Patents that Match your Standards: Firm-level Evidence on Competition and Innovation

Antonin Bergeaud Julia Schmidt Riccardo Zago*

July 2023

Abstract

When a technology becomes the new standard, the firms that are leaders in producing this technology gain a competitive advantage. Matching the semantic content of patents to standards, we show that firms closer to the new technological frontier increase their market share and sales. In addition, if they operate in a competitive market, these firms also increase their R&D expenditure. Yet, these effects are temporary since standardization creates a common technological basis for everyone, which allows followers to catch up and the economy to grow.

JEL classification: L15, O31, O33
Keywords: Standardization, Patents, Competition, Innovation, Text mining.

*Contact addresses: Antonin Bergeaud: HEC Paris, CEP-LSE and CEPR (bergeaud@hec.fr); Julia Schmidt: Banque de France (julia.schmidt@banque-france.fr); Riccardo Zago: Banque de France (riccardo.zago@banque-france.fr). The authors want to thank without implicating Philippe Aghion, Simon Bunel, Maarten De Ridder, Michele Fioretti, Stéphane Guibaud, Jorge Lemu, Dimitris Papanikolaou, Gianluca Violante and seminar participants at Northwestern University, ESSIM 2022, BSE Summer Forum 2022, SED 2023, Collège de France and INSEAD for helpful comments.

1 Introduction

The development and production of goods and services is often subject to a myriad of technical standards. From payments systems to specifications for doorframes or autonomous vehicles, industrialized societies rely heavily on technical standards in every sector of the economy. By defining a common set of rules, guidelines and specifications, standardization guarantees the interoperability of devices, compatibility of inputs, or the safety and quality of products at the benefit of both producers and consumers. Technological standardization inherently involves the selection of one technology over others as it aims to ensure the widespread proliferation of the best technologies and practices within each industry. This is achieved by requiring industry participants to align on a common set of rules, thereby limiting the proliferation of potentially incompatible alternatives.

Yet, a firm's ability to adapt to a new standard, which we refer to as the new *technological frontier*, hinges on its past strategic decisions. Certain firms, given their innovation history, may be better technologically equipped to implement the new requirements and processes described in the recently issued standard. Consequently, firms that are close to the new frontier could gain an immediate competitive advantage. This could result in a shift in market power in their favor, thereby presenting a well-known trade-off between rewarding successful innovations and preventing the emergence of monopolies. This paper contributes to this ongoing debate. By introducing a new measure of firm-level proximity to the technological frontier, we show how the selection of one technology among competing ones through standardization affects competition, innovation and growth.

To address the above question empirically, we combine (i) knowledge of which technologies have been adopted by an entire industry and (ii) the innovative activity of individual firms. For the former, we rely on the fact that large-scale technology adoption requires industry participants to coordinate on a set of common rules, a process formally known as *standardization*. For this we use documents approved by industry experts from Standard-Setting Organizations (SSOs) that describe the basic features of the selected technology (known as *standards*). Prominent examples are mobile telecommunication standards (such as the 5G standard family) or Internet protocols. For the latter, we use patent data which is a widely used measure of innovation at the firm-level (see Hall et al., 2005). Hence, we match the semantic content of patents to standard documents and introduce a novel measure of the proximity of a firm to the new technological frontier. This allows us to characterize in detail firms' responses to standardization and to provide new evidence on its macroeconomic implications.

Our results show that, in response to the release of a new standard, firms that own patents closer to the newly defined frontier have an immediate competitive advantage that translates into higher sales and market shares. We also find that, if the market is competitive, frontier firms invest more in R&D while this is not the case if the level of competition is low. These results are consistent with the interpretation of standardization as a reduction in the level of competition in the short-run, benefiting technological leaders. However, this advantage is not permanent. Indeed, the objective of standardization is to establish a common framework that allows the rest of the industry to catch up over time through spillovers. We show that this mechanism has a positive long-term effect on sectoral growth. Contrary to creating enduring monopolies or fostering rent-seeking behavior, standards stimulate innovation across the industry and contribute to long-term economic growth.

Our analysis proceeds as follows. First, we apply a semantic algorithm to measure the proximity of a patent to a standard. In particular, we use the fact that each standard is associated with a set of relevant keywords that can be directly compared to the information in patent abstracts. From this procedure, we are able to link 21.5 million patents to over 0.6 million standards and measure the semantic similarity between each pair. This new measure represents a substantial novelty as most of the literature focuses on either patent data to measure innovation at the firm-level (e.g. see Griliches, 1990 and Hall et al., 2005), or standards data to measure technology adoption at the industry-level (e.g. see Baron and Schmidt, 2014). Our measure correlates with the economic value of patents (defined as in Kogan et al., 2017), their scientific value (measured by forward patent citations) and their private value (patent holders are more likely to pay renewal fees). Yet, it spans beyond those attributes: it identifies patents that, within comparable technologies and of similar quality, describe innovations better suited to adapt to the new frontier. For the firms owning these patents, this entails both the diffusion of their technologies as well as their capacity to quickly deploy, scale and market new applications of the standard.

In the second part of our investigation, we analyse the firm-level implications of owning patents close to new standards. Specifically, we use the crosswalk from Kogan et al. (2017) to match firm-level quarterly data from Compustat, Crisp and Ibes to patent data and to our new measure of technological proximity that we aggregate at the firm-quarter level.

As a first exercise, we analyze the validity of our firm-level proximity measure. In fact, if a certain technology is already widely used in an industry (ex-post standardization), our measure would correctly capture successful innovations, but not the additional impact of standardization. To investigate this point, we study the information content of our proximity measure by looking at the stock market reaction that follows the release of a new standard. We find that financial market participants indeed respond to standardization by revising upward earnings expectations for firms closer to the frontier. The revision occurs only at the time when the content of the standard is made public and despite the fact that the firm's patent stock is already known. This result suggests that the specific content of a standard and –consequently– the relative proximity of each firm to the frontier represent a surprise for markets. This is an important point as it implies that –despite the process of standardization being endogenous– the approval of a new standard is a necessary step for firms to deploy, scale and market their innovations. Stock markets pick up these effects by revising growth opportunities for firms closer to the frontier.

As a second exercise, we show that firms closer to the new frontier indeed gain both in terms of sales and market shares only after the publication of the standard. These effects are not anticipated and last for roughly five consecutive quarters. In particular, we estimate that –for these firms– this translates into an (average) increase of sales and market share respectively of 5.6% and 5.2% in the first year following the standard's release. To shed light on the interplay between competition and innovation, we consider the responses of investment in R&D following standardization, which depend on market structure. Specifically, we find that if a firm is operating in a competitive (non-competitive) market and is close to the technological frontier, it will invest more (less) after the release of the standard. This is consistent with the literature that has emphasized a non-linear relationship between competition and innovation (see Aghion et al., 2005). Overall, for the entire sample of firms, the expansion of R&D is the prevailing effect. We estimate that frontier firms (on average) increase their investment in innovation by 4.7% in the first year following the standard's release. In light of this, we can interpret the release of a standard as a temporary shift in competition in favor of those firms better equipped to immediately adopt and deploy the new technological frontier.

Yet, our identification could be affected by various potential threats. Most importantly, it is possible that some firms lobby or exercise undue influence over SSOs to push for the inclusion of their patents into the new standard (Bekkers et al., 2011; Farrell and Simcoe, 2012). If this is the case, our proximity measure would capture the lobbying power of a firm rather than its innovative capacity. To check this, we perform two tests. First, we use SSO membership data to control that results are not driven by the direct participation of firms in the standardization process. Second, we exclude standards developed by US SSOs from the calculation of our proximity measure as it is more likely that US firms have stronger influence at national level.

Finally, we delve into the broader implications of standardization for growth. Our findings suggest that while standardization confers a significant advantage to firms closer to the new frontier, this effect is short-lived. In the long run, standardization enhances knowledge spillovers and fosters broader technology diffusion, as evidenced by the flow of patent citations. This mechanism ultimately enables other firms to catch up, thereby promoting overall growth in the industry. Specifically, four years post the standard's release, sectoral growth increases by 0.11 percentage points, driven by the industry-wide catch-up process.

In light of this evidence, this paper contributes to the policy debate on the link between competition and innovation and its implications for economic growth. Standardization and its consequences represent an important and overlooked dimension to study this question. On the one hand, proponents of standardization argue that it is both an acknowledgment of the best technology among competing ones, and also a way to speed up the diffusion of this technology and subsequent improvements. On the other hand, the release of a standard can lock a certain industry in the chosen technology. This might prevent the emergence of competing technologies by transferring substantial market power to firms that have a considerable stake in the standardized technology. Not surprisingly, the policy debate among regulators and standard-setting organizations has centered around this complex trade-off (Lerner and Tirole, 2015).

Related literature. Our study relates to different strands of the literature.

The first one is on technological standardization which has received much attention in the industrial organization (IO) literature, but remains largely overlooked in macroeconomics despite the omnipresence of standards in every aspect of economic activity (see Kindleberger, 1983 for an historical overview). The IO literature has identified a wide range of benefits of standardization. By allowing for interoperability, compatibility and network effects (Katz and Shapiro, 1985; Farrell and Saloner, 1985), lower transaction costs and the reduction of information asymmetries (Leland, 1979), standardization is especially important for the large-scale deployment of inventions and technologies. In order to reap the benefits of standardization, technological specifications and details must be agreed upon by industry participants. Standard-setting organizations (SSOs) are fundamental in that process (Rysman and Simcoe, 2008).

Consequently, standardization is an essential prerequisite for the industry-wide adoption of new technologies, especially in the case of general-purpose technologies (Basu and Fernald, 2008; Jovanovic and Rousseau, 2005). This has macroeconomic implications (see Baron and Schmidt, 2014, who exploit the timing of standard releases to study the business cycle implications of technology adoption).

The benefits of standardization notwithstanding, several concerns have been highlighted in the literature. With the arrival of new technologies, the optimality of the incumbent standard is called into question. However, high switching costs may prevent the adoption of new technologies such that industries become "locked in" a certain standard (Farrell and Klemperer, 2007; Farrell and Saloner, 1986). The QWERTY keyboard is an often cited example of such a lock-in effect as consumer habits and compatibility prevent the adoption of more efficient keyboards such as DVORAK (David, 1985).

Another related concern is that standards, by favoring one technology over another, can give too much market power to the owners of the technology in question, especially if its use is safeguarded by patent protection. It is for this reason that SSOs insist that holders of so-called standard-essential patents (SEPs) respect fair, reasonable and non-discriminatory (FRAND) licensing principles. This loose prescription has led to an intense debate among regulators, economists and lawyers, and to a theoretical literature on the optimal design of rules on standard development, SEP licensing or voting procedures (Lerner and Tirole, 2015; Schmalensee, 2009; Llanes and Poblete, 2014; Spulber, 2019).[1] Contrary to this microeconomic literature, our paper analyzes the macroeconomic effects of standardization. We aim to evaluate whether the beneficial aspects of the large-scale diffusion of new technologies through standardization can outweigh the potentially detrimental aspects that may arise when some firms acquire too much market power and future innovation is impeded. While we measure the proximity between patents and standards, our measure is not confined to the standard essentiality of patents. More generally, firms whose patents are close to a new standard can be expected to have an immediate competitive advantage in deploying the technology described therein.

The second strand of literature this paper speaks to concerns the link between innovation and competition. In endogenous growth models (in particular Romer, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1991), an increase in the level of competition should reduce the incentive to innovate as it also reduces future rents. However, as surveyed in Aghion and Griffith (2005), this prediction is not very clear in the data. This motivates the authors to emphasize the non-linear relationship between competition and innovation: while competition can still dampen innovation, it also induces firms to intensify their innovation activities in order to escape competition. Empirically, a number of papers have looked at the reaction of innovative firms to competition shocks, often using trade shocks (Autor et al., 2020; Aghion et al., 2018; Bloom et al., 2016; Iacovone et al., 2011; Aghion et al., 2021; Akcigit et al., 2018). To the extent that patents give a temporary monopoly power to its assignee and that standards lock a whole industry in a given technology, then standardization leads to a reduction of competition if the underlying technology is owned by a small number of firms. Our paper therefore contributes to this empirical literature by considering a more direct measure of competition and allows to look at the impact of a change in the degree of competition at the firm and aggregate level.

We also study the information content of standardization. In particular, we relate to a literature that studies how financial markets react to innovation-related corporate events. For example, Eberhart et al. (2004), Chan et al. (1990) and Szewczyk et al. (1996) show that firms exhibit positive abnormal returns and higher share value when the management announces an unexpected R&D investment plan. Similar results are found in Kogan et al. (2017), Pakes (1985), Nicholas (2008) and Austin (1993), who

[1]While empirical studies have used data for selected SSOs for which SEP declarations are available (Bekkers et al., 2017; Baron and Pohlmann, 2018), true standard essentiality is often questioned and problems of both over-declaration and under-declaration may arise (see the discussion in Brachtendorf et al., 2022).

show that markets positively reacts to news on patenting activity.[2] All these papers demonstrate that the market efficiency hypothesis (see for example Daniel et al., 1998, Mitchell and Stafford, 2000) holds also when information on corporate innovation activity is disclosed: markets are able to correctly understand and discount what the future benefits of innovation will be. Our paper shows that this is the case also when information on a new standard is released.

Finally, our work contributes to the literature on text-mining applied to the semantic analysis of patents and standards.[3] Text mining methods are increasingly used in economics and in particular in innovation economics, notably for the analysis of patent data (see Abbas et al., 2014 for an overview). For example, the semantics of patent documents can be used to measure patent similarity (Arts et al., 2018; Kuhn et al., 2020), to select patents in specific technologies (Bergeaud and Verluise, 2022; Dechezleprêtre et al., 2021; Bloom et al., 2021) or to classify patents (Bergeaud et al., 2017; Webb et al., 2018; Argente et al., 2020). The content of patent publications has also been used to construct measures of novelty based on the amount of textual dissimilarity from previous patents and high similarity with subsequent ones as done by Kelly et al. (2021).

The paper is organized as follows: Section 2 briefly describes the matching procedure and the construction of the data; Section 3 looks at how standardization relates to indicators of patent quality; Section 4 presents our firm-level results and link our results with the theoretical literature on innovation and competition; Section 5 discusses the aggregate implications of our results; Section 6 concludes.

2 Data construction and matching

Our aim is to construct a measure of proximity between patents and standards. Contrary to Brachtendorf et al. (2022) who use text-mining techniques to investigate the standard essentiality of patents, our goal is not to measure this dimension of proximity. Rather, we want to pick up the ability of firms to adapt to the new standard and to quickly develop applications of a new standard and deploy the new technology. It is for this reason that we use the universe of standard and patent documents and not just those that are linked through declared standard essentiality.

[2]In a similar vein, Ma (2021) uses patent data to construct measures of technological obsolescence and analyzes earnings forecasts and stock returns in response to the obsolescence of a firm's innovation portfolio.

[3]Brachtendorf et al. (2022) also consider the link between standards and patents. Specifically, they use SEP declarations for one specific SSO, namely ETSI (European Telecommunications Standards Institute), to evaluate the true standard essentiality of patents. Contrary to their paper, we concentrate on the universe of standards released by a large variety of SSOs and are interested in how standardization affects real outcomes at the firm- and macroeconomic level. The *Semantic similarity of patent–standard pairs database* described in Brachtendorf et al. (2020) considers standards not only from ETSI, but also from IEEE and ITUT.

2.1 Data sources

Patent data. A patent is an exclusive right granted to an inventor or an assignee for an invention in exchange for the disclosure of technical information. For the matching procedure, we use all priority applications that are available in the `IFI CLAIMS` database from 1980 to 2020, without restrictions on the technological field.[4]

The `IFI CLAIMS` database contains most of the information we need about patents. In particular, we extract the abstract, the technological field (through the International Patent Classification code, or IPC) and the filing date of the patent application. We restrict our sample to patents filed between years 1980 and 2010. This corresponds to over 21.5 million observations on the patent-level.

Standard data. A standard, similar to a patent, is a document that describes certain features of a product, a production process or a protocol. Contrary to patents that are filed by individual inventors or firms, standards are developed by standard-setting organizations (SSOs) which gather industry experts from both the private and public sector in working groups and technical committees. Well-known examples are international SSOs such as ISO (International Organization for Standardization), national standard bodies such as DIN (Deutsches Institut für Normung) or industry associations such as IEEE (Institute of Electrical and Electronics Engineers). Most standards are considered public goods and many SSOs are non-profit organizations. Requiring approval by all stakeholders involved in the development of standards, they are often called *consensus standards*.

To collect information on standards, we use the Searle Centre Database on Technology Standards and Standard Setting Organizations (see Baron and Spulber, 2018 for more details). This data is largely based on `Perinorm`, a bibliographical database of product standards whose purpose is to provide subscribers (usually professionals) with basic information on the standard and the possibility to purchase the access to individual standard documents. Our database covers all types of standards that have been released in a large number of industrialized countries. The `Perinorm` database also contains keywords describing each standard. These keywords were provided by `Perinorm` experts when including standards into their database to facilitate the search for specific standards by its users. They represent one of the main ingredients for our matching procedure.

[4] Patents are grouped into families which include different publications that are more or less related to the same invention. More precisely, during a 12-month period following the filing of an application, the applicant has a *right of priority* meaning that during this period, she can file a similar patent in a different patent office and *claim the priority* of the first application. If the priority claim is valid, the date of filing of the first application is considered to be the effective legal date for all subsequent applications. All the patents sharing a similar *priority application* define a family. The priority application is the first patent in a family (see Martinez, 2010 for more details).

We clean the standards data as follows. First, we regroup standard documents that are equivalent. Indeed, a single standard can be released several times, for example once by a French SSO and once by a German SSO. To avoid keeping duplicates, we regroup those standards and create a database in which we store the standards group identifiers, the standards contained in the group, their ICS (International Classification of Standards) and the earliest date of publication. Finally, we store the keywords associated to the standards of the group. More details are provided in Appendix A.

