EPO-ACADEMIC RESEARCH PROGRAM FINAL REPORT

Government-sponsored research and technical standards: Evidence from standard-essential patents

PI: Emilio Raiteri
Co-Investigators: Carlo Bottai, Rudi Bekkers

October 1, 2023

TECHNICAL REPORT

This report describes the work conducted between April 2021 and September 2022 on the EPO-ARP funded project *Government-sponsored research and technical standards: Evidence from standard-essential patents*. The document provides a broad overview of the project; Section 1 describes the data collection and data construction process and the empirical strategy; Section 2 describes the data; Section 3 presents the main empirical results of the project; Section 4 concludes.

Introduction

Today, we live in an increasingly interconnected world. For such a complex system to work effectively, we heavily rely on crucial technologies that not only allow individuals to exchange information with one another but also allow for communication between technologies (Bekkers and Martinelli, 2012; Bekkers et al., 2020a). This type of communication is made possible by a set of common rules voluntarily adopted by producers to ensure interoperability that goes by the name of *technical standards*. Formal decisions about the design and evolution of a technical standard are often taken in the framework of Standard Setting Organizations (SSOs). SSO membership is voluntary, and SSOs' activities could be seen as a self-governance effort made by private firms (Simcoe, 2012). In the last 15 years, economics and management scholars have devoted growing attention to the role of SSOs in the standardization process and in determining the success of a technical standard (Lerner and Tirole, 2006; Chiao et al., 2007; Baron et al., 2014, among others).

In most cases, these works consider technical standards as the outcome of a purely private and market-driven process. However, several factors suggest that the contribution of governmental agencies to standards development might be underestimated. First, even though their direct involvement within SSOs is limited, governmental bodies may actively participate in SSOs. Second, and more importantly, public agencies may support the development of technologies that end up in technical standards. It is well-known that the technical foundations of the modern Internet were laid down by the creation of the ARPANET and the adoption of the packet-switching technology by the US Advanced Research Project Agency in the 1970s (Ruttan, 2006). Mazzucato (2013) reports that cellular communication technology received enormous government support in its early days. Indeed, the research grants provided by the US National Science Foundation (NSF) greatly contributed to the development of key technologies such as spectrum auctions, spectrum sharing, and massive MIMO antennas.[1] Even Qualcomm, one of the most influential contributors to mobile telecommunication standards such as CDMA, in its early days, benefited from several contracts awarded by the US Department of Defense (DoD) and National Science Foundation (NSF) in the context of the Small Business Innovation Research (SBIR) program.[2] The National Standard Strategy issued by the US administration in 2023 stresses that, historically, the US Government has facilitated vital innovation in technical standards through public R&D investments, shaping successful standards that include Wi-Fi, the C computer programming language, and the suite of technologies comprising cellular communications (Executive Office of the President, 2023).

Yet, the contribution of government-sponsored research to scientific discoveries on which technical standards build is scarcely acknowledged and studied in the economic and man-

[1]See `https://www.nsf.gov/cise/advancedwireless/`.

[2]See `https://www.sbir.gov/success-story/qualcomm-inducted-sbir-hall-fame`.

agement literature. This lack of attention is quite striking and might have far-reaching implications. A failure in tracking and quantifying the actual impact of publicly funded research may lead to underestimating the value of such investments and, thus, insufficient public support for basic and applied research. This issue is especially concerning in an era in which the private sector is reducing its investment in *basic science*, while the relevance of fundamental research for private innovation is not declining (Arora et al., 2018).

The main objective of this project is to fill the literature gap and shed light on the potential relevance of the link between publicly-funded research and technical standards. To address this challenge empirically, we mainly rely on the abundant information provided in patent data and, in particular, on about 19,000 patents disclosed as potentially standard-essential (declared SEP) to the European Telecommunications Standards Institute (ETSI). A growing number of studies have focused on SEPs and showed that they appear to have a particular economic and technological value as measured by conventional patent metrics, such as citation, claim counts, and renewal (Rysman and Simcoe, 2008; Bekkers et al., 2017, 2020a,b). The main question we aim to answer in this paper is whether this group of patents disproportionately relies on publicly funded science compared to a set of similar inventions never disclosed as SEPs.

Working with patent data also allows us to follow the potential trail between public funding and technology development. In recent years, several research teams systematically collected large databases of government-funded corporate patents (Rai and Sampat, 2012; Li et al., 2017; de Rassenfosse et al., 2019; Fleming et al., 2019; Argente et al., 2020). Most of these works look for the government's direct involvement in developing an invention, and only a few devoted attention to the indirect links between patents, science, and public funding. To do so, they exploit the references to the non-patent literature (NPL) available in patent documents: a patent is deemed as building on publicly-funded research if it cites, as relevant *prior art*, a scientific article that reports the support of a government award (grant or procurement contract) in its acknowledgment section. In this work, we adopt this approach and classify a declared SEP as linked to public funding based on its references to the NPL.

Clearly, to assess whether SEPs disproportionately rely on public science, we need a suitable reference point, i.e., a group of patented inventions that are technically similar to our focal SEPs but were not disclosed as potentially essential to ETSI. To identify such inventions, we mainly rely on text similarity between patent documents, as done in de Rassenfosse et al. (2020), and identify about 27,000 similar inventions. We then rely on econometric analysis to analyze the potential differences between the set of disclosed SEPs and similar inventions.

Our main results show that, on average, disclosed SEPs that were applied during the early development of mobile telecommunication standards are 10 percent more likely to cite a scientific article as relevant prior art and about 14 percent more likely to build on publicly funded research compared to similar inventions not disclosed as SEP. When we consider patents applied from 2003 onward, we find no significant difference in the average share of patents building on scientific articles supported by public agencies.

To present our work and findings systematically, this report is structured as follows: Section 1 delves into the data collection, construction processes, and empirical strategy; Section 2 provides an in-depth description of the data; Section 3 presents the main results from our research.

1 Data and method

A necessary condition to carry out our study and to answer whether patents connected to technical standards disproportionately rely on knowledge funded by the public purse entails constructing a novel database linking patents to potential governmental funding. Such a process involves three main steps. First, we must identify and reconstruct the patent family relations of patented inventions connected to technical standards. Second, we must find a way to unambiguously link these patents to knowledge generated via direct public funding. Third, we need to identify a group of patents not connected to technical standards but highly similar to the patents in that group to establish a reference point.

Upon completing the data construction process, we will be able to offer descriptive econometric insights highlighting potential disparities in the dependence on publicly funded science between declared SEPs and similar inventions. The rest of the section discusses the details of each of the steps in the data construction process.

1.1 Identifying declared-Standard Essential Patents

To identify patents connected to technical standards, we build on an extensive literature that uses patents disclosed as potentially standard-essential to Standards Developing Organizations (SDO)(Baron et al., 2014; Bekkers et al., 2017, 2020b) To maximize the completeness and reliability of the disclosure data, we decided to focus on patents disclosed to one specific SDO, the European Telecommunication Standardization Institute (ETSI).

ETSI has been instrumental in the evolution of mobile communication standards. One of its most notable contributions is the development of the Global System for Mobile Communications (GSM), which became the de facto standard for mobile communications and paved the way for further advancements like 3G, 4G, and 5G technologies. ETSI maintains a public and complete database of patents voluntarily disclosed by the patent owners as potentially essential to an ETSI standard.

We build on the work of Bekkers et al. (2020b) and focus on 19,118 patent families

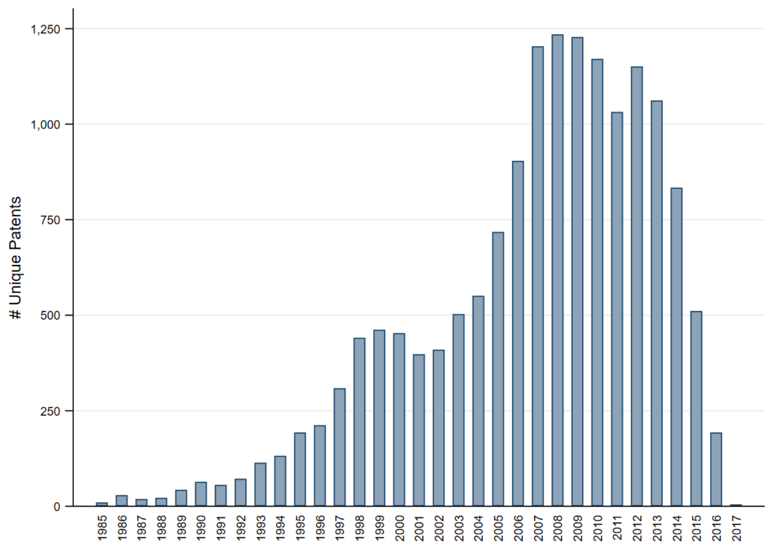


Figure 1: Distribution of declared-SEPs over time

disclosed to ETSI until March 2018 that include at least one granted USPTO patent.[3] We then primarily focus on the USPTO family members, as this facilitates the next steps of the data construction process. Figure 1 reports the distribution of declared-SEPs over time.

