts to classify them. To satisfy the second condition, we exploited the fact that award recipients are required to report information about the award in the patent text, including the identification number (ID) of the award. We extracted IDs from the patent text and linked each patent to the award-level information taken from the U.S. government archives. On the other hand, we leveraged the fact that patents are frequently referenced by their issuance number on the patent owner's corporate website–such as on Virtual Patent Marking pages–to link each patent to potentially relevant web pages, where available. The remainder of this section outlines our approach to assembling the database.

Appendix A.2.1. Linking awards to patents

Retrieving information on the federal awards. Given that the DoD classifies all SBIR- and STTR-related contractual actions as procurement contracts (and not as research grants), to recover complete information about federal awards, we rely on the DCADS and USAspending.gov databases. As explained by de Rassenfosse and colleagues, the Federal Funding Accountability and Transparency Act (FFATA), approved in September 2006 by the U.S. Congress, required federal contracts, grants, loans, and other financial assistance awards to be displayed on a searchable, publicly accessible website in order to give the American public access to information on how tax dollars are being spent (de Rassenfosse et al., 2019). In 2014, the Digital Accountability and Transparency Act (DATA Act) further expanded the transparency efforts of FFATA. In December 2007, the U.S. government launched the USAspending.gov website to comply with the FFATA's requirements. About federal procurement contracts, USAspending.gov includes the full data from the Federal Procurement Data System (FPDS) database from the fiscal year 2000 (starting October 1999) to the present. The FPDS tracks every U.S. federal procurement contract whose estimated value is above $3,000, and every modification to that contract, regardless of the dollar value.

For awards signed after October 2000, we fully rely on data provided by USAspending.gov. For contracts signed before October 2000, we rely instead on the Defense Contract Action Data System (DCADS), retrieved through the National Archives and Records Administration (NARA) website. The DCADS data go back to the fiscal year 1976 and cover contracts with a value above $25,000. The information content of DCADS is, to a large extent, coherent with the data included in USAspending.gov.

For each DoD contract signed between the fiscal year 1983 and 2018, we download the complete data of contractual actions from USAspending.gov or DCADS, and retain information about the contract's identification number, Procurement Instrument Identifier (PIID); the signed date; the contract start date; the contract (potential) end date; the awarding sub-agency and office; the recipient name and DUNS number; the total dollar amount awarded on the contract; the product or service code; and the SBIR/STTR Phase (if any).

Once all the available DoD awards have been identified, we exclusively retain the ones classified as SBIR/STTR Phase I/II actions only. To verify the completeness and reliability of the selected data sources, we cross-reference the information extracted from USAspending.gov or DCADS using data from `https://www.sbir.gov/`. This exercise confirms that the extracted data correspond almost perfectly with the data available on the SBIR website for the whole time period we consider. Therefore, the awards' information we consider, even though limited to

the Department of Defense, covers most of the SBIR (and STTR) program history.

The Bayh-Dole Act and the Federal Acquisition Regulation. The next step in our data construction effort is to unambiguously link the retrieved SBIR/STTR awards to the patented inventions they supported.

As explained by de Rassenfosse and colleagues, the Bayh–Dole Act, approved in December 1980 by the U.S. Congress, and the U.S. Federal Acquisition Regulation (FAR) made the information we need publicly available (de Rassenfosse et al., 2019; Sharp, 2003). Under the Bayh–Dole Act and its subsequent modifications (35 U.S.C. § 202(c)(6)), private entities must acknowledge federal support and rights to an invention—funded, at least in part, by a federal research grant or procurement contract—in the written specification of the invention for all non-provisional U.S. patent applications. Furthermore, the FAR, in its Subparts 27.3 and 52.2, requires including, in the text of the patent, a statement reporting also an identification of the specific governmental agency and the identification number of the relevant contract. Therefore, these requirements allow us to identify the patented inventions produced in the performance of work under a government contract and to link them to the specific award connected with their production.

Award-IDs extraction. The USPTO includes information about government interest statements disclosed in U.S. patent documents in its PatentsView database (Jones and Madhavan, 2020). More specifically, the `government_interest` table of PatentsView reports the full-text of the government interest statement, as extracted from U.S. patents when available. Using a random sample of the PatentsView data, we verified the quality of this extraction. Quality is high, with a negligible number of errors.

To connect each patent to a specific award, we use the full-text of the government interest statements as provided by PatentsViewand extract the Procurement Instrument Identifier (PIID) from the statements. The PIID is the official contract identification number that uniquely identifies each procurement contract and allows connecting the data with the federal procurement data system.

To extract this information from the text of the government interest statement, we developed a Python script that uses several regular expressions exploiting the standardized structure of the PIID. The script:

1. Cleans the text from poor-formatting (some due to OCR mistakes). For example, HTML codes are converted to the corresponding Unicode characters. So `&mdash;` becomes `-`; `&num;` becomes `#`; etc.

2. Removes several substrings that can be easily confused with a PIID code. Among others, ZIP and post-box codes; dates; patent numbers; and references to laws.
3. Modifies recurring partial patterns to facilitate the identification of the full PIID. For instance, the script transforms a common way of reporting a list of contracts such as '`AA123, 234, and 345`' into '`AA123, AA234, and AA345`'.
4. Tokenizes the text using relevant punctuation characters such as [`.,;:-&`].
5. Preserves, from the list of tokens, only strings that:
    - Are not acronyms of U.S. Federal Agencies.
    - Are longer than one character.
    - If longer than three characters, contain at least one digit.
    - Their basic form (singular for nouns; present tense for verbs) is not in the English dictionary.
6. Joins tokens shorter than four characters with the one that follows or precedes them, and, then, drops the tokens shorter than four characters not containing digits.

The tokens resulting from the extraction process described above are considered as potentially valid PIIDs. Figure A.9 reports some examples of the PIID extraction process. They have been chosen to illustrate the different challenges that the PIID-extraction script deals with. For each patent, the figure reports the patent number, the full-text included in the government interest statement, and the potential PIID the script extracts from the statement, if any. As a final step in this process, we link each potential patent-PIID pair to the award information, matching the patent information with the contract-level data described in the previous section. The script is available at `https://github.com/n3ssuno/iris-utils/blob/75ab02daf038af72d3c48a5d4ee0b04b9735c7be/award_id.py`; refer to the commented lines in the code itself for a detailed explanation of the procedure followed.

Matching Phase I to Phase II contracts. As discussed in the article, the SBIR/STTR programs have two 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, in which they receive larger funding to pursue the research started in Phase I. Therefore, each SBIR/STTR Phase II contract is, by definition, linked to a Phase I contract. However, two connected Phase I and Phase II contracts have different contract identification numbers (PIID) and this may lead to double counting. To take this issue into account, we develop a strategy

Patent num.	Government interest statement	PIID
US8059142	Some aspects of this invention were made with Government support under contract FA8650-04-M-5443 awarded by the United States Air Force Research Laboratory. The Government has certain rights in the invention.	FA8650-04-M-5443
US8431390	STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT This invention was made with government support under Grant No. 5R01HG003583-01 awarded by the NIH; Project No. W911 SR-04-P-0047 awarded by the Department of Defense; Contract No. NBCHCO50133 awarded by HSARPA, and Agreement No. W81XWH-04-9-0012 (Order No. TTA-1-0014) awarded by the Department of Defense. The government has certain rights in the invention.	5R01HG003583-01 NBCHCO50133 TTA-1-0014 W81XWH-04-9-0012 W911+SR-04-P-0047
US9981980	This invention was made with Government support under Award Nos. NIH/NIHMD P20GM103475; NIH/NIGMS SC3GM116713; NIH/NIMHD G12MD007600 and NIH/NCI U54 CA096297 awarded by The National Institutes of Health. The U.S. Government has certain rights in the invention.	NIHMD+P20GM103475 SC3GM116713 G12MD007600 U54+CA096297
US10398742	This invention was made with Government Support under NIH grants R01HD062844, R21/R33AI094519, and R21/R33AI079740 by the National Institutes of Health. The Government has certain rights in the invention.	R01HD062844 R21/R33AI079740 R21/R33AI094519
US6838301	STATEMENT AS TO FEDERALLY SPONSORED RESEARCH The invention described herein was made in the performance of work under a NASA contract, and is subject to the provisions of Public Law 96-517 (35 U.S.C. 202) in which the Contractor has elected to retain title.	
US9945114	FEDERALLY-SPONSORED RESEARCH AND DEVELOPMENT A System and Method for the Rapid Installation of a Portable Structure in a Confined Vertically Inaccessible Location is assigned to the United States Government and is available for licensing for commercial purposes. Licensing and technical inquiries may be directed to the Office of Research and Technical Applications, Space and Naval Warfare Systems Center, Pacific, Code 72120, San Diego, Calif., 92152; voice (619) 553-5118; email ssc_pac_T2@navy.mil. Reference Navy Case Number 103010.	

Figure A.9: Examples of Government Interest Statements and PIIDs as extracted by the script described in Appendix A.2.1. In the first line, only one *award id* is clearly stated. The second line is more complicated. First, multiple PIIDs must be extracted. Second, the PIIDs are composed of multiple, disconnected substrings. To form a unique string for each award id, the script adds a '+' character between the substrings. In the third, each award ID is preceded by the name of the funding agency. This is a particularly hard task, and indeed the script partly fails to provide all the cleaned PIIDs (see the first identifier, where also part of the agency's name is reported). This is because (a) the agencies' names (highlighted by red squares) are shortened, which can be confounded with a sub-portion of a PIID; (b) both the agencies' and sub-agencies' names are reported, linked with a "/" sign. This as well can be easily confused with a PIID component (*e.g.*, see the example following). In the fourth line, the award IDs are partly shortened. Indeed, the three PIIDs extracted actually represent five awards `R01-HD062844`, `R21-AI079740`, `R33-AI079740`, `R21-AI094519`, and `R33-AI094519`. If thought together with the previous one, this example well illustrates the difficulties the script must overcome trying to balance between including the PIIDs and excluding other elements that can be confounded with them. The last two are cases where no award ID is provided in the statement. However, several elements can be easily confused with a potential PIIDs: the legal reference in the first; the post-code, the phone number, etc. in the second.

to link each DoD-SBIR Phase I contract included in the database to a Phase II contract, either linked or not linked to a patented invention. To do so, we use the DoD data and match a DoD-SBIR Phase I contract with any Phase II contract with the same recipient DUNS number (Dun & Bradstreet's Data Universal Numbering System number), and we only consider Phase II contracts that were first signed at least three months after the start date of the focal Phase I contract and no more than 30 months after it. Then, we evaluate each paired couple's plausibility in three steps. First, we exploit the awards data provided on the SBIR/STTR program's website (`https://www.sbir.gov/`) which provides, for some contracts, a *tracking number* that uniquely identifies Phase I–Phase II couples. We flag any contract pair sharing the same *tracking number* as an actual Phase I-Phase II link. Second, for contracts for which the tracking number is not available, we focus on pairs of contracts awarded by the same office *and*, at the same time, share the same PSC code. We employed the `maximum_bipartite_matching` algorithm, from the SciPy library (`https://scipy.org/`), to maximize the number of couples and the number of linked contracts of both kinds preserved. Lastly, after removing the Phase II contracts matched in the previous steps, we repeat step two but relaxing the matching conditions. More specifically, we consider contract pairs either signed with the same awarding office *or* belonging to the same PSC code. Respectively, 196 couples have been established through the first step, 243 through the second, and 178 through the third. To evaluate the quality of our matches, we test the approach used in steps two and three on the contract-pairs formed on the basis of the *tracking number*. Our approach correctly identifies 84.3 percent of the contract pairs for which the tracking number is available.

Appendix A.2.2. Linking patents to commercial products

Once we identified the connection between a DoD's SBIR/STTR contract and one or more patented inventions, we need to link SBIR-related inventions to commercial products. To do so, we adopt an approach inspired by the IPRoduct project (`https://iproduct.io/`), and look for patent-protected products on the websites of the companies owning the patents.

As discussed by de Rassenfosse (2018), companies have a number of legal and economic incentives to make public that one of their products is patent-protected. In the U.S. in particular, the 2011 Leahy-Smith America Invents Act (AIA), by extending the 35 U.S.C. § 287(a), encourages patentees to provide constructive notice to the public that an article is patented by allowing them to affix the word "patent" or "pat." on the article along with a URL of a web page that associates the patented article with the patent number(s); a practice known as 'virtual patent

marking' (VPM). Patentees have incentives to disclose information accurately, as virtual marking allows the recovery of damages prior to notice of infringement.

Web pages that comply with the AIA marking requirements—*i.e.*, VPM pages, properly said—provide a clear link between a patent and a product commercialized by the patent owner. Our web-based approach entails searching for the existence of such pages associated with any of the DoD's SBIR/STTR patented inventions identified in the previous steps. However, since we are not exclusively interested in AIA-complying VPM pages, we interpret the patent marking idea more broadly and consider any kind of web page reporting an explicit patent-product connection. For instance, brochures used by companies to provide detailed information about their products often disclose the existence of patents covering the products, even if the aim is not necessarily to comply with 35 U.S.C. § 287(a). Therefore, our strategy involves searching for relevant web pages providing a direct patent-to-product path for our DoD's SBIR/STTR patents.

Figures A.11–A.13 display the content of three relevant web pages linked to DoD's SBIR/STTR patents. They have been chosen as they cover the most common kinds of web pages linked to the patents in our sample: properly-defined VPM pages making explicit reference to the U.S. virtual marking regulation (A.11), web pages reporting information that complies with the regulatory requirements but do not make explicit reference to the regulation (A.12), and product brochures (A.13).

Building an indirect path. As mentioned in the introduction, we are not exclusively interested in identifying a *direct path* between DoD's SBIR/STTR patent and a commercialized product, but also *indirect paths*. An *indirect path* to commercialization exists when a SBIR-funded patent is cited by another patented invention as relevant prior-art, and the citing patent is then mentioned on a relevant web page. Therefore, we also extract the full list of patents citing any of the SBIR/STTR patents in our sample from the `uspatentcitation` table and use the process described in the next section to determine whether they are linked to commercial products. All in all, our working sample is composed of two groups: the 3,070 SBIR-funded granted patents (of which 2,304 received at least one citation from another USPTO utility patent) and the 40,020 granted patents citing those in the first group.

Identifying relevant web pages. To search for the existence of relevant web pages for the patents in our sample, we adopt a web-based multi-step approach.

Figure A.10 depicts the steps of our search process. The first step correctly identifies `https://www.immersion.com/` as the website of Immersion Corporation. With the second step, we have been able to identify, within Immersion's corporate website, the VPM page, `https://www.immersion.com/legal/trademarks-and`

`-patent-markings`, where US8502792 is claimed as covering the TS1000™ software developed by the company.

Identifying a company's website. The first step (top part of Figure A.10) requires the identification of the website (*domain*) of the companies that own the patents in our data. To do so, we start by harmonizing the assignee names provided in PatentsView so as to have a unique company name including the abbreviation indicating the business legal structure (*e.g.*, '*INC*' or '*Corp*'). We harmonize for two reasons. First, because, ideally, we would like to narrow the scope of our search as much as possible. In fact, querying a search engine for the term '*Immersion*' will return an entirely different set of results than querying for '*Immersion Corp*'. Second, as business designations could be abbreviated in different forms, we would like to search for both '*Immersion Corp*' and '*Immersion Corporation*', as the two queries would not lead to identical results on web search engines.

Then, we write a scraper that searches for the full company names on Google Search. Moreover, since Google Search also returns aggregators that mention the searched corporation, but that are not the corporate website of the searched patent assignee, we increase the precision of the results by searching for the assignee's website on Bloomberg and the official SBIR/STTR program's website (`https://www.sbir.gov`).