2.2 Semantics-based matching of patents to standards

Matching procedure. We start by processing the keywords that have been provided by `Perinorm` experts for each standard. We first clean these keywords using common techniques used in text-mining (such as removing upper-case letters, special symbols, punctuation or stop words such as *the*, *at*, *from*, etc.). We then form k-grams, i.e. a sequence of k words that we consider as a unique entity (i.e. the 2-gram *air condition* is not the same as considering *air* and *condition* separately). We stem these k-grams which consists in only keeping the "root" of the keyword (i.e. *fertilizing* and *fertilizer* become both *fertiliz*). As a result, we obtain a database where each standard is associated with a list of k-grams.

Then, we proceed similarly and extract keywords from the patent abstracts, form and stem k-grams, and keep those that are in the list of standards keywords. Thus, we obtain a database where each patent and standard is listed with their associated k-grams. We calculate the so-called *inverse document frequencies* for each k-gram in our respective database of extracted standard and patent k-grams to assign them an importance weight.[5] We only keep k-grams that do not appear in more than 1 out of 1000 (5000) standard (patent) documents. Then, we register all patent-standard combinations which share the same k-gram on the k-gram-level. A score is then calculated by summing the importance weights across all patent-standard combinations and normalizing the score by the number of k-grams that were extracted from the patent abstract. This score forms the basis of our analysis and measures the semantic proximity of each patent to standard. The exact definition of this measure is presented in Appendix B.1.2.

This matching procedure results in more than 1.6 billion patent-standard combinations. For reasons of computational power, we need to restrict the number of patent-standard matches that we use for our empirical analyses in Sections 3 and 4. We therefore extract only the first 100 million best matches (based on the highest score). This choice of 100 million is admittedly arbitrary, but is in line with the highly skewed distribution of the scores. Appendix B describes the matching procedure in detail.

[5] The inverse document frequency is based on a measure of how often a word shows up in a database of documents. See Appendix B for details.

Sample selection. Based on the extraction of the first 100 million matches, we report descriptive statistics of our score in panel (A) of Table 1. The first row reports the distribution of the score. The second row shows the number of standards that a patent is matched to: the median patent is closely linked to 8 standards, but the distribution is highly skewed, with the majority of patents only being matched to one or a few standards and 1% to more than 400 standards.

For the econometric analysis on the patent- and firm-level (respectively Sections 3 and 4), we consider both patents that are matched and those that are not matched to a standard. The descriptive statistics for this sample can be found in panel (B) of Table 1. There, we also report the time lag between the release of the patent and the release of the matched standard for this sample. On average, the release of a matched standard occurs 2.6 years before the filing date of the patent, thus indicating that standards more often lead than lag an associated patent. In other words, standardization may lead to more patenting if the standardized technology leads to follow-up innovation. Actually, such standard-induced innovation is a specific aim of the standardization process: by defining common rules for the design and use of an *underlying* technology, firms are incentivized to invest into the technology and develop marketable applications and products.[6]

However, for our analysis, we are interested in the firm-level effects of standardization events that relate to a firm's patent portfolio. Therefore, in panel (C), we restrict the sample to only those matches where the release of the standard occurs the same year or subsequent to the filing of the patent application. For this sub-sample, the median time lag for this restricted sample is 8.0 years while the average is slightly higher, at 10.1 years.

Finally, in panel (D), we report the aggregated score, summing all scores across all matched standards on the patent-level. Mirroring the distribution of zero matches, we note once again a highly skewed distribution.

To evaluate the quality of our matching procedure we verify how individual patent-standard matches relate broad categories of the IPC (patents) and ICS (standards) classifications. Essentially, we are linking the two classification systems on the basis of the individual matches obtained in our matching procedure. The results of that exercise can be found in Appendix B where Table B.1 lists the closest IPC class for every ICS field. Across the board, the matching seems reasonable and confirms our approach.

[6]Patenting activity might also increase following standardization if firms patent for strategic purposes (Hall and Ziedonis, 2001; Choi and Gerlach, 2017, see also Kang and Bekkers, 2015 for a discussion of "just-in-time" patenting).

Table 1: DESCRIPTIVE STATISTICS OF THE MATCHING PROCEDURE

	Mean	SD	Min	Max	p1	p5	p25	p50	p75	p95	p99	N
(A) Keyword matching sample												
Score	715.7	1,766.6	138.8	658,691.2	141.3	151.8	211.8	315.2	638.7	2,345.7	6,289.0	100,000,000
Standards	41.9	87.5	1.0	1,233.0	1.0	1.0	2.0	8.0	35.0	217.0	471.0	2,389,251
(B) All patents (matched and unmatched)												
Score	599.6	1,634.6	0.0	658,691.2	0.0	0.0	166.4	262.4	543.5	2,026.9	5,622.0	113,427,683
Time lag	-2.6	15.6	-50.0	38.0	-40.0	-31.0	-13.0	-1.0	8.0	21.0	30.0	95,201,007
Standards	4.6	31.9	0.0	1,233.0	0.0	0.0	0.0	0.0	0.0	10.0	136.0	20,506,259
(C) Restricted sample: excl. matches with patent filing year > standard release year												
Score	505.6	1,549.8	0.0	658,691.2	0.0	0.0	0.0	227.1	480.4	1,799.8	5,139.7	64,574,039
Time lag	10.1	7.7	0.0	38.0	0.0	0.0	4.0	8.0	15.0	26.0	32.0	46,347,363
Standards	2.6	163.7	0.0	681,495.0	0.0	0.0	0.0	0.0	0.0	4.0	72.0	17,596,230
(D) Aggregated sample												
$\sum$ score	1,592.2	30,814.2	0.0	27,657,164.0	0.0	0.0	0.0	0.0	0.0	1,495.9	31,087.2	20,506,259

Notes: The table reports descriptive statistics for the score, the number of matched standards per patent and the time lag (in years) between the release of the standard and the filing year of the patent. The keyword matching sample comprises the extraction of the first 100 million scores of our matching procedure. The sample of utility patents discards design patents and also includes unmatched patents which receive a score of zero. The restricted sample only comprises utility patents, matched and unmatched, for which the patent filing year does not exceed the standard release year. The aggregated sample sums all scores on the patent-level for the restricted sample.

2.3 Firm-level data

Aggregation of scores at the firm-level. We use the mapping provided by Kogan et al. (2017) to associate each patent to the Compustat firm that filed it.[7] Given the mapping between patents and standards, we can aggregate scores at the firm-quarter level by weighting the sum of patents' scores with the relative importance of each 3-digit IPC class in the firm's initial (pre-sample) stock of patents. Formally, define J as the set of all IPC classes such that $j \in J$ is a specific IPC class, and call $Score_{i,p,j,t}$ the score obtained by firm i when matching patent p that belongs to the IPC class j and was issued up to $t-4$ to a standard published at time t. Then, the weighted aggregation of scores over IPC classes can be seen as a measure of the proximity of firms to the new technological frontier (defined by standard released at t). Formally, we define the proximity as:

$$Tech.Prox_{i,t} = \sum_{j \in J} \omega_{j,t_0} \sum_{p \in j} Score_{i,p,j,t}$$

where ω_{j,t_0} is the share of patents in the IPC class j measured in t_0, i.e. before 1980. We do this weighting for two reasons: first, the weighting reduces the role of those patents in IPCs that are not at the core of the firm's research activity and technological field; second, computing the weights in a pre-sample periods reduces the problem of firm self-selection into a specific IPC, which they anticipate to become important for a potential standard at some point in time. We come back to this in Section 4.5 where we validate the goodness of our measure.

In conclusion, the variable $Tech.Prox_{i,t}$ is a firm-quarter level continuous variable ex-

[7]More precisely, we use an updated version of Kogan et al. (2017) taken from their Github repository.

pressing the (IPC-weighted) proximity of the stock of patents (accumulated until $t-4$) of a firm to the standard released in quarter t. This variable can be either equal to zero, if the patents of a firm do not map into a new standard, or positive. In this case, the bigger the variable $\mathtt{Tech.Prox}_{i,t}$ the closer the firm's portfolio of patents to the newly released standard.[8] In Section 4.5, we provide evidence that this variable is not just a by-product of the type of the firm innovation activity (for example favoring some IPC classes against other), or the quantity of innovation, but it is indeed capturing the effective proximity of firms' patents to the new standard.

Balance sheet data. We use firms' balance sheet data from Standard&Poor's Compustat to build all (real) dependent and control variables used in the empirical analysis of Section 4. The dependent variables under consideration are: sales, R&D investment and market share. Sales are the revenues of the firm as reported at the end of the quarter in the income statement. Since it is usually under-reported, R&D expenditure is measured as a 4-quarter moving average. For comparability across firms, we normalize these two variables by the (mean) level of fixed assets (property, plant, equipment).[9] The last variable of interest is the market share of the firm, defined as the ratio of firm-level sales on the total volume of sales in a NAICS 3-digit industry (NAICS3).

Along with these variables, we consider also the following characteristics: the age of the firm (expressed in quarters), the q-value of investments (built as the book value of liabilities plus the market value of common equity divided by the book value of assets), leverage (as debt over the book value of assets), market capitalization (expressed in billions of USD) and a dummy taking value one if the firm is operating in a high-tech industry (i.e. drugs, office equipment and computers, electronic components, communication equipment, scientific instruments, medical instruments, and software) as defined in Chan et al. (1990). Finally, we follow De Loecker et al. (2020) to construct an estimate of the average markup at the 3-digit NAICS industry-level. This information allows to understand which industry is (on average) less or more competitive and –therefore– which firms operate in a less or more competitive market. We define a firm as belonging to a non-competitive market if the average markup of its industry is above the 75th percentile of the distribution.

Financial market data. As explained in Mitchell and Stafford (2000), abnormal returns are useful to study short-term market reactions to corporate events. Following

[8] We normalize this measure by its standard deviation such that it ranges from 0 to over 6. It is equal to 0 for more than half of the sample. See Table 2 for more details.

[9] As we show in Bergeaud et al. (2022), the value of assets is sensitive to the proximity measure. For this reason, we prefer to normalize sales, capital investment and R&D with the mean-level of fixed assets rather then with the contemporaneous level or some lag. By doing so, the change in the numerator of the index is not influenced by the change in the denominator.

this line, we want to evaluate how markets interpret the standardization event. Since our analysis focuses on the real effects of technological proximity on competition and sales within a NAICS3 industry, we calculate abnormal returns at that level of disaggregation. Here, we describe the procedure of extrapolation. First, we match Compustat with data from the Center for Research in Security Prices (CRSP). Then, for each NAICS3 industry, we build the returns of a portfolio composed of all firms listed in that industry. Formally, given the number of firms I_t belonging to the NAICS3 industry s at time t, the return on the industry s portfolio can be written as $r^s_t = \sum_{i=1}^{I_t} \omega^s_{i,t} r_{i,t}$. Notice that $\omega^s_{i,t}$ is the weight of each firm i in the industry-specific portfolio s, and it is equal to the relative market capitalization of firm i in industry s at that moment in time. Hence, we estimate a statistical model which differs from the baseline Capital Asset Pricing Model (see Jensen et al., 1972) only for the definition of the market portfolio, here defined at industry-level. Formally –given information on the 3-month t-bill rate (r^f_t) and the return on each industry portfolio (r^s_t)– for every firm i belonging to industry s and 10-year rolling window with ending period τ, our asset pricing model is:

$$r_{i,t} - r^f_t = \alpha_{i,\tau} + \beta_{i,\tau}(r^s_t - r^f_t) + \varepsilon_{i,t}, \;\; \forall t \in (\tau - 10yrs, \tau]$$

where $r_{i,t} - r^f_t$ is the excess return of firm i, $r^s_t - r^f_t$ is the excess return of industry s portfolio, $\varepsilon_{i,t}$ is the error term. Then, we use the OLS estimates $\widehat{\alpha}_{i,\tau}$ and $\widehat{\beta}_{i,\tau}$ to predict the firm's (excess) return one quarter after the end of each 10-year estimation window, i.e. in period $\tau + 1$. Finally, we define the abnormal return ($ar^s_{i,t}$) of a firm i from industry s as the difference between the observed (excess) return and the predicted one:

$$ar^s_{i,\tau+1} = (r_{i,\tau+1} - r^f_{\tau+1}) - \left(\widehat{\alpha}_{i,\tau} + \widehat{\beta}_{i,\tau}(r^s_{\tau+1} - r^f_{\tau+1})\right).$$

We repeat this procedure for every firm i in the sample and for all available 10-year rolling windows with ending period equal to τ, $\tau+1$, $\tau+2$, ..., $\tau+T$.

In order to look at markets' reaction beyond abnormal returns, we match Compustat to data from the Institutional Brokers' Estimate System (IBES). From this dataset, we collect professional analysts' expectations over the future Earning-Per-Share (EPS) ratio of the firm. In particular, we look at how forecasters expect the EPS to be at the end of the following fiscal year. In fact, by considering a fixed forecasting horizon, we can study how expectations change over time as the end of the fiscal year approaches. Therefore, for each firm and quarter, we take the mean of the 1-year EPS forecast across all professional forecasters, and obtain a measure of market expectations over the future economic performance of the firm.

Sample selection. Once equipped with these firm-level variables, we follow Brown et al. (2009) and exclude all regulated utility and financial firms as well as firms with missing assets. Then, we match the remaining sample of Compustat firms with patent

Table 2: DESCRIPTIVE STATISTICS FOR THE FIRM-LEVEL DATA

	Mean	SD	p1	p5	p25	p50	p75	p95	p99	N
(A) Proximity Measure										
Tech.Prox	0.34	2.02	0.00	0.00	0.00	0.00	0.10	1.27	6.24	24,162
$\mathbb{I}$[Tech.Prox >0]	0.48	0.49	0.00	0.00	0.00	0.00	1.00	1.00	1.00	24,162
(B) Firm Characteristics										
Sales	0.62	0.72	0.01	0.08	0.25	0.47	0.78	1.60	2.99	24,162
R&D	0.04	0.26	0.00	0.00	0.00	0.01	0.02	0.14	0.56	24,162
Market Share (NAICS3)	0.05	0.10	0.00	0.00	0.00	0.01	0.05	0.21	0.49	24,162
Age (quarters)	98.99	49.92	21.00	21.00	53.00	110.00	137.00	171.00	181.00	24,162
Q	1.93	2.15	0.74	0.90	1.17	1.49	2.12	4.43	8.69	24,162
Leverage	0.19	0.15	0.00	0.00	0.06	0.17	0.27	0.45	0.65	24,162
Market Cap. (Billion$)	9.17	28.99	0.00	0.02	0.19	1.27	5.61	42.22	139.89	24,162
$\mathbb{I}$(*Tech-firm*)	0.30	0.45	0.00	0.00	0.00	0.00	1.00	1.00	1.00	24,162
Industry Markup (NAICS3)	1.50	0.30	1.05	1.13	1.25	1.40	1.75	1.92	2.43	24,162
$\mathbb{I}$(*Non-Competitive Industry*)	0.25	0.43	0.00	0.00	0.00	0.00	1.00	1.00	1.00	24,162
(C) Financial Mkts										
ar^{NAICS3}	0.00	0.40	-0.58	-0.32	-0.09	0.00	0.08	0.27	0.56	18,531
1yr EPS Forecast ($)	1.43	0.96	0.06	0.19	0.67	1.25	1.99	3.40	3.99	15,766

Notes: The variable Tech.Prox measures the proximity of the portfolio of patent of the firm to the standard. $\mathbb{I}$[Tech.Prox >0] is a dummy that takes value one for positive values of the variable Tech.Prox. Sales is the firm-level of sales normalized by the mean-level of fixed assets (property, plant, equipment). The *Market Share* is constructed at the NAICS 3-digit level. R&D is the level of R&D expenditure normalized by the mean-level of fixed assets. Age is the number of quarters the firm is active. Q is the q-value of investments, and is built as the value of liabilities plus the market value of common equity divided by the book value of assets. Leverage is debt over the book value of assets. *Market Capitalization* is expressed in Billion of US dollars. The dummy variable $\mathbb{I}$(*Tech-firm*) takes value one if the firm operate in one of the following industries: drugs, office equipment and computers, electronic components, communication equipment, scientific instruments, medical instruments and software. The NAICS 3-digit industry markup is constructed following De Loecker et al. (2020). $\mathbb{I}$(*Non-Competitive Industry*) is a dummy that takes value one if a firm is operating in a NAICS 3-digit industry with markup above the 75^{th} percentile. ar^{NAICS3} is a measure of stock market abnormal return built from a standard CAPM model with a NAICS 3-digit index as market portfolio. The *1yr EPS Forecast* is the mean forecast across all professional forecasters of the earning-per-share expected by the end of the following fiscal year, and it is expressed in dollars.

data and our proximity measure. Then, in order to implement our identification strategy (see Section 4), we keep only firms that are publicly listed, for which all constructed variables are jointly available (except abnormal returns and EPS forecasts), and that have registered at least one patent in their life. By doing so, we end up with a sample of 24,162 firm-quarter observations spanning from 1984 to 2010.

Table 2 reports descriptive statistics for this sample. As from panel (A), the proximity measure at the firm-quarter level has a mean equal to 0.34 and a standard deviation equal to 2.02. In our sample, 48% of firms have a positive proximity value. As from panel (B), the mean level of sales is 62% of the value of fixed assets. The mean (flow) investment in research and development (R&D) is equal to 4% of the value of fixed assets. Within NAICS3 industry, the average firm has a market share equal to 5%. The average age of the firm is roughly 25 years, with a q-value equal to 1.93, 19% of its balance sheet is composed by debt, it has a market capitalization of 9.17 billion USD and a 28% probability to be in a high-tech industry. The average firm operates in a NAICS3 industry with a markup of 1.5. 25% of firms are from industries with markups above or equal to 1.75, and we define these industries as non-competitive. When matching this data with information on abnormal returns and EPS forecast, the sample is reduced. As from panel (C), our sample contains 18,531 observations on abnormal returns and 15,766 observations on EPS forecasts. The average abnormal return is zero while the average 1-year EPS forecast is 1.43 dollars per share.