1.2 Linking declared-SEPs and publicly-funded science

Once we identify the list of about 19-thousand US patents disclosed as potentially standard-essential to ETSI, we follow the approach used by Rai and Sampat (2012); Li et al. (2017) to trace their potential connections to publicly-funded science. This method identifies the path linking a patented invention to research funded by a public agency through the references to the non-patent literature (NPL) included in the patent document. A patent is considered linked to public funding if it cites at least one scientific paper that acknowledges direct funding provided by a public funding agency. It is worth emphasizing here that while governments may fund scientific research through different tools, our primary focus is on pinpointing direct funding where a governmental agency actively influenced the selection and direction of the projects that received support. We thus narrow our attention to project grants while acknowledging that block grants also hold significant potential in fostering scientific advancements.

We follow a two-step strategy to identify the link between declared-SEPs and project grants. First, we link our disclosed SEPs to the *Reliance on Science* database, as collected by Marx and Fuegi (2022). This database includes information about in-text and front-page citations from patents to scientific articles. Using this resource, we connect 4,178 disclosed SEPs (27 percent of the sample) to more than 5,400 unique scientific articles with

[3]Even if the initial number of disclosed patents is 19,118, our working sample is composed of 15,362 declared SEPs for which we identified at least one similar invention.

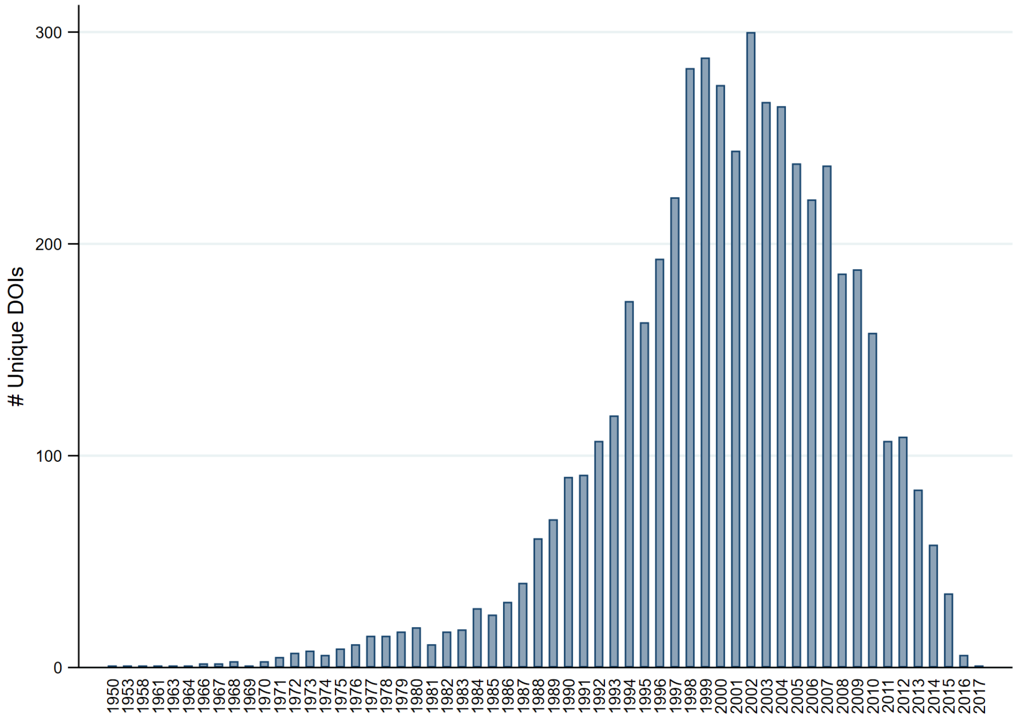


Figure 2: Distribution over time of unique DOIs cited by declared-SEPs

a Digital Object Identifier (DOI).[4] Figure 2 reports the distribution of unique DOIs cited by disclosed SEPs by year of publication. Interestingly, SEPs cite scientific works spanning an extended period, from the 1950s up to 2017, with the median papers being published in 2001. Figure 3 reports instead the distribution of the time lag between the SEPs' priority date and the publication year of the scientific articles they cite as NPL. As the figure shows, the modal patent-paper lag is just one year, and the median- and the average lag are 3 and 4.5 years, respectively, suggesting a pretty fast integration of scientific works into marketable innovations.

Second, we need to establish whether the research effort that produced a scientific publication cited by our patents benefited from direct public funding. The extant literature aimed at the same objective mainly relied on commercially available databases. For instance, Fleming et al. (2019) used the Web of Science (WoS) database to establish the existence of a citation link between a US patent and a scientific article that received funding from the US government. In principle, WoS collects information from the acknowledgment section of scientific articles and report funding information in an organized fashion, listing the source of the funding and, if available, the contract or the grant identification number. Initially, we intended to adopt the approach of Fleming et al. (2019), leveraging WoS data to pinpoint public support for scientific works. However, we found that funding details are

[4]A DOI is a unique alphanumeric string assigned to a document (such as a journal article or a report) to provide a persistent identifier for it.

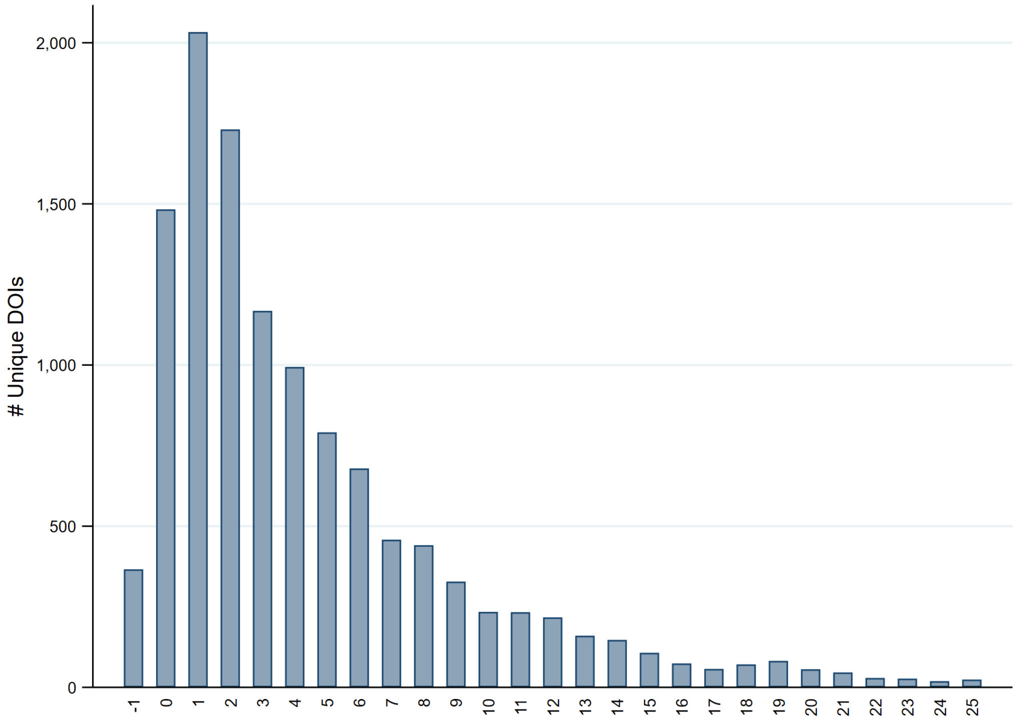


Figure 3: Distribution of the lag between the disclosed-SEPs priority date and year of publication of cited DOIs

Rank	Publication
1	IEEE International Conference on Acoustics, Speech, and signal processing
2	IEEE Global Telecommunications Conference
3	IEEE Vehicular Technology Conference
4	IEEE International Conference on Communications

Table 1: Most cited conference proceedings by declared SEPs

often missing for a significant portion of papers published by the Institute of Electrical and Electronics Engineers (IEEE). Additionally, while WoS may omit conference proceedings, their significance in disciplines such as computer science and engineering is universally recognized. This lack of coverage has serious implications for constructing our database, as about 81 percent of the DOIs cited by ETSI-disclosed SEPs are IEEE publications. Moreover, more than 25 percent of the articles cited by declared SEPs are published exclusively as proceedings of the conferences listed in table 1.

To quantify the significance of this issue, we implemented a test using a random sample of 100 papers with a DOI cited by our focal patents. We use the WoS database to retrieve funding information and obtained the following results:

1 Out of the 100 DOIs we searched, only 37 are included in the WoS database

2 For none of the 37 papers retrieved, any information about the research funding is

available

We then performed a manual check and searched for these 100 papers on the website of their respective IEEE journal. We could download the full version of the searched article in PDF format for 99 of the searched papers. Parsing the acknowledgment section of these articles, we determined that 27 of them actually acknowledge direct financial support from one or more funding agencies. Notably, among these, 11 articles that recognized government support were part of the 37 indexed in WoS; however, the WoS data omitted any funding details. These outcomes highlighted the impracticality of depending on a pre-existing database to identify public support for scientific publications. This posed a notable data challenge, considerably prolonging the data construction phase of our project.[5]

Consequently, we chose to extract funding information directly from the PDFs of the targeted scientific publications. We develop a Python script that autonomously locates the relevant paper online and parses the article's content to detect potential government support for the research. The script entails two main steps. First, it downloads and parses each PDF document to extract the *funding statement* of the scientific paper. Second, the script parses each extracted *funding statement* and identifies the names and the country of origin of the *funding agencies* whose support is mentioned in the statement. More specifically, the first step consists of:

1. Opening the PDF file and scanning the first two pages (the *text header*) and/or the *full text* of the scientific paper.

2. Searching several keywords (supported,ing, sponsored,ing, funded,ing, grant(s), financial support, carried out within) in the *text header* and extracting the 600 characters around each of them, if any.