The scraper, written in JavaScript using Google's Puppeteer library (`https://pptr.dev/`), uses Google and Bing search engines. Our use of the script complies with all the requirements and limitations imposed by the websites employed. For each assignee, a query for the patent assignee name, like:

> ("IMMERSION CORP" OR "IMMERSION CORPORATION")
> -site:gov -site:edu -site:mil -site:int -site:bloomberg.com

is searched on `https://www.google.com`. If the patent assignee's name is different from the name of the award recipient's name, we also include the name of the award recipient in the query with an *OR* operator. The queries return a list of web pages that are likely to be associated with the company names. For each query, we then retain up to ten of the most relevant results.

At the same time, a query like

> ("IMMERSION CORP" OR "IMMERSION CORPORATION")
> site:bloomberg.com/profile/company

is searched on `https://www.bing.com`. Also in this case, we retain up to ten of the most relevant results, but we then look into the retrieved Bloomberg pages and

Figure A.10: Schematic example of a relevant web page (like a VPM page) searching process. First, the assignee's name is searched on a search engine. Second, all the results are collected. Third, the patent of interest is searched within each scraped website on a search engine. Lastly, the results are collected as potentially relevant web pages.

further search for the websites of the searched company as reported in the Bloomberg company database. Lastly, we performed a query like

("IMMERSION CORP" OR "IMMERSION CORPORATION")
site:sbir.gov/sbc

using the search engine of the SBIR program's website (`https://www.sbir.gov`). For each page retrieved, we further look for the website of the searched company as reported in the SBIR database.

From each of the web pages recovered through the automated search described above, we then extract the part of the website that can be considered the domain of the page. For example, for a website like `https://www.immersion.com/news/some-news.html` we preserve the immersion.com part.

In this step, we searched for 6,647 distinct patent assignees and retrieved 11,731 unique web domains.

We then filtered out as many domains as possible, among those that clearly were not the searched corporate website. This cleaning mostly consists of removing information aggregators like govtribe.com or rocketfinancial.com, and manually inspecting domains retrieved more often than ten times. After this cleaning step, we obtained 9,411 unique domains linked to at least one SBIR-funded patent through the name of the assignee.

Identifying patent-product links on a company's website. Once we obtain the list of potential domains for each searched assignee, we need to iteratively parse them, looking for relevant web pages that would provide information about patent-products links for our DoD's SBIR/STTR patented inventions.

To do so, we develop a scraper whose main tasks are largely similar to the ones of the scraper used in the previous step (bottom part of Figure A.10). This time, the scraper searches on Google (`https://www.google.com`) a query like

(site:immersion.com) AND (8502792 OR 8898242)

where *immersion.com* is the potential website retrieved in the previous step, and the numbers in the second search block correspond to the patent registration numbers of the patented inventions owned by *Immersion Corporation* and generated with the SBIR program support. The scraper retrieves all the pages on the searched websites that mention the patent numbers in question. The scripts are available at `https://github.com/n3ssuno/iris-scraper/blob/42077301e412b72ed9d94f3bc37e432d6f7c7652/scrape-for-websites.js` and `https://github.com/n3ssuno/iris-scraper/blob/42077301e412b72ed9d94f3bc37e432d6f7c7652/scrape-for-vpm-pages.js`, respectively.

From the Google Search's results pages, we extracted the first n hits, or less; where n equals the cardinality of the list of searched domains (right head side of the query) times the cardinality of the list of searched patents (left head side of the query).

Classifying detected pages as relevant web pages. Using the scraping process described above, we collect 3,131 web pages containing a string of characters compatible with one (or more) of the patent numbers of interest. However, this is not enough to establish a clear patent-product link. Figures A.14–A.17 report the most common examples of web pages retrieved by the scraper, which do not identify a link between a product and a patent, but yet report the relevant patent number(s). As the figure shows, we have cases in which the web page simply includes the PDF version of a legal document connected to the patent. Oftentimes, it is the patent document itself (A.14), as granted by the USPTO, or an official form submitted to the Securities and Exchange Commission disclosing a company's IP assets (A.15). In other cases, the web page reports a list of patents owned by the company, but without making any explicit connection to specific products commercialized by the company (A.16). Finally, the web page may report a number that is identical to the patent registration number, but which is actually something else, as for instance a catalog number, or a telephone number (A.17). To assess whether each collected web page is providing an actual patent-product link, we adopt a classification strategy that involves both an automatic and a manual step.

Automatic classifier. As a first step, we develop a script able to automatically classify pages that (a) are very unlikely to be actual relevant web pages, or (b) are very likely to be relevant web pages. This script is available at `https://github.com/n3ssuno/iris-classifier/blob/8f311b18a869e4c08b65cced8c3029ba617d49a5/pre-classify.py`. To specify the rules followed by the script, we use a subset of pages that we manually inspect and classify. Specifically, we determine that pages matching the following characteristics belong to group (a):

1. Pages exclusively including the PDF file of the patent document as released by the USPTO. To identify these pages, the script downloads the PDF and parses it to check if the first 125 characters match with one of the following regular expressions:
   - `United States Patent([^s].*|$)`;
   - `United States.*Patent Application Publication`;
   - `The Director of the United States.*Patent and Trademark Office .*Has received an application for a patent.`

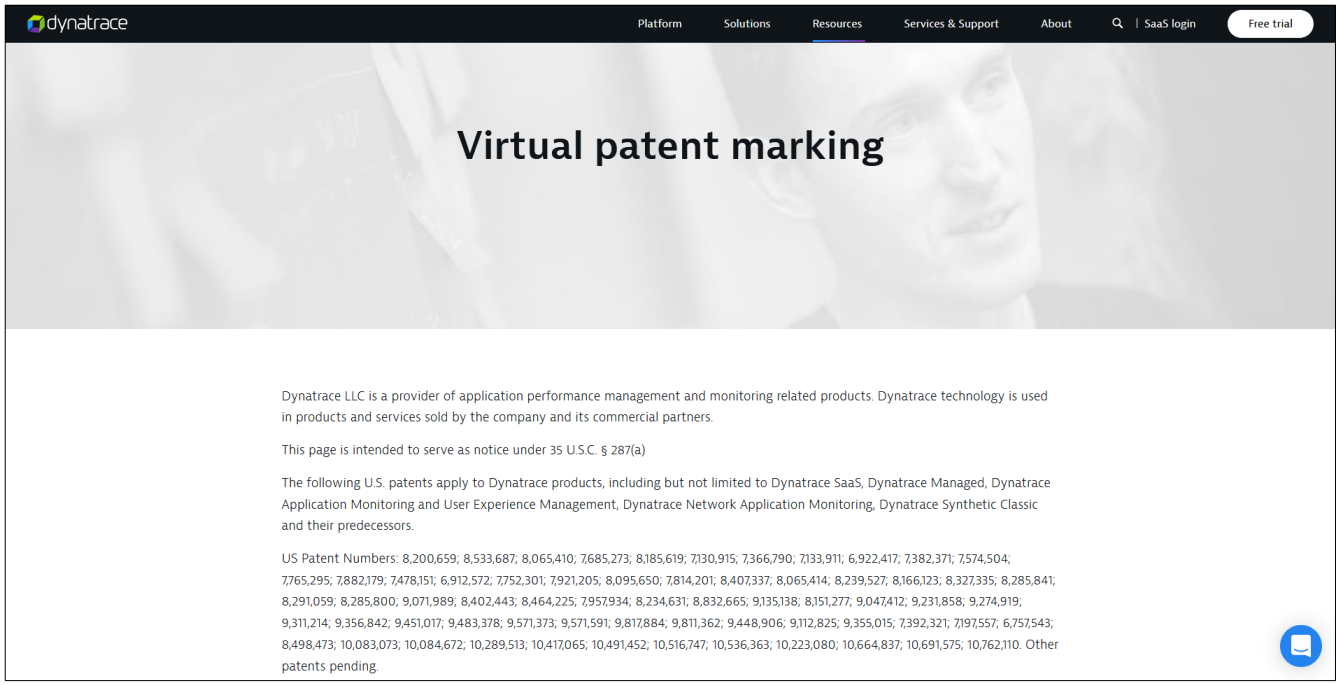
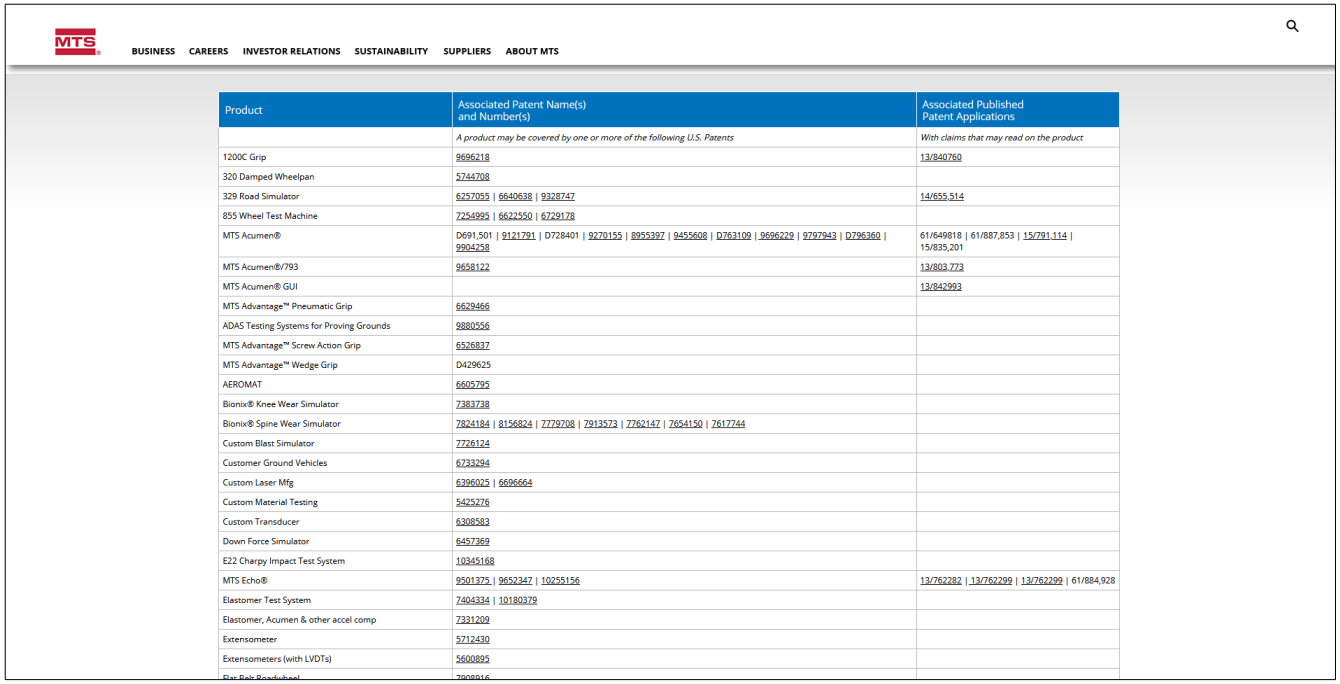


Figure A.11: Example of a properly said VPM page. This is spotted not only by the title of the web page, but also by the sentence reporting the legal reference to '35 U.S.C. § 287(a)'. Therefore, this is an example of one of these pages that we have been able to classify automatically as a correctly-detected VPM page (see Appendix A.2.2).



Product	Associated Patent Name(s) and Number(s)	Associated Published Patent Applications
	A product may be covered by one or more of the following U.S. Patents	*With claims that may read on the product*
1200C Grip	9696218	13/840760
320 Damped Wheelpan	5744708	
329 Road Simulator	6257055 | 6640638 | 9328747	14/655,514
855 Wheel Test Machine	7254995 | 6622550 | 6729178	
MTS Acumen®	D691,501 | 9121791 | D728401 | 9270155 | 8955397 | 9455608 | D763109 | 9696229 | 9797943 | D796360 | 9904258	61/649818 | 61/887,853 | 15/791,114 | 15/835,201
MTS Acumen®/793	9658122	13/803,773
MTS Acumen® GUI		13/842993
MTS Advantage™ Pneumatic Grip	6629466	
ADAS Testing Systems for Proving Grounds	9880556	
MTS Advantage™ Screw Action Grip	6526837	
MTS Advantage™ Wedge Grip	D429625	
AEROMAT	6605795	
Bionix® Knee Wear Simulator	7383738	
Bionix® Spine Wear Simulator	7824184 | 8156824 | 7779708 | 7913573 | 7762147 | 7654150 | 7617744	
Custom Blast Simulator	7726124	
Customer Ground Vehicles	6733294	
Custom Laser Mfg	6396025 | 6696664	
Custom Material Testing	5425276	
Custom Transducer	6308583	
Down Force Simulator	6457369	
E22 Charpy Impact Test System	10345168	
MTS Echo®	9501375 | 9652347 | 10255156	13/762282 | 13/762299 | 13/762299 | 61/884,928
Elastomer Test System	7404334 | 10180379	
Elastomer, Acumen & other accel comp	7331209	
Extensometer	5712430	
Extensometers (with LVDTs)	5600895	
Flat Belt Roadwheel	7908816	

Figure A.12: Example of a list of patents, each linked to a specific product code commercialized by the company owning the web page. Even though the page does not provide any clear reference to the VPM legislation, it perfectly plays the role required by the U.S. federal statute.