3 Innovation and standardization: patent-level results

In this section, we look at the characteristics of patents that are associated with a high score, i.e. patents semantically close to a specific standard. In particular, we compare the computed score with measures of patent quality or economic value.

3.1 Economic value of a patent

Kogan et al. (2017) compute the financial value of a patent based on the stock market reaction to the news of a patent application being granted. This is a forward-looking measure of economic agents' evaluation of the granted patent. While we expect our score to correlate with this measure, there are conceptual differences. While both measures are indicative of the economic value of a patent, our score captures the underlying technology's potential for market-wide adoption or its adaptability to the new standard. It is therefore particularly meaningful to study questions of market share and competition. The economic value à la Kogan et al. (2017) measures markets' perception of the future value of the technology at the time of the patent grant, but potentially abstracts from any future developments and spillovers that are not known at the time of the grant (standardization being one of them).

To relate our score with the economic value of a patent, we sum the score across all associated standards at the patent-level (unmatched patents have a zero score). We then run the following patent-level regression:

$$\log\left(\mathrm{value}_i\right) = c + \alpha \log\left(1 + \mathrm{score}_i\right) + \beta \log\left(1 + \mathrm{cit}_i\right) + f_{t(i),k(i)} + \varepsilon_i \qquad (1)$$

where value_i is the economic value of patent i (in millions USD) from Kogan et al. (2017) and score_i is the normalized sum of scores across all associated standards of patent i. We include the number of forward citations cit_i, also taken from Kogan et al. (2017), as a control variable as well as fixed effects $f_{t(i),k(i)}$, namely the interaction of the year and quarter of the grant date of the patent $t(i)$ and its 3-digit IPC class $k(i)$.

Table 3 summarizes the results: column 1 corresponds to the empirical specification that does not control for the number of citations whereas it is added to the model in column 2. Whether we control for the number of forward citations received or not, our aggregated score is positively associated with a higher financial value of the patent and is statistically significant, even within a given technological class in a given year. In order to translate these results into quantitative numbers, we run regression specification (1) with a dummy indicating whether a patent is matched to at least one standard or not, adding fixed effects as in Table 3. The coefficient for the dummy for a non-zero score ranges between 0.047 and 0.081 for the different specifications, implying that a close link with at least one standard is associated with a 4.7–8.1% higher patent valuation. The median (mean) patent being valued at 9.6 (27.0) mio USD in the sample

Table 3: REGRESSION RESULTS FOR FINANCIAL, SCIENTIFIC AND PRIVATE PATENT VALUE

	(1)	(2)	(3)	(4)	(5)
	Kogan et al. (2017)		Forward citations	Expiration	
Score	0.0062***	0.0050***	0.0035***	-0.0030***	-0.0024***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]
Observations	1,163,913	1,163,913	20,479,110	5,441,961	5,441,961

Notes: Patent level regressions. "Score" is the log of 1 plus the sum of scores across all associated standards of a given patent (see Section 3.1). The dependent variables are: columns 1 and 2: the economic value of the patent as computed by Kogan et al. (2017); columns 3: the number of forward citations received by the patent over a 10 year window; columns 4-5: maintenance decisions for patent i. Columns 1-2 use an OLS estimator, column 3 uses a Poisson model and columns 4 and 5 use a Cox hazard model. Columns 1-2 include a technological class (3-digit IPC) interacted with the granting year/quarter fixed effect as in Kogan et al. (2017). Column 3 includes the 3-digit IPC class interacted with the filing year/quarter fixed effect. Similarly, the Cox model in columns 4 and 5 stratifies the data by the interaction of the 3-digit IPC class with the filing year/quarter. Columns 2 and 5 additionally control for the logarithm of the number of forward citations received by the patent. The sample of patents includes: columns 1 and 2: priority patents from IFI claims matched to the Kogan et al. (2017) sample with filing year between 1980 and 2010; column 3: priority patents from IFI claims published over the period 1980-2010; columns 4 and 5: priority patents from IFI claims matched to USPTO patents with information on renewal fees for the period 1981-2015. In the last two columns, a patent can be included several times in the sample due to different schedules of renewal (see Section 3.3). "*", "**" and "***" designate significance at the 1%, 5% and 10% level.

where we match our database to the Kogan et al. (2017) data, this amounts to raising its value by 452,000–779,000 (1.3–2.2 mio) USD.

Results do not change when using the real instead of nominal value of patents or using unweighted counts, i.e. by simply counting the number of associated standards per patent.

3.2 Scientific value of a patent: forward citations

A popular measure of the scientific value of a patent are forward citations (Hall et al., 2005), i.e. citations of the patent in question by subsequent patents. A highly cited patent is used by a larger number of future inventions and therefore signals high technological content and to a certain extent also high economic value.

We extract forward citations from `IFI CLAIMS` and concentrate on the number of forward citations received within ten years after publication. As is common in the literature (see e.g. Hausman et al., 1984), we use a Poisson regression model approach to take into account the discrete nature of the dependent variable and the large number of zeros. In all other respects, the regression setup follows equation (1).

The results are presented in Table 3, column 3, and mirror the ones from columns 1 and 2. There is a clear positive relation between our aggregated score and the number of citations a patent receives. Once again, results are robust to using unweighted counts of the number of standards associated to a patent.

3.3 Private value of patent protection: renewals

As a last exercise, we look at how patent owners themselves value their patents. In fact, patent holders have to pay maintenance or renewal fees to keep a patent in force.[10] Pakes and Schankerman (1984) and Pakes (1986) have argued that these expenses for the renewal of patents is an indicator of the private return of holding a patent. The duration of effective patent protection is therefore an indicator of the economic value of a patent, either for the purpose of extracting royalties or to hinder competitors from using the technology.

In the US, renewal decisions are due 3.5, 7.5 and 11.5 years after the grant date and patent holders have to pay a maintenance fee in order to benefit from intellectual property rights. We obtain data on the payment of maintenance fees for USPTO patents for the period 1981–2015 and match these data to our dataset. Patent renewal decisions are a function of the age and cohort of the patent and the discounted value of the net economic benefit of holding the patent (see Schankerman, 1998 for a discussion). Empirical analyses therefore turn to survival models where the exit of a patent (i.e. the non-payment of maintenance fees) is described by observable explanatory variables. As is common in the literature, we adopt a Cox hazard model specification (Cox, 1972) to investigate the "survival" of a patent as a function of the sum of scores that a patent "received" between its filing date and the due date of the maintenance fee. Similar to the approach taken in Sections 3.1 and 3.2, we stratify the data by the interaction of the filing year and quarter and the 3-digit IPC class and include 10-year forward citations as a control.

Columns 4 and 5 of Table 3, respectively excluding and including citations as a control, report the beta coefficients of the Cox model. The higher is the score of a patent, the lower is the hazard-probability that the owner will let that patent expire. In other words, patent holders who observe standardization events that are associated with their patent (over the time window during which the decision has to be taken) are more willing to pay the maintenance fee in order to renew their patent rights. This indicates that our score is positively related to the private value of patents. Results do not change when including grant lags as additional controls or stratifying the data by grant year and quarter.

We therefore conclude that our proximity measure is related to various dimensions of patent quality. However, it is the first that captures a patent's potential for widespread market adoption and/or its utility in deploying new technologies. In the following section, we will illustrate that standardization offers a distinct value-added that extends beyond the conventional measures of patent quality.

[10] After 20 years, patent protection cannot be renewed.

4 The implications of standardization: firm-level results

In this section, we move from patent- to firm-level data. We first relate the firm-level aggregation of patent-to-standard scores, $\texttt{Tech.Prox}_{i,t}$, to financial market data in order to study its information content. In a second step, we relate our proximity measure to real variables to study the implications of standardization in terms of sales, markets shares and R&D expenditures.

4.1 Empirical strategy

In order to explain clearly our empirical strategy, it is first important to understand how the standardization procedure works in practice, what is the timing of events from the conception to approval of a standard, and why firms have the incentive to comply with the new standard.

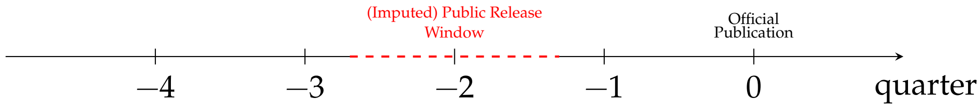
We can briefly summarize the publication path of a standard as follows.[11] Standards are developed in working groups and technical committees of SSOs. All stakeholders, be they private firms or public sector representatives, can contribute to the standard development – a process that often spans several years. Once the standard is proposed and drafted, it goes under the scrutiny of a committee. This first phase concludes with a vote. If the committee's vote is positive, then the draft of the standard is publicly released and circulated to other sub-committees, external committees of experts, other national or international standard-setting organizations for comments. This corresponds to the very moment that information on the content of the standard becomes available to the public for the first time. In the following phase –which lasts 3 months– suggestions and comments are collected. If no substantial critique is raised, the final version of the draft will be immediately approved and published within the next 6 weeks. On the contrary, if some revision is needed or further analysis is required, then the process is extended in order to give the proposing organ some extra-time (2 to 3 months) to comply with the specific requests. Then, the committee has 2 months to judge the revision to the document. If the new draft of the standard is satisfying, then it is approved and published within the following 6 weeks.

As we observe only the official publication date of approved standards, knowledge of the administrative procedure of approval allows to back up for each standard the time-window in which the first draft became public knowledge, i.e. roughly between (minimum) 4 and (maximum) 8 months before the final publication date. Figure 1 sketches the timeline (in quarters) of the administrative procedure of standards' approval along with the official publication date in black and the imputed time-window of public circulation of the first version of the standard in red. As shown, if publi-

[11] As a reference, see the International Organization for Standardization website.

cation occurs at time 0, the first (imputed) public release of the standard occurs in a time-window around quarter -2.

Figure 1: THE TIMING OF STANDARDS APPROVAL



Notes: This figure sketches the administrative procedure of standards' approval. The official date of publication of the standard by the standard-setting organization is known and occurs at quarter 0. Given information on the administrative procedure of approval and publication of a standard, we back up the (imputed) time-window in which the judging committee voted in favor of the standard and made the standard's draft publicly available. This happens roughly around −2, i.e. 2 quarters before the official publication date.

What happens when a standard is finally published? From the moment of the publication onwards, firms are free to chose whether to apply the new standard to their products on a voluntary basis. In fact, standards are not legally binding[12] (unless they are referenced by government regulation, as for example in health or environmental legislation). Yet, it is difficult for an individual firm to not comply with what becomes *standard* in its industry as its products would be at a considerable disadvantage compared to those that follow the standard specifications.[13] Consumers and producers value products or inputs that are compatible, have a certain quality level or are less subject to information asymmetries. Indeed, interoperability and network effects are one of the main reasons SSOs take on the coordinating role of standard development among industry stakeholders.[14] In a similar manner, value chains (both domestic and across borders) require that downstream and upstream producers agree on common specifications to allow for compatibility. Therefore, unless a firm is the first to market an entirely new, independent product, market forces and demand effects can render a standard *de facto* binding.

Given the procedural path of approval and firms' incentives to comply, we can now introduce our empirical model to assess the impact of standardization on firm dynamics. Yet, it is important to stress that different standards can be released in subsequent periods. Therefore, in order to better isolate the effect of the introduction of a new standard, we resort to a distributed lead-lag model. The main interest of this approach with respect to a static analysis is that it allows to capture the full dynamics of the response. In particular, in our setting, we know that a static model would be biased since

[12]They are referred to as *voluntary consensus standards*. See e.g. the general description by ISO.

[13]See Schmidt and Steingress (2022) for the role of harmonized standards for international trade integration. They argue that the benefits of standardization are a major driver of standard adoption by firms when adoption costs are lowered through the cross-country harmonization of standards, thus increasing trade among countries whose SSOs agree on the same (voluntary) standards.

[14]Early examples of network effects are railway gauges (Gross, 2020), shipping containers (Bernhofen et al., 2016) or the QWERTY keyboard (David, 1985).

the firm's response could be affected also by subsequent and previous releases. Our generic model is described in equation (2):

$$Y_{i,t} = \alpha_i + \phi_{s(i),t} + \sum_{n=-16}^{N=12} \beta_n \mathrm{Tech.Prox}_{i,t+n} + X'_{i,t-1}\eta + \varepsilon_{i,t}, \tag{2}$$

where $Y_{i,t}$ is the firm-level dependent variable under consideration. α_i is a firm fixed effect, $\phi_{s(i),t}$ a NAICS 3-digit industry fixed effect interacted with a time fixed effect. This controls for any time effect that might differ across industries (e.g. because of sector-specific demand variation, seasonality, changes in legislation at the industry-level, momentum, etc.). $\mathrm{Tech.Prox}_{i,t}$ expresses the proximity of the stock of patents of firm i at time $t-4$ to the standard publicly released at t. We include 16 lags and 12 leads of the proximity measure (recall that the time unit here is a quarter). Finally, $X_{i,t-1}$ is a vector of control variables (which we discuss later) and $\varepsilon_{i,t}$ is the error term, which we assume to be normally distributed (conditional on all our covariates) and to be independent across different i.

In this model, β_n measures the effect of the introduction of a new standard at $t+n$ on the value of Y measured at t, controlling for the effect of all previous and future standards' releases. We will check that the response of the firm to future releases remains insignificant and will present our results by plotting the values of $\hat{\beta}_n$ for all n, along with its 95% confidence interval.

4.2 Information effects of standardization

Our goal is to use the proximity measure $\mathrm{Tech.Prox}_{i,t}$ to study the effect of standardization on firm-level economic performances and market structure. Although standards are defined to foster the adoption of the best technology, it might be the case that the technology selected by the SSO has already been broadly adopted by the time of the standard release. In this case of ex-post standardization, our proximity measure would correctly pick up the quality of firms' patents, but not the additional impact of standardization. Our goal is to capture the selection of a (not yet adopted) technology and the relative proximity of firms with respect to this new frontier. By studying the information content of the variable $\mathrm{Tech.Prox}_{i,t}$, we show that this is the case.

To do so, we look at how financial markets react when the content of a standard becomes public. In the near future following a standard release, firms that already own technologies that allow them to deploy and scale their know-how should perform better than their peers. If markets are efficient (e.g. see Eberhart et al., 2004, Daniel et al., 1998, Mitchell and Stafford, 2000), they should update their expectations when new information is disclosed as they discount firms' future performance.

In order to test this, we consider our baseline lead-lag model of equation (2) using two alternative dependent variables aimed at capturing markets' reaction:

1. the abnormal return over a NAICS3-industry portfolio, i.e. $ar_{i,t}^{NAICS3}$;
2. the change in the 1-year EPS forecast from professional agencies, i.e. $\Delta\mathbb{E}[EPS_{i,t+4}] = \mathbb{E}[EPS_{i,t+4}|\mathbb{I}_t] - \mathbb{E}[EPS_{i,t+4}|\mathbb{I}_{t-1}]$, where $\mathbb{I}_t$ is the information set available to professional forecasters in that period.[15]

The vector of controls $X_{i,t-1}$ includes age, q-value of investment, leverage and market capitalization of firm i along with a dummy variable taking value one if the firm is operating in a high-tech industry. We consider these variables to take into account respectively for how long a firm has been listed, its growth opportunities, its capital structure, market value and whether it is already working in an innovative sector. As explained in Chan et al. (1990) and Szewczyk et al. (1996), these characteristics are important for the magnitude of the stock market reaction following abnormal R&D activity or other innovation-related events.

Figure 2a plots all estimated β_n (along with 95% confidence intervals) for the dependent variable $ar_{i,t}^{NAICS3}$. Standard errors are double-clustered at NAICS3 level and date since the release of a new standard has implications at industry-level, with contemporaneous effects on all firms operating in the same industry and period. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication period of the standard, as reported by the standard-setting organization.

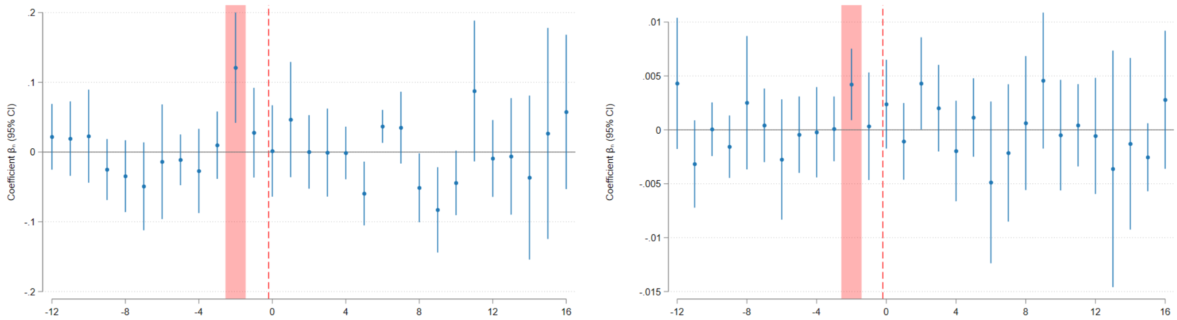
Until the (imputed) public release of the standard, the estimated coefficients are not significantly different from zero, i.e. there is no common pre-trend across firms. At $t = -2$, the estimated β is positive and significantly different from zero, which indicates that firms whose patents are closer to the standard over-perform on the stock market and exhibit unprecedented returns. This proves that, although the process of standardization is endogenous and markets may know which firms have innovated well in their past, they perceive the standardization event as good news for innovators close to the frontier only at the moment the standard becomes public.

In Figure 2b, we use the change in the 1-year EPS forecast as dependent variable. Also in this case, we do not observe any pre-trend, but we find that professional forecasters indeed updated their expectations over the future EPS precisely at the time of the public release of the standard. In words, once the information is public, firms whose portfolio of patents is closer to the standard are now expected to have a higher EPS in one year.