3. Searching the *acknowledgments* paragraph in the *full text.* This paragraph is identified as the portion of text starting with the word "Acknowledg(e)ment(s)" and ending with one word among "Reference(s)", "Bibliography", or "Author(s)".

4. Identifying which of the text segments, procured from steps 2 and 3, constitutes a *funding statement.* For this purpose, the script examines each segment for two sets of primary-contextual keyword pairs..

   (a) The first group is composed of three main keywords –suppor(t), suppor(t)ed,ing, sponsored,ing, funded,ing– and eight contextual keywords –work, article, project, program, research, plan, grant(s), grateful(ly).

[5]Automating data collection from online PDF documents demanded intensive and sustained efforts from our entire research team. This several-month process encompassed multiple iterations on expanding sample sizes, refining scripts and classifiers, and continuous quality checks.

(b) The second group of two main keywords –grant(s), contract(s)– and five contextual keywords –work, article, project, research, plan.

The script subsequently filters out segments where: (a) for each primary-contextual keyword pair, none of the keywords are within ten (for the first group) or twenty (for the second group) words of each other; and (b) neither the phrases *financial support* nor *carried out within* appear within the segment.

The second step of the script uses the Named-Entity Recognition (NER) component of the Spacy library (`https://spacy.io`). The script executes the following steps:

1. Uses the NER to analyze the *funding statements* and extracts any entity classified as an organization (`ORG`) by the library.

2. Sequentially processes this list, maintaining a reference to the last organization reviewed in a specific variable. For each organization, it then searches for the subsequent two RegEx patterns:

   - `(by|from)(( the)?( [A-Z]?[a-z]+( of)?( the)?)? PO,? and (in part )?(by)?)?( the)?( [A-Z][a-z]+)? OO`;
   - `(by|from) a OO ([A-Z][a-z]+ )?(grant|fellowship|scholarship)`,

   where `OO` is substituted with the name of the currently searched organization and `PO` is substituted with the name of the previously searched organization. If one of the two RegEx hits a match with the *funding statement*, it classifies the organization as a *funding agency.*

3. Searches in the *funding statements* a few specific words that we know are important and not well classified by the Spacy's NER, and adds them to the *funding agency* list. Among others, "863", "DFG", "AFOSR", and "ESA".

The process of determining the country of origin for each *funding agency* was executed semi-automatically through multiple iterations. This led to the creation of a comprehensive dictionary, cataloging organizations and funding programs, which facilitated associating each organization with a specific country.

The script is used mainly to parse IEEE publications (81 percent of the sample), which are the ones that are less represented in the WoS database. For the remaining 19 percent of the publications, we rely on the information available on WoS. Adopting this method, we were able to link 1,208 disclosed SEPs, i.e., about 29 percent of the declared patents with at least one DOI associated with them, to a scientific publication that acknowledges direct support from a public agency.[6] As figure 4 shows, the vast majority of disclosed SEPs

[6] We conducted extensive manual random checks on the data obtained through this automatic process, and they all confirmed the high reliability of the method.

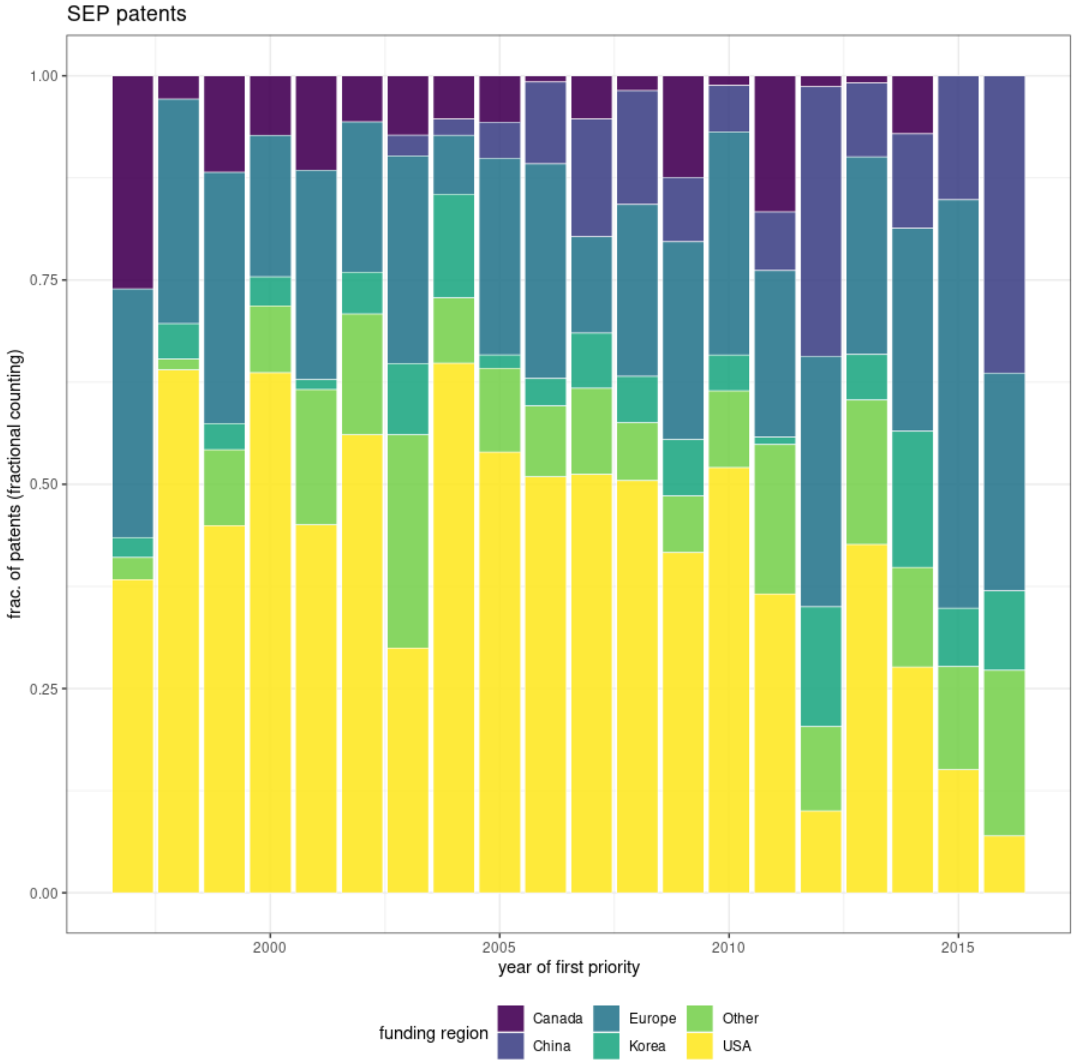


Figure 4: Distribution of disclosed SEPs by region of origin of the funding organization and year of first priority

citing publicly funded science cites research funded by US funding agencies. Table 2 reports the top 20 funders of SEP-related science: the US National Science Foundation and the Department of Defense are the top funders of the science relevant to patents declared as SEP. Nevertheless, 4 shows that the prevalence of US funding in SEP-related science seems to be decreasing over time, with science funded by Chinese and European agencies becoming more relevant in the last part of the period taken into account.

Our data also allow us to identify the link between the country of origin of the funding organization and the nationality of the patent applicant that building on government-funded knowledge. Therefore, we can check whether companies mainly exploit the knowledge generated with the support of the government of their own country or if they tap into knowledge funded by agencies of other nations. Figure 5 reports the knowledge flows between the country of origin of the funding organizations and of the patent applicants in two distinct time periods: 1995-2005 and 2006-2016. As the figure shows, while in both periods, patent