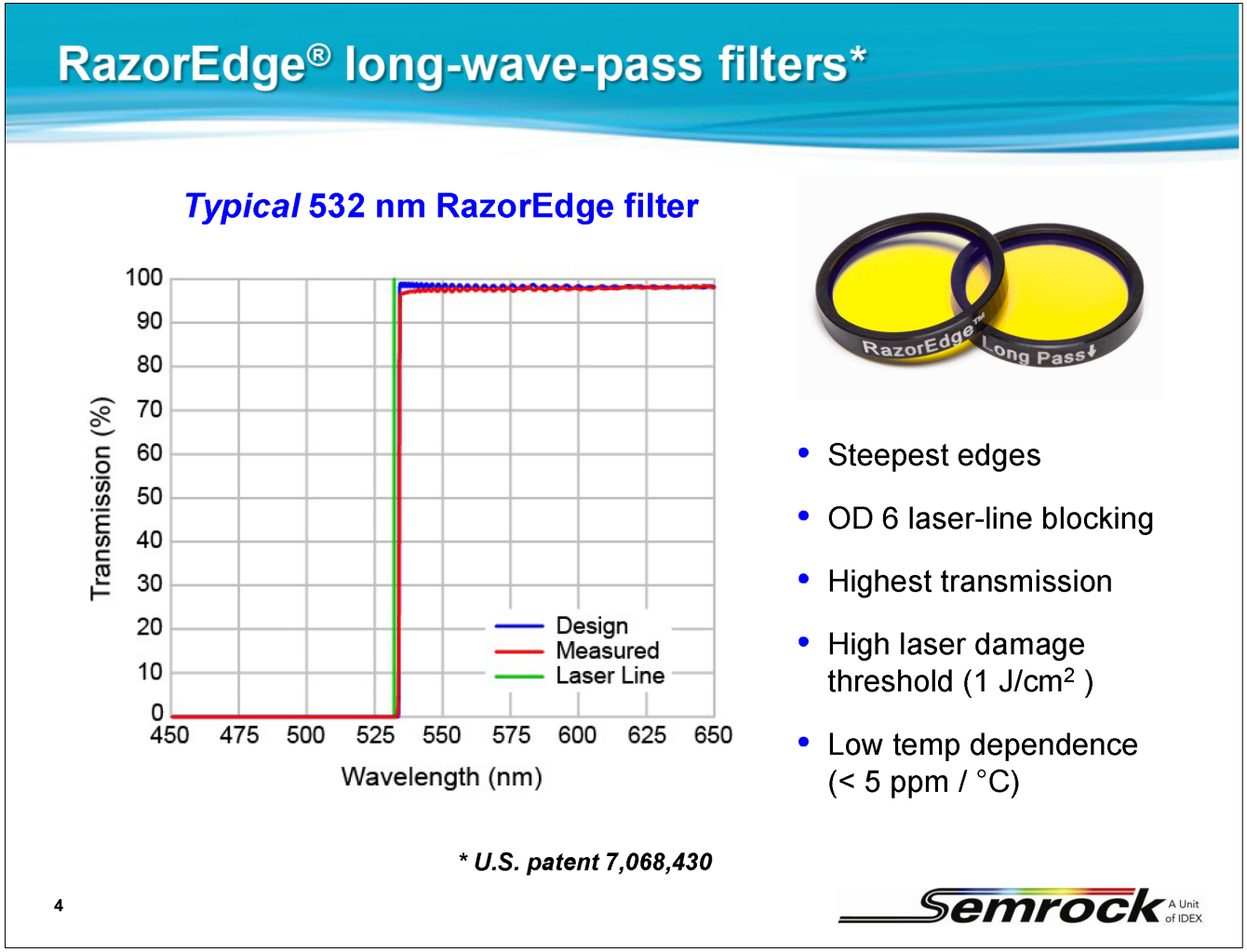
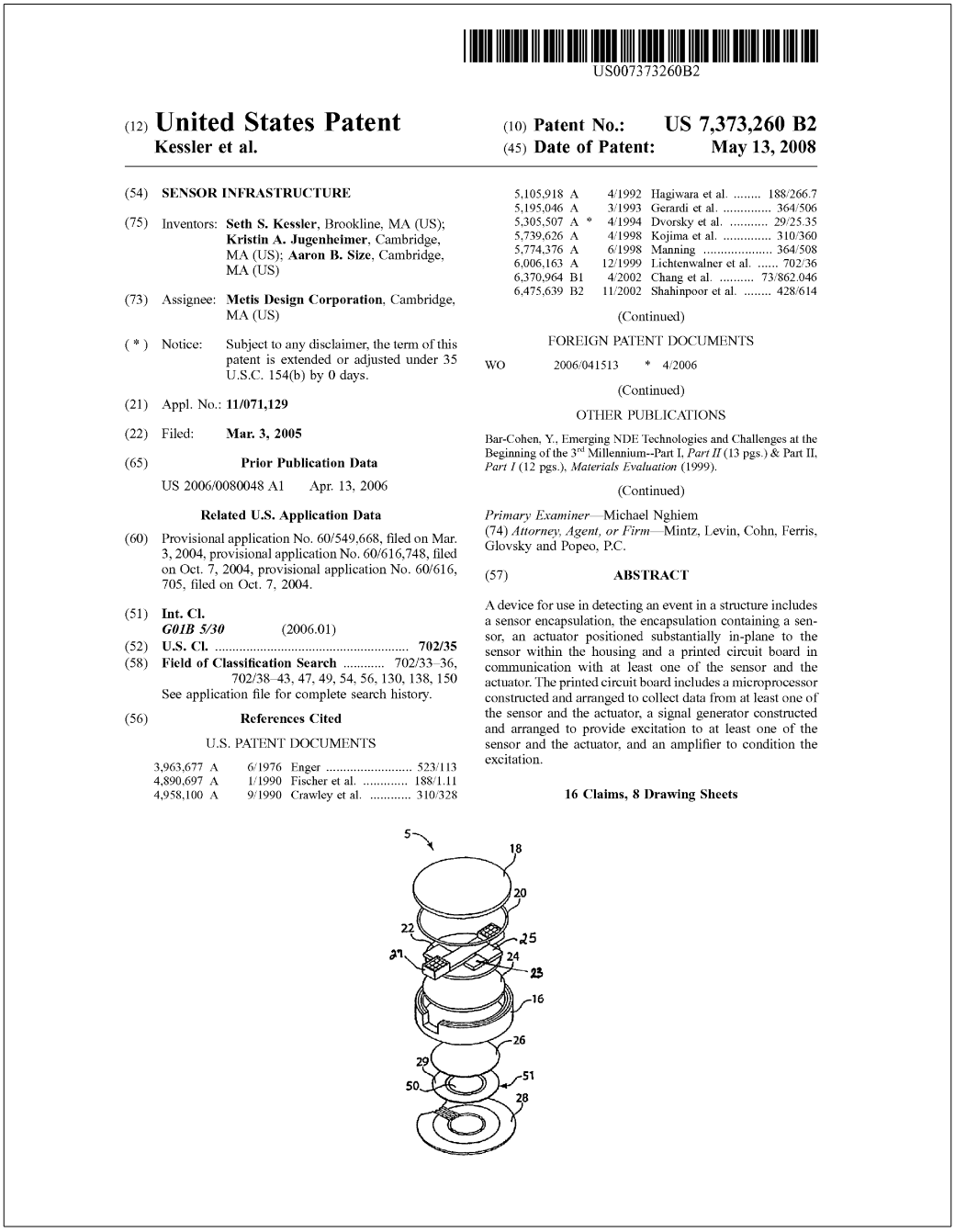


Figure A.13: Example of a product information brochure. The main goal of this web page is to illustrate the characteristics of a product commercialized by the company owning the web page. However, in the description of the product, the company also specified that it is covered by a patent it owns. Either that the company did so to comply with the requirements of the VPM legislation, or to advertise about quality and technologically advanced content, the web page clearly highlights the existence of a patent-product link.



US007373260B2

(12) **United States Patent**
Kessler et al.

(10) **Patent No.: US 7,373,260 B2**
(45) **Date of Patent: May 13, 2008**

(54) **SENSOR INFRASTRUCTURE**

(75) Inventors: **Seth S. Kessler**, Brookline, MA (US); **Kristin A. Jugenheimer**, Cambridge, MA (US); **Aaron B. Size**, Cambridge, MA (US)

(73) Assignee: **Metis Design Corporation**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/071,129**

(22) Filed: **Mar. 3, 2005**

(65) **Prior Publication Data**

US 2006/0080048 A1 Apr. 13, 2006

Related U.S. Application Data

(60) Provisional application No. 60/549,668, filed on Mar. 3, 2004, provisional application No. 60/616,748, filed on Oct. 7, 2004, provisional application No. 60/616,705, filed on Oct. 7, 2004.

(51) **Int. Cl.**
G01B 5/30 (2006.01)

(52) **U.S. Cl.** **702/35**

(58) **Field of Classification Search** 702/33–36, 702/38–43, 47, 49, 54, 56, 130, 138, 150
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,963,677	A	6/1976	Enger	523/113
4,890,697	A	1/1990	Fischer et al.	188/1.11
4,958,100	A	9/1990	Crawley et al.	310/328
5,105,918	A	4/1992	Hagiwara et al.	188/266.7
5,195,046	A	3/1993	Gerardi et al.	364/506
5,305,507	A *	4/1994	Dvorsky et al.	29/25.35
5,739,626	A	4/1998	Kojima et al.	310/360
5,774,376	A	6/1998	Manning	364/508
6,006,163	A	12/1999	Lichtenwalner et al.	702/36
6,370,964	B1	4/2002	Chang et al.	73/862.046
6,475,639	B2	11/2002	Shahinpoor et al.	428/614

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2006/041513	*	4/2006

(Continued)

OTHER PUBLICATIONS

Bar-Cohen, Y., Emerging NDE Technologies and Challenges at the Beginning of the 3rd Millennium--Part I, *Part II* (13 pgs.) & Part II, *Part I* (12 pgs.), *Materials Evaluation* (1999).

(Continued)

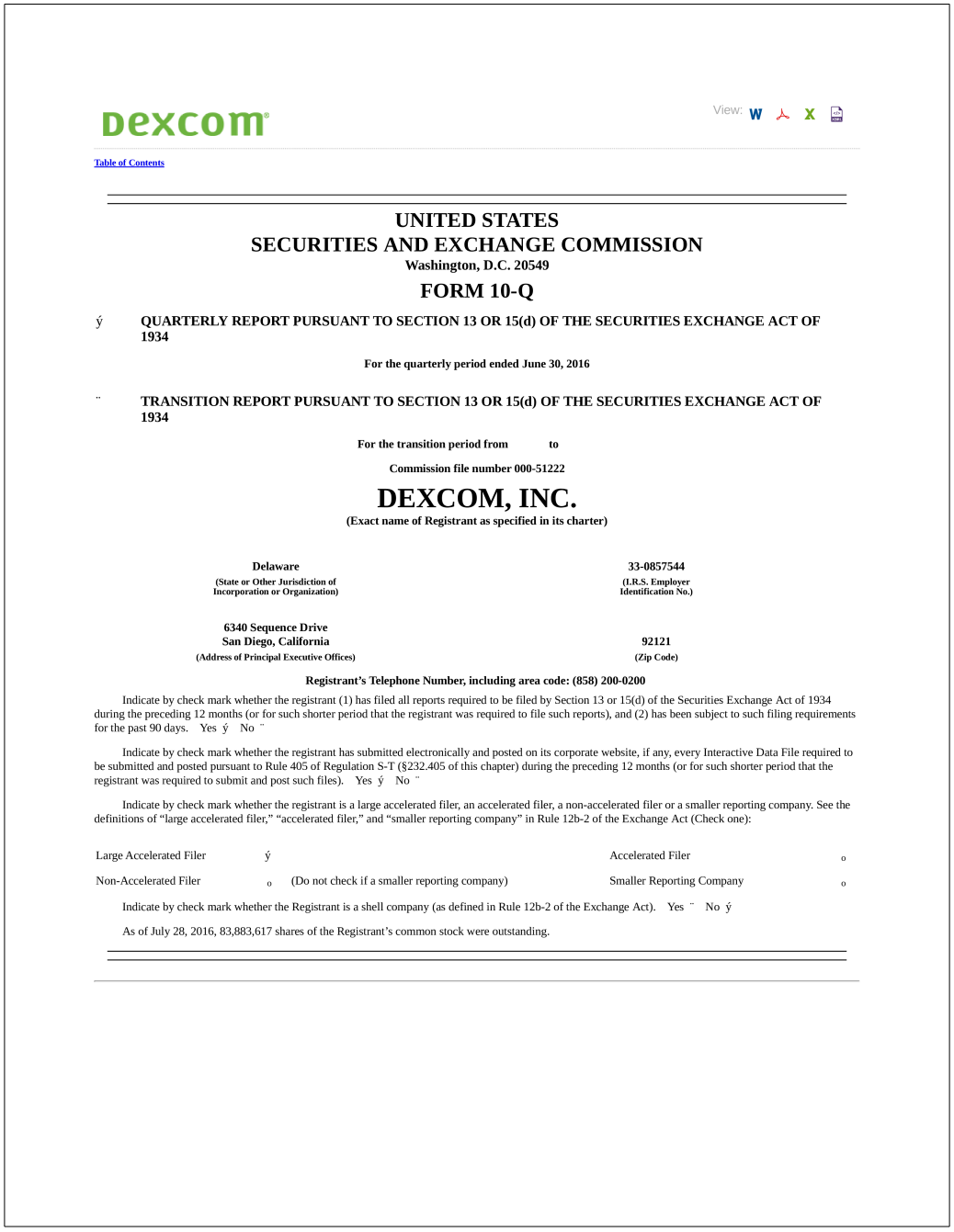
Primary Examiner—Michael Nghiem
(74) *Attorney, Agent, or Firm*—Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C.

(57) **ABSTRACT**

A device for use in detecting an event in a structure includes a sensor encapsulation, the encapsulation containing a sensor, an actuator positioned substantially in-plane to the sensor within the housing and a printed circuit board in communication with at least one of the sensor and the actuator. The printed circuit board includes a microprocessor constructed and arranged to collect data from at least one of the sensor and the actuator, a signal generator constructed and arranged to provide excitation to at least one of the sensor and the actuator, and an amplifier to condition the excitation.

16 Claims, 8 Drawing Sheets

Figure A.14: Example of a web page collected through the multi-step approach described in Appendix A.2.2, but that cannot be considered a relevant web page. Specifically, this is a web page pointing to the PDF version of the patent itself (`https://www.metisdesign.com/docs/US09839073.pdf`).



dexcom

View:

Table of Contents

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 10-Q

ý **QUARTERLY REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the quarterly period ended June 30, 2016

¨ **TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the transition period from to

Commission file number 000-51222

DEXCOM, INC.

(Exact name of Registrant as specified in its charter)

Delaware	**33-0857544**
(State or Other Jurisdiction of Incorporation or Organization)	**(I.R.S. Employer Identification No.)**
6340 Sequence Drive **San Diego, California**	**92121**
(Address of Principal Executive Offices)	**(Zip Code)**

Registrant's Telephone Number, including area code: (858) 200-0200

Indicate by check mark whether the registrant (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months (or for such shorter period that the registrant was required to file such reports), and (2) has been subject to such filing requirements for the past 90 days. Yes ý No ¨

Indicate by check mark whether the registrant has submitted electronically and posted on its corporate website, if any, every Interactive Data File required to be submitted and posted pursuant to Rule 405 of Regulation S-T (§232.405 of this chapter) during the preceding 12 months (or for such shorter period that the registrant was required to submit and post such files). Yes ý No ¨

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer or a smaller reporting company. See the definitions of "large accelerated filer," "accelerated filer," and "smaller reporting company" in Rule 12b-2 of the Exchange Act (Check one):

Large Accelerated Filer	ý	Accelerated Filer	o
Non-Accelerated Filer	o (Do not check if a smaller reporting company)	Smaller Reporting Company	o

Indicate by check mark whether the Registrant is a shell company (as defined in Rule 12b-2 of the Exchange Act). Yes ¨ No ý

As of July 28, 2016, 83,883,617 shares of the Registrant's common stock were outstanding.

Figure A.15: Example of a web page collected through the multi-step approach described in Appendix A.2.2, but that cannot be considered a relevant web page. Specifically, this is a form required by the U.S. Securities and Exchange Commission (SEC) that reports one of the targeted patent numbers, but for reasons different from those of interest for the present study. (`https://investors.dexcom.com/node/16146/html`).

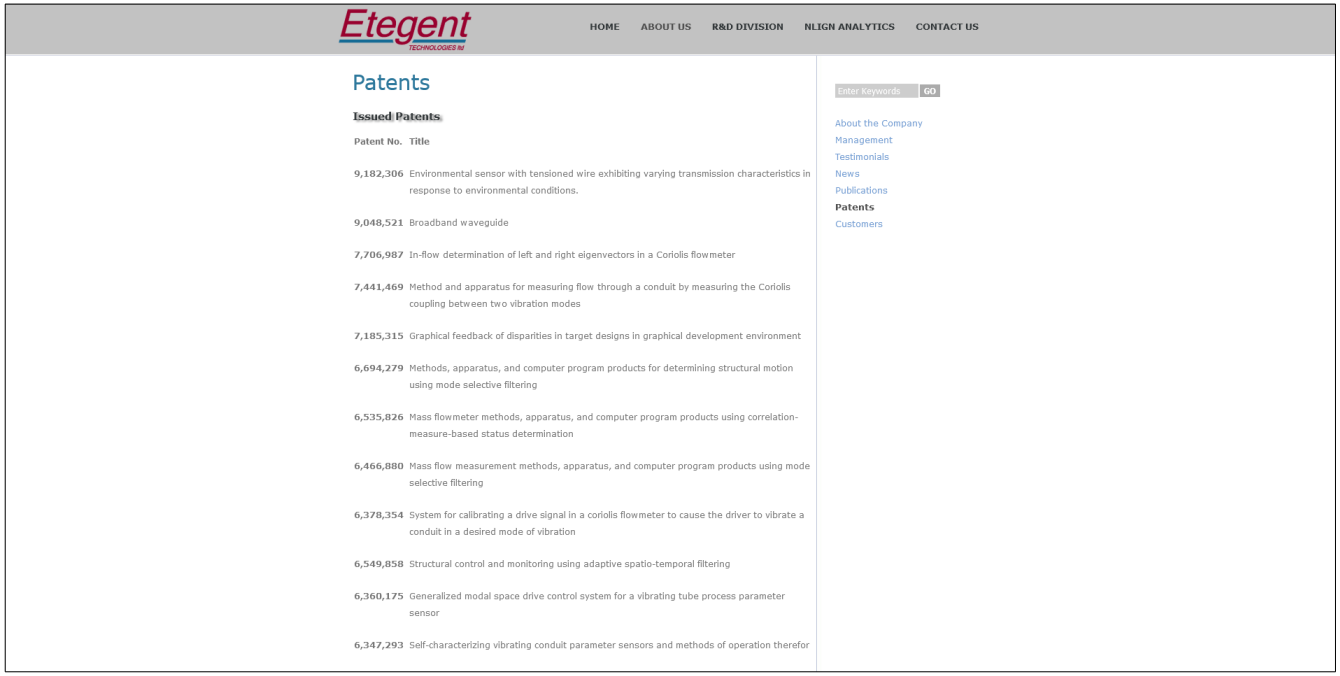


Figure A.16: Example of a web page collected through the multi-step approach described in Appendix A.2.2, but that cannot be considered a relevant web page. Specifically, this is a web page listing several patents without any connection to products commercialized by the company that owns the web page (`http://www.etegent.com/about-us/patents.html`).