[15]Since the release of a new standard can affect returns and expectations of all firms in the same industry and period, we normalize both dependent variables respectively by the volatility of the NAICS3-industry portfolio and EPS forecast in that period.

Figure 2: TECHNOLOGICAL PROXIMITY AND FINANCIAL MARKETS' REACTION

(a) ar^{NAICS3} (b) $\Delta\mathbb{E}(EPS)$



Notes: Figure 2a plots the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is the firm-level abnormal return computed through the CAPM model with market portfolio defined at the NAICS3 industry-level. Figure 2b plots the estimated coefficients when the dependent variable is the change in the 1-year EPS forecast. See Section 2.3 for more information on variables construction. In both figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.

In Section 4.5, we test these results (as all other in the following sections) through a large list of robustness checks. Moreover, in Appendix C.1 we show that these results also hold when abnormal returns are extracted with other methodologies (e.g. using the SP500 as measure of market portfolio or through the French-Fama 3-factor model). On the other hand, we do not find that professional forecasters review their EPS expectations over a longer horizon.[16] In Appendix C.2-C.4, we run a number of robustness checks and show that these results hold when considering our proximity measure at the intensive margin (which demonstrates that proximity to the new standard really matters), when clustering errors at the firm-level, and when using alternative measures of scores for the computation of the proximity measure (see Appendix C.4 for more details on these alternative measures).

The above evidence suggests that the timing of approval of the standard and its specific content represent a surprise for markets beyond what could already be inferred prior to the standard release. In fact, despite knowledge of firms' portfolio of patents, the release of a standard contains additional information that allows markets to re-evaluate firms' future performance. We thus conclude that standardization events have a non-negligible information content and trigger a meaningful economic mechanism: only through standardization can promising technologies unfold their full potential. In the absence of standardization, industry-wide adoption might be impaired, thus preventing firms close to the frontier to deploy, scale and market their innovations. Stock markets pick up these effects at the time of information release. It is important to note,

[16] This is consistent with the dynamics of sales and its persistence observed after the publication of the standard. See Section 4.3.

however, that we are not able to exclude that ex-post standardization nevertheless occurs in the data. We therefore interpret our results in the remainder of the analysis as a lower bound.

4.3 Implications for sales and market shares

In this section, we investigate whether the release of a standard indeed changes future cash-flows as expected by financial markets. In particular, we study what are the real effects of standardization on sales and market shares.

To do so, we reconsider our baseline lead-lag model of equation (2), but with the normalized value of sales as dependent variable. As from Figure 3a, after the official date of publication of the standard, firms with a portfolio of patents closer to the new technological frontier start to sell more. This increase of sales is positive and significantly different from zero (at the 95% level of significance) for five consecutive quarters. In other words, the firm that is closer to the new technological frontier generates higher cash-flows through higher sales.

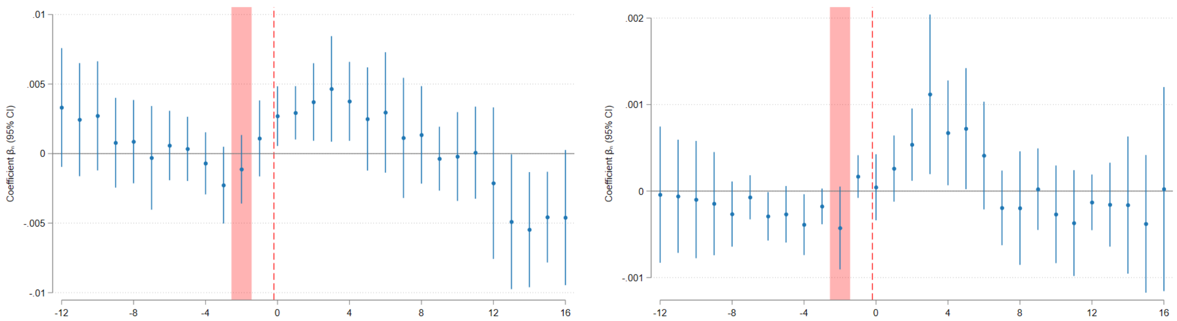
Now, it is important to understand if the increase in sales is due to an overall expansion of the market following the standard introduction (*demand effect*) or whether proximity to the new frontier also leads to gains in terms of market shares (*competition effect*). To check this, we reconsider the same model but with the firm-level market share – defined at NAICS3 level– as dependent variable. As shown in Figure 3b, firms that are closer to the frontier experience also a significant –but temporary– expansion of their market share. Standardization can therefore affect competition and market concentration for roughly one year and a half. As shown in Appendix C.2-C.4, these results hold in light of the same robustness checks that were discussed in Section 4.2 (focus on the intensive margin of the proximity measure, alternative clustering and different construction of the variable $\mathtt{Tech.Prox}$), and also when including non-listed firms in the sample (Appendix C.5).

Can we better quantify the effect of standardization? Given how we constructed the proximity measure, it is hard to interpret the estimated coefficients of Figure 3a and 3b in an economically meaningful way. For this reason, instead of the continuous variable $\mathtt{Tech.Prox}_{i,t}$, we re-estimate equation (2) with the dummy $\mathbb{I}[\mathtt{Tech.Prox}_{i,t} > 0]$ as explanatory variable. We can thus measure the (average) effect of standardization on sales and market shares (now both in logs) for frontier firms vis-à-vis firms not affected by standardization at all. By summing up the estimated β_n for the first four quarters after the standardization event, we find that frontier firms increase sales and market share respectively by 5.6% and 5.2% by the end of the first year after the publication of the standard.

To conclude, the above evidence suggests that the publication of a standard (which proxies technology adoption at the industry-level) attributes a competitive advantage

Figure 3: TECHNOLOGICAL PROXIMITY, SALES AND MARKET SHARE

(a) Sales (b) Market Share



Notes: Figure 3a and 3b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is respectively the level of sales (normalized by the mean-level of fixed assets) and the firm-level market share defined at NAICS3 industry-level. See Section 2.3 for more information on variables construction. In both figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.

to those firms with a portfolio of patents closer to the new technological frontier. This advantage translates into higher sales and higher market shares. For this reason, we claim that standardization operates in the market as a (negative) temporary shift in competition.

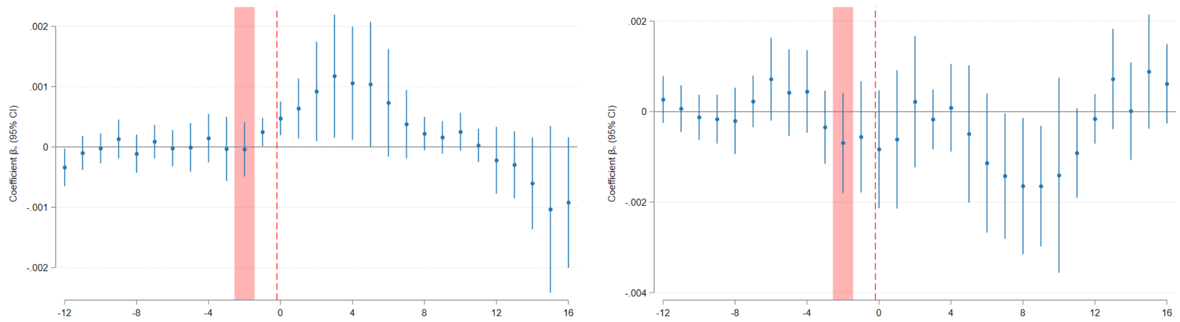
4.4 Implications for R&D expenditure

If standardization leads to higher sales and market shares, it may also affect firm-level incentives to innovate in the future. However, the incentives to do so should depend on competition. Indeed, if firms are operating in a competitive market, then the short-run advantage that standardization gives to frontier firms is very large relative to others. Consequently, given a highly competitive market structure, frontier firms will strive to keep the lead in the future by investing more into R&D. On the contrary, if the level of competition is too low, i.e. if leading firms were already far ahead of others, then the introduction of a new standard gives lower incentives to innovate as the non-competitive market structure will protect them from future competition. Consistently with the theoretical literature on Schumpeterian growth and competition (Aghion et al., 1997, 2005), we look in this section at whether we observe heterogeneous investment responses to standardization depending on the degree of competition in different sectors.

To investigate this, first we need to define competitive and non-competitive markets. We follow the work of De Loecker et al. (2020), who study markups across industries (see data description in Section 2.3). Then, we split industries in those that historically have a markup above the 75$^{\text{th}}$ percentile (non-competitive industries) and those below (competitive industries). We then use our lead-lag model to study the impact of

Figure 4: TECHNOLOGICAL PROXIMITY AND R&D

(a) R&D (Competitive Ind) (b) R&D (Non-Competitive Ind)



Notes: Figure 4a and 4b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is the 4-quarter moving average of R&D expenditure (normalized by the mean-level of fixed assets) and the sample is composed respectively by firms operating in a competitive and non-competitive industry. See Section 2.3 for more information on variables construction. In all figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.

technological proximity on R&D expenditure in competitive and non-competitive industries. If standardization decreases competition, we should find asymmetric results across the two groups of industries.

As shown in Figure 4a, firms operating in a competitive industry and closer to the technological frontier invest more in R&D following a standardization event. This effect starts already in the same quarter of the official publication of the standard and lasts one year and a half. Conversely, when considering non-competitive industries, as in Figure 4b, we find that firms significantly cut R&D expenditures starting from six quarters after the publication of the standard.[17] As shown in Appendix C.2-C.4, these results hold to the same robustness checks previously listed, and also when including non-listed firms in the sample (Appendix C.5).

All in all, these asymmetric responses corroborate the idea that the introduction of a new standard leads to a temporary (negative) shift in competition that gives a competitive advantage to frontier firms. Since their portfolio of patents better complies with the standard, they are able to expand their market share and –if the market was very competitive before standardization– they invest more in R&D in order to reinforce and protect their position from future competition.

Yet, it is important to mention that the increase in R&D is the dominating effect when we consider all firms in the sample. In order to quantify the effect of standardization on these variables, we repeat the same analysis as at the end of Section 4.3, i.e. we compare frontier firms to firms not directly affected by standardization. In this case,

[17] In Appendix C.6, we show that the same dynamics are observed when considering capital investment (CapX) as dependent variable.

we find that frontier firms increase R&D by 4.7% by the end of the first year following the publication of the standard.

4.5 Validation of the proximity measure

In this section, we want to provide evidence that the above results are not driven by confounding factors that affect our proximity measure. Our goal is to analyze the impact of standardization that can be traced back to "good" innovations and their selection for industry-wide adoption via standardization. In light of this, we want to make sure that i) our results are not driven by firms with direct influence on SSOs and the standard development process and ii) that our proximity measure is not merely capturing the extensive margin of past innovation activity but actually the quality and proximity of firms' innovations relative to the new frontier.

4.5.1 Lobbying and participation in the standard development process

Our identification relies on the fact that, conditional on all observables, firms are similar except for the nature of their patent portfolio. While our models control for a sector fixed effect and firm characteristics, there might be some unobserved features that explain why a firm's patents receive a higher score than others. Here, we address this issue from different angles.

Firms join SSOs as members for two main reasons. First, firms participate to SSOs to acquire information on the standard development process. This can guide their R&D decisions, potentially equipping them with the know-how needed to adapt to future standard releases. To the extent that knowledge about SSOs' activities is public and participation in SSOs is open to all stakeholders, a level playing field is provided such that our proximity measure correctly picks up firms' innovative capacity. However, this might not always be the case if participation represents a fixed cost that not all industry participants are able to pay (Fiedler et al., 2023). Therefore, our measure could potentially be plagued by differential access to private information produced in SSOs.[18] The second reason for participation is that firms own technologies that can be potentially included in the new standard. If their technologies are selected for their quality, our measure correctly captures the innovative capacity of these firms. On the other hand, it would be biased if the same firms exercise lobbying and undue influence in order to include their patents into the new standard– although the role of SSOs is to

[18]If firms are *ex ante* excluded from observing and possibly contributing to the standard development process, our measure would also pick up the lack of a level playing field across firms. As we only analyze publicly listed firms, we are confident that this issue is only a minor one, and can in any case be addressed by firm fixed effects as well as by controlling for SSO membership. Also note that many small firms participate in SSOs (Waguespack and Fleming, 2009; Gupta, 2017).

ensure that the new standard reflects the current consensus in the industry and not the influence of a few players.[19]

While it is impossible to distinguish standards that were subject to undue influence versus those that effectively chose the best available technology, we are able to show that our results hold when controlling for SSO membership using data from Baron and Pohlmann (2018). This database, starting in 1996, reports the name and year of membership of firms when belonging to a SSO. Then, we match this list of firms to Compustat using their names and years of membership and keep only observations from 1996 onwards. We end up with a sample in which 29% of firms were members of a SSO for at least one year,[20] and therefore potentially able to influence the standardization process in their favour around that period. To check whether this is the case, we include a membership dummy in the lead-lag operator of equation (2). This dummy takes value one if the firm is a SSO member in the same year of the standard release or the previous two.[21] This allows us to check if the results of Sections 4.2–4.4 hold or are due to membership around the standardization event. As shown in Appendix D.1, the results are very similar when this additional control is included. In addition, the sum of lagged coefficients of the membership dummy are not significantly different from zero.

To further corroborate this point, we tackle the problem of lobbying in an alternative way. SSOs can be organized on the national level (i.e. US-based such as ANSI, the American National Standards Institute) or on the supranational or international level (i.e. European SSOs such as CENELEC, the European Committee for Electrotechnical Standardization, or ISO, the International Organisation for Standardization). We suppose that US firms have more lobbying power within American SSOs, whereas less so within international ones. In light of this, we go back to our patent-level data and re-build the variable $\mathtt{Tech.Prox}_{i,t}$ considering scores obtained from matching patents only to standards issued by international SSOs (which represent 85% of all standards in our data). As shown in Appendix D.2, the results from the previous sections still hold: it is the release of a standard from an international SSO –where US firms have smaller lobbying influence– and the firm-proximity to this standard that explains the results.[22]

Second, besides network and lobbying activity, some firms might always be more in-

[19]Spulber (2019) shows in a theoretical model that the voting process in SSOs assures that standards are defined efficiently, as a sufficiently large number of industry participants share its economic benefits. This outweighs the detrimental impact of conveying too much market power to firms that might profit from the chosen standard.

[20]The average membership duration in the data is 2.5 years.

[21]See Appendix D.1 for details.

[22]On the other hand, when considering only standards issued by US SSOs to build the firm-level proximity measure, the effects of standardization on each dependent variable is null or a pre-trend can be detected.

novative and successful than others due to some unobserved characteristics such as the quality of their management or their innovation culture. Although these (supposedly) time-invariant characteristics are captured by the firm-level fixed effect in equation (2), we show in Appendix D.3 that the results also hold when removing the top 25% of most innovative firms (i.e. those firms that always have larger and more frequent values of $\mathtt{Tech.Prox}$).

Third, in Appendix D.4 we show that –once controlling for firm fixed effects– firms receiving a positive value for the proximity measure ($\mathbb{I}[\mathtt{Tech.Prox} > 0]$) do not significantly differ ex-ante[23] from others in several dimensions such as: q-value of investment, leverage, market capitalization, return on equity (ROE), price/earning ratio, cost of capital, size, age. More interestingly, the number of newly issued patents does not explain the proximity measure (see column 9 of Table D.1). This corroborates the idea that patenting activity alone, or the strategic issuance of patents just before the standard publication, does not guarantee that firms receive a positive value for $\mathtt{Tech.Prox}$.[24]

4.5.2 Measurement issues

By definition, only patenting firms can have positive values for $\mathtt{Tech.Prox}_{i,t}$. Yet, it can be that firms with a large stock of patents are more likely to receive a higher value than others. At the same time, since there is substantial heterogeneity in patenting activity across industries and technological classes, it can be that our measure actually captures only differences in innovation and patenting intensity across technological fields. In this section, we run some robustness analyses to show that our proximity measure is not driven by the specific technological field in which the firm is operating, nor is merely capturing the extensive margin of past innovation activity.

Our first exercise consists in constructing groups of firms that are similar in terms of their patent portfolio. To do so, we construct a network of firms based on the co-occurrence of the IPC classes of their patents. We then use a k-mean clustering algorithm to construct 100 groups of firms. Within these groups, firms are therefore similar in terms of the technological classes of their patents. We then augment equation (2) by adding a dummy for each of these groups interacted with time fixed effects. This controls for technology-specific dynamics that could be correlated with our measure of standardization. If our results hold, then this would corroborate that even within firms that patent in similar technologies, those that are closer to a new standard increase their sales, market share and R&D. Results and details are presented in Appendix D.6 and

[23]Four quarters before the release of a new standard.

[24]In addition, we checked that, controlling for the number of patents, there is no correlation between the lagged market share and the magnitude of $\mathtt{Tech.Prox}$. In other words, market power does not explain why a firm receives a positive value for $\mathtt{Tech.Prox}$.

show that our baseline results are indeed robust to this test. Note that this also ensures that our matching procedure captures more than just the nature of the technology and is able to differentiate between patents of similar technologies.

In another robustness test, we include the number of patents filed during one year prior to the standardization event in the lead-lag operator of equation (2).[25] This additional control will capture changes in the dependent variable that could result from a more intensive innovation activity in periods just before the release of the standard (for example through strategic patenting because they exploit private information obtained through their participation in SSOs). Indeed, if firms file more patents, they are more likely to receive a positive value for $\mathtt{Tech.Prox}$ all other things equal. As discussed in Appendix D.5, our results are robust to the addition of such a control,[26] i.e. they are are not only capturing the quantity dimension of a firm's innovation activity.

Finally, in Appendix D.4 we show also that, firms with a positive proximity value ($\mathbb{I}[\mathtt{Tech.Prox}_{i,t} > 0]$) do not significantly differ ex-ante from others also in terms of the (cumulative) number of patents filed until one year before the standardization event (see column 10 of Table D.1). This corroborates the idea that our standardization measure –which is constructed by matching each standard to the portfolio of patents held by the firm– does not merely reflect the process of patent accumulation along the life of the firm. This is important as it is suggests that our measure is not a by-product of long-term patenting activity.