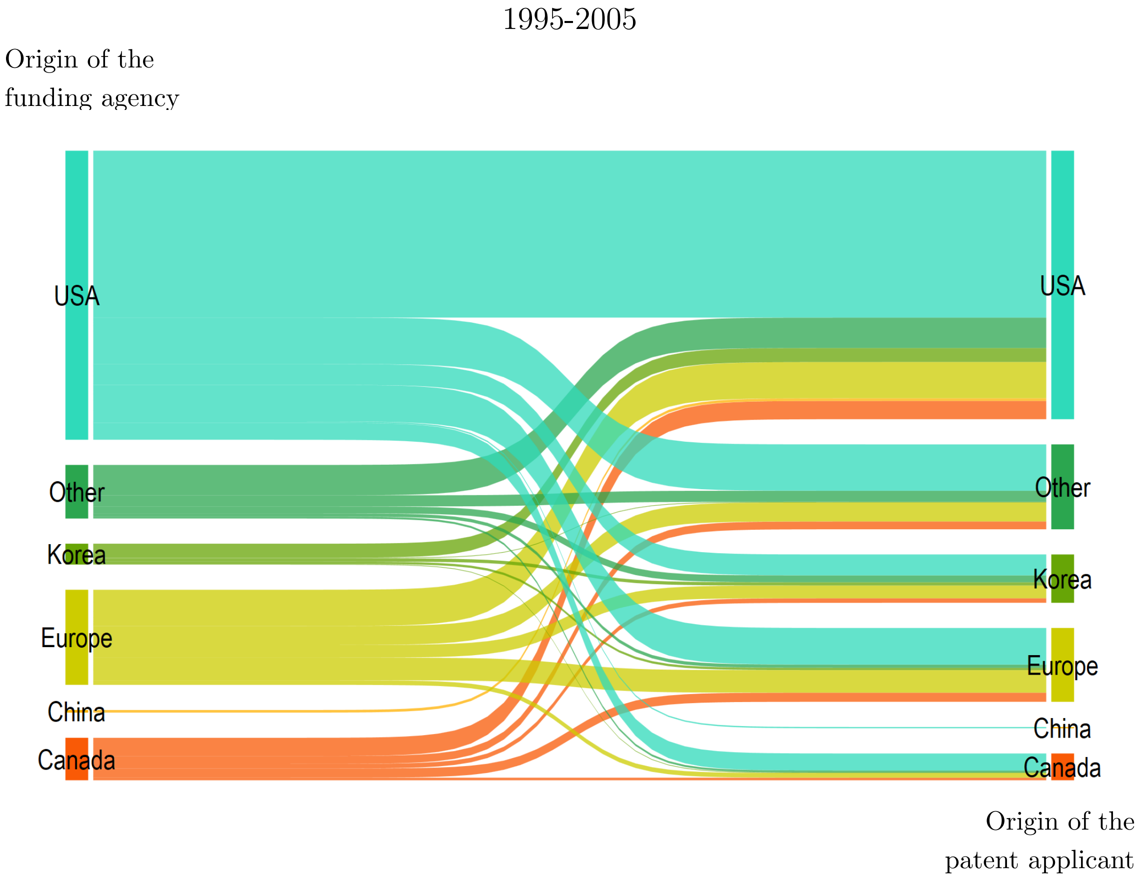
Rank	Organization	Country	N
1	National Science Foundation	USA	300
2	Department of Defense	USA	216
3	European Union	EU	99
4	Natural Sciences and Engineering Research Council	Canada	55
5	National Natural Science Foundation of China	China	46
6	National Science Council	Taiwan	35
7	Ministry of Science and Technology	China	31
8	University Grants Committee	Hong Kong	31
9	National Research Foundation	Korea	21
10	German Research Foundation	Germany	19
11	Institute of Information Technology Advancement	Korea	17
12	National Aeronautics and Space Administration	USA	14
13	Australian Research Council	Australia	13
14	Ministry of Information and Communication	Korea	12
15	UK Research and Innovation	UK	12
16	Ministry of Trade, Industry and Energy	Korea	11
17	Ministry of Education, Science, and Technology	Korea	10
18	National Research Council	Italy	8
19	Ministry of Education and Research	Germany	7
20	Ministry of University and Research	Italy	7

Table 2: Top 20 funding organizations for disclosed SEPs (number of times an organization is listed as funding agencies in the cited papers)

holders heavily rely on the knowledge generated with the support of US funding agencies, the prevalence of US funding seems to decrease in relative terms in the later period, mainly due to the increasing relevance Chinese and European agencies.

1.3 Identifying similar inventions

The last step in defining our empirical strategy to assess the reliance of disclosed SEPs on publicly funded research is constructing a suitable reference point. To do so, we combine information coming from the Google Patents platform (through Google BigQuery) and the EPO PATSTAT database. For each of the granted patents disclosed as potentially essential to ETSI, we identify the set of USPTO-granted patents that are assigned to the same technology class (CPC, subclass level) and that share the year of first priority with the disclosed SEPs. We also ensure that the patents in this set are not part of an INPADOC patent family with any members declared as essential to ETSI. From this group of potentially similar patents, we select up to five inventions for each disclosed SEP based on their textual similarity with the focal SEP. We exploit data from the Google Patent project to measure textual similarity between patents. This data is the output of a model that has learned



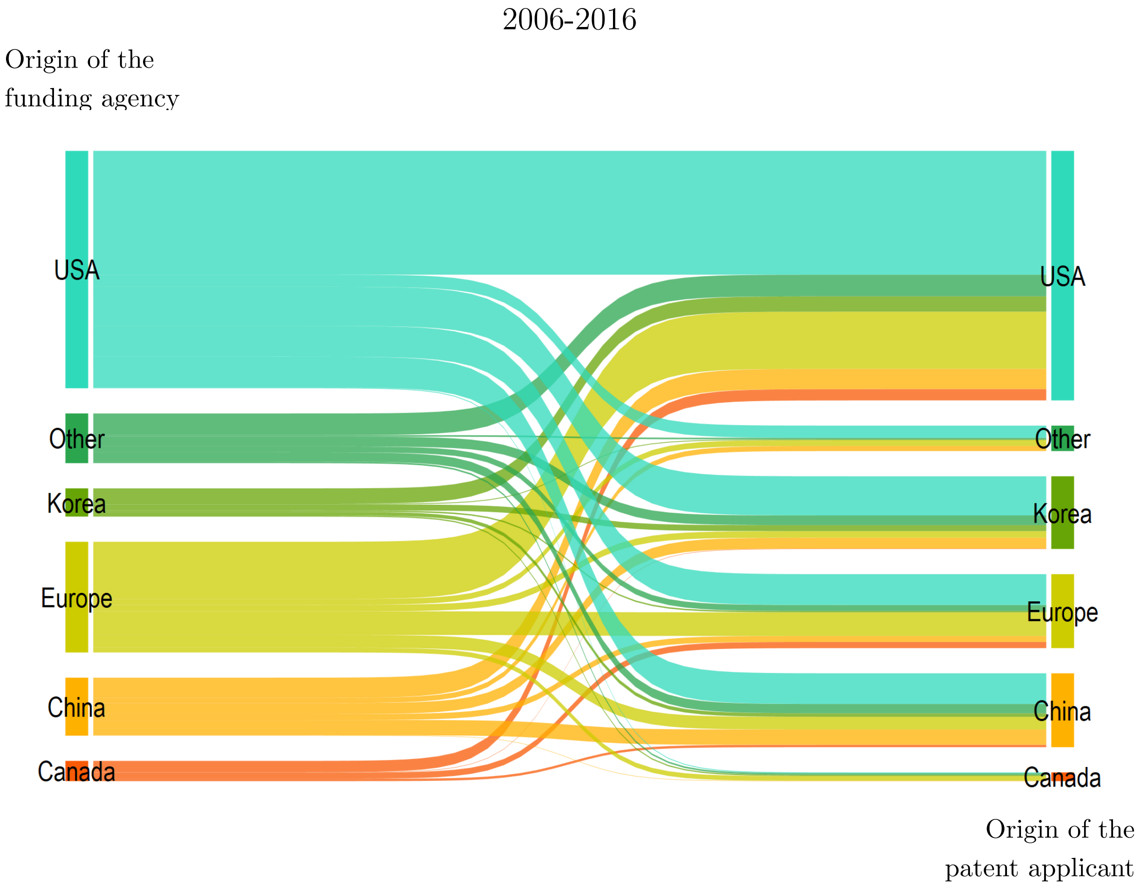


Figure 5: Fractional count of patent-paper associations by country of origin of the applicant and country of origin of the funding agency (only SEP with a funding link)

a set-of-words embedding using the WSABIE embedding algorithm (Weston et al., 2011).[7] We use the embedding vectors from Google Patents and compute the *cosine similarity* measuring the textual affinity between two patents. For each disclosed SEP, we select the five patents with the highest level of textual similarity in the set of potentially similar patents previously identified.[8] In this way, we identify a sample of 58,034 patents that we will use as a comparison point for our SEPs.[9] In addition, to avoid limiting ourselves exclusively to patents that cover technical and scientific ground that is very similar to the one covered by the disclosed SEPs, we also construct an alternative version of our reference group, populated with 23,038 patents. This group excludes patents that cite any patent document also cited by a declared SEP in our dataset.

We then adopt the empirical approach we applied to the declared SEPs and described in section 1.2 to determine whether our non-SEP inventions build on scientific publications and, in case they do, whether this publication received direct funding from a public agency. Matching the similar inventions set with the Marx and Fuegi (2020) database, we link 16,922 patents (29%) to at least one scientific publication through their NPL references. Applying the Python script described above, we then link 5,491 of these patents, i.e., about 32 percent of the similar inventions linked to at least one DOI, to a scientific publication that acknowledges direct support from a public agency. As figure 6 shows, also in the case of the similar inventions set, the vast majority of publicly funded scientific papers cited as relevant prior-art received support from a US funding agency. Table 3 reports the list of the top 20 funding organizations and shows, as in the case of disclosed SEPs, the US National Science Foundation and the US DoD as the main funders of the science linked to non-SEPs.

1.4 Empirical analysis

Once we have identified the declared SEP and non-SEP groups, we need to implement an empirical analysis that allows us to test for the existence of potential differences in the reliance on scientific knowledge and, in particular, on government-sponsored scientific knowledge between the two sets. To do so, we estimate the following basic *linear probability model*:

$$y_{it} = \beta \cdot \text{Declared\_SEP}_i + \delta_{\text{year, applicant, country, class}} + \gamma \cdot X_i + \epsilon_{it}$$

where y_{it} represents the two main outcome variables. The first one, $Cites\_science$, takes the value one if a patent cites as relevant prior art at least one scientific publication identified

[7]See, *e.g.*, `https://patents.google.com/?q=~patent\%2fUS7945525B2`. More details on the similarity algorithm are available at `https://media.epo.org/play/gsgoogle2017`.

[8]It is important to note that we consider a patented invention to be similar to a disclosed SEP only if their cosine similarity is above 0.7.

[9]In an earlier phase of the project, we worked with a reference set of limited size (about 26,000 patents) due to capacity constraints in the data collection process. Enlarging the non-SEP group required a substantial additional effort in terms of data collection but allowed us to improve the accuracy of our analysis.