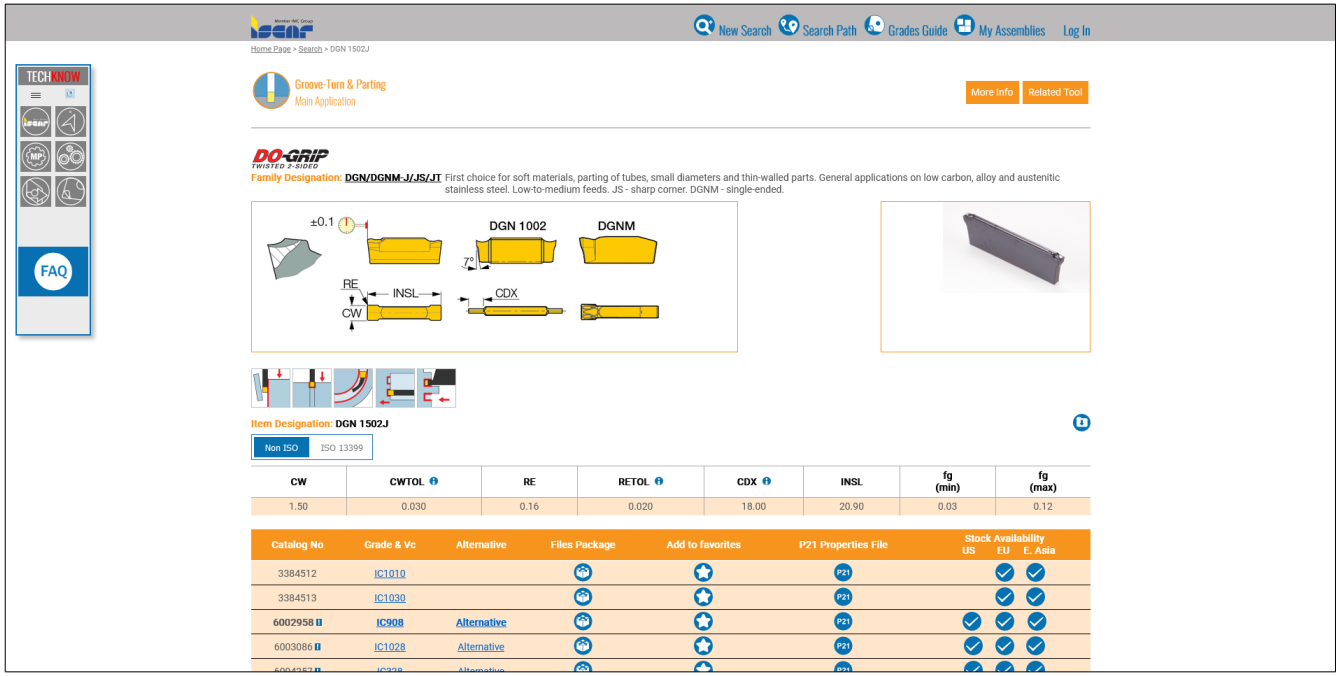


Figure A.17: Example of a web page collected through the multi-step approach described in Appendix A.2.2 but that cannot be considered a relevant web page. Specifically, this is a web page listing several products commercialized by the company that owns the web page, one of which has a *catalog number* compatible with a *patent number*: here the `6002958` (`https://www.iscar.com/ecat/e.asp?cat=6002958`).

2. Pages exclusively including the PDF file of legal forms required by the U.S. Securities and Exchange Commission (SEC). To identify these pages, the script downloads the PDF and parses it to check if the first 125 characters match one of the following regular expressions:
    - `United\W*States\W*Securities\W*and\W*Exchange\W*Commission\W*Washington`;
    - `\W*D\W?C\W*[0-9]+\W*(Form|Schedule)\W*([A-Z]|[0-9]1,2)-?([A-Z]|[0-9]1,2)`.

The logic behind the rules described above is quite straightforward. The likelihood that a web page exclusively displaying the PDF version of official documents granted by the USPTO or addressed to the SEC is actually a relevant web page is extremely low.

Pages matching the following characteristics belong instead to group (b):

1. Pages whose URL contains the sentence *virtual patent marking*. Specifically, the following regular expression has been used:
    - `(virtual|patents?).?marking`.
2. Pages matching one of the following regular expressions:
    - `America Invents Act`,
    - `35 U\.?S\.?C\.?(\ssect)?\W*287`.
    - `287\(a\) of Title 35 of the United States Code`
3. Pages in which a (registered) trademark symbol—®; (r); or ™—has been identified, in the surrounding (500 characters) of one of the patent numbers considered.
4. Pages whose text, in the surrounding of one of the patent numbers considered, contains expressions, such as *covered by* or *employs our patent*, frequently associated with patent-protection of a product. More specifically, the script looks for the following regular expressions:
    - `(^|\s)(protect|cover)[a-z]* (by|under|our)`;
    - `(^|\s)manufactur[a-z]* under`;
    - `patent\W*protected`;
    - `our patented`;
    - `(^|\s)(((emplo|appl(y|ie))[a-z]*|uses?) (the|a|our|a number|several|some)|using our) .{0,50}patent`.

The motivation behind the choice of the first two rules is quite clear. The reference to the VPM-related legislation provides a clear signal that the page in question is a relevant web page. The identification of the last two rules requires instead some additional explanation. Rule (3) builds on the idea that commercial products are often protected by trademarks. By convention, the ® and ™ symbols indicate that the preceding mark is a trademark. It follows that if we find one of these symbols in the proximity of the patent number we are searching for, we can expect that the web page in question provides a linkage between the patent and a commercial product. Finally, rule (4) builds on the observation that companies use a limited and recurring pattern of expression to describe a patent-product link on their corporate website. For instance, firms often use sentences like "our product XYZ is protected by patent 123" or "XYZ employs our 123 patent." Therefore, web pages that use variations of these expressions are likely to provide an actual patent-product link.

The automatic classifier processes 3,131 potentially relevant web pages. It identifies 713 pages (22.8 percent) as actually relevant web pages, 365 (11.7 percent) as pages that are not relevant, and marks the remaining 2,053 pages as 'uncertain.'

To validate the output of the automatic classifier, we visually inspect and assess the relevance of a subset of 500 web pages. The automatic classifier identifies 254 of these web pages as either a relevant web page or as an irrelevant one, whereas the remaining 246 web pages are labeled as 'uncertain'. To assess the performance of the automatic classifier, we focus on the first two groups. Table A.5 reports the results of the assessment in a confusion matrix contrasting the actual and the predicted outcomes. As the table shows, the script automatically classifies 44 web pages as irrelevant web pages and 210 as relevant ones, whereas the actual numbers (based on a manual inspection by the authors) are 70 and 184 web pages, respectively. Even if not perfect, the automatic classification reaches an overall *accuracy* of 89.8 percent, as it correctly identifies 228 true positives or true negatives out of a sample of 254 occurrences. In a further evaluation including potentially relevant web pages for both SBIR-funded and benchmark patents, we visually inspect and assess the relevance a subset of 1,109 web pages. The automatic classifier identifies 555 of these web pages as either relevant web pages or irrelevant ones, whereas the remaining 554 web pages are labeled as 'uncertain'. The script automatically classifies 141 web pages as irrelevant web pages and 414 web pages as relevant ones, whereas the actual numbers are 189 and 365 web pages, respectively, with an *accuracy* of 90.9 percent.

Manual classifier. The automatic classifier marks 2,053 web pages as 'uncertain'. To classify these last ones, we develop a manual classifier. You can find its code at `https://github.com/n3ssuno/iris-classifier/blob/8f311b18a869e4c08b65cced8c3029ba617d49a5/classify.py`.

Table A.5: Confusion matrix that reports the results of the assessment of the performance of the automatic classifier.

	Predicted		
Ground truth	Non rel.	Rel.	Tot.
Non relevant	44	26	70
Relevant	0	184	184
Total	44	210	254

Notes. Web pages are classified as containing relevant or non-relevant commercialization information. The ground truth (yellow) is based on a visual inspection of each web page by the authors. The predicted values (blue) are based on the results of the automatic classifier. The numbers exclude what the automatic classifier includes in the 'uncertain' category (*i.e.*, roughly two-thirds of the potentially relevant web pages).

As briefly discussed in the main text, Figure 1 shows the interface of the classifier and highlights its main features. The tool opens the web page in the left panel (blue box (7)) that must be classified using a browser-like interface. As the figure shows, the web page's URL appears in the address bar in the top-left corner (blue box (1)). The scroll bar on the left (2) allows navigating the page, whereas the arrow buttons in the top-left corner (3) allow the user to navigate back and forth between the web pages. Once the user has manually inspected the page and determined whether it qualifies as relevant or not, she can use the buttons in the right panel to classify it into several categories (6). The first three buttons indicate a positive patent-product link, whereas the others indicate that the page does not report an actual patent-product link. If the user needs to further explore the web page, she can open it in a standard web browser by clicking the button in the top-right corner of the figure (box (5)). To facilitate patent identification, the terminal that launches the classifier (not included in the figure) also prints out the patent numbers that are supposed to be found on a given page and the name of the patent assignee. Once the user decides on the page type, the classifier stores the information and proceeds to the following web page on the list. The box in the bottom-left corner reports the number of pages left to classify (4).

Classification results. The classification process described in this section led to the removal of 1,743 web pages out of the 3,131 collected through the scraping process. Therefore, according to our classification, about 44.3 percent of the web pages collected by the scraper are actually relevant web pages linked to at least one of the SBIR-funded patents previously identified.

Appendix A.2.3. Determining the time-to-market of a patented invention

Finally, we also estimate the time-to-market of patented inventions linked to commercial products.

To do so, we establish a date of creation for each relevant web page, and, for each patent, we consider the creation date of the earliest relevant web page associated with that patent as the commercialization date. We then measure the time-to-market of a patented invention as the difference between the commercialization date and the patent filing date.

We used two alternative methods to determine the creation date of each web page, and we preserved the earliest. First, we query the Wayback Machine of the Internet Archive (see `http://web.archive.org/`) for the earliest snapshot they have on record for each of the relevant web pages that we identify. Specifically, due to the way in which the Wayback Machine's APIs work, we search for the closest to January 1st, 1994. However, since the Web started to diffuse in the mid-1990s, this choice does not affect the results since it is virtually impossible to retrieve records earlier than this date. Second, we search for the URL of the relevant web page on Google Search, asking for results between January 1st, 1994, and December 31, 2021. We identify the earliest result available and extract the exact date provided by Google as the date of creation, if available.

Using this method, we are able to provide a time-to-market estimate for 209 patents out of the 225 for which we identified an actual direct commercialization path and for 466 out of the 498 with an *indirect path* to a commercial product.

Appendix A.3. Description of the database

This section presents each of the tables available in the database and the variables they include.

Appendix A.3.1. Main tables

Award table.

`award_id` It is the unique identifier of an award. It corresponds to the Procurement Instrument Identifier (PIID) of the award.

`date_start` It is the starting date of the award, as declared in the first observed transaction. This information is not present in DCADS. Therefore, the variable is set to NA for awards that have no transaction after October 2000.

`date_end` It is the (expected) ending date of the award, as declared in the first observed transaction.

`date_action_first` It is the date of the first transaction observed in the data.

`date_action_last` It is the date of the last transaction observed in the data.

`awarding_agency_id` It is the identification number (ID) of the awarding agency, as declared in the first observed transaction. Note that, for the database here described, it is always equal to `97` (Dept of Defense).

`awarding_subagency_id` It is the identification number (ID) of the awarding sub-agency, as declared in the first observed transaction. For data from DCADS, the raw codes have been translated into those used by USAspending.gov according to the following schema: `1` is `2100`; `2` is `1700`; `3` is `5700`; `4` is `97AS`. The code `5` has not been mapped to any code—but no contract matched with any patent; therefore, this fact is irrelevant to our purpose. The most represented sub-agencies are `2100` (Army); `1700` (Navy); `5700` (Air Force).

`awarding_office_id` Awarding office ID

`funding_agency_id` Funding agency ID, as declared in the first observed transaction. Present only for awards with a transaction from October 2000 onward. The same applies, logically, also for funding sub-agency and office.

`funding_subagency_id` Funding sub-agency ID

`funding_office_id` Funding office ID

`obligated_amount` It represents the sum of the monetary value (in dollars) of all the observed transactions of a given award.

`psc_id` It is the Product and Service Code ID of the first observed transaction. If it starts with `A`, it is an R&D contract and the last digit represents the R&D stage (`1` for *basic research*; `2` for *applied research*; `3-5` for *development*; `6-7` for *other*).

`sbir_sttr` It is the SBIR/STTR Phase of the award. In the (infrequent) case in which different transactions of the same award show different Phases, Phase II was preferred to Phase I, and SBIR to STTR.

Patent table.

`patent_id` It is the unique identifier of the patent and corresponds to the USPTO's patent number.

`application_id` It is the patent's application number.

`date_application` It is the patent's application date.

`date_grant` It is the patent's grant date.

`uspc_id` It is the first United States Patent Classification (USPC) class code attributed to the patent. This information is extracted from USPTO's PatEx database—See `https://www.uspto.gov/sites/default/files/documents/Appendix%20A.pdf` for a description of such data.

`is_federally_funded` This variable is TRUE if—according to PatentsView—the patent acknowledged a government interest on itself, and FALSE otherwise.

`cbsa_id` U.S. Core-Based Statistical Area (CBSA) where the majority of the inventors of the focal patent are located. A CBSA is a geographic area defined by the U.S. Office of Management and Budget (OMB) that consists of several *counties* anchored by an urban center of at least 10,000 people, plus adjacent counties that are socioeconomically tied to the urban center by commuting.

`family_id` It is the unique ID of the patent's INPADOC family, as attributed by PATSTAT. Note that there are missing data (NA) in this variable.

`date_priority` It is the priority date of the patent INPADOC family, as declared in PATSTAT.

`count_claim` It is the number of claims contained in the patent.

`count_bwd_pat_cit` It is the number of references to other patents listed in the patent.

`count_bwd_npl_cit` It is the number of non-patent literature (NPL) references contained in the patent.

`count_fwd_patent_cit` It is the number of citations received by the patent from other U.S. utility patents.

`count_fwd_pat_cit_10y_from_application` It is the number of citations received by the patent from other U.S. utility patents in the first 10 years from its application date.

`count_fwd_pat_cit_5y_from_appln` It is the number of citations received by the patent from other U.S. utility patents in the first 5 years from its application date.

`count_fwd_pat_cit_3y_from_appln` It is the number of citations received by the patent from other U.S. utility patents in the first 3 years from its application date.

`appln_fam_size` It is the number of different patent applications belonging to the INPADOC family of the patent.

`geo_fam_size` It is the number of different national patent offices at which at least one of the patent applications belonging to the INPADOC family of the patent has been applied.