All in all, this further evidence confirms that our proximity measure is meaningful in terms of measuring the quality of a firm's patent portfolio with respect to the newly chosen frontier at the time the new standard is released.

5 Aggregate effects and implications

The aggregate impact of standardization can be analyzed by separately considering its short-term and long-term effects. As discussed in the previous section, standardization quickly provides an advantage to firms that are close to the frontier. However, the primary objective of standardization is to foster the diffusion and large-scale adoption of industry best practices. Over the long term, this can have a positive aggregate effect by mitigating information frictions and facilitating technology transfer.

In this section, we test this mechanism of technology diffusion for growth. Our analysis

[25] Recall that the variable $\mathtt{Tech.Prox}_{i,t}$ is built by aggregating individual scores for all patents filed up to $t-4$, where t is the quarter in which the standard is published. Therefore, in the robustness check described in this section, we control for the number of new patents issued between $t-8$ and $t-4$. See Appendix D.5 for details.

[26] Results also hold when controlling for the total number of patents accumulated up to one year before the standardization event.

is motivated, among others, by the findings of Bloom et al. (2013): R&D efforts yield benefits for the innovating firms, but in the long run, technology spillovers to other firms become the dominant effect. Indeed, as demonstrated in Rysman and Simcoe (2008), the goal of standardization is to enhance and incentivize this diffusion process by creating a shared knowledge base for industry participants to build upon. Therefore, standardization could be a pivotal driver of economic growth, as knowledge diffusion not only allows for catch-up but also encourages new innovation, as suggested by Hegde et al. (2022) and Furman et al. (2021).

In light of this, the aggregate effect of standardization should be viewed as a combination of a short-term effect of increasing the advantage of leaders, and a long-term effect where other firms benefit from technological spillovers and increase their research effort. Which one of these two competing effects on aggregate growth dominates?

To answer this question, we first study if standardization leads to higher growth at industry-level. In particular, we study how much of the change in growth due to standardization can be explained by leaders and by the rest of the industry (the followers). Second, we analyse the dynamics of these followers in response to standardization in terms of innovation activity and economic performance. Third, we provide evidence of knowledge spillovers to explain the catching-up process of followers.

Sectoral growth through standards. Our empirical analysis of Section 4 shows that when a firm is close to the frontier ($\mathtt{Tech.Prox}_{i,t} > 0$), then it has an immediate competitive advantage compared to its competitor in the same market that translates into higher sales and market share. This positive effect implies that these firms can be, at least temporarily, seen as the leader in their industry as they are now closer to the newly defined technological frontier. In light of this, we split the Compustat sample of firms used in the previous sections into two groups. For every industry and quarter, we define as *leaders* those firms with $\mathtt{Tech.Prox}_{i,t} > 0$, and *followers* all the others. Then, we aggregate and build an industry-level panel dataset where sectoral sales and their growth rate are decomposed between leaders and followers. As shown in Table 4, the average industry grows by 1.64% per quarter. With a rate of 1.04% (0.60%), leaders (followers) explain 63% (37%) of sectoral growth.

Given these figures, we now study the (cumulative) effect of standardization on sectoral growth, and by how much the change in growth is explained by leaders and followers. For this, we consider again model (2), but now defined for our industry-level panel dataset.[27] We estimate this model with the dependent variable being the indus-

[27]Since we are now dealing with a panel where the dependent variables and covariates are defined at the industry-level, we drop the interaction between industry and time fixed effects from model (2) as this would capture all the within-industry variation over time. The set of control variables remains the same as in the firm-level exercise, but they are here aggregated at NAICS3 level. Appendix E explains

Table 4: AGGREGATE EFFECTS ON GROWTH

	Industry	Leaders	Followers
Mean Sectoral Growth Rate (%)	1.64	1.04	0.60
(1yr-Cumulative) Change in Growth due to Mean Sectoral Shock (pp)	-0.03 (0.06)	0.08 (0.02)	-0.11 (0.07)
(4yr-Cumulative) Change in Growth due to Mean Sectoral Shock (pp)	0.11 (0.05)	0.02 (0.02)	0.09 (0.04)

Notes: The first line of this table shows the average sectoral growth and its decomposition between leaders and followers. The second and third line show the cumulative effect of the introduction of a standard on sectoral growth respectively one and four year after the official publication of the standard. Standard errors are in parenthesis. See Appendix E for details on data and estimation.

try growth rate as well as the growth rate of leaders and followers. The explanatory variable is the average value of our proximity measure for leaders for each quarter and industry. This captures to which extent the average leader in the industry can adapt to the new technology frontier at the moment of the release of the standard. We estimate the model and sum the coefficients over the first year and first-to-forth year after the publication of the standard (see Figure E.1 in Appendix E). This allows us to quantify the short- and long-run effect of standardization on sectoral growth along with the contribution of followers and leaders.

As reported in the second line of Table 4, in the first year after the introduction of the standard, sectoral growth is not significantly different from zero. Yet, when looking at the decomposition, we find that the growth rate of leaders increases significantly more in industries where leaders are already very close to the new technology frontier. The percentage increase of leaders' growth is 0.08pp for the average value of our (mean) proximity measure. This effect is counterbalanced by the negative growth rate of followers. In fact, since by definition followers are far away from the frontier, the more leaders in the same industry are mastering the new technology the less followers grow in the short-run. For the average of our (mean) proximity measure on leaders, followers' growth rate diminishes (although not significantly) by 0.11pp. Over the four years following the introduction of the standard, the contribution among leaders and followers reverses. In fact, in the long-run, the industry starts growing. The more leaders were near the technological frontier at the moment of standard release, the more sectoral sales increase (by 0.11pp for the average value of $\mathtt{Tech.Prox}$ after four years). This result is mostly explained by followers –for which the growth rate increases (significantly) by 0.09pp– and not by leaders whose contribution is small and insignificant.

in detail the construction of the industry-level data and the empirical model used in this section.

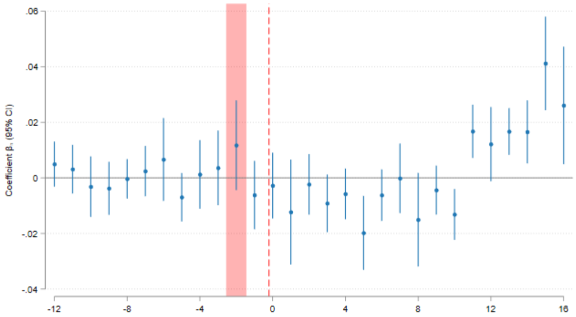
Catching-up effects. In line with the evidence from Section 4.3, these results corroborate the idea that the gains for leaders are only temporary. On the other hand, it seems that it is followers that drive sectoral growth in the long-run. If the catching-up motive is in action, we should observe a bigger increase in followers' sales, R&D investment in sectors in which the distance from the frontier of the (average) leader and follower is larger, i.e. in industries where the introduction of a new standard can potentially generate stronger spillovers. To check this, we use our industry-level panel and relate leaders to followers by constructing the following industry-level variables: (i) the industry-level market share of followers, (ii) the followers' share of total R&D expenditure in the industry, (iii) the followers' share of total patents issued in the industry. Then, we use (i)-to-(iii) as dependent variables in model (2). Hence, by doing so, it is possible to understand which group drives sales and innovation in the industry, by how much and when. Figure 5a shows that, in industries where the average leader is closer to the frontier and the average distance between leaders and followers is larger, followers' aggregate market shares slightly decline in the first 2.5 years after the introduction of the standard. However, from the end of the third year, this dynamic is reverted as the market share of followers increases persistently for the remaining periods.

Figure 5b shows that –within the first two years after the introduction of the standard– aggregate R&D activity is explained by leaders in the industry, but this effect is also reverted thereafter. In fact, in the long-run, followers increase their R&D expenditure relatively more and more persistently. This pattern is confirmed by Figure 5c: if sectoral research output is mostly explained by leaders in the short-run, it is followers that drive patenting activity in the economy in the long-run. Also in this case, the long-run effect is stronger and it lasts longer.[28]

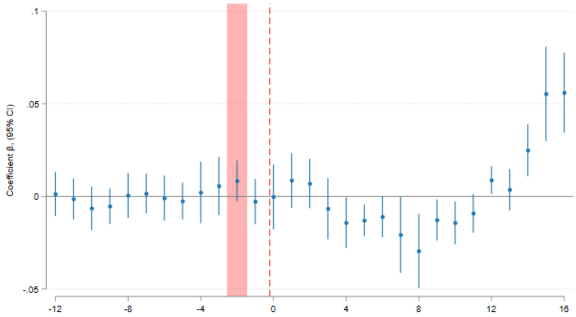
[28]Using levels as dependent variable does not substantially change the results. We therefore conclude that the results are not driven by composition effects: the increase in followers' shares is not an artefact of leaders' dynamics.

Figure 5: Aggregate Effects of Standardization for Followers

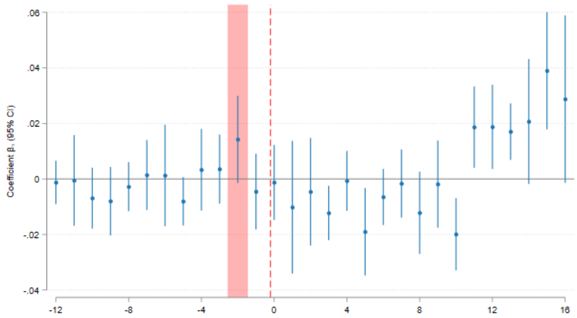
(a) Market Share (Followers)



(b) Share of R&D (Followers)



(c) Share of New Patents (Followers)



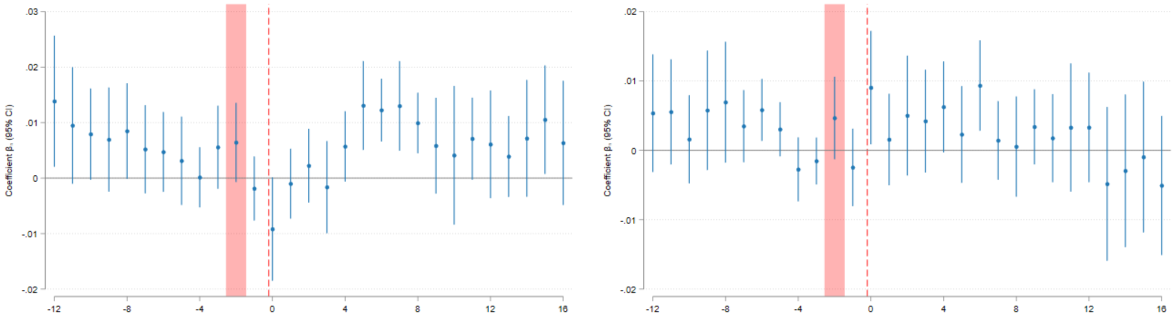
Notes: Figures 5a-5c plot the estimated coefficients when the dependent variable is respectively: the market share of followers in the NAICS3 industry, the followers' share of total R&D expenditure in the NAICS3 industry, the followers' share of total patents issued in the NAICS3 industry. See Appendix E for more details on data construction and estimation. In all figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.

Knowledge spillovers. As final evidence that the catching-up process is indeed driven by spillovers, we study the evolution of citations of the patents of a firm received from other firms. To do so, we use the network of citations from our patent data and the patent-to-firm crosswalk from Kogan et al. (2017), such that we can pin down the number of (contemporaneous) citations a patent receives from either a laggard or a leader. Then, we aggregate the number of citations at the firm-level and re-run model (2) with two different dependent variables: the logarithm of the number of citations received from followers and (ii) from leaders. Figure 6a shows that firms whose portfolio of patents match well the new standards are cited more by laggard firms. In particular, followers start citing the best patents roughly one year after the introduction of the standard and for five consecutive periods. On the other hand, as shown in Figure 6b, also leaders cite more the patents that matched the standard the most. Yet, such effect is not persistent.

Figure 6: PATENT CITATIONS

(a) Citations from Followers

(b) Citations from Leaders



Notes: Figure 6a plots the estimated coefficients of equation (2) when the dependent variable is the log of the number of citations of the patents of a firm received from laggard firms. Figure 6b when the dependent variable is the log of the number of citations of the patents of a firm received from laggard firms. In both figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.

6 Conclusion

This paper studies how standardization –i.e. the selection and adoption of a new technology at the industry-level– affects competition, innovation and growth at the firm and sectoral level. The contribution of the paper is threefold.

First, we use semantic algorithms to match the content of patents and standards. This methodology allows to measure the proximity of each patent to the new technological frontier imposed by the standard, and –therefore– the capacity of firms to market their products in line with the new standard. We show that the information retrieved from the semantic matching is meaningful as patents closer to the content of a standard are associated with greater economic, scientific and private value.

Second, we cross this novel measure with firm-level data to study (i) to which extent standardization is distinct from just observing patent quality, and (ii) how firm dynamics change depending on the proximity of the firm's portfolio of patents to the new standard.

We address these questions through a dispersed lead-lag model, which captures the entire response following the release of a new standard. Under this strategy, we show that financial markets do not anticipate the timing and content of a standard. In fact, markets react only at the very moment that information on the new standard becomes public. Thereafter, we show that firms closer to the new standard gain temporarily in terms of sales and market shares once the standard is published. This suggests that standardization affects market structure since it gives a temporary competitive advantage to those firms that have the technology and knowledge to immediately adjust to the standard specifications. In addition, we also observe heterogeneous reactions across markets. In markets with high levels of competition, firms closer to the new technological frontier do more R&D after the release of the standard.

In the final part of the paper, we investigate the aggregate implications of standardization at the industry-level. We find that sectors in which the potential for knowledge spillovers is higher exhibit higher growth in the long-run. This is only partially explained by the gains of the leaders in the industry. Actually, sectoral long-term growth is mostly explained by followers. In fact, in those industries, followers invest more in R&D and their research output is higher. This allows them to catch up and the industry to grow more in the long-run.

In light of these results, this paper not only sheds light on the effect of standardization on competition and innovation, but it has a clear policy implication as it proves that, under a competitive market structure, standardization rewards frontier firms while stimulating further investment by followers and –ultimately– economic growth.

References

Abbas, Assad, Limin Zhang, and Samee U. Khan, "A literature review on the state-of-the-art in patent analysis," *World Patent Information*, 2014, *37*, 3–13.

Adams, Stephen, "The text, the full text and nothing but the text: Part 1–Standards for creating textual information in patent documents and general search implications," *World Patent Information*, 2010, *32* (1), 22–29.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–351.

_ and Rachel Griffith, *Competition and growth: reconciling theory and evidence*, MIT press, 2005.

_ , Antonin Bergeaud, Matthieu Lequien, and Marc J Melitz, "The heterogeneous impact of market size on innovation: evidence from French firm-level exports," Technical Report w24600, National Bureau of Economic Research 2018.

_ , _ , _ , Marc Melitz, and Thomas Zuber, "Opposing firm-level responses to the China shock: horizontal competition versus vertical relationships?," Technical Report w29196, National Bureau of Economic Research 2021.

_ , Christopher Harris, and John Vickers, "Competition and growth with step-by-step innovation: An example," *European Economic Review*, 1997, *41* (3-5), 771–782.

_ , Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: an Inverted-U Relationship," *The Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Akcigit, Ufuk, Sina T Ates, and Giammario Impullitti, "Innovation and trade policy in a globalized world," Technical Report w24543, National Bureau of Economic Research 2018.

Argente, David, Salomé Baslandze, Douglas Hanley, and Sara Moreira, "Patents to products: Product innovation and firm dynamics," FRB Atlanta Working Paper 2020-4, Federal Reserve Bank of Atlanta 2020.

Arts, Sam, Bruno Cassiman, and Juan Carlos Gomez, "Text matching to measure patent similarity," *Strategic Management Journal*, 2018, *39* (1), 62–84.

Austin, David H, "An event-study approach to measuring innovative output: The case of biotechnology," *The American economic review*, 1993, *83* (2), 253–258.

Autor, David, David Dorn, Gordon Hanson, Gary Pisano, and Pian Shu, "Foreign competition and domestic innovation: Evidence from US patents," *American Economic Review: Insights*, 2020, *2* (3), 357–74.

Baron, J. and Julia Schmidt, "Technological Standardization, Endogenous Productivity and Transitory Dynamics," Working papers 503, Banque de France 2014.

Baron, Justus and Daniel F. Spulber, "Technology Standards and Standard Setting Organizations: Introduction to the Searle Center Database," *Journal of Economics & Management Strategy*, 2018, *27* (3), 462–503.

— **and Tim Christoph Pohlmann**, "Mapping Standards to Patents Using Declarations of Standard-Essential Patents," *Journal of Economics and Management Strategy*, 2018, *forthcoming.*

Basu, Susanto and John G. Fernald, "Information and Communications Technology as a General Purpose Technology: Evidence from U.S. Industry Data," *Federal Reserve Bank of San Francisco Economic Review*, 2008, pp. 1–15.

Bekkers, Rudi, Christian Catalini, Arianna Martinelli, Cesare Righi, and Timothy Simcoe, "Disclosure Rules and Declared Essential Patents," Working Paper 23627, National Bureau of Economic Research July 2017.

— **, René Bongard, and Alessandro Nuvolari**, "An empirical study on the determinants of essential patent claims in compatibility standards," *Research Policy*, 2011, *40* (7), 1001–1015. Issue dedicated to Chris Freeman, Founding Editor.

Bergeaud, Antonin and Cyril Verluise, "The rise of China's technological power: the perspective from frontier technologies," Discussion Paper 1876, Centre for Economic Performance 2022.

— **, Julia Schmidt, and Riccardo Zago**, "The Implications of Technological Standardization for the Value of the Firm," *mimeo*, 2022.