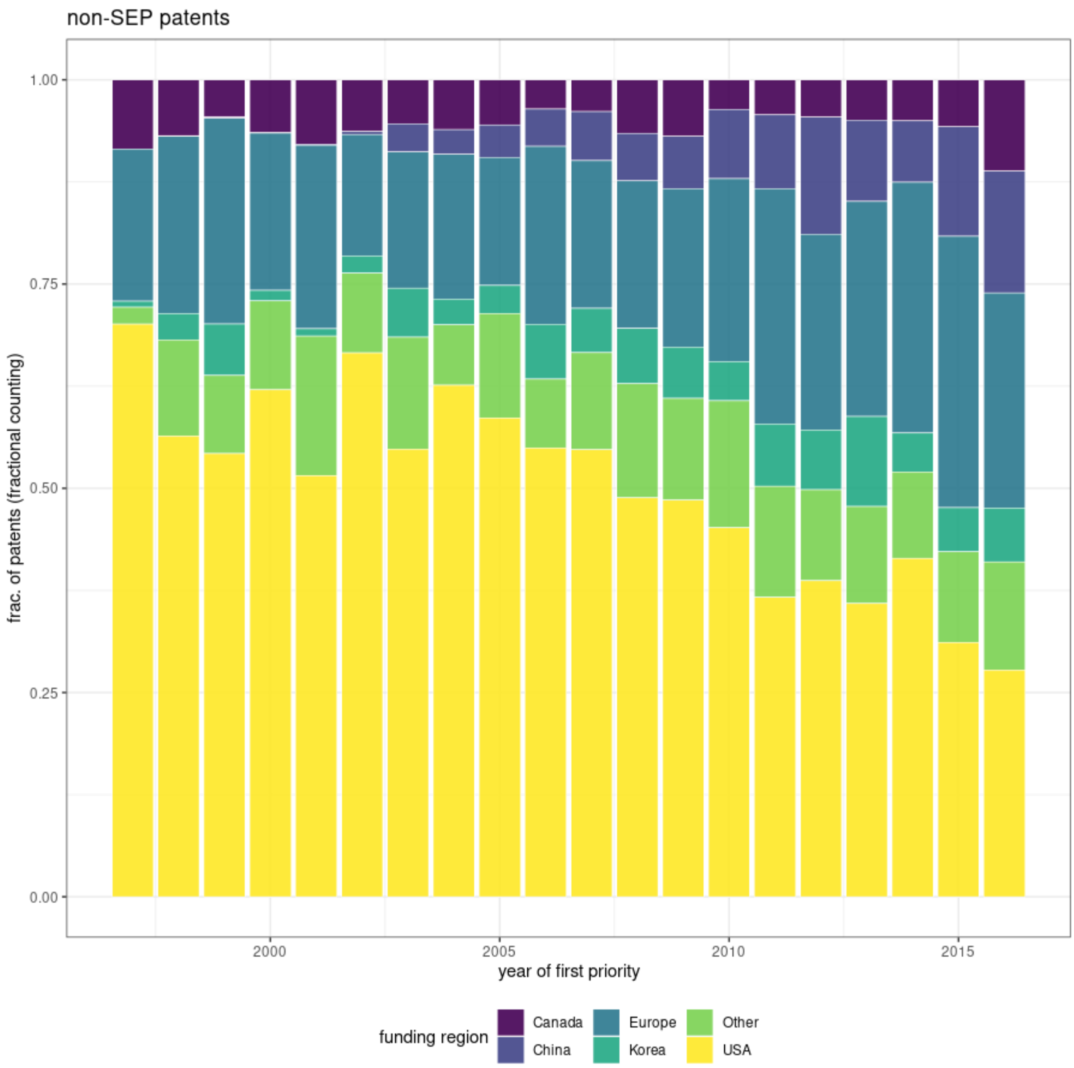


Figure 6: Distribution of non-SEP by region of origin of the funding organization and year of first priority

Rank	Organization	Country	N
1	National Science Foundation	USA	1593
2	Department of Defense	USA	1072
3	European Union	EU	444
4	Natural Sciences and Engineering Research Council	Canada	262
5	National Natural Science Foundation of China	China	230
6	National Science Council	Taiwan	196
7	University Grants Committee	Hong Kong	101
8	National Research Foundation	Korea	90
9	National Aeronautics and Space Administration	USA	88
10	UK Research and Innovation	UK	74
11	Ministry of Science and Technology	China	71
12	German Research Foundation	Germany	61
13	Ministry of University and Research	Italy	47
14	Ministry of Trade, Industry and Energy	Korea	47
15	Australian Research Council	Australia	45
16	Ministry of Education and Research	Germany	41
17	Institute of Information Technology Advancement	Korea	41
18	Ministry of Information and Communication	Korea	35
19	European Space Agency	EU	33
20	Ministry of Education, Science, and Technology	Korea	32

Table 3: Top 20 of funding organizations for non-SEP patents

with a DOI, and 0 otherwise. The second one, $Cites\_funded\_science$ takes the value one if a patent cites as relevant prior art at least one scientific publication that acknowledges the support of direct public funding, and 0 otherwise. $Declared\_SEP$ is our main variable of interest and identifies patents declared as potentially essential to an ETSI standard before March 2019. A significant coefficient for this variable would suggest a difference in the degree of reliance on science for potentially essential patents. The vector δ includes a battery of *fixed effects* for the year of first priority, the technology class (CPC group level), the patent applicant, and the applicant's country of residence. Finally, using the EPO PATSTATd database, we construct the vector X, which includes a set of patent-specific characteristics that may affect the probability of an invention to build on publicly funded knowledge. Specifically, we include, for each patent, the number of independent claims, the number of inventors, the number of patents it cites as relevant prior art (backward citations), the number of different CPC codes assigned to it, the number of citations it receives by other patents in the five years after being filed (forward citations). Exploiting information on the authors' affiliation included in Marx and Fuegi (2020) database, we are also able to create two additional variables, the variable *Academic* and the variable *Industry* that report, respectively, whether a patent cites at least one paper produced by an author with an academic affiliation, or if a patent cites at least one paper produced by an author with a corporate affiliation. The latter variables are available only for the analyses carried out on the subsample of patents associated with at least one DOI.

In addition to the model described above, which is aimed at answering the main research question of the project, we also implement a supplemental analysis to investigate whether science-based disclosed SEPs, and in particular, disclosed SEPs building on publicly funded science, have a higher value than non-SEP with similar characteristics. To do so, we follow the literature that uses patent citations as a proxy for the economic value of a patent and create the variable *Top_Cit* that takes the value one if a patent belongs to the top decile in terms of patent citations received in a time window of five years after application and 0 otherwise. We then estimate the following linear probability model:

$$TopCit_{it} = \beta_0 \text{Declared\_SEP}_i + \beta_1 \text{Cites\_Science}_i + \beta_2 \text{Declared\_SEP}_i \cdot \text{Cites\_Science}_i + \delta_{\text{year, applicant, country, class}} + \gamma \cdot X_i + \epsilon_{it}$$

where the interaction term $Declared\_SEPXCites\_Science$ is the main variable of interest.

2 Descriptive Statistics

The data construction process described in section 1.2 leads to a sample of 73,396 patents, of which 15,362 are patented inventions declared as potentially essential to ETSI and 58,034 are technically similar non-SEP inventions. As discussed, using the *Reliance on Science* database (Marx and Fuegi, 2020, 2022) and collecting information automatically via a Python script, we determine that about 29 percent of the patents (21,100 patents) in our sample builds on knowledge embedded in scientific publications. About 31.6 percent of the patents linked to at least one scientific publication build on scientific works that acknowledge direct public support. The script also collected information about the funding organization and the country of origin of the funding organization. The *Reliance on Science* database also allows us to determine which of the patents in our sample builds on scientific articles authored by scientists affiliated with a higher education institute and which do not. About 45 percent of the patents associated with a DOI cite at least one article authored by a researcher affiliated with an academic institution.

Table 4 reports the descriptive statistics for the patents declared as potentially essential to ETSI standards and the non-SEP group. Quite interestingly, the share of patents citing a scientific publication as relevant prior art (*Cites_science*) is similar in the declared SEP and in the non-SEP group, and the same holds true for the average number of scientific articles cited in a patent (# DOIs). Non-SEPs cite scientific works authored by academics more frequently (Cites Academics), whereas both groups are, on average, equally likely to cite science that received direct government funding ($Cites\_funded\_science$). As the table shows, declared SEPs make fewer references to the patent literature (*#bwdcites*) but receive more forward citations from subsequently patented inventions ($\#fwdcites$) in the five-year time window after the filing date.