Webpage table.

`url_id` It is the unique identifier of a web page.

`url` It is the URL of the web page where a number compatible with one of the patents of interest has been identified. The variable `is_relevant` in the patent-to-patent-marking associative table allows filtering only these pages later classified as true VPM pages.

`is_relevant` This variable is TRUE if the web page in question has been classified, following the procedure described in Appendix A.2.2 as providing information about actual patent-coverage of a product. This variable is useful since we decided to report all the web pages retrieved by the scraper in this table. Only the ones for which this variable is TRUE, are the ones later classified as actually providing evidence of a link between a patent and consumer goods.

`date_url_archives` This is the oldest date on which the web page has been archived by the Internet Archive initiative (`https://web.archive.org/`).

`date_url_google` This is the date provided by Google if we search for the detected URL, by filtering for results between 1994 and 2021, and select the first result.

`date_url` This is the minimum between the previous two variables. It is our best guess for the page creation date. We consider this a proxy of the commercialization of the product protected by the patent.

Legal entities table. This table joins and elaborates information from several PatentsView tables (`assignee`, `inventor`, `non_inventor_applicant`), USAspending.gov, DCADS, the SBIR website, and the USAspending.gov APIs.

`legalentity_id` It is the unique identifier of a legal entity. There is a one-to-one correspondence between a legal entity's ID and its `name`.

`name` It is the name of the legal entity. If this is a physical entity, the variable follows the *First Last* name convention. The names have been cleaned and standardized. Moreover, for cases with more than one DUNS number corresponding to one common name, the variable has been checked on USAspending.gov through the API provided—see `https://api.usaspending.gov/api/v2/recipient/duns/`.

`is_person` This variable is TRUE if it is a natural person, and FALSE if it is a legal person. We did our best to assign either a TRUE or a FALSE to this variable. However, some cases are unknown and, therefore, are left as NA, even though it is highly likely that these should be flagged as FALSE.

`is_sbir_sttr_recipient` This variable is TRUE if the legal entity is a recipient of at least one SBIR/STTR award, and FALSE otherwise.

Appendix A.3.2. Associative tables

Award to patent table.

`award_id` It is the unique identifier of an award.

`patent_id` It is the unique identifier of a patent.

Patent to agency table.

`patent_id` It is the unique identifier of a patent.

`awarding_agency_name` It is the name of the awarding agency of the patent. The information is provided as is in PatentsView. There can be more than one agency for each patent.

Patent to assignee table. This table contains information about the patent assignee(s) of each patent included in the data. For the vast majority of patents, the assignee is the organization (in general, a private corporation) that applied for the patent. There are a few cases in which none of the assignees is an organization; in this case, the variable `assignee_is_inventor` is equal to TRUE. Moreover, in a few cases, the assignee is not specified in PatentsView; in this case, we filled the gaps using information about the non-inventor applicants of the patent and the variable `assignee_is_non_inventor_applicant` is equal to TRUE.

`patent_id` It is the unique identifier of a patent.

`legalentity_id` It is the unique identifier of a legal entity.

`assignee_id` It is the unique identifier of the assignee linked, in PatentsView, to the name of the legal entity. For convenience, the variable refers to the 'assignee.' However, when no assignee is present on PatentsView, we used information about the non-inventor applicant (preferably) or the inventor of the patent. While in PatentsView each `assignee_id` corresponds to only one `name`, this is no longer the case here. Indeed, in a few cases, it has been necessary to clean the data in such a way that this bijective correspondence has been lost (*i.e.*, some `assignee_id` correspond to more than one `name`).

`assignee_is_non_inventor_applicant` This variable is TRUE if, according to PatentsView, the `assignee_id` is of a non-inventor applicant, and FALSE otherwise.

`assignee_is_inventor` This variable is TRUE if, according to PatentsView, the `assignee_id` is of an inventor, and FALSE otherwise.

Patent to patent forward citation table. This table provides a many-to-many relationship between patents and patents' forward citations—where with *forward citation* we mean a patent citing the *focal* patent. Both variables correspond, therefore, to a patent number provided by the USPTO. They can be used to link this table to any table containing information about the patents, first and foremost the `patent` table.

`patent_id` It is the unique identifier of the cited patent.

`citation_id` It is the unique identifier of the citing patent.

Patent to web page table. This table provides a many-to-many relationship between patents and web pages.

`patent_id` This is the unique identifier of each patent and corresponds to the patent number provided by the USPTO. The variable can be used to link this table to any table containing information about the patents, first and foremost the `patent` table.

`url_id` This is a unique identifier of each web page detected through the scraping process described in Appendix A.2.2. The page detected should contain a sequence of numbers identical to the patent number in question. The variable can be used to link this table with the `patent_marking` one, where further information about the web page in question is stored.

Appendix A.4. Data access and reproducibility

The data are available at `https://doi.org/10.5281/zenodo.16779954`, under a Creative Commons Attribution 4.0 International license—see `https://creativecommons.org/licenses/by/4.0/`.

The scraping part of the project is subject to changes in the evolution of the Web and its content; as such, its reproducibility is hardly possible. However, the code provided at `https://github.com/n3ssuno/iris-scraper/releases/tag/v0.9-rc` and `https://github.com/n3ssuno/iris-classifier/releases/tag/v0.9-rc` can be used to produce similar data. The code contained in these repositories is provided under an MIT license.

Appendix B. Supplementary Descriptive Statistics

As discussed in the main text, the aim of our paper is to introduce a novel web-based method for measuring invention commercialization by leveraging targeted web searches, applying this approach to assess commercialization outcomes in the U.S. DoD's SBIR/STTR program. To do so, we identify the universe of USPTO patents that acknowledge support by the program and establish whether these patents are linked to commercial products via the web-based approach discussed in Appendix A. To establish a benchmark against which to evaluate the program, we construct a benchmark group composed of patents with similar characteristics to the SBIR-funded patents in the sample. For each SBIR-funded patent, we randomly select up to three benchmark patents that match the following requirements: (i) they share the same filing year with the SBIR-funded patent; (ii) they belong to the same main USPC technological class as the SBIR-funded patents; (iii) they are assigned to a private company classified as a small business by the USPTO. This classification is based on the maintenance fee paid, as small enterprises pay a reduced fee. We retrieve this piece of information from the USPTO's Patent Examination Research Dataset (PatEx) database (Graham et al., 2015).

The final dataset consists of 2,896 SBIR-funded patents, assigned to 1,060 distinct companies, and 4,622 benchmark patents, assigned to 3,892 distinct companies. By design, SBIR-funded and benchmark patents have filing years distributed within the same time frame, ranging from 1984 to 2019. We further discuss the similarities and differences between the two sets below.

Appendix B.1. Comparison of SBIR-funded and benchmark patents

Figures B.18–B.19, complementing Figures 4–5 in the main text, report the distribution of SBIR-funded and benchmark patents by application year and NBER technological category. The figure displays the share of patents with a *direct* or *indirect path* to a commercial product. As the figure shows, the temporal and technological distributions of the two groups are, by construction, very similar. However, the commercialization rate of SBIR-funded patents appears to be higher than for benchmark patents. The gap between the two groups widens over time and seems to be more pronounced in specific fields such as Chemicals and Electronics.

Figure B.20 complements the evidence provided by Table 3 in the main text. The panels show, for each of the control variables included in the regressions (see the main text and Appendix C), two boxplots, one for the SBIR-funded and the other for the benchmark patents. The jittered points on the back, in light gray, help to visualize the actual distribution of the variables. Overall, the SBIR-funded patents seem to make more claims and cite more non-patent literature citations to the relevant prior

art, compared to the benchmark patents. However, the mean difference is less than a claim and about one NPL citation more for the first group of patents and the significance for the NPL citations is mild. About the citations to other prior-art patents, the difference between SBIR-funded and benchmark patents is rather in favor of the benchmark patents. Instead, it appears that the SBIR-funded patents have been extended to a smaller number of countries, compared to the benchmark patents. When it comes to citations received from other patents in the first three years of patent life, we observe no structural differences between the two groups.

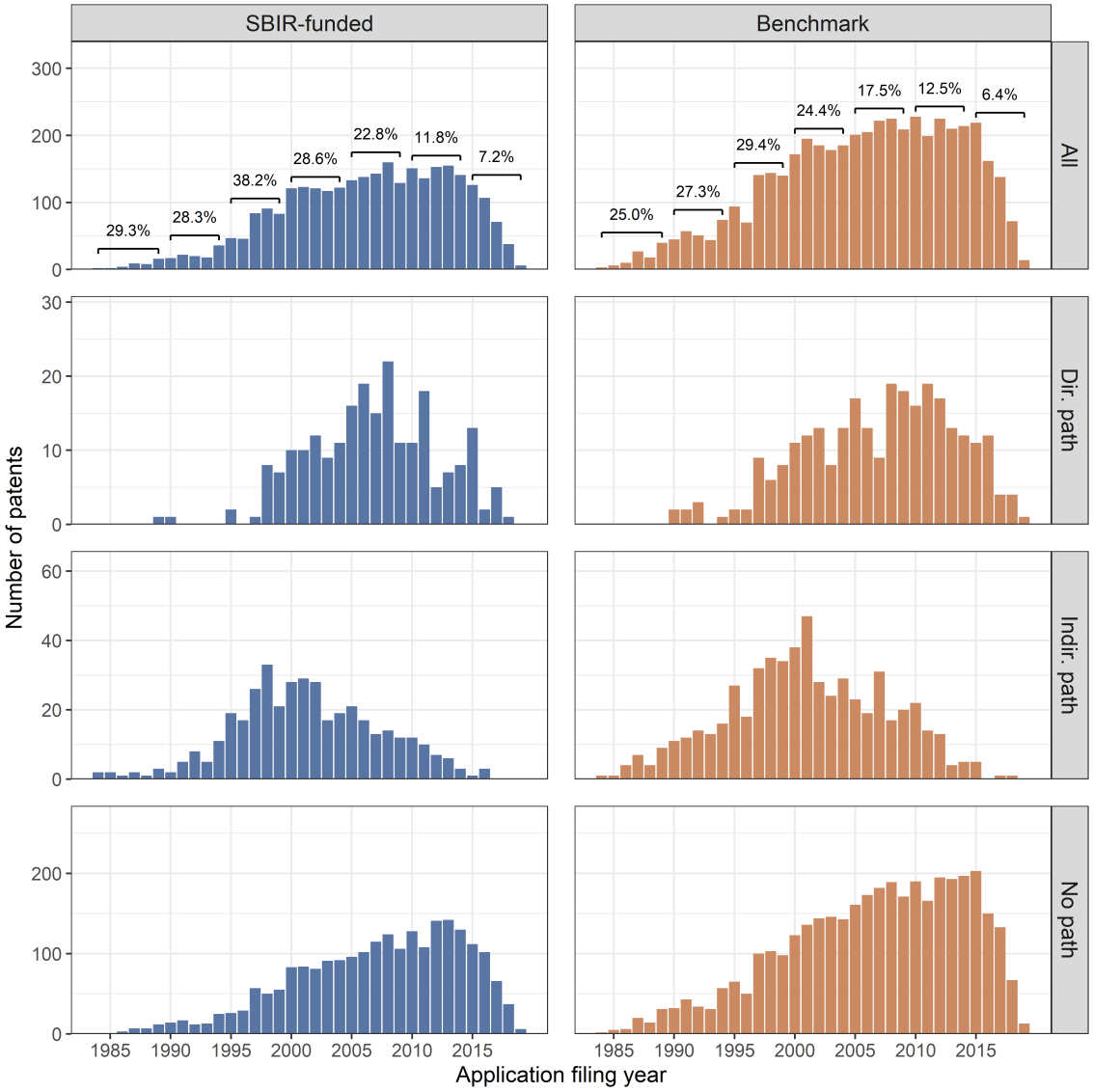


Figure B.18: Distribution of SBIR-funded (blue) and benchmark (orange) patents by patent's application year. The top panels show all the patents included in the data set. The numbers on top of the bars report the percentage of patents for which we detected a *path* to a product, for each five-year group. The rest of the figure distinguishes between patents for which we did not find any commercialization trace (No path), those directly protecting a product (Dir. path), and those cited by a product-protecting patent (Indir. path). Notice that a patent both directly and indirectly linked to a relevant page is counted among the *direct paths*.

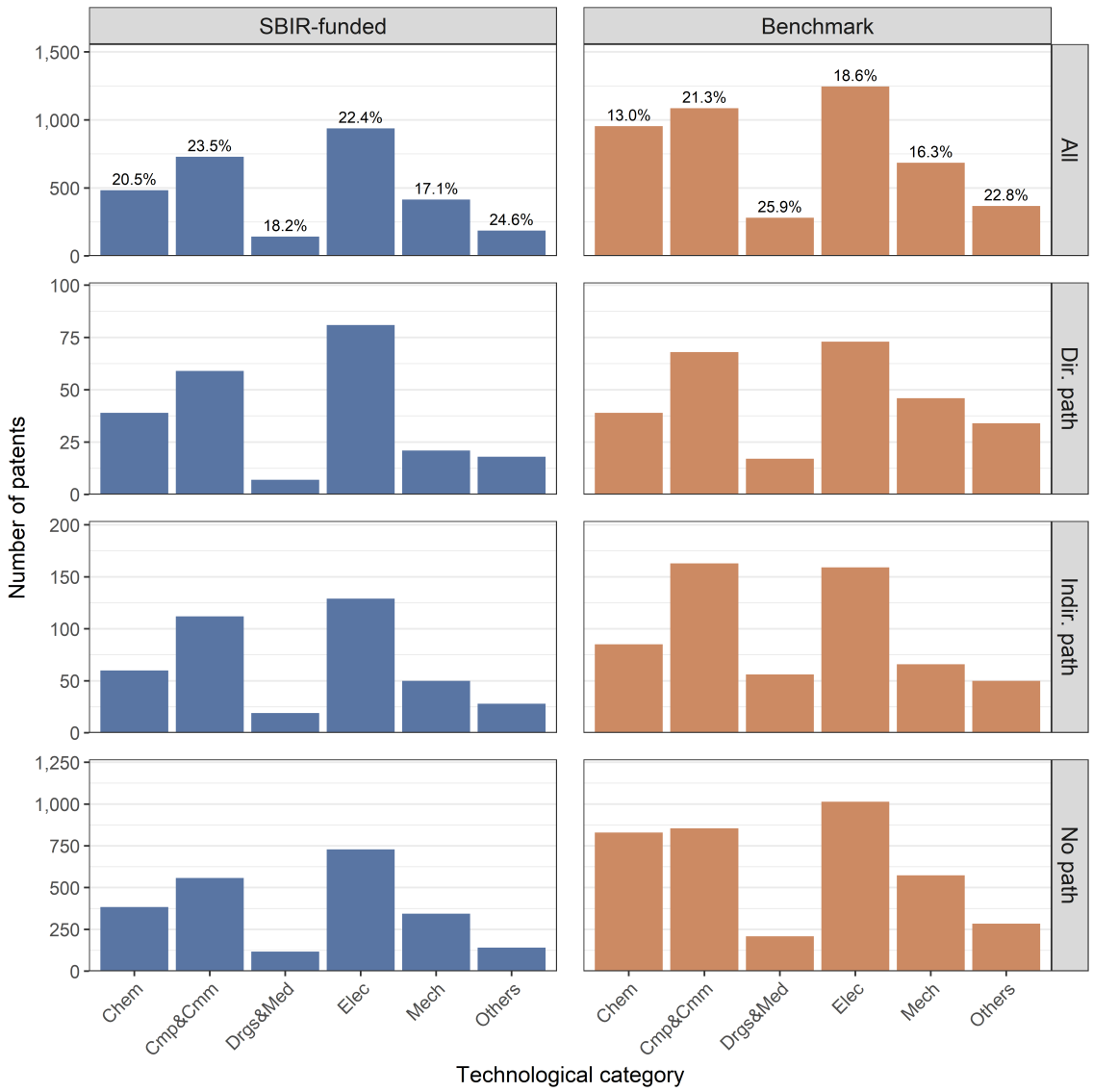


Figure B.19: Distribution of SBIR-funded (blue) and benchmark (orange) patents by patent's technological category (NBER). The top panels show all the patents included in the data set. The numbers on top of the bars report the percentage of patents for which we detected a *path* to a product. The rest of the figure distinguishes between patents for which we did not find any commercialization trace (No path), those directly protecting a product (Dir. path), and those cited by a product-protecting patent (Indir. path). Notice that a patent both directly and indirectly linked to a relevant page is counted among the direct paths.