— **, Yoann Potiron, and Juste Raimbault**, "Classifying patents based on their semantic content," *PloS one*, 2017, *12* (4), e0176310.

Bernhofen, Daniel M., Zouheir El-Sahli, and Richard Kneller, "Estimating the effects of the container revolution on world trade," *Journal of International Economics*, 2016, *98* (C), 36–50.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

— **, Mirko Draca, and John Van Reenen**, "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *The Review of Economic Studies*, 2016, *83* (1), 87–117.

— **, Tarek Alexander Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun**, "The Diffusion of Disruptive Technologies," Working Paper 28999, National Bureau of Economic Research July 2021.

Brachtendorf, Lorenz, Fabian Gaessler, and Dietmar Harhoff, "Approximating the standard essentiality of patents – A semantics-based analysis," Technical Report, Report for the EPO Academic Research Programme 2020.

— **, — , and —** , "Truly Standard-Essential Patents? A Semantics-Based Analysis," *Journal of Economics & Management Strategy*, 2022, *forthcoming*, 1–26.

Brown, James R, Steven M Fazzari, and Bruce C Petersen, "Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom," *The Journal of Finance*, 2009, *64* (1), 151–185.

Chan, Su Han, John D Martin, and John W Kensinger, "Corporate research and development expenditures and share value," *Journal of Financial Economics*, 1990, *26* (2), 255–276.

Chavalarias, David and Jean-Philippe Cointet, "Phylomemetic patterns in science evolution—the rise and fall of scientific fields," *Plos One*, 2013, *8* (2), e54847.

Choi, Jay Pil and Heiko Gerlach, "A Theory of Patent Portfolios," *American Economic Journal: Microeconomics*, February 2017, *9* (1), 315–51.

Cox, D. R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1972, *34* (2), 187–202.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 1998, *53* (6), 1839–1885.

David, Paul A, "Clio and the Economics of QWERTY," *American Economic Review*, May 1985, *75* (2), 332–337.

Dechezleprêtre, Antoine, David Hémous, Morten Olsen, and Carlo Zanella, "Induced automation: evidence from firm-level patent data," Technical Report 384, University of Zurich, Department of Economics 2021.

Eberhart, Allan C, William F Maxwell, and Akhtar R Siddique, "An examination of long-term abnormal stock returns and operating performance following R&D increases," *The Journal of Finance*, 2004, *59* (2), 623–650.

Farrell, Joseph and Garth Saloner, "Standardization, Compatibility, and Innovation," *RAND Journal of Economics*, Spring 1985, *16* (1), 70–83.

__ **and** __ , "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *American Economic Review*, December 1986, *76* (5), 940–955.

__ **and Paul Klemperer**, "Coordination and Lock-In: Competition with Switching Costs and Network Effects," in Mark Armstrong and Robert Porter, eds., *Handbook of Industrial Organization*, Vol. 3, Elsevier, 2007, chapter 31, pp. 1967–2072.

__ **and Timothy Simcoe**, "Choosing the rules for consensus standardization," *The RAND Journal of Economics*, 2012, *43* (2), 235–252.

Fiedler, Clemens, Maria Larrain, and Jens PrÃŒfer, "Membership, governance, and lobbying in standard-setting organizations," *Research Policy*, 2023, *52* (6), 104761.

Furman, Jeffrey L., Markus Nagler, and Martin Watzinger, "Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program," *American Economic Journal: Economic Policy*, November 2021, *13* (4), 239–70.

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, December 1990, *28* (4), 1661–1707.

Gross, Daniel P., "Collusive Investments in Technological Compatibility: Lessons from U.S. Railroads in the Late 19th Century," *Management Science*, 2020, *66* (12), 5683–5700.

Grossman, Gene M. and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 1991, *58* (1), 43.

Gupta, Kirti, "How SSOs Work: Unpacking the Mobile Industry's 3GPP Standards," 2017.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, Spring 2005, *36* (1), 16–38.

Hall, Bronwyn H and Rosemarie Ham Ziedonis, "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *RAND Journal of Economics*, Spring 2001, *32* (1), 101–128.

Hausman, Jerry A, Bronwyn H Hall, and Zvi Griliches, "Econometric models for count data with an application to the patents-R&D relationship," 1984.

Hegde, Deepak, Kyle F Herkenhoff, and Chenqi Zhu, "Patent Publication and Innovation," Working Paper 29770, National Bureau of Economic Research February 2022.

Iacovone, Leonardo, Wolfgang Keller, and Ferdinand Rauch, "Innovation Responses to Import Competition," 2011. Unpublished manuscript.

Jensen, Michael C, Fischer Black, and Myron S Scholes, "The capital asset pricing model: Some empirical tests," in "Studies in the Theory of Capital Markets," Praeger Publishers Inc, 1972.

Jovanovic, Boyan and Peter L. Rousseau, "General Purpose Technologies," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, pp. 1181–1224.

Kang, Byeongwoo and Rudi Bekkers, "Just-in-time patents and the development of standards," *Research Policy*, 2015, *44* (10), 1948–1961.

Katz, Michael L. and Carl Shapiro, "Network Externalities, Competition, and Compatibility," *American Economic Review*, June 1985, *75* (3), 424–440.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, September 2021, *3* (3), 303–20.

Kindleberger, Charles P., "Standards as Public, Collective and Private Goods," *Kyklos*, 1983, *36* (3), 377–396.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Kuhn, Jeffrey, Kenneth Younge, and Alan Marco, "Patent citations reexamined," *The RAND Journal of Economics*, 2020, *51* (1), 109–132.

Leland, Hayne E, "Quacks, lemons, and licensing: A theory of minimum quality standards," *Journal of Political Economy*, 1979, *87* (6), 1328–1346.

Lerner, Josh and Jean Tirole, "Standard-Essential Patents," *Journal of Political Economy*, 2015, *123* (3), 547–586.

Llanes, Gastón and Joaquín Poblete, "Ex ante agreements in standard setting and patent-pool formation," *Journal of Economics & Management Strategy*, 2014, *23* (1), 50–67.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Ma, Song, "Technological Obsolescence," Working Paper 29504, National Bureau of Economic Research November 2021.

Martinez, Cataline, "Insight into Different Types of Patent Families," Working Paper 2010/02, OECD Science, Technology and Industry 2010.

Mitchell, Mark L and Erik Stafford, "Managerial decisions and long-term stock price performance," *The Journal of Business*, 2000, *73* (3), 287–329.

Nicholas, Tom, "Does innovation cause stock market runups? Evidence from the great crash," *American Economic Review*, 2008, *98* (4), 1370–96.

Pakes, Ariel, "On patents, R & D, and the stock market rate of return," *Journal of political economy*, 1985, *93* (2), 390–409.

— **and Mark Schankerman**, "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, University of Chicago Press, 1984, pp. 73–88.

Pakes, Ariel S, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, July 1986, *54* (4), 755–784.

Romer, Paul M, "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), 71–102.

Rysman, Marc and Timothy Simcoe, "Patents and the Performance of Voluntary Standard-Setting Organizations," *Management Science*, November 2008, *54* (11), 1920–1934.

Schankerman, Mark, "How Valuable is Patent Protection? Estimates by Technology Field," *The RAND Journal of Economics*, 1998, *29* (1), 77–107.

Schmalensee, Richard, "Standard-setting, innovation specialists and competition policy," *The Journal of Industrial Economics*, 2009, *57* (3), 526–552.

Schmidt, Julia and Walter Steingress, "No double standards: Quantifying the impact of standard harmonization on trade," *Journal of International Economics*, 2022, *137*, 103619.

Spulber, Daniel F, "Standard setting organisations and standard essential patents: Voting and markets," *The Economic Journal*, 2019, *129* (619), 1477–1509.

Szewczyk, Samuel H, George P Tsetsekos, and Zaher Zantout, "The valuation of corporate R&D expenditures: Evidence from investment opportunities and free cash flow," *Financial Management*, 1996, pp. 105–110.

Waguespack, David M. and Lee Fleming, "Scanning the Commons? Evidence on the Benefits to Startups Participating in Open Standards Development," *Management Science*, 2009, *55* (2), 210–223.

Webb, Michael, Nick Short, Nicholas Bloom, and Josh Lerner, "Some facts of high-tech patenting," Working Paper w24793, National Bureau of Economic Research 2018.

Online Appendix

A Data

A.1 Standards data

Variables used. We rely on the following information from a `Perinorm` dataset, which is part of the Searle Centre Database on Technology Standards and Standard Setting Organizations (see Baron and Spulber, 2018). In particular, we use the following information:

- *Identifier*: Each standard document is registered with a unique identifier from `Perinorm`.
- *Publication date*: The date of the release (publication) of the standard by the respective SSO.
- *Equivalences*: A standard can be released by several SSOs. Indeed, the internationalization of the standard-setting process where the bulk of standards originates in supranational SSOs such as European SSOs (ETSI, CEN, CENELEC) or international SSOs (ISO, ITU, IEC) results in the co-existence of equivalent standards in `Perinorm`. A standard developed by an international SSO is often accredited by national SSOs to include it in the national standard catalogue. Similarly, accreditations by several SSOs in the same country can be observed, often due to the standard being developed jointly by two or more SSOs. Two standards can be considered equivalent if their content are the same, but they often differ with respect to the release date and the language used in the standard document.
- *Version history*: Standards are constantly updated and several versions can succeed or supersede a previous version. In the latter case, a subsequent standard explicitly replaces a former version whereas the former case implies just a simple update. SSO-specific norms determine the details. Given some of the technical complexities, it is also possible that several standards share a common previous version because standard projects are split into different directions.
- *ICS classification*: The International Classification of Standards is a classification system maintained by the International Organization for Standardization, aimed at covering all possible technical or economic sectors that standards are govering. The ICS classes are composed of three levels, the first one (two digits) designating a general field such as 49 – Aircraft and space vehicle engineering, followed by a second level (three digits) such as 49.030 – Fasteners for aerospace construction, and sometimes a third level (two digits) such as 49.030.10 – Screw threads.
- *Keywords*: `Perinorm` is a bibliographical database, which allows subscribers to search for a standard and to purchase the standard document. To facilitate the search, keywords have been assigned to each standard document. These comprise both 1-grams such as “automation” or 3-grams such as “internal combustion engine”.

Cleaning. We clean the standards data, in particular with respect to the publication dates, the equivalences, the version history, ICS classification as well as the keywords.

For some publication dates, the month or the day of the date are missing in which case we assume December for the month and 28 for the day, thus implicitly favoring standards for which the date information is complete.

For some of the equivalences, there is additional information on whether a standard is identical/equivalent or not equivalent. As we want to regroup only those standards that are identical, we correct the list of equivalences and exclude non-equivalent standards. Du to misreporting or chronological reporting, a single standard observation does not necessarily reveal all equivalences. In the case of chronological reporting, only equivalences known at the time of the release are listed and subsequent equivalences are only reported for newly released standards. The identification of equivalent standards is implemented with the algorithm described below.

We take the list of standard identifiers that constitute the version history of each standard document and identify prior versions by comparing the publication dates of these identifiers with the standard document in question. If there is at least one standard with prior publication date in the version history, the standard is not considered a first version.

ICS classifications can be erroneous and are cleaned to only include official codes, respecting the format designed by the ICS.

Keywords are cleaned and processed as described in Appendix B below.

Identifying equivalences. We use graph theory to identify all standards that belong to one group by assigning them the same group identifier. In particular, we use the following breadth-first search algorithm (which we specifically adapt to the structure of the dataset) to connect all standards by exploring their equivalences:

1. Initialize the group identifier, equal to a standard's row number in the dataset, for each standard.
2. Starting with $n = 1$, store the group identifier of standard n in the database (i.e. A).
3. Add the group identifiers of the equivalent standards, i.e. B, to the vector of stored group identifiers.
4. Note the smallest element of the vector of stored group identifiers.
5. Modify the group identifiers of standard n and its equivalent standards by assigning them the value identified in step 4 (i.e. A and B will have the same group identifier).
6. Delete the stored group identifiers.
7. Go on to the next standard $n + 1$ and repeat from step 2 onwards.

In order to minimize the computing power needed to run the algorithm, we use a simple hash function to build a dictionary of all standards whose IDs, which are strings, are mapped one-to-one to numeric values.

Relevant subset and grouping of keywords. For each group of standards (defined as regrouping all equivalent standard documents), we exclude within-country duplicate standard releases, only keeping the earliest standard release. We then restrict the sample to first versions only. All ICS and keywords are aggregated on the level of the group identifier. Only unique keywords are kept to avoid double counting due to the fact that a group includes a large number of individual, equivalent standard documents.

B Matching

B.1 Matching procedure

B.1.1 Brief outline of the matching procedure

Our goal is to find the patents that are the "closest" to a given standard. Our approach relies on the set of keywords associated with a standard, which we take to be a sufficient information set to describe the standard, and on the abstract of patents. More specifically, for each standard, we scan our patent database and give a score for each patent that reflects how relevant these standard's keywords are to describe the patent's abstract. One of the main challenge with this type of large scale data mining approach is to design a method that is suitable for big data (there are around 0.8m standards and 1.9m patents in our dataset). We briefly present our approach below.

The standard database includes, among others, a standard identifier, the title, a release date and a number of keywords that were manually provided by `Perinorm` staff when incorporating a standard into the database. For example, the Austrian standard AT98957039 with the title "OENORM Aerospace series - Nickel base alloy NI-B15701 (NiPd34Au30) - Filler metal for brazing - Wire" is included in the database with the following keyword information:

```
standard id   date        ICS          keywords
AT98957039    01/07/1997  49.025.15    Aerospace transport⋆Air
                                       transport⋆Brazing
                                       alloys⋆Nickel base
                                       alloys⋆Space transport⋆Wires
```

We process these keywords as follows.

1. **Stemming and cleaning keywords**: this first step consists in "normalizing" the set of keywords contained in each standard by removing upper-case letter, punctuation and "stop-words" (*the*, *at*, *from* etc...). We then keep only the stem of each word.[29]

2. **Constructing k-grams**: the second step consists in associating successive stems into one unique semantic unit. These "multi-stems", or *k-grams* are constructed

[29] Families of words are generally derived from a unique root called stem (for example *compute*, *computer*, *computation* all share the same stem *comput*).

as groups of size k, with $k \leqslant 3$. The rationale from considering group of words can be illustrated with the example of a standard containing "air conditionning" as one of its keywords. If we do not consider k-grams in addition to single stems, then we would be screening the patent database for the stems *air* and *condition*, which are clearly irrelevant in that case. Thus, at the end of this procedure, we can associate for each standard j a set $\mathcal{A}(j)$ of 1-grams, 2-grams and 3-grams taken from its keywords.[30]

3. **Computing Inverse Document Frequency**: we then associate for each k-grams $l \in \bigcup_{j \in \mathcal{J}} \mathcal{A}(j)$ a quantity that seeks to measure how frequent this k-gram is. This is known as the inverse document frequency and is defined as follow:

$$\text{IDF}(l) \equiv \log \left(\frac{1 + |\mathcal{J}|}{1 + \sum_{j \in \mathcal{J}} \mathbb{1}(l \in \mathcal{A}(j))} \right)$$

   Where $\mathbb{1}(X)$ is equal to 1 if X is true and $|\mathcal{J}|$ is the cardinal of $\mathcal{J}$ (the number of standards). In other words, $\text{IDF}(l)$ is calculated from the inverse of the share of standards that contains k-gram l.

4. **Removing uninformative k-grams**: from the set of k-grams l and their associated IDF, we further restrict the sample by removing k-grams whose IDF is below a given threshold T. The choice of such a threshold will be discussed below and results from a trade-off between efficiency and exhaustiveness (see Chavalarias and Cointet, 2013 and Bergeaud et al., 2017 for a discussion).

Whereas we have keywords already provided in the standards database, this is not the case for the patents where we rely on their abstracts to extract keywords as described further below. The EPO patent EP0717749A4 with the title "Self-addressable self-assembling microelectronic systems and devices for molecular biological analysis and diagnostics" is included in the database with the following information:

patent id	date	IPC	abstract
49188362	25/01/2000	G01/C40	A self-addressable, self-assembling microelectronic device is designed and fabricated to actively carry out and control multi-step and multiplex molecular biological reactions ...

We use these abstracts to form k-grams contained in the abstract of patents by considering all possible combinations of words in these continuous up to k-grams of 3 words.

[30]One might wonder why we do not consider groups of words as they appear in the standards' keywords list. The reason is that we believe that matching part of a k-gram still brings some information. Consider the (real) case of a keyword "ISO screw thread", then a patent containing the 2-gram "screw thread" is still highly relevant.

We proceed to the same cleaning and stemming procedure as for standards' keywords. Note that contrary to other studies that have used semantic analysis on patents' abstract (see e.g. Bergeaud et al., 2017 or more generally regarding patents Adams, 2010), we are not doing anything to select words based on their grammatical functions in the abstract. This is because the number of standards' keywords is limited and there is no need to reduce the size of the patents' abstracts to improve the performance of the algorithm.

B.1.2 Measuring proximity

Once the procedure detailed above is done, we are left with a set of patent $i \in \mathcal{P}$ and a set of standards $j \in \mathcal{J}$. For each patent i, we denotes the set of extracted k-grams by $\mathcal{B}(i)$ while for each standards j, we denotes the set of k-grams by $\mathcal{J}(j)$. We want to compute a score $S(i,j)$ for each pair of a patent and a standard based on the semantic proximity between $\mathcal{B}(i)$ and $\mathcal{A}(j)$. In constructing this score, we keep several criteria in mind:

- We want to give more weight to keywords that have a high IDF since they are more likely to be useful in describing the specificity of a given standard.
- We want to favor a patent whose abstract matches different keywords rather than a patent that matches the same keyword several times.[31]. We therefore only consider keywords once even if they show up several times in a patent abstract.
- We want to value the length of the matched k-grams (i.e. a matching 3-gram will have more relevance than a matching 1-gram).