Table 4: Descriptive statistics

	Non-SEPs	Declared-SEPs	Total
Cites science	0.292	0.272	0.287
Cites funded science	0.0946	0.0770	0.0909
Cites Academics	0.137	0.111	0.132
Cites Industry	0.931	0.941	0.933
# DOIs	1.063	0.913	1.032
# CPC codes	2.346	2.784	2.438
# claims	3.645	4.003	3.720
# inventors	2.852	3.126	2.910
# bdw cites	8.295	6.937	8.011
# fwd cites	2.132	3.143	2.344
N	58034	15362	73396

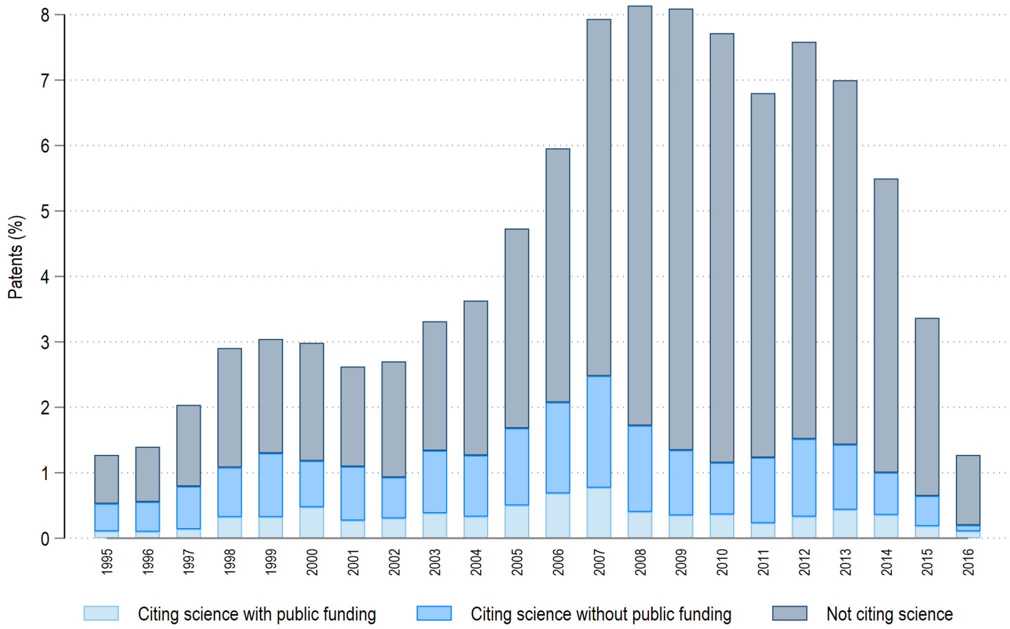
Mean coefficients;

Figure 7 displays the distribution of patents by year of first priority and the relative share of patents that build on scientific literature and on publicly funded science. As expected, the vast majority of the patented inventions in our sample were first introduced in the second part of the period we take into account, i.e., after 2007. Interestingly, while the number of patents declared as potentially essential substantially increased in particular between 2007 and 2012, the number of patents relying on knowledge disclosed in scientific publications remained relatively stable over time for both the declared SEP and the non-SEP group, suggesting a relative decrease in the reliance on science in the telecommunication sectors in more recent years.

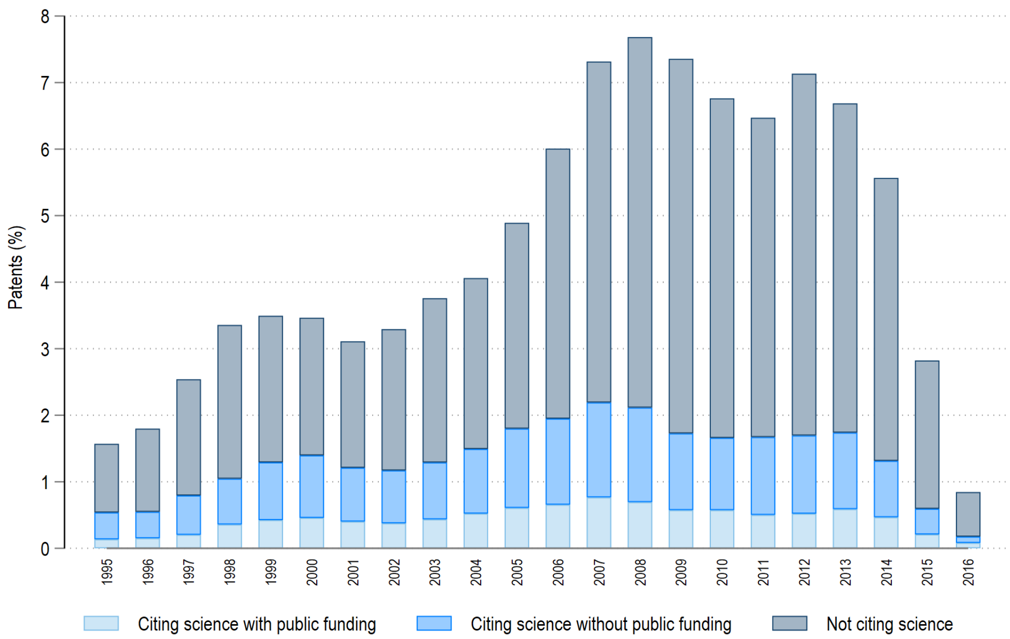
From the technology viewpoint, unsurprisingly, the patents in our working sample are concentrated in a small number of CPC classes. Class *H04W: Wireless Communication Networks*, class *H04L: Transmission of digital information*, and class *H04B: Transmission* account respectively for about 51, 26, and 9 percent of the patents, jointly covering 86 percent of the inventions in our sample. Figure 8 reports the distribution of patents by the ten most common CPC groups associated with declared SEP and non-SEP inventions instead. As the figure shows, we have a high level of variability in the degree of reliance on science in the different technology classes. For instance, in the group *G10L19* that identifies *Speech*

Figure 7: Distribution of patents by citing status and year of first priority

Declared SEPs



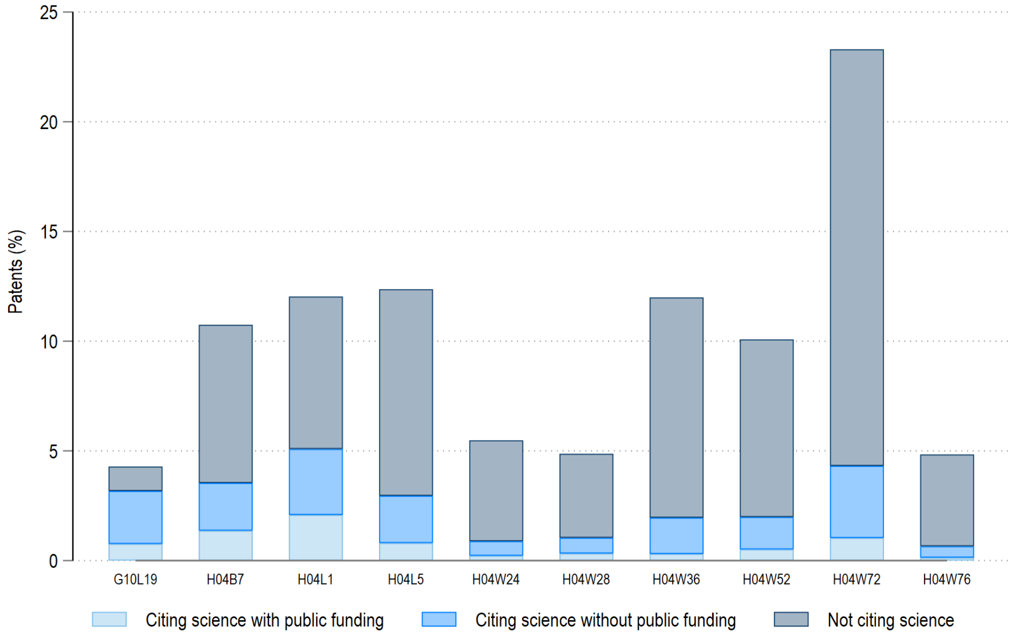
Non-SEPs



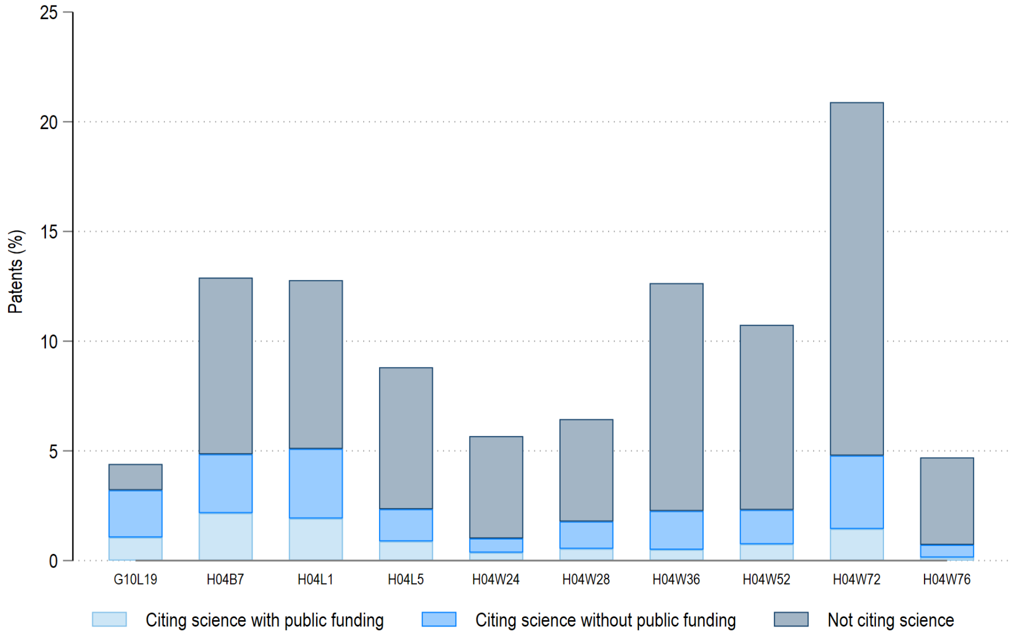
or audio signals analysis-synthesis techniques for redundancy reduction, e.g. in vocoders; Coding or decoding of speech or audio signals, using source filter models or psychoacoustic analysis, about three out of four patents cite at least a scientific paper as relevant prior art, whereas in other CPC groups such as *H04W24*, identifying *Supervisory, monitoring or testing arrangements*, the degree of reliance on science appears to be substantially lower.