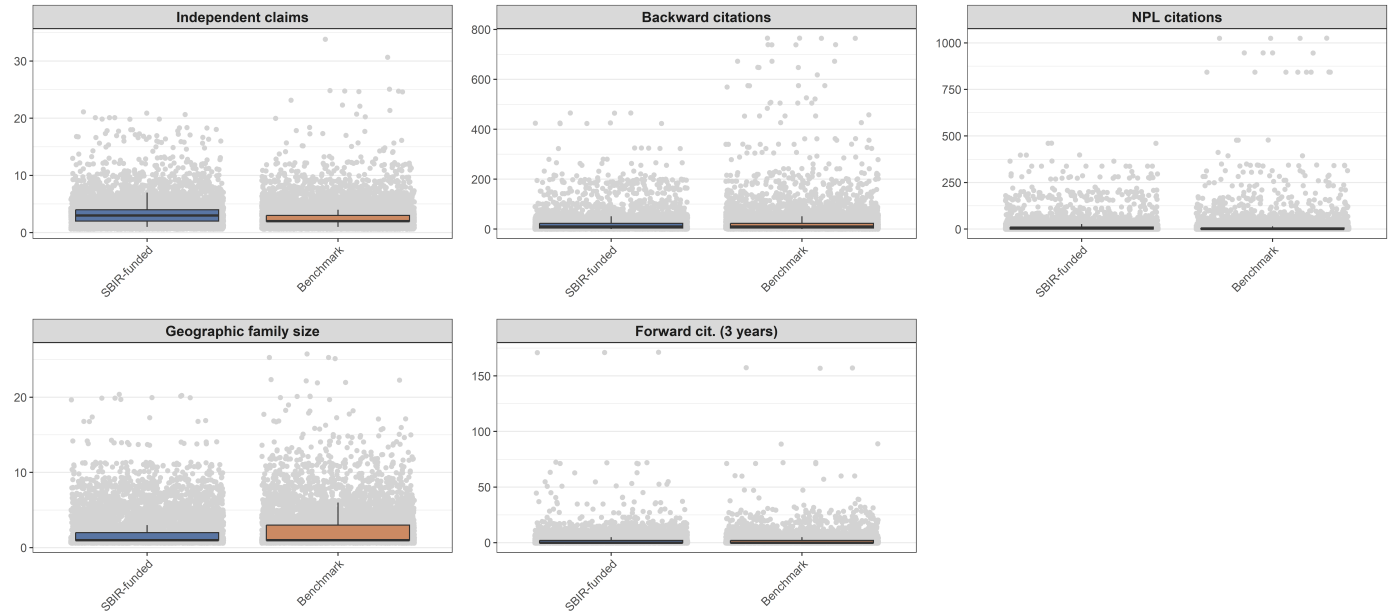


Figure B.20: Descriptive statistics for patent-quality indicators for SBIR-funded and benchmark patents (Squicciarini et al., 2013).

Appendix C. Supplementary Results

As discussed in the main text, in this paper, we compare the commercialization performance of SBIR-funded and benchmark patents. We adopt standard regression analysis to contrast the commercialization probability of SBIR-funded and benchmark inventions. More specifically, we estimate the following linear probability model (LPM):

$$\Pi_i = \beta_0 + \beta_1 \cdot \text{SBIR}_i + \mathbf{X_i} \cdot \beta + \gamma_i + \delta_i + \varepsilon_i$$

where SBIR_i is an indicator function equal to 1 if patent i acknowledges funding from the SBIR program of the Department of Defense, and 0 otherwise. To account for patent heterogeneity, we also included several patent-level control variables (in logarithmic scale) in the equation ($\mathbf{X_i}$): namely, the number of independent claims made by the patent (`claims`); the number of citations to other patents (`bwd_cit`) and to the non-patent literature (`npl_cit`) made by patent i; the number of countries in which patent i has been applied (`geo_fam`); the number of citations received by patent i in the first three years after its application date (`fwd_cit`). Lastly, we include dummy variables for the year of first priority γ_i and for the USPC patent class δ_i of patent i to control for some time- or technology-dependent specific factor. Moreover, Π_i is an indicator function equal to 1 if patent i is linked to a product, and 0 otherwise. We will distinguish between three kinds of commercialization path: any, direct, and indirect. A path is *direct* if patent i is listed in one of the web pages we classified as providing information about product-coverage of a patent. Instead, a path is *indirect* if a patent found in one of these pages cites patent i. Finally, *any* includes both *direct* and *indirect paths*.

For each LPM proposed, we also estimated a corresponding Probit model for which we report the marginal effects.

To estimate the parameters of the empirical models considered, each observation has been weighted so that, for each SBIR-funded–benchmark group (*i.e.*, patents with the same USPC patent class and same application year), the sum of the weights of SBIR-funded and benchmark patents is the same. As discussed, we select up to three benchmark patents for each SBIR-funded patent in our sample. To make sure our sample is balanced, we group together SBIR-funded and benchmark patents based on their application year and USPC main patent class. For each of these groups, we count SBIR-funded and benchmark patents it contains, and we assign to the benchmark patents a weight equal to the fraction between these two values ($w_g = {}^{|\mathbb{S}_g|}/_{|\mathbb{B}_g|}$), where w_g is the weight assigned to each benchmark patent in group g, $|\mathbb{S}_g|$ the number SBIR-funded patents and $|\mathbb{B}_g|$ that of the benchmark patents in group g. Figure C.21 describes the distribution of benchmark weights by SBIR-funded–

benchmark group. The weight assigned to each SBIR-funded patent is, instead, always equal to one. In other words, the weights assigned to the benchmark patents sum to the number of the SBIR-funded patents they are linked with.

We also computed several alternative weights for the benchmark patents, based on the characteristics of the awards acknowledged in the connected SBIR-funded patents. For instance, to run a regression including these patents acknowledging Phase I procurement contracts only, we proceeded in two steps. First, we zero-weighted all the SBIR-funded patents acknowledging no Phase I awards. Second, we looped through each SBIR-funded–benchmark group and re-compute the weights of each benchmark patent in the group with the usual formula, $w_g = {}^{|\mathbb{S}_g|}/_{|\mathbb{B}_g|}$ where $|\mathbb{S}_g|$ is, this time, equal to the number of positive-weighted SBIR-funded patents only. We did so for several awards' characteristics: among others, the award SBIR/STTR Phase and the R&D stage and kind of activity carried out in compliance with the contract.

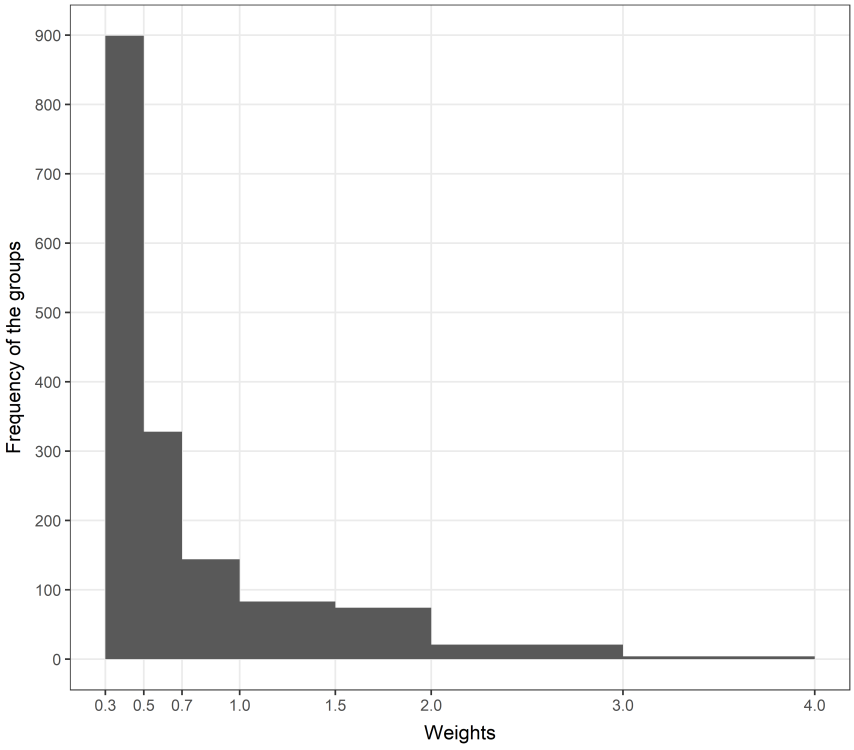
In the main text, we reported only the estimated value of β_1 for some key LPMs. The first part of this appendix provides additional details on the results presented in the main text. In the second part, we report the results from additional model specifications, which are not included in the main text.

Appendix C.1. Extended Findings and Complete Model Estimates

The following tables are reported:

Tab. C.6 Baseline models. The first model includes patents with any possible year of first priority. The others include, respectively, only patents with year of first priority earlier than 2000, between 2000 and 2009, and later than 2009. Looking at the first sub-table, the association between the SBIR program and commercialization outcome is stronger for older patents, and not even significant for the more recent period. Instead, for *direct paths*, this association is only slightly significant for the pre-2000 period. For *indirect paths*, the relation is positive and significant only for the pre-2000 period, while it becomes non-significant for the other two sub-periods considered. Taken together, these observations suggest that the lack of a significant relationship between recent patents and commercialization via *indirect paths* may reflect the fact that newer inventions have not yet had sufficient time to reach the consumer market. On the other hand, given that our approach relies on web searches, it is reasonable to assume that traces of *direct paths* for some of the oldest patents may never have appeared online, either because they predate the widespread adoption of web-based documentation or because they have since disappeared due to the assignee no longer maintaining its corporate website.

Figure C.21: Frequency distribution of the weights by SBIR-funded–benchmark group.



The x-axis reports the values $w_g = {}^{|\mathbb{S}_g|}/_{|\mathbb{B}_g|}$ used to weight the benchmark patents of each SBIR-funded–benchmark group in the regression models, where $|\mathbb{S}_g|$ is the number of SBIR-funded patents in group g and $|\mathbb{B}_g|$ the number of benchmark patents in the same group. The height of the bars represents the frequency of the groups using a given weight for their benchmark patents. Therefore, there are about 900 groups in which there are three benchmark patents for each SBIR-funded patent; about 300 in which the ratio is two benchmark patents for each SBIR-funded patent, and so on.

Tab. C.7 Models including all the patents acknowledging a Phase II award or a Phase I award that has been lately *extended* to a Phase II award—It is worth noticing that this last Phase II award did not necessarily lead to a patent.

Tab. C.8 Models that include all patents linked to a Phase I award that was never subsequently *extended* to a Phase II award.

Tab. C.9 Models including all patents acknowledging awards performing, respectively, *Basic Research*; *Applied Research*; or *Development* activities.

Tab. C.10 Additional details for the difference-in-differences (DiD) estimator that focuses on SBIR-funded patents awarded in the years immediately before and after the policy change (1996–2005) described in the main text.

Fig. C.22 Comparison of SBIR-funded and benchmark patents with respect to the time it takes for a patent to reach the consumer market.

Appendix C.2. Additional regression tables

The following tables are reported:

Tab. C.11 Models exploring in more detail the results of the *indirect paths* reported in Tab. C.6. The first sub-table shows results where the dependent variable is equal to one if there exists an *indirect path* that consists of a product-protecting patent citing another SBIR-funded or benchmark patent of the same assignee; *i.e.*, a *self-citation*. Instead, the second sub-table includes only patents never cited by other patents of the same assignee (patents without *self-citations*), and the dependent variable is equal to one if any of these citing patents is listed on a relevant web page. Lastly, the third sub-table includes only patents that received citations by other patents of the same assignee, but the dependent variable is equal to one only if the citing patent listed on a relevant web page is not the 'self-citation' itself (patents without product-protecting *self-citations*). For the patents indirectly linked to a product through a *self-citation*, 41 percent of them are linked to a product through at least a federally-funded patent; more specifically, on average, 16 percent of the citing patents are federally-funded.

Tab. C.12 Models estimated exclusively using commercialization as observed from 'proper' VPM pages (*i.e.*, excluding product brochures and other web pages).

Tab. C.13 Models including only one benchmark patent for each SBIR-funded patent, selected at random among the available ones.

Tab. C.14 Models including only SBIR-funded patents, strictly speaking (not STTR-funded ones).

Tab. C.15 Models including only STTR-funded patents (not SBIR-funded ones, strictly speaking)

Table C.6: Baseline models.

Dep. var.:	OLS				Probit			
Any path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.035***	0.060**	0.035**	0.015	0.032***	0.066**	0.035**	0.009
	(0.010)	(0.025)	(0.016)	(0.017)	(0.010)	(0.028)	(0.016)	(0.014)
log(`claims`)	0.017**	0.016	0.023*	0.003	0.017**	0.020	0.025*	0.003
	(0.009)	(0.019)	(0.013)	(0.016)	(0.008)	(0.021)	(0.013)	(0.013)
log(`bwd_cit`)	0.007	0.019	0.010	-0.003	0.008	0.024	0.011	-0.003
	(0.005)	(0.015)	(0.008)	(0.009)	(0.005)	(0.017)	(0.008)	(0.007)
log(`npl_cit`)	-0.007	-0.017	-0.008	-0.003	-0.008*	-0.020	-0.009	-0.003
	(0.004)	(0.011)	(0.007)	(0.007)	(0.004)	(0.013)	(0.007)	(0.006)
log(`geo_fam`)	0.019**	0.044**	0.005	0.011	0.020***	0.051***	0.007	0.008
	(0.008)	(0.018)	(0.012)	(0.015)	(0.008)	(0.019)	(0.012)	(0.011)
log(`fwd_cit`)	0.106***	0.116***	0.101***	0.093***	0.091***	0.126***	0.092***	0.066***
	(0.008)	(0.016)	(0.012)	(0.019)	(0.007)	(0.017)	(0.010)	(0.011)
Constant	0.977***	0.487***	0.517***	0.545				
	(0.218)	(0.188)	(0.197)	(0.382)				
Observations	7216	1699	3385	1514	7216	1699	3385	1514
R^2	0.151	0.205	0.115	0.126				
Pseudo R^2					0.150	0.171	0.111	0.163

Dep. var.:	OLS				Probit			
Direct path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.021***	0.029*	0.030**	0.003	0.015**	0.013	0.025**	0.002
	(0.007)	(0.017)	(0.012)	(0.018)	(0.006)	(0.012)	(0.011)	(0.015)
log(`claims`)	0.001	-0.001	0.004	-0.006	0.001	0.001	0.004	-0.006
	(0.006)	(0.014)	(0.010)	(0.017)	(0.005)	(0.009)	(0.009)	(0.014)
log(`bwd_cit`)	0.007*	0.023**	0.005	0.005	0.006*	0.019**	0.004	0.005
	(0.004)	(0.010)	(0.006)	(0.010)	(0.003)	(0.008)	(0.006)	(0.008)
log(`npl_cit`)	-0.001	-0.003	-0.002	-0.002	-0.002	-0.004	-0.003	-0.003
	(0.003)	(0.008)	(0.005)	(0.007)	(0.003)	(0.005)	(0.005)	(0.006)
log(`geo_fam`)	0.005	0.006	-0.000	0.000	0.005	0.003	0.002	-0.001
	(0.006)	(0.011)	(0.011)	(0.016)	(0.005)	(0.008)	(0.009)	(0.012)
log(`fwd_cit`)	0.014***	0.003	0.019**	0.024	0.012***	0.001	0.017**	0.020*
	(0.005)	(0.010)	(0.009)	(0.016)	(0.004)	(0.007)	(0.007)	(0.011)
Constant	0.505***	0.017	0.508**	0.532				
	(0.192)	(0.101)	(0.201)	(0.382)				
Observations	6552	1072	2850	1178	6552	1072	2850	1178
R^2	0.053	0.162	0.068	0.086				
Pseudo R^2					0.087	0.221	0.087	0.120