One natural way to do this would be to consider the following score:

$$S_1(i,j) = \sum_{l \in \mathcal{A}(j)} \frac{n(k,i)}{|\mathcal{B}(i)|} \mathrm{IDF}(l) \tag{B.1}$$

where we have denoted

$$n(l,i) \equiv \sum_{k \in \mathcal{B}(i)} \mathbb{1}(l = k) \tag{B.2}$$

the number of times k-gram l appears in $\mathcal{B}(i)$. This score simply counts the number of times a k-gram in $\mathcal{A}(j)$ appears in patent i's abstract, weighted by the inverse document frequency of this k-gram and standardized by the length of patent i's abstract $|\mathcal{B}(i)|$. However, such a score does not fully take into account the length of the different k-grams, the number of common k-grams between $\mathcal{A}(j)$ and $\mathcal{B}(i)$. We therefore introduce

[31]Indeed, a patent abstract $\mathcal{B}(i)$ can contain the same k-gram several times.

a more complete structure:

$$S_2(i,j) = \sum_{l \in \mathcal{A}(j)} \sqrt{\left(\frac{n(k,i)}{|\mathcal{B}(i)|}\right)^{s(l)}} \text{IDF}(l) \left(|\mathcal{A}(j) \cap \mathcal{B}(i)|\right) \tag{B.3}$$

which compared to S_1: (1) adds a multiplicative term for the number of common k-grams between $\mathcal{A}(j)$ and $\mathcal{B}(i)$; (2) adds a power terms $s(l)$, which returns the length of the k-gram l ($s(l) = 1, 2$ or 3) to the number of concurrences between $\mathcal{A}(j)$ and $\mathcal{B}(i)$ so as to give more weights to longer k-grams and (3) adds a concave function to reduce the impact of the term frequency in the patent to increase the impact of the number of distinct common keywords. In the paper, we consider S_2 as our main measure but we also report results using S_1 in Appendix C.4, as robustness.

B.1.3 Implementation in practice

The size of the databases poses technical difficulties. Because there are more than 21 million priority patents and over 640,000 unique standard documents, we are faced with over 1.4×10^{13} possible matches. We proceed as follows. We first extract all the cleaned and stemmed k-grams from the standards keywords and store these as a dictionary with which all patent abstracts are compared in the next step. When extracting k-grams from the patent abstract, we do not store any k-grams that do not appear in our dictionary of admissible keywords obtained from the standards keywords. We do so for two reasons. First, as the goal of the keyword extraction from patent abstracts is to match those to standard keywords, we do not need to store redundant keywords as they do not match with anything that is in our standards database. Second, the keyword extraction proceeds in forming k-grams from a continuous text that has been stemmed, thus building a large number of k-grams void of sense. For example, from the sentence "The authentication procedure allows for personal data protection." which becomes "authenticat proced allow personal data protect" after stemming, the following 3-grams are extracted from the text: "authenticat proced allow", "proced allow personal", "allow personal data", "personal data protect" as well as the corresponding 2-grams. Only the 3-gram "personal data protect" as well as the 2-grams "authenticat proced", "personal data" and "data protect" are probably meaningful, which is why the use of a pre-defined dictionary as a benchmark is warranted.

After extracting all keywords for each standard, we regroup all associated standard identifiers. We store for each unique keyword in the standards database its associated IDF and a list of all standard ids that correspond to this keyword. We do so similarly for the patent database and store additionally for each associated patent id the number of occurrences of the keyword in the patent abstract as well as the total number of keywords per patent id. Equipped with these two lists, we can match patents to standards by simply building the Carthesian product of the associated standard identifiers and the associated patent identifiers of each keyword. We then add up all patent-standard combinations across all common keywords to compute the scores as described above.

B.2 Matching of ICS and IPC classes

One way to evaluate the quality of our matching procedure is to verify how individual patent-standard matches relate broad categories of the IPC (patents) and ICS (standards) classifications. Essentially, we are linking the two classification systems on the basis of the individual matches obtained in our matching procedure. For the IPC classification, we consider the second hierarchical level, which is the IPC class, and for which 122 classes exist (for example C06 – Explosives; matches.). For the ICS classification, we consider the two-digit level which comprises 40 different ICS fields (for example 49 – Aircraft and space vehicle engineering). Summing the score over all patent-standard combinations that belong to the same IPC-ICS combinations; we obtain a concordance between the two classification systems. Table B.1 lists the closest IPC class for every ICS field.

Table B.1: ICS-IPC CONCORDANCE

ICS	ICS description	IPC	IPC description
1	Generalities. Terminology. Standardization. Documentation	E04	Building
3	Services. Company Organization, Management And Quality. Administration. Transport. Sociology	G06	Computing; calculating; counting
7	Mathematics. Natural Sciences	C12	Biochemistry; beer; spirits; wine; vinegar; microbiology; enzymology; mutation or genetic engineering
11	Health Care Technology	A61	Medical or veterinary science; hygiene
13	Environment. Health Protection. Safety	C02	Treatment of water, waste water, sewage, or sludge
17	Metrology And Measurement. Physical Phenomena	G01	Measuring; testing
19	Testing	G01	Measuring; testing
21	Mechanical Systems And Components For General Use	F16	Engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
23	Fluid Systems And Components For General Use	F16	Engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
25	Manufacturing Engineering	B23	Machine tools; metal-working not otherwise provided for
27	Energy And Heat Transfer Engineering	G21	Nuclear physics; nuclear engineering
29	Electrical Engineering	H01	Basic electric elements
31	Electronics	H01	Basic electric elements
33	Telecommunications. Audio And Video Engineering	H04	Electric communication technique
35	Information Technology. Office Machines	H04	Electric communication technique
37	Image Technology	G03	Photography; cinematography; analogous techniques using waves other than optical waves; electrography; holography

Continuation of Table B.1			
ICS	ICS description	IPC	IPC description
39	Precision Mechanics. Jewellery	A44	Haberdashery; jewellery
43	Road Vehicles Engineering	B60	Vehicles in general
45	Railway Engineering	B64	Aircraft; aviation; cosmonautics
47	Shipbuilding And Marine Structures	B63	Ships or other waterborne vessels; related equipment
49	Aircraft And Space Vehicle Engineering	B64	Aircraft; aviation; cosmonautics
53	Materials Handling Equipment	B66	Hoisting; lifting; hauling
55	Packaging And Distribution Of Goods	B65	Conveying; packing; storing; handling thin or filamentary material
59	Textile And Leather Technology	D01	Natural or artificial threads or fibres; spinning
61	Clothing Industry	A44	Haberdashery; jewellery
65	Agriculture	A01	Agriculture; forestry; animal husbandry; hunting; trapping; fishing
67	Food Technology	A23	Foods or foodstuffs; their treatment, not covered by other classes
71	Chemical Technology	F42	Ammunition; blasting
73	Mining And Minerals	E21	Earth or rock drilling; mining
75	Petroleum And Related Technologies	C07	Organic chemistry
77	Metallurgy	C23	Coating metallic material; coating material with metallic material; chemical surface treatment; diffusion treatment of metallic material; coating by vacuum evaporation, by sputtering, by ion implantation or by chemical vapour deposition, in general; inhib
79	Wood Technology	B27	Working or preserving wood or similar material; nailing or stapling machines in general
81	Glass And Ceramics Industries	C03	Glass; mineral or slag wool
83	Rubber And Plastic Industries	C08	Organic macromolecular compounds; their preparation or chemical working-up; compositions based thereon
85	Paper Technology	D21	Paper-making; production of cellulose
87	Paint And Colour Industries	B05	Spraying or atomising in general; applying liquids or other fluent materials to surfaces, in general
91	Construction Materials And Building	E04	Building
93	Civil Engineering	E02	Hydraulic engineering; foundations; soil-shifting
95	Military Engineering	F41	Weapons
97	Domestic And Commercial Equipment. Entertainment. Sports	A63	Sports; games; amusements

C Robustness checks

C.1 Other measures for abnormal returns and EPS forecasts

Here we consider two alternative statistical models to build abnormal returns. First, we consider the baseline CAPM model, with the SP500 as market portfolio. Second, we use the French-Fama 3-factor model[32] which augments the baseline CAPM model by considering also the excess returns of small-cap companies over large-cap companies, and the excess returns of value stocks (high book-to-price ratio) over growth stocks (low book-to-price ratio).

We follow the methodology explained in Section 2.3, and estimate these two models over 10-year rolling windows. Hence, we define the abnormal return as the difference between the observed excess return of the company in this period and the one predicted from the model whose estimation window ends in the previous period. Hence, we end up with two different measures: (i) $ar_{i,t}^{CAPM}$, i.e. the abnormal return measured through the CAPM model, and (ii) $ar_{i,t}^{French\text{-}Fama}$, i.e. the abnormal return measured through the French-Fama 3-factor model.

When using these measures as dependent variables in the empirical model of equation (2), we confirm the results of Section 4.2. As shown in Figures C.1a and C.1b, firms whose portfolio of patents is closer to the new standard experience a significant abnormal return at the (imputed) time of public release of the content of the standard.

Finally, we look at the EPS forecast over a 2-year horizon instead of the 1-year horizon considered in Section 4.2. We define $\Delta\mathbb{E}[EPS_{i,t+8}] = \mathbb{E}[EPS_{i,t+8}|\mathbb{I}_t] - \mathbb{E}[EPS_{i,t+8}|\mathbb{I}_{t-1}]$ as the change in the 2-year EPS forecast from professional agencies. As shown in Figure C.1c, we do not find any effect in this case. This implies that professional forecasters do not significantly change their expectations about the EPS two fiscal years ahead. This view is consistent with the dynamics of sales observed after the publication of the standard: as explained in Section 4.3, sales increase only for five consecutive quarters.

C.2 The intensive margin of the variable $\mathsf{Tech.Prox}$

As from Table 2, we know that 50% of firms receive a positive value for $\mathsf{Tech.Prox}$, i.e. they have patents whose content can be matched to a newly released standard. Here, we exploit this fact to understand if the intensive margin of $\mathsf{Tech.Prox}$ really matters.

To show this, we re-estimate the empirical specifications of Sections 4.2–4.4 using only the sample of firm-quarter observations with positive values of $\mathsf{Tech.Prox}$. As shown in Figure C.2, the intensive margin matters for our results to hold. Overall, this evidence corroborates the idea that the size of $\mathsf{Tech.Prox}$ –i.e. the intensity of technological proximity– is important.

C.3 Main results with an alternative clustering procedure

Since standards have an impact at the industry-level, we chose to double-cluster errors at the (NAICS3) industry- and date-level in Section 4.2–4.4 in order to account for correlation of the error term for firms belonging to the same industry and affected by the standard release in the same period. Here, instead, we cluster errors at the firm-level thus taking into account that residuals may correlate within each firm. As shown in

[32] Data on SMB_t and HML_t is available on the data library of Kenneth French's website.

Figure C.3, results do not change with the only exception that the effect of standardization on the change in EPS forecast (Figure C.3b) is significant at the 90% confidence level.

C.4 Main results with a different definition of technological proximity

We want to check whether our results differ if we use another methodology to compute scores in the process of matching patents to standards. As explained in Appendix B.1.2, there are multiple features that we want to consider in constructing the score at the patent-standard level that all capture the idea of semantic proximity. Here, we re-estimate the results of Section 4.2–4.4 when using a different definition of the score to build the firm-level measure of technological proximity. In particular, we use the definition of $S_1(i,j)$ in equation (B.1) of appendix B.1.2 as an alternative which consists in dropping the power term $s(l)/2$ (the term $s(l)$ corresponds to the length of the k-gram l). As Figure C.4 shows, results do not change substantially. Using other scores yields similar results, they are available upon request to the authors.

C.5 Main results including the sample of non-listed firms

In Sections 4.2–4.4, we consider only a sample of firms for which stock market data is available, i.e. publicly listed firms. Here, we add to the sample also firms that are not listed on the stock market. Then, we reconsider model (2) but without market capitalization and q-value of investment as control variables (they depend on stock market prices, which are available only for listed firms) and re-estimate our empirical specifications. Figure C.5 shows these results. Using this augmented sample and a different set of controls, our results do not change.

C.6 The effect of technological proximity on capital investment (CapX)

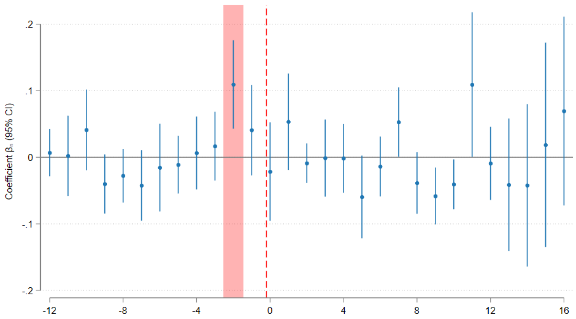
Here, we analyse the impact of standardization on capital investment (CapX).[33] As shown in Figure C.6a, firms operating in a competitive industry and closer to the new technological frontier significantly increase capital investment four quarters after the official publication of the standard. In contrast, when considering non-competitive industries, as in Figure C.6b, we find that technological proximity leads to a decline in capital investment already around the imputed date of release of the first version of the standard.

However, when considering all firms in the sample, the increase in CapX is the dominating effect. In order to quantify the effect of standardization on these variables, we repeat the same analysis as at the end of Section 4.4, i.e. we compare frontier firms to firms not directly affected by standardization. In this case, we find that frontier firms increase CapX by 11.3% by the end of the first year following the publication of the standard.

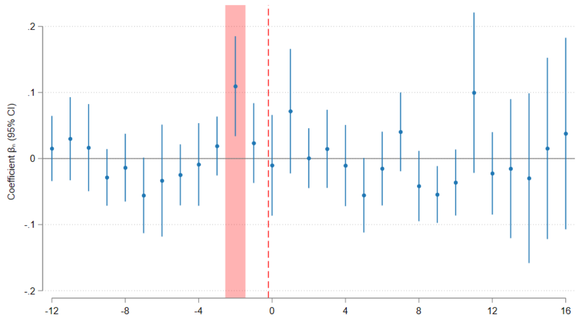
[33] Capital investment is the gross (flow) expenditure for new capital net of depreciation.

Figure C.1: Technological Proximity and Financial Markets' Reaction

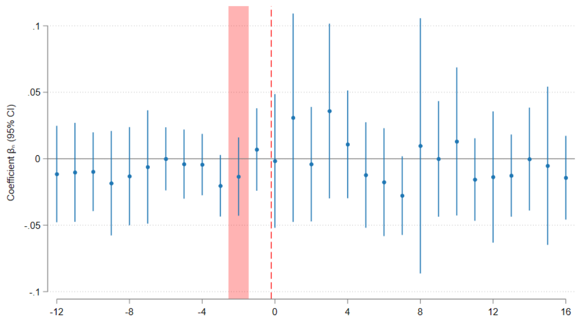
(a) ar^{CAPM}



(b) $ar^{French-Fama}$



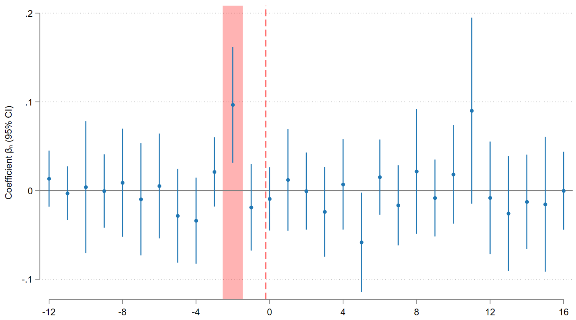
(c) $\Delta\mathbb{E}(EPS)$



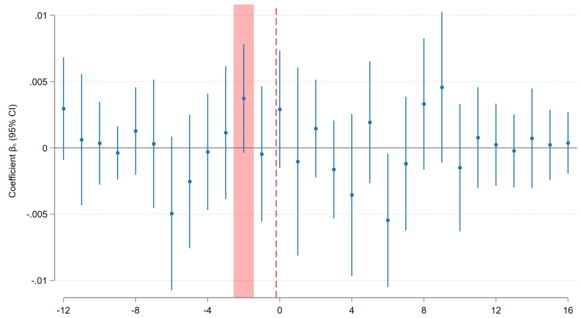
Notes: Figures C.1a and C.1b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is the firm-level abnormal return computed through the CAPM model and French-Fama 3-factor model. Figure C.1c plots the estimated coefficients when the dependent variable is the change in the 2-year EPS forecast.