Figure 8: Distribution of patents by citing status and CPC group

Declared SEPs



Non-SEPs



3 Results

As discussed above, our main objective is determining whether patents declared as potentially essential to ETSI disproportionately rely on science, particularly government-sponsored science. Section 2 described how we constructed our dataset and presented the linear probability model we use to evaluate the presence of significant differences in the degree of reliance on science between the declared SEP and the non-SEP groups. In this section, we present the results of our estimations.

We start by discussing the results retrieved using the variable *Cites_science* as the outcome variable. This variable takes the value 1 when a patent cites a scientific publication as relevant prior art and 0 otherwise. Table 5 reports the result. As the table shows, declared SEPs appear to have a slightly lower likelihood of building on scientific publications. However, when we split the results by time windows, this result seems to be driven by patents filed in the later part of the period taken into account. Column (2) shows that, when we only consider patents filed during the development of 2G and the early stages of development of the 3G standard (Baron and Gupta, 2018), we find that patents declared as essential to ETSI have a substantially higher likelihood to build on the scientific literature.

Table 5: Patents citing scientific publications

	(1)	(2)	(3)	(4)
	All	<=2001	2002-2009	2010-2016
Declared SEP	-0.008*	0.030***	-0.003	-0.039***
	(0.004)	(0.010)	(0.007)	(0.006)
# CPC codes	0.005***	0.016***	0.010***	0.004**
	(0.001)	(0.003)	(0.003)	(0.002)
# claims	0.004***	0.006***	0.002	0.003
	(0.001)	(0.002)	(0.002)	(0.002)
# bdw cites	0.003***	0.002***	0.003***	0.005***
	(0.000)	(0.000)	(0.000)	(0.001)
# inventors	0.006***	0.004	0.008***	0.003*
	(0.001)	(0.003)	(0.002)	(0.002)
Constant	0.209***	0.251***	0.210***	0.179***
	(0.006)	(0.012)	(0.010)	(0.011)
Fixed-effects				
CPC class	Yes	Yes	Yes	Yes
Priority year	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
Country	Yes	Yes	Yes	Yes
Observations	71809	17133	27370	25290
R^2	0.181	0.209	0.171	0.196

Robust standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 6 reports the estimated coefficients of our linear probability model when we use the variable *Cites_funded_science* as the outcome variable. As discussed above, this variable takes the value 1 when a patent cites at least one scientific work that received direct support from a public funding agency and the value 0 otherwise. Given that patents that are not associated with at least one DOI have, by definition, no chance of relying on publicly funded science based on our definition, we run this model exclusively on the subsample of patents that cite at least one scientific work as relevant prior art. As the table shows, patents declared as potentially essential to ETSI standards appear to be overall marginally more likely to build on scientific articles that received direct public funding. Again, the results seem to be driven by patents belonging to a specific time window. Declared SEPs filed before 2003 are 4.6 percentage points more likely to build on publicly funded science compared to the patents in the non-SEP group. Considering that, in this time window, 32 percent of the patents associated with at least one DOI cite a publicly funded work, declared SEPs appear to be 14 percent more likely to cite funded science. All in all, a sizable difference. Nevertheless, this difference fades in the later stages of technological development,

Tables 7 and 8 display the results of the analysis conducted using the alternative non-SEP group discussed in section 1.4. This different reference group is populated exclusively with patents with no citations to the patent literature in common with the disclosed SEPs. The main rationale for conducting this analysis is to remove non-SEPs that cover technical grounds too closely related to the ones covered by declared patents.

As the table shows, the differences between declared patents and non-SEPs found in the focal analysis are magnified by the exclusion of non-SEPs that build on scientific knowledge very close to the one declared patents build upon. Disclosed SEPs seem to be 6.3 percentage points more likely to build on a scientific publication and 3.8 percentage points more likely to cite publicly funded science compared with the patents in the reference set.

As discussed in section 1.4, we also try to assess whether science-based disclosed SEPs, and in particular disclosed SEPs that build on publicly funded science, have higher economic value than non-SEP inventions with similar characteristics. To do so, we run a distinct linear probability model presented in section 1.4, where the dependent variable is a binary indicator reporting whether a given patent belongs to the top decile of the distribution of forward citations and the main variable of interest is the interaction term between the variable $DeclaredSEP$ and the variable $Citescience$. Table 9 reports the analysis results. As expected and in accordance with previous literature Bekkers et al. (2020b, 2017), patents declared as SEP appear more likely to belong to the set of heavily cited patents. Also coherent with the extant literature, we find that patents that cite scientific works as relevant prior art appear to be more valuable than their non-science-based counterparts (Arora et al., 2022). Interestingly, the interaction term is also positive and significant, suggesting that patents disclosed as potentially essential to ETSI that build on scientific literature are likely more valuable than similar non-SEP inventions.

Table 6: Patents citing publicly-funded scientific publications

	(1)	(2)	(3)	(4)
	All	<=2002	2003-2009	2010-2016
Declared SEP	0.017*	0.046**	0.007	-0.015
	(0.009)	(0.019)	(0.018)	(0.023)
Cites Industry	-0.037***	-0.068	-0.018	-0.037
	(0.010)	(0.047)	(0.032)	(0.038)
Cites Academics	0.164***	0.116***	0.179***	0.172***
	(0.009)	(0.038)	(0.026)	(0.030)
# CPC codes	-0.000	-0.012**	0.005	0.005
	(0.002)	(0.006)	(0.007)	(0.005)
# claims	0.003***	0.003	0.001	0.004
	(0.001)	(0.003)	(0.004)	(0.006)
# bdw cites	-0.001***	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)
# inventors	0.001	0.005	-0.004	0.005
	(0.002)	(0.006)	(0.005)	(0.008)
# DOIs	0.027***	0.031***	0.022***	0.032***
	(0.001)	(0.006)	(0.003)	(0.005)
Constant	0.144***	0.175***	0.142***	0.119**
	(0.015)	(0.048)	(0.041)	(0.052)
Fixed-effects				
CPC group	Yes	Yes	Yes	Yes
Priority year	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
Country	Yes	Yes	Yes	Yes
Observations	18425	5245	7456	5170
R^2	0.269	0.362	0.245	0.281

Standard errors in parentheses

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

Table 10 reports instead the results of a model where we interact the variable $DeclaredSEP$ and the variable $Citefundedscience$. Once again, given that the variable $Citefundedscience$ can only take the value one for patents associated with scientific publications, we run this analysis on the subsample of patents associated with at least one DOI. As the table shows, also in this setting, patents declared as SEP to an ETSI standard are more likely to belong to the top decile of the citation distribution. Patents citing publicly funded science are marginally more likely to belong to the high-value group, but the interaction term is not significantly different than zero. Therefore, we find no evidence that declared SEP building on publicly funded science may be, on average, more valuable than non-SEP with similar characteristics.

However, the simple citation link between a patent and a publicly funded article may not say much about the actual proximity between the patented invention and the scientific articles it cites. It might still be the case that disclosed SEPs that are more closely related

Table 7: Patents citing scientific publications (alternative non-SEP group)

	(1) All	(2) <=2002	(3) 2003-2009	(4) 2010-2016
Declared SEP	0.063***	0.149***	0.001	0.010
	(0.011)	(0.027)	(0.018)	(0.016)
# CPC codes	0.001	0.015***	0.001	0.002
	(0.002)	(0.005)	(0.004)	(0.002)
# claims	0.005***	0.008***	0.003	0.002
	(0.001)	(0.003)	(0.002)	(0.003)
# bdw cites	0.003***	0.002***	0.003***	0.007***
	(0.000)	(0.000)	(0.000)	(0.001)
# inventors	0.007***	0.003	0.012***	0.004
	(0.002)	(0.005)	(0.003)	(0.003)
Constant	0.143***	0.143***	0.208***	0.138***
	(0.014)	(0.030)	(0.022)	(0.021)
Fixed-effects				
CPC group	Yes	Yes	Yes	Yes
Priority year	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
Country	Yes	Yes	Yes	Yes
Observations	34529	4123	12645	16871
R^2	0.195	0.224	0.180	0.228

Robust standard errors in parentheses

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

to the publicly funded science they cite could be more valuable than non-SEP inventions. To test for this possibility, we computed a measure of textual similarity between the title of the patent and the title of the publicly funded article they are associated with, using latent semantic analysis. We then re-run the same analysis as in 10 but dropping patents that cite a publicly funded article whose title has a low value of textual similarity with the patent title (i.e., cosine similarity below 0.60). Table 11 reports the results of this analysis. As the table shows, declared SEPs citing closely related scientific prior art appear to be more likely to attract more forward citations in this case.