Dep. var.:	OLS				Probit			
Indirect path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.025**	0.060**	0.018	0.008	0.019**	0.066**	0.017	0.001
	(0.010)	(0.025)	(0.015)	(0.015)	(0.009)	(0.028)	(0.014)	(0.009)
log(`claims`)	0.017**	0.017	0.021*	0.001	0.015**	0.021	0.023**	0.001
	(0.008)	(0.019)	(0.011)	(0.016)	(0.007)	(0.021)	(0.011)	(0.008)
log(`bwd_cit`)	0.006	0.016	0.008	-0.006	0.007	0.022	0.009	-0.004
	(0.005)	(0.015)	(0.007)	(0.008)	(0.005)	(0.017)	(0.007)	(0.004)
log(`npl_cit`)	-0.005	-0.014	-0.005	-0.003	-0.005	-0.017	-0.005	-0.000
	(0.004)	(0.011)	(0.006)	(0.006)	(0.004)	(0.013)	(0.006)	(0.004)
log(`geo_fam`)	0.016**	0.043**	-0.001	0.020	0.017**	0.051***	-0.000	0.010
	(0.008)	(0.018)	(0.011)	(0.016)	(0.007)	(0.019)	(0.011)	(0.007)
log(`fwd_cit`)	0.120***	0.123***	0.119***	0.117***	0.092***	0.135***	0.099***	0.044***
	(0.008)	(0.015)	(0.011)	(0.020)	(0.006)	(0.017)	(0.009)	(0.008)
Constant	0.786***	0.473**	0.267	0.433**				
	(0.216)	(0.186)	(0.186)	(0.194)				
Observations	6769	1676	3263	977	6769	1676	3263	977
R^2	0.192	0.210	0.138	0.192				
Pseudo R^2					0.211	0.177	0.151	0.310

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.7: Models including only patents funded through at least a Phase II award or a Phase I award that has been lately *extended* to a Phase II award.

Dep. var.:	OLS				Probit			
Any path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.039***	0.073**	0.043**	0.018	0.036***	0.079**	0.043***	0.010
	(0.011)	(0.029)	(0.017)	(0.019)	(0.011)	(0.032)	(0.017)	(0.015)
log(claims)	0.017*	0.022	0.022	0.001	0.016*	0.025	0.024*	0.002
	(0.009)	(0.022)	(0.014)	(0.018)	(0.009)	(0.024)	(0.014)	(0.014)
log(bwd_cit)	0.005	0.035**	0.006	-0.005	0.006	0.043**	0.006	-0.003
	(0.006)	(0.018)	(0.008)	(0.010)	(0.006)	(0.019)	(0.009)	(0.008)
log(npl_cit)	-0.003	-0.022	-0.005	-0.002	-0.004	-0.026*	-0.006	-0.002
	(0.005)	(0.014)	(0.008)	(0.007)	(0.005)	(0.015)	(0.007)	(0.006)
log(geo_fam)	0.017*	0.038*	0.008	0.005	0.018**	0.045**	0.010	0.004
	(0.009)	(0.021)	(0.013)	(0.016)	(0.008)	(0.022)	(0.013)	(0.012)
log(fwd_cit)	0.107***	0.125***	0.103***	0.095***	0.091***	0.136***	0.094***	0.068***
	(0.009)	(0.018)	(0.012)	(0.020)	(0.007)	(0.020)	(0.011)	(0.011)
Constant	0.941***	0.166	0.542***	0.578				
	(0.230)	(0.231)	(0.197)	(0.379)				
Observations	6084	1290	2966	1245	6084	1290	2966	1245
R^2	0.156	0.222	0.123	0.136				
Pseudo R^2					0.155	0.187	0.118	0.174

Dep. var.:	OLS				Probit			
Direct path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.027***	0.039*	0.042***	-0.002	0.021***	0.021	0.036***	-0.004
	(0.008)	(0.022)	(0.013)	(0.020)	(0.007)	(0.015)	(0.012)	(0.016)
log(claims)	0.000	-0.001	0.004	-0.011	0.001	0.000	0.006	-0.011
	(0.007)	(0.017)	(0.011)	(0.019)	(0.006)	(0.011)	(0.009)	(0.015)
log(bwd_cit)	0.007	0.045***	0.002	0.004	0.006	0.036***	0.001	0.006
	(0.004)	(0.013)	(0.007)	(0.011)	(0.004)	(0.010)	(0.007)	(0.009)
log(npl_cit)	0.000	-0.011	0.001	-0.003	-0.001	-0.011	-0.000	-0.004
	(0.004)	(0.011)	(0.006)	(0.008)	(0.003)	(0.008)	(0.005)	(0.007)
log(geo_fam)	0.005	0.011	0.000	-0.007	0.005	0.004	0.003	-0.006
	(0.007)	(0.014)	(0.012)	(0.017)	(0.006)	(0.010)	(0.010)	(0.014)
log(fwd_cit)	0.017***	0.006	0.024**	0.025	0.014***	0.001	0.021***	0.021*
	(0.006)	(0.014)	(0.010)	(0.017)	(0.005)	(0.009)	(0.008)	(0.011)
Constant	0.574***	0.157	0.517***	0.579				
	(0.213)	(0.214)	(0.199)	(0.388)				
Observations	5334	717	2479	981	5334	717	2479	981
R^2	0.064	0.187	0.079	0.103				
Pseudo R^2					0.097	0.244	0.099	0.140

Dep. var.:	OLS				Probit			
Indirect path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.026***	0.076***	0.022	0.009	0.020**	0.084***	0.019	0.003
	(0.010)	(0.029)	(0.016)	(0.018)	(0.010)	(0.032)	(0.015)	(0.010)
log(claims)	0.017*	0.020	0.017	0.005	0.015**	0.025	0.020*	0.005
	(0.009)	(0.022)	(0.012)	(0.019)	(0.008)	(0.024)	(0.012)	(0.009)
log(bwd_cit)	0.005	0.029	0.006	-0.007	0.006	0.037*	0.006	-0.005
	(0.005)	(0.018)	(0.008)	(0.009)	(0.005)	(0.020)	(0.008)	(0.005)
log(npl_cit)	-0.003	-0.017	-0.003	0.000	-0.003	-0.022	-0.003	0.002
	(0.005)	(0.014)	(0.007)	(0.007)	(0.004)	(0.015)	(0.007)	(0.004)
log(geo_fam)	0.015*	0.037*	0.002	0.017	0.015**	0.045**	0.003	0.011
	(0.008)	(0.021)	(0.012)	(0.017)	(0.007)	(0.022)	(0.012)	(0.008)
log(fwd_cit)	0.120***	0.133***	0.119***	0.119***	0.090***	0.146***	0.100***	0.047***
	(0.008)	(0.018)	(0.012)	(0.022)	(0.006)	(0.019)	(0.009)	(0.009)
Constant	0.746***	0.229	0.287	0.435**				
	(0.231)	(0.225)	(0.192)	(0.196)				
Observations	5700	1257	2833	783	5700	1257	2833	783
R^2	0.196	0.223	0.143	0.192				
Pseudo R^2					0.217	0.190	0.154	0.308

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.8: Models including only patents funded through at least a Phase I award that has never been lately *extended* to a Phase II award.

Dep. var.:	OLS				Probit			
Any path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.010	0.026	0.004	0.022	0.004	0.032	-0.007	-0.001
	(0.021)	(0.043)	(0.036)	(0.056)	(0.021)	(0.048)	(0.034)	(0.054)
log(`claims`)	0.001	-0.014	-0.004	0.021	0.003	-0.015	-0.001	0.020
	(0.017)	(0.032)	(0.028)	(0.055)	(0.017)	(0.036)	(0.026)	(0.048)
log(`bwd_cit`)	0.000	-0.025	0.029	0.003	0.005	-0.028	0.034*	-0.001
	(0.011)	(0.024)	(0.019)	(0.023)	(0.011)	(0.029)	(0.019)	(0.020)
log(`npl_cit`)	-0.020**	-0.021	-0.021	-0.005	-0.026***	-0.032	-0.026*	-0.000
	(0.009)	(0.019)	(0.015)	(0.023)	(0.009)	(0.023)	(0.015)	(0.022)
log(`geo_fam`)	0.033**	0.054*	0.008	0.072	0.036**	0.061*	0.015	0.068*
	(0.016)	(0.029)	(0.029)	(0.044)	(0.015)	(0.033)	(0.027)	(0.037)
log(`fwd_cit`)	0.111***	0.096***	0.118***	0.100**	0.109***	0.115***	0.114***	0.090***
	(0.015)	(0.026)	(0.026)	(0.045)	(0.014)	(0.029)	(0.022)	(0.033)
Constant	0.926***	0.960***	0.013	0.053				
	(0.230)	(0.298)	(0.173)	(0.239)				
Observations	1790	588	640	243	1790	588	640	243
R^2	0.205	0.257	0.183	0.221				
Pseudo R^2					0.197	0.220	0.181	0.206

Dep. var.:	OLS				Probit			
Direct path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	-0.009	0.025	-0.039	0.035	-0.016	0.002	-0.045*	0.006
	(0.017)	(0.041)	(0.036)	(0.055)	(0.012)	(0.029)	(0.026)	(0.049)
log(`claims`)	0.002	-0.012	-0.021	0.053	0.001	-0.007	-0.015	0.055
	(0.014)	(0.033)	(0.028)	(0.055)	(0.010)	(0.023)	(0.019)	(0.047)
log(`bwd_cit`)	0.003	-0.025	0.018	0.016	0.003	-0.021	0.012	0.011
	(0.009)	(0.020)	(0.018)	(0.024)	(0.006)	(0.014)	(0.014)	(0.019)
log(`npl_cit`)	-0.005	0.010	-0.020*	0.003	-0.005	0.008	-0.018*	0.010
	(0.006)	(0.018)	(0.012)	(0.022)	(0.005)	(0.012)	(0.009)	(0.019)
log(`geo_fam`)	0.012	0.003	0.020	0.043	0.010	-0.002	0.014	0.036
	(0.011)	(0.029)	(0.026)	(0.042)	(0.008)	(0.021)	(0.019)	(0.034)
log(`fwd_cit`)	0.010	-0.013	-0.003	0.024	0.009	-0.016	-0.003	0.021
	(0.011)	(0.021)	(0.024)	(0.045)	(0.008)	(0.017)	(0.018)	(0.036)
Constant	0.228	0.346	0.139	-0.020				
	(0.158)	(0.211)	(0.163)	(0.243)				
Observations	1315	247	409	186	1315	247	409	186
R^2	0.091	0.218	0.132	0.198				
Pseudo R^2					0.139	0.256	0.175	0.193

Dep. var.:	OLS				Probit			
Indirect path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.013	0.019	0.015	0.025	0.007	0.025	0.013	-0.004
	(0.021)	(0.042)	(0.034)	(0.066)	(0.019)	(0.047)	(0.028)	(0.025)
log(`claims`)	0.001	-0.006	0.012	-0.105	0.005	-0.006	0.017	-0.049**
	(0.017)	(0.032)	(0.026)	(0.070)	(0.015)	(0.036)	(0.021)	(0.023)
log(`bwd_cit`)	0.001	-0.017	0.032*	-0.021	0.008	-0.020	0.042***	-0.012
	(0.011)	(0.024)	(0.019)	(0.024)	(0.011)	(0.028)	(0.016)	(0.010)
log(`npl_cit`)	-0.016*	-0.021	-0.008	-0.011	-0.020**	-0.031	-0.008	-0.002
	(0.009)	(0.018)	(0.014)	(0.023)	(0.009)	(0.022)	(0.012)	(0.010)
log(`geo_fam`)	0.029*	0.056*	-0.001	0.120*	0.029**	0.063*	0.011	0.044**
	(0.016)	(0.030)	(0.028)	(0.068)	(0.014)	(0.032)	(0.023)	(0.020)
log(`fwd_cit`)	0.127***	0.103***	0.143***	0.194***	0.112***	0.123***	0.125***	0.060**
	(0.015)	(0.026)	(0.024)	(0.058)	(0.013)	(0.029)	(0.018)	(0.029)
Constant	0.878***	0.894***	-0.025	0.712***				
	(0.229)	(0.299)	(0.175)	(0.235)				
Observations	1692	588	603	145	1692	588	603	145
R^2	0.240	0.256	0.227	0.367				
Pseudo R^2					0.251	0.220	0.258	0.466

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.9: Models including patents acknowledging awards performing activities at different R&D stages.

Dep. var.:	OLS			Probit		
Any path	Basic	Applied	Devel.	Basic	Applied	Devel.
SBIR	-0.018	0.066***	0.066***	-0.028	0.059***	0.068***
	(0.020)	(0.018)	(0.024)	(0.019)	(0.018)	(0.025)
log(`claims`)	0.017	-0.015	0.023	0.018	-0.017	0.025
	(0.018)	(0.016)	(0.021)	(0.017)	(0.015)	(0.021)
log(`bwd_cit`)	0.005	0.009	0.026**	0.004	0.009	0.028**
	(0.010)	(0.009)	(0.012)	(0.010)	(0.010)	(0.013)
log(`npl_cit`)	-0.007	-0.011	-0.006	-0.008	-0.012	-0.008
	(0.009)	(0.008)	(0.011)	(0.009)	(0.008)	(0.011)
log(`geo_fam`)	0.017	0.051***	-0.003	0.018	0.049***	0.000
	(0.015)	(0.016)	(0.018)	(0.014)	(0.014)	(0.019)
log(`fwd_cit`)	0.095***	0.109***	0.125***	0.086***	0.099***	0.119***
	(0.016)	(0.014)	(0.016)	(0.013)	(0.012)	(0.015)
Constant	0.607***	0.075	0.184			
	(0.165)	(0.372)	(0.435)			
Observations	1834	2230	1604	1834	2230	1604
R^2	0.193	0.232	0.211			
Pseudo R^2				0.190	0.225	0.196