Figure C.2: MAIN RESULTS: THE INTENSIVE MARGIN OF THE TECHNOLOGICAL PROXIMITY

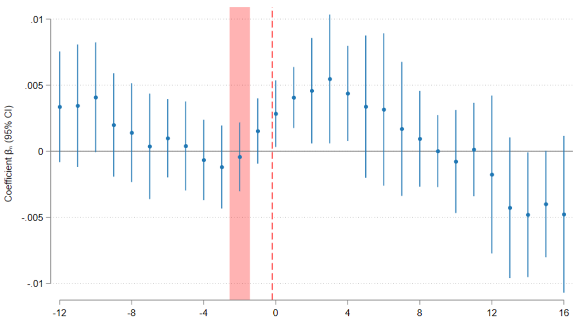
(a) $\text{ar}^{\text{NAICS3}}$



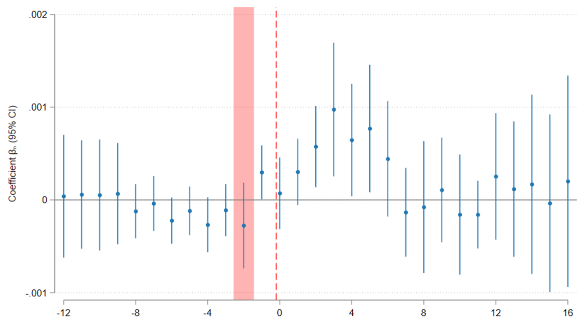
(b) $\Delta\mathbb{E}(\text{EPS})$



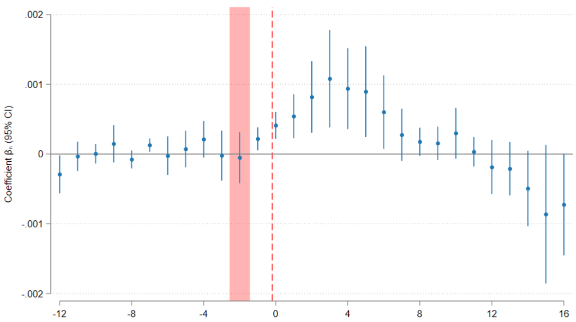
(c) Sales



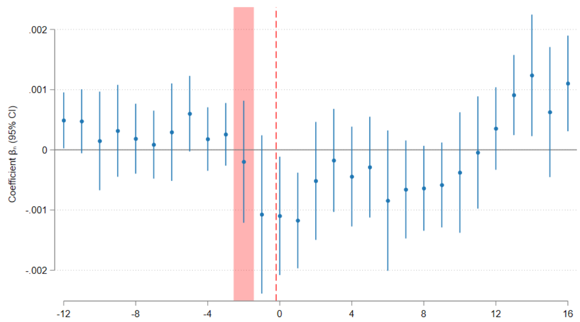
(d) Market Share



(e) R&D (Competitive Ind)



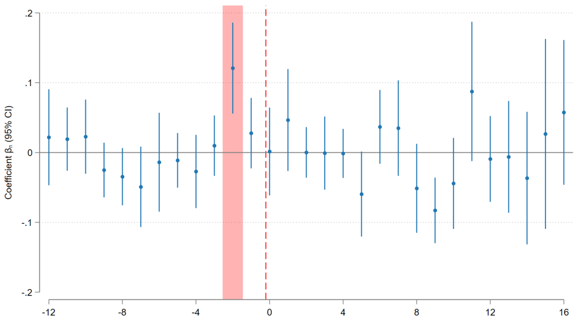
(f) R&D (Non-Competitive Ind)



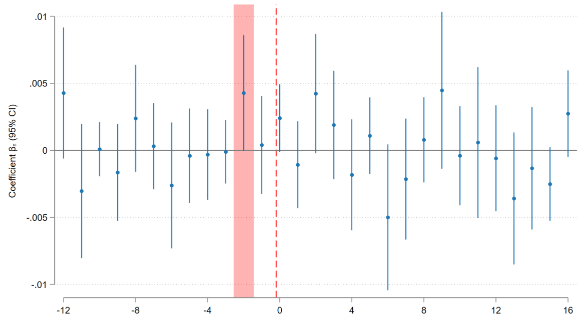
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They differ in the sample of firms included as explained in Appendix C.2: only firm-quarter observations with a positive values for `Tech.Prox` are included.

Figure C.3: MAIN RESULTS WITH ALTERNATIVE CLUSTERING

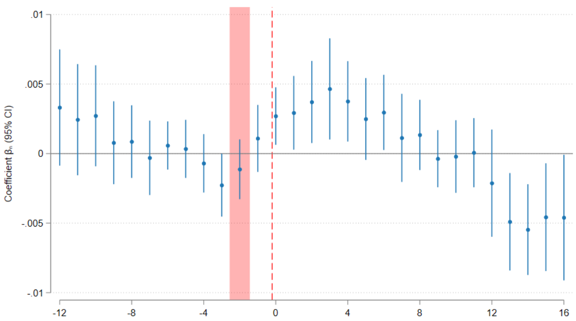
(a) $\alpha r^{\mathrm{NAICS3}}$



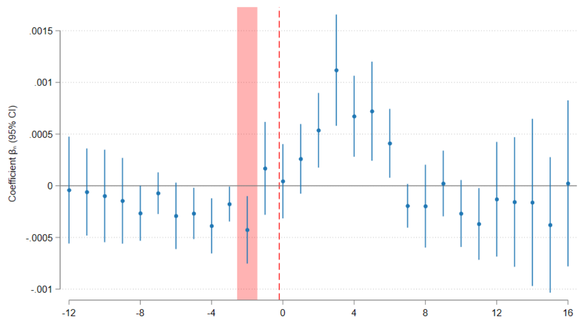
(b) $\Delta\mathbb{E}(\mathrm{EPS})$



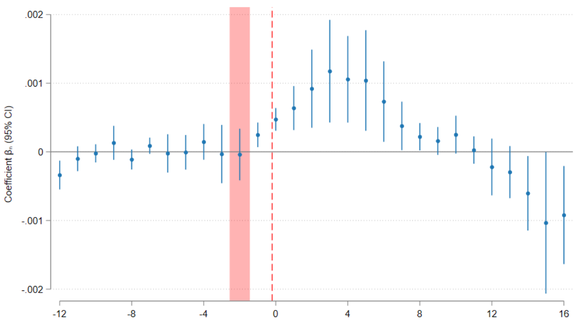
(c) Sales



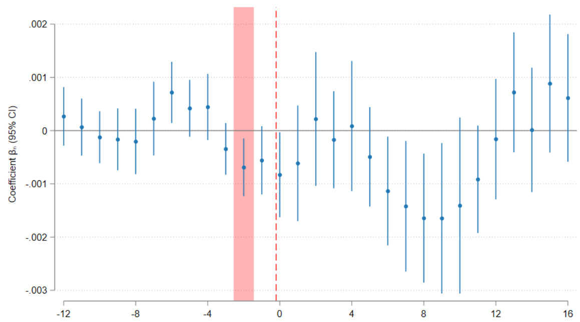
(d) Market Share



(e) R&D (Competitive Ind)



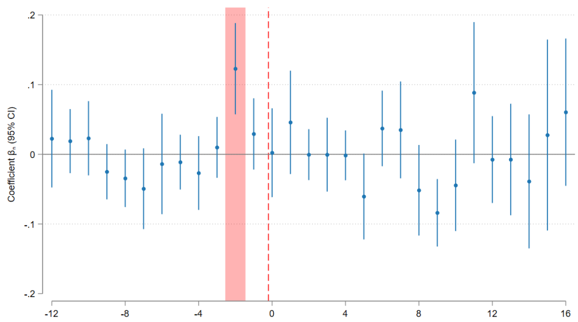
(f) R&D (Non-Competitive Ind)



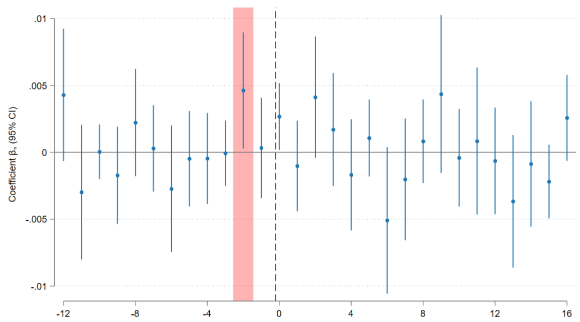
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They differ in clustering of the standard errors as explained in Appendix C.3: standard errors are clustered at the firm-level.

Figure C.4: Main Results with a Different Definition of Tech.Prox

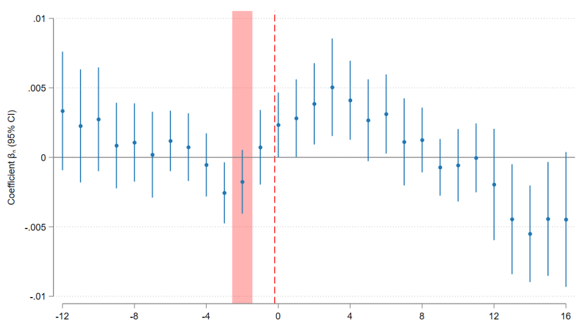
(a) ar^{NAICS3}



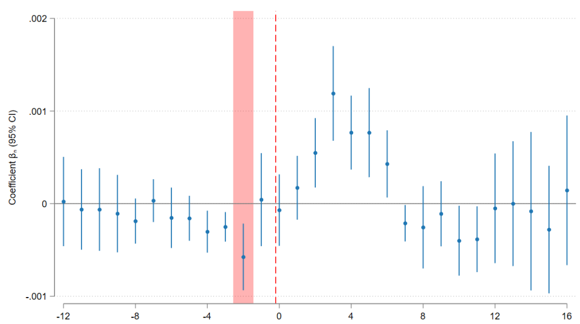
(b) $\Delta\mathbb{E}(EPS)$



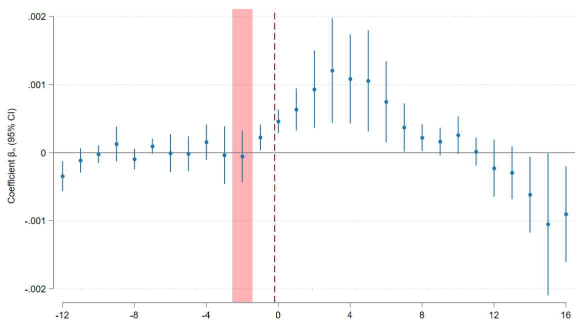
(c) Sales



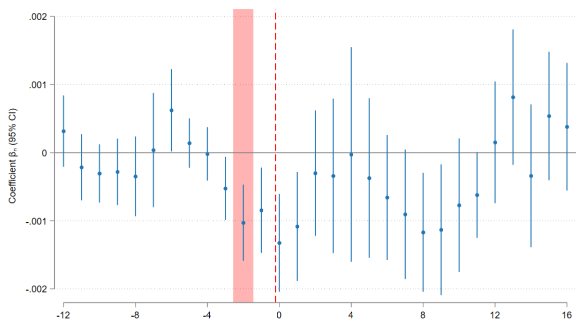
(d) Market Share



(e) R&D (Competitive Ind)



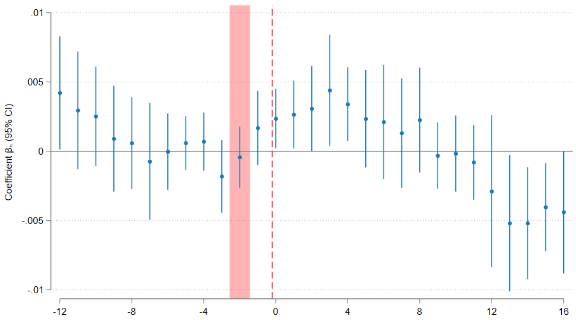
(f) R&D (Non-Competitive Ind)



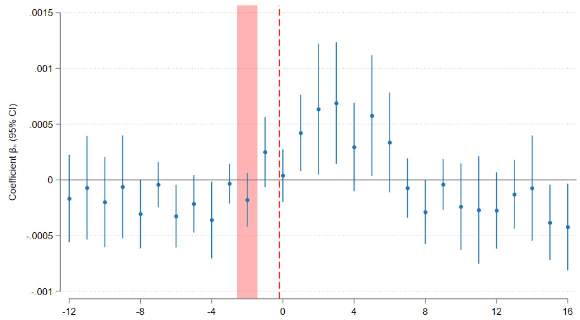
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They differ in the definition of the score that is aggregated at the firm-level as explained in Appendix B.1.2: with respect to the baseline exercises, we use a score that does not include a power term that is applied to the number of times the k-gram l appears in $B(i)$ (see equation (B.1).

Figure C.5: Main Results with Non-listed Firms Included

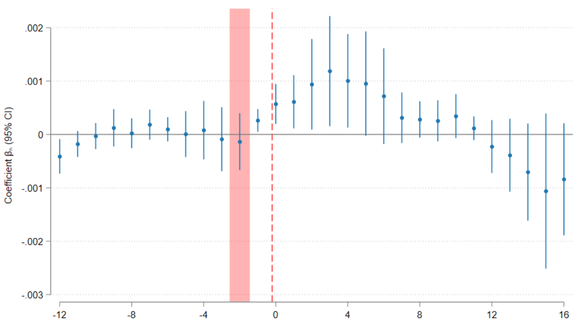
(a) Sales



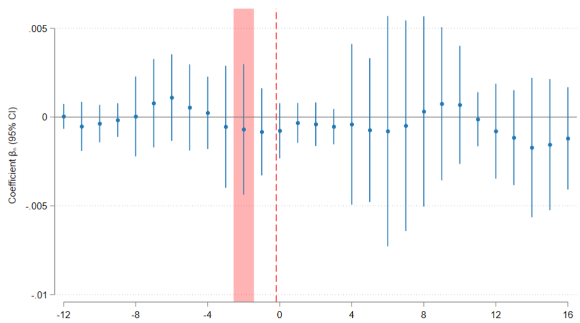
(b) Market Share



(c) R&D (Competitive Ind)



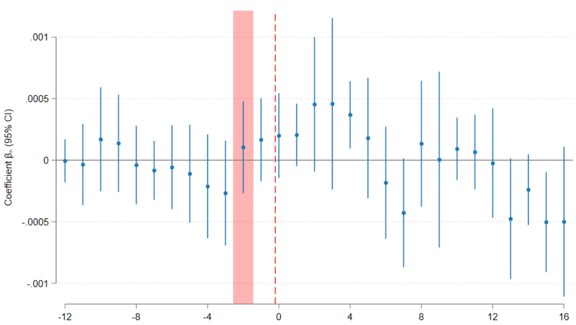
(d) R&D (Non-Competitive Ind)



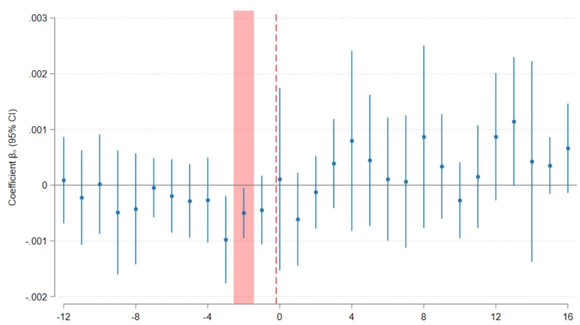
Notes: These figures replicate Figures 3a, 3b, 4a and 4b respectively. They differ in the sample of firms included as explained in Appendix C.5: Compustat firms that are not listed are also included.

Figure C.6: The Effect of Technological Proximity on Capital Investment (CapX)

(a) CapX (Competitive Ind)



(b) CapX (Non-competitive Ind)



Notes: Figure C.6a and C.6b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is capital expenditure (normalized by the mean-level of fixed assets) and the sample is composed respectively by firms operating in a competitive and non-competitive industry.

D Validation of the proximity measure

D.1 Main results controlling for SSO membership

We use data from Baron and Spulber (2018) on SSO membership such that we are able to identify those firms that are active in working groups and technical committees of a SSO. The data also include information on the year of the firm's membership. The data start in 1996, and we match it to our firm-level dataset. We find that 29% of the firms in our sample are a member of a SSO at some point between 1996 and 2010. Then, we build a dummy variable $\mu_{i,t}$ equal to one if the firm i is a member of a SSO during the current and two previous years. We then augment our baseline model as follows:

$$Y_{i,t} = \alpha_i + \phi_{s(i),t} + \mu_{i,t} + \sum_{n=-16}^{N=12} [\beta_n \mathrm{Tech.Prox}_{i,t+n} + \gamma_n \mu_{i,t+n}] + X'_{i,t-1}\eta + \varepsilon_{i,t}. \tag{D.1}$$

This model allows us to check whether the effect of standardization (captured by the β_n-coefficients) remains significant. As Figure D.1 shows, this is indeed the case. In Figure D.2 we plot the γ_n-coefficients which are statistically not significant. We conclude that the results of Section 4.2–4.4 are not explained by SSOs members.[34]

D.2 Main results excluding standards from American SSOs

Among all standards in our database, 15% of them are issued by American SSOs. We believe that US firms may have bigger influence on these SSOs rather than on international ones. For this reason, we exclude these American standards from the computation of the score. Then, we build a new variable $\mathrm{Tech.Prox}$ that excludes the scores from patents matched to US standards. Finally, we re-estimate the empirical models of Section 4.2–4.4. As shown in Figure D.3, results still hold. On the other hand, when considering only standards issued by US SSOs to build the proximity measure, the effect of standardization on each dependent variable is null or it exhibits a pre-trend.

D.3 Main results excluding the most innovative firms

It is possible that it is always the same few firms that experience positive values of technological proximity ($\mathrm{Tech.Prox}_{i,t} > 0$) in a specific industry. In this section, we control that our results are not driven by this group of firms. This is important as we are aware that –within an industry– some firms could be dis-proportionally more innovative than others and more capable to lobby for their patents to become a standard.

To do this check, we study first how much the technological proximity of a single firm explains the sum of values of $\mathrm{Tech.Prox}$ of the entire NAICS3 industry. Formally, for a

[34] Results do not qualitatively change when defining the dummy $\mu_{i,t}$ equal to one only in the year of membership, or when considering a time-invariant dummy taking value one if a firm has been a member at some point.

firm i belonging to NAICS3 industry s, we define:

$$[\textit{Tech.Prox Concentration}]_{i,s} = \frac{\sum_t \mathsf{Tech.Prox}_{i,t}}{\sum_{i \in S} \sum_t \mathsf{Tech.Prox}_{i,t}}$$

as a concentration measure capturing by how much a single firm explains the total value of technological proximity in its industry across time. This variable has a mean of 0.9% (median equal to 0% and standard deviation equal to 6%), which implies that the average firm explains alone only 0.9% of the technological proximity realized in its corresponding industry. Then, within each NAICS3 industry, we drop the top 25th percentile of firms that explain the technological proximity at sectoral level. We assume that these are the firms that might have the innovative capacity, lobbying power or other time-invariant characteristic that allow their patents to match very well to the new standards. Hence, we re-estimate the empirical specifications of Section 4.2-to-4.4.

Figure D.4 shows these results. Using this restricted sample, we find that our baseline results are not driven by firms that have a consistently higher score in their industry.

D.4 Differences across "treated" and "untreated" firms

In this section, we study whether there are significant (pre-standardization) differences across firms that do receive positive values of $\mathsf{Tech.Prox}$ ($\mathbb{I}[\mathsf{Tech.Prox} > 0]$) and those that do not ($\mathbb