Conclusions

The primary objective of our project, *Government-sponsored research and technical standards: Evidence from standard-essential patents*, was to uncover and better understand the linkage between government-sponsored research and the development of technical standards, vital in today's interconnected digital age. Our main findings show that SEPs disclosed during the nascent phases of mobile telecommunication standards are more likely to build on the scientific literature and have a considerably higher likelihood of being grounded in publicly

Table 8: Patents citing publicly-funded scientific publications (alternative non-SEP group)

	(1) All	(2) <=2002	(3) 2003-2009	(4) 2010-2016
Declared SEP	0.038*	0.067	0.054*	0.016
	(0.021)	(0.051)	(0.032)	(0.034)
Cites Academics	0.148***	0.095*	0.160***	0.165***
	(0.020)	(0.057)	(0.031)	(0.033)
Cites Industry	-0.070***	-0.099	-0.042	-0.081*
	(0.025)	(0.070)	(0.036)	(0.041)
# DOIs	0.028***	0.032***	0.023***	0.036***
	(0.003)	(0.010)	(0.003)	(0.005)
# CPC codes	-0.002	-0.018**	0.004	0.005
	(0.004)	(0.009)	(0.008)	(0.006)
# claims	0.004	0.003	0.002	0.004
	(0.002)	(0.004)	(0.004)	(0.007)
# bdw cites	-0.001***	-0.001	-0.001*	-0.004
	(0.000)	(0.001)	(0.000)	(0.003)
# inventors	-0.004	-0.004	-0.004	0.006
	(0.004)	(0.008)	(0.006)	(0.008)
Constant	0.167***	0.218***	0.129**	0.111*
	(0.034)	(0.083)	(0.053)	(0.060)
Fixed-effects				
CPC group	Yes	Yes	Yes	Yes
Priority year	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
Country	Yes	Yes	Yes	Yes
Observations	5058	1199	2147	1606
R^2	0.255	0.362	0.244	0.272

Robust standard errors in parentheses

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

funded research than their non-SEP counterparts. However, for patents applied after 2003, this differential appears to diminish, indicating a shift in the relationship between public funding and its realized outcomes in patent disclosures.

Yet, it is important to recognize some potential caveats inherent to the methodology adopted in our study. One notable issue concerns potential reporting bias. As discussed in the report, we exploit the information that authors voluntarily disclose in the acknowledgment section of their published articles to identify the trail of public support to scientific contributions. Even if, in principle, all researchers should have similar incentives to disclose funding sources, some agencies might mandate explicit acknowledgment in all resulting publications, while others may adopt a less strict approach. This potential issue implies that we might have overestimated or underestimated the relevance of the support offered by specific countries.

Furthermore, our research focused on direct government support to research activities

Table 9: Patents in the top decile of forward citations (1)

	(1)	(2)	(3)
Declared SEP	0.040***	0.040***	0.036***
	(0.003)	(0.003)	(0.002)
Cites science		0.026***	0.019***
		(0.003)	(0.003)
Cites science X Declared_SEP			0.015***
			(0.004)
# CPC codes	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
# claims	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.000)
# bdw cites	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
# inventors	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
Constant	0.059***	0.053***	0.055***
	(0.004)	(0.004)	(0.003)
Fixed-effects			
CPC group	Yes	Yes	Yes
Priority year	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Country	Yes	Yes	Yes
Observations	71809	71809	71809
R^2	0.377	0.379	0.379

Robust standard errors in parentheses

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

afforded through project grants. This type of support does not entirely capture the significant influence of other modes of funding, notably block grants allocated to academic institutions. While perhaps less direct than project grants, such grants might play a pivotal role in shaping the technological landscape.

In sum, our findings offer first insights into the interplay between public investment in research and the development of key technical standards. Nevertheless, this multi-faceted relationship would benefit from further detailed explorations to better delineate its contours and implications.

Table 10: Patents in the top decile of forward citations (2)

	(1)	(2)	(3)
Declared SEP	0.049***	0.050***	0.051***
	(0.006)	(0.006)	(0.006)
Cites funded science		0.009*	0.011*
		(0.005)	(0.006)
Declared SEP X Funded Science			-0.006
			(0.011)
# CPC codes	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
# claims	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
# bdw cites	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
# inventors	0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)
Cites Academics	0.005	0.003	0.002
	(0.006)	(0.006)	(0.006)
Cites Industry	0.009	0.009	0.009
	(0.007)	(0.007)	(0.007)
# DOIs	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
Constant	0.088***	0.087***	0.086***
	(0.009)	(0.009)	(0.009)
Fixed-effects			
CPC group	Yes	Yes	Yes
Priority year	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Country	Yes	Yes	Yes
Observations	18425	18425	18425
R^2	0.338	0.338	0.338

Robust standard errors in parentheses

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

Table 11: Patents in the top decile of forward citations (3)

	(1)	(2)	(3)
Declared SEP	0.059***	0.059***	0.057***
	(0.008)	(0.008)	(0.008)
Cites funded science		0.030*	-0.010
		(0.018)	(0.029)
Declared SEP X funded science			0.060*
			(0.034)
# CPC codes	0.004***	0.004***	0.004***
	(0.002)	(0.002)	(0.002)
# claims	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
# bdw cites	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
# inventors	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
Cites Academics	-0.010	-0.010	-0.010
	(0.009)	(0.009)	(0.009)
Cites Industry	0.014	0.015	0.015
	(0.010)	(0.010)	(0.010)
# DOIs	0.005***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
Constant	0.104***	0.105***	0.106***
	(0.014)	(0.014)	(0.014)
Fixed-effects			
CPC group	Yes	Yes	Yes
Priority year	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Country	Yes	Yes	Yes
Observations	12898	12898	12898
R^2	0.397	0.397	0.397

Robust tandard errors in parentheses

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

References

Argente, D., S. Baslandze, D. Hanley, and S. Moreira (2020): "Patents to Products: Product Innovation and Firm Dynamics," Tech. Rep. 2020-4, FRB Atlanta Working Paper.

Arora, A., S. Belenzon, and A. Patacconi (2018): "The decline of science in corporate R&D," *Strategic Management Journal*, 39, 3–32.

Arora, A., S. Belenzon, and J. Suh (2022): "Science and the Market for Technology," *Management Science*, 68, 7176–7201.

Baron, J. and K. Gupta (2018): "Unpacking 3GPP standards," *Journal of Economics & Management Strategy*, 27, 433–461.

Baron, J., Y. Ménière, and T. Pohlmann (2014): "Standards, consortia, and innovation," *International Journal of Industrial Organization*, 36, 22–35.

Bekkers, R., C. Catalini, A. Martinelli, C. Righi, and T. Simcoe (2017): "Disclosure rules and declared essential patents," Tech. rep., National Bureau of Economic Research.

Bekkers, R., J. Henkel, E. M. Tur, T. van der Vorst, M. Driesse, B. Kang, A. Martinelli, W. Maas, B. Nijhof, E. Raiteri, and L. Teubner (2020a): "Pilot study for essentiality assessment of standard essential patents," Tech. rep., European Commission and Joint Research Centre.

Bekkers, R. and A. Martinelli (2012): "Knowledge positions in high-tech markets: Trajectories, standards, strategies and true innovators," *Technological Forecasting and Social Change*, 79, 1192–1216.

Bekkers, R., E. Raiteri, A. Martinelli, and E. M. Tur (2020b): "Landscape study of potentially essential patents disclosed to ETSI," in *Digital Economy Working Paper 2020-06; JRC Technical Report, JRC120137*, European Commission, Joint Research Centre.

Chiao, B., J. Lerner, and J. Tirole (2007): "The rules of standard-setting organizations: an empirical analysis," *The RAND Journal of Economics*, 38, 905–930.

de Rassenfosse, G., A. Jaffe, and E. Raiteri (2019): "The procurement of innovation by the US government," *PloS one*, 14, e0218927.

de Rassenfosse, G., G. Pellegrino, and E. Raiteri (2020): "Do patents enable disclosure? Evidence from the Invention Secrecy Act," *Evidence from the Invention Secrecy Act (March 26, 2020)*.

Executive Office of the President (2023): "National Standards Strategy for Critical and Emerging Technology," United States Government Document.

Fleming, L., H. Greene, G. Li, M. Marx, and D. Yao (2019): "Government-funded research increasingly fuels innovation," *Science*, 364, 1139–1141.

Lerner, J. and J. Tirole (2006): "A model of forum shopping," *American economic review*, 96, 1091–1113.

Li, D., P. Azoulay, and B. N. Sampat (2017): "The applied value of public investments in biomedical research," *Science*, 356, 78–81.

Marx, M. and A. Fuegi (2020): "Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles," *Strategic Management Journal*, 41, 1572–1594.

——— (2022): "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations," *Journal of Economics and Management Strategy*, 31, 369–392.

Mazzucato, M. (2013): *The entrepreneurial state*, Penguin books.

Rai, A. K. and B. N. Sampat (2012): "Accountability in patenting of federally funded research," *Nature biotechnology*, 30, 953–956.

Ruttan, V. W. (2006): *Is war necessary for economic growth?: military procurement and technology development*, Oxford University Press.

Rysman, M. and T. Simcoe (2008): "Patents and the performance of voluntary standard-setting organizations," *Management science*, 54, 1920–1934.

Simcoe, T. (2012): "Standard Setting Committees: Consensus Governance for Shared Technology Platforms," *American Economic Review*, 102, 305–336.

Weston, J., S. Bengio, and N. Usunier (2011): "Wsabie: Scaling up to large vocabulary image annotation," in *Twenty-Second International Joint Conference on Artificial Intelligence*.