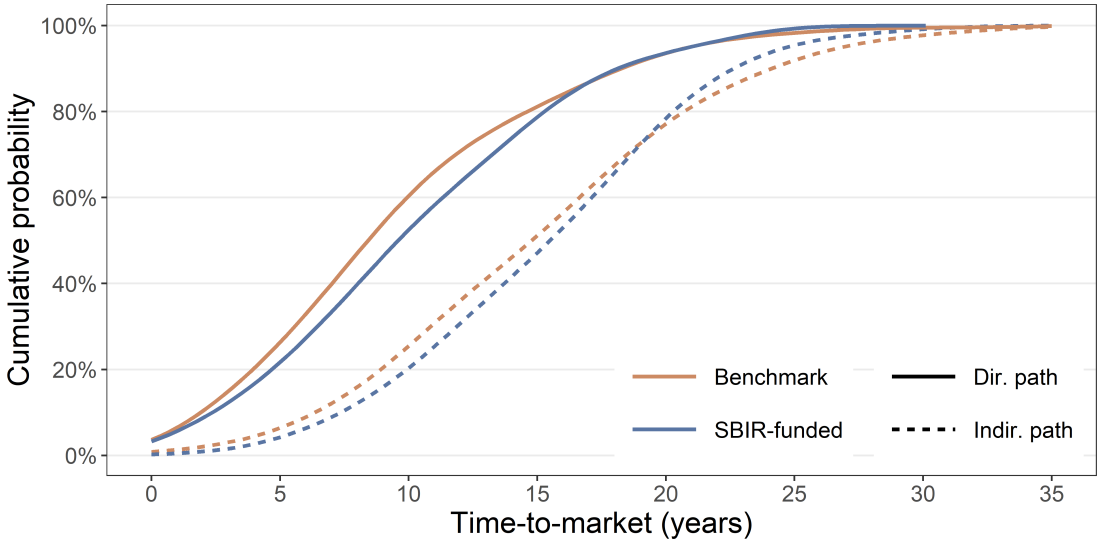
Dep. var.:	OLS			Probit		
Direct path	Basic	Applied	Devel.	Basic	Applied	Devel.
SBIR	0.001	0.016	0.041**	-0.017	0.007	0.027*
	(0.019)	(0.016)	(0.020)	(0.015)	(0.012)	(0.015)
log(`claims`)	0.001	-0.022	-0.001	0.001	-0.019*	-0.005
	(0.017)	(0.014)	(0.016)	(0.014)	(0.011)	(0.012)
log(`bwd_cit`)	0.006	0.012	0.026**	0.005	0.008	0.020**
	(0.010)	(0.008)	(0.010)	(0.008)	(0.006)	(0.008)
log(`npl_cit`)	0.008	-0.005	-0.006	0.007	-0.005	-0.006
	(0.008)	(0.007)	(0.009)	(0.006)	(0.005)	(0.007)
log(`geo_fam`)	-0.021	0.046***	-0.008	-0.020*	0.031***	-0.007
	(0.013)	(0.015)	(0.014)	(0.012)	(0.009)	(0.011)
log(`fwd_cit`)	0.012	0.006	0.017	0.007	0.007	0.016*
	(0.016)	(0.011)	(0.012)	(0.012)	(0.008)	(0.009)
Constant	-0.081	0.320	0.298			
	(0.072)	(0.376)	(0.221)			
Observations	1262	1633	1211	1262	1633	1211
R^2	0.140	0.122	0.154			
Pseudo R^2				0.178	0.173	0.203

Dep. var.:	OLS			Probit		
Indirect path	Basic	Applied	Devel.	Basic	Applied	Devel.
SBIR	-0.010	0.053***	0.055**	-0.011	0.038**	0.059***
	(0.019)	(0.017)	(0.023)	(0.016)	(0.016)	(0.022)
log(`claims`)	0.004	-0.000	0.023	0.004	0.001	0.024
	(0.017)	(0.015)	(0.020)	(0.014)	(0.014)	(0.018)
log(`bwd_cit`)	0.003	0.012	0.022*	0.003	0.011	0.024**
	(0.010)	(0.009)	(0.012)	(0.008)	(0.010)	(0.012)
log(`npl_cit`)	-0.011	-0.009	-0.005	-0.013*	-0.006	-0.008
	(0.008)	(0.008)	(0.010)	(0.007)	(0.007)	(0.010)
log(`geo_fam`)	0.023	0.037**	-0.010	0.021*	0.032**	-0.005
	(0.015)	(0.015)	(0.018)	(0.012)	(0.013)	(0.017)
log(`fwd_cit`)	0.121***	0.125***	0.147***	0.093***	0.103***	0.128***
	(0.016)	(0.014)	(0.016)	(0.012)	(0.011)	(0.014)
Constant	0.933***	0.130	0.174			
	(0.274)	(0.374)	(0.399)			
Observations	1651	1990	1513	1651	1990	1513
R^2	0.232	0.269	0.233			
Pseudo R^2				0.259	0.283	0.238

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Figure C.22: Years that it takes to at least $X\%$ of the SBIR-funded or benchmark patents to reach the market through a *direct* or *indirect path* (time-to-market) for the patents for which we managed to estimate a commercialization date.



	$\geq 1\%$	$\geq 5\%$	$\geq 10\%$	$\geq 50\%$	$= 100\%$
Direct path					
SBIR-funded	0	1	3	9	25
Benchmark patents	0	1	2	8	34
Indirect path					
SBIR-funded	1	5	7	15	32
Benchmark patents	1	4	6	14	35

Notes. We have been able to date 193 SBIR-funded and 216 benchmark patents directly linked to a relevant web page (see Appendix A.2.3). For patents indirectly linked to a relevant web page, we attributed a date to 455 SBIR-funded and 641 benchmark patents. The curves represent kernel-smoothed versions of the empirical cumulative distribution functions (ECDFs) using Gaussian kernel density estimation to help visualize the underlying continuous distribution. The chart shows no striking differences between SBIR-funded and benchmark inventions in terms of time-to-market. Looking at direct paths, it takes about nine years for the average SBIR-funded invention to reach the final consumers, whereas it takes eight years for benchmark inventions. However, this difference is not statistically significant (a t-test for a difference in means has a p-value of 0.24). The picture is very similar for the indirect paths, for which the commercialization path is 15 years long for the SBIR-funded and 14 for the benchmark patents, on average.

Table C.10: Policy change.

Dep. var.:	OLS			Probit		
Direct path	(1)	(2)	(3)	(4)	(5)	(6)
Phase II	0.045	0.042	-0.040	0.046	0.041	-0.041
	(0.031)	(0.031)	(0.053)	(0.035)	(0.035)	(0.051)
Post 2000		0.054*	-0.054		0.060**	-0.048
		(0.030)	(0.066)		(0.028)	(0.067)
Phase II × Post 2000			0.128*			0.128*
			(0.069)			(0.071)
log(`claims`)	0.003	0.006	0.005	0.001	0.004	0.004
	(0.020)	(0.021)	(0.020)	(0.019)	(0.019)	(0.019)
log(`bwd_cit`)	0.007	0.005	0.006	0.006	0.004	0.005
	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
log(`npl_cit`)	0.020*	0.019	0.020*	0.018*	0.018*	0.019*
	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
log(`geo_fam`)	0.003	0.007	0.008	0.001	0.004	0.005
	(0.022)	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)
log(`fwd_cit`)	0.035**	0.038**	0.036**	0.032**	0.035**	0.033**
	(0.016)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)
Constant	0.342	0.313	0.387			
	(0.341)	(0.324)	(0.319)			
Observations	809	809	809	809	809	809
R^2	0.130	0.134	0.138			
Pseudo R^2				0.131	0.137	0.140

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.11: Models considering only *indirect paths* as dependent variable. The first group focuses on *paths* involving a self-citation. The second, on *paths* involving a non-self-citation, and including only patents receiving no self-citations. The third, on *paths* involving a non-self-citation, but including only patents receiving at least a self-citation.

Dep. var.:	OLS				Probit			
Self-cit. path	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.014**	0.042	0.019*	-0.016	0.007*	0.025	0.013**	-0.008
	(0.006)	(0.025)	(0.010)	(0.023)	(0.004)	(0.020)	(0.006)	(0.008)
log(`claims`)	0.004	0.003	0.009	-0.022	0.003	0.007	0.009*	-0.004
	(0.006)	(0.022)	(0.008)	(0.025)	(0.003)	(0.016)	(0.005)	(0.005)
log(`bwd_cit`)	0.001	0.014	-0.003	0.000	0.001	0.007	-0.002	-0.002
	(0.004)	(0.020)	(0.006)	(0.012)	(0.003)	(0.014)	(0.004)	(0.003)
log(`npl_cit`)	0.006**	-0.004	0.011**	-0.005	0.003*	-0.005	0.006**	-0.003
	(0.003)	(0.011)	(0.005)	(0.009)	(0.002)	(0.008)	(0.003)	(0.002)
log(`geo_fam`)	0.007	0.035*	-0.012	0.034	0.005	0.029**	-0.005	0.008
	(0.006)	(0.019)	(0.008)	(0.026)	(0.003)	(0.013)	(0.005)	(0.006)
log(`fwd_cit`)	0.036***	0.023*	0.043***	0.100***	0.020***	0.019**	0.026***	0.022**
	(0.005)	(0.013)	(0.008)	(0.025)	(0.003)	(0.009)	(0.004)	(0.009)
Constant	0.977***	0.617***	0.258	0.195				
	(0.246)	(0.177)	(0.202)	(0.170)				
Observations	4938	775	2222	412	4938	775	2222	412
R^2	0.082	0.155	0.083	0.173				
Pseudo R^2					0.172	0.187	0.171	0.360

Dep. var.: *Non-self-cit.*	OLS				Probit			
path with some self-cit.	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	0.038***	0.092***	0.025**	0.004	0.019***	0.078***	0.015**	-0.001
	(0.007)	(0.020)	(0.011)	(0.033)	(0.004)	(0.016)	(0.006)	(0.004)
log(`claims`)	0.011**	-0.002	0.020**	0.059*	0.007**	-0.001	0.014***	0.014
	(0.005)	(0.013)	(0.008)	(0.033)	(0.003)	(0.011)	(0.005)	(0.009)
log(`bwd_cit`)	0.003	0.006	0.005	0.015	0.003	0.011	0.004	0.002
	(0.003)	(0.011)	(0.006)	(0.018)	(0.002)	(0.009)	(0.004)	(0.002)
log(`npl_cit`)	-0.003	-0.002	-0.005	-0.008	-0.002	-0.006	-0.002	-0.003
	(0.003)	(0.009)	(0.004)	(0.014)	(0.002)	(0.007)	(0.003)	(0.002)
log(`geo_fam`)	0.014**	0.051***	-0.009	0.037	0.007**	0.045***	-0.008	0.008
	(0.006)	(0.015)	(0.008)	(0.042)	(0.003)	(0.010)	(0.005)	(0.006)
log(`fwd_cit`)	0.064***	0.074***	0.073***	0.077***	0.028***	0.064***	0.039***	0.010*
	(0.006)	(0.012)	(0.009)	(0.029)	(0.003)	(0.009)	(0.005)	(0.006)
Constant	0.410**	0.428**	0.044	-0.069				
	(0.182)	(0.214)	(0.110)	(0.092)				
Observations	5746	1497	2443	206	5746	1497	2443	206
R^2	0.128	0.184	0.113	0.205				
Pseudo R^2					0.239	0.228	0.214	0.432

Dep. var.: *Non-self-cit.*	OLS				Probit			
path w/o any self-cit.	All	< 2000	$2000 - 2009$	$>= 2010$	All	< 2000	$2000 - 2009$	$>= 2010$
SBIR	-0.016**	-0.044**	-0.011	0.021	-0.015**	-0.048**	-0.013	0.012
	(0.008)	(0.021)	(0.013)	(0.017)	(0.006)	(0.020)	(0.011)	(0.009)
log(`claims`)	0.006	0.010	0.005	-0.009	0.005	0.013	0.008	-0.006
	(0.007)	(0.016)	(0.010)	(0.018)	(0.005)	(0.015)	(0.008)	(0.008)
log(`bwd_cit`)	0.007*	0.021*	0.010*	-0.013	0.006*	0.023*	0.010*	-0.009*
	(0.004)	(0.012)	(0.006)	(0.009)	(0.003)	(0.012)	(0.006)	(0.004)
log(`npl_cit`)	-0.008**	-0.014	-0.011**	-0.003	-0.007**	-0.014	-0.010**	-0.000
	(0.004)	(0.010)	(0.006)	(0.007)	(0.003)	(0.009)	(0.005)	(0.004)
log(`geo_fam`)	0.003	-0.015	0.018*	0.004	0.003	-0.017	0.015*	-0.000
	(0.006)	(0.015)	(0.010)	(0.015)	(0.005)	(0.014)	(0.008)	(0.007)
log(`fwd_cit`)	0.044***	0.050***	0.042***	0.070***	0.029***	0.046***	0.032***	0.028***
	(0.007)	(0.014)	(0.010)	(0.022)	(0.004)	(0.012)	(0.006)	(0.007)
Constant	0.004	-0.127	0.077	0.512***				
	(0.083)	(0.100)	(0.053)	(0.192)				
Observations	6356	1580	2898	632	6356	1580	2898	632
R^2	0.101	0.133	0.076	0.143				
Pseudo R^2					0.159	0.139	0.115	0.273

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.12: Main regressions estimated exclusively using commercialization as observed from 'proper' VPM pages.

Dep. var.:	Any path OLS	Direct path OLS	Indirect path OLS
SBIR	0.020**	0.023***	0.011
	(0.008)	(0.007)	(0.008)
log(`claims`)	0.003	-0.002	0.005
	(0.007)	(0.006)	(0.007)
log(`bwd_cit`)	0.011***	0.007*	0.010**
	(0.004)	(0.004)	(0.004)
log(`npl_cit`)	-0.001	0.003	-0.002
	(0.004)	(0.003)	(0.003)
log(`geo_fam`)	0.009	0.005	0.004
	(0.007)	(0.006)	(0.006)
log(`fwd_cit`)	0.084***	0.016***	0.086***
	(0.007)	(0.005)	(0.007)
Constant	0.341**	-0.072***	0.376**
	(0.150)	(0.017)	(0.151)
Observations	6413	3922	6152
R^2	0.137	0.050	0.156

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.13: Main regressions using one benchmark patent only for each SBIR-funded patent.

Dep. var.:	*Any path* OLS	*Direct path* OLS	*Indirect path* OLS
SBIR	0.036***	0.020**	0.022**
	(0.012)	(0.009)	(0.011)
log(`claims`)	0.018*	0.003	0.017*
	(0.010)	(0.008)	(0.009)
log(`bwd_cit`)	0.009	0.010**	0.010*
	(0.006)	(0.005)	(0.005)
log(`npl_cit`)	-0.004	0.001	-0.005
	(0.005)	(0.004)	(0.005)
log(`geo_fam`)	0.034***	0.010	0.029***
	(0.009)	(0.008)	(0.009)
log(`fwd_cit`)	0.112***	0.018**	0.126***
	(0.009)	(0.007)	(0.008)
Constant	1.013***	0.714***	0.826***
	(0.273)	(0.241)	(0.283)
Observations	4793	4032	4529
R^2	0.167	0.076	0.200

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.14: Main regressions using SBIR-funded patents only (not STTR-funded ones).

Dep. var.:	*Any path* OLS	*Direct path* OLS	*Indirect path* OLS
SBIR	0.025**	0.014*	0.017*
	(0.011)	(0.008)	(0.010)
log(`claims`)	0.017**	0.004	0.014*
	(0.008)	(0.006)	(0.008)
log(`bwd_cit`)	0.009*	0.007*	0.009*
	(0.005)	(0.004)	(0.005)
log(`npl_cit`)	-0.003	0.005	-0.005
	(0.004)	(0.003)	(0.004)
log(`geo_fam`)	0.014*	0.007	0.008
	(0.008)	(0.006)	(0.007)
log(`fwd_cit`)	0.110***	0.009*	0.127***
	(0.008)	(0.005)	(0.007)
Constant	0.648***	0.275	0.593***
	(0.226)	(0.174)	(0.205)
Observations	6935	6246	6547
R^2	0.150	0.052	0.193

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$

Table C.15: Main regressions using STTR-funded patents only (not SBIR-funded ones).

Dep. var.:	*Any path* OLS	*Direct path* OLS	*Indirect path* OLS
SBIR	-0.011	-0.020	0.010
	(0.044)	(0.047)	(0.058)
log(`claims`)	0.044	0.080*	-0.052
	(0.044)	(0.044)	(0.052)
log(`bwd_cit`)	0.024	0.037**	0.007
	(0.017)	(0.017)	(0.022)
log(`npl_cit`)	-0.004	-0.014	0.011
	(0.015)	(0.017)	(0.018)
log(`geo_fam`)	0.044	0.049	0.016
	(0.035)	(0.038)	(0.047)
log(`fwd_cit`)	0.163***	0.070*	0.214***
	(0.042)	(0.038)	(0.044)
Constant	0.260	-0.032	0.816***
	(0.237)	(0.219)	(0.293)
Observations	330	252	180
R^2	0.235	0.149	0.404

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.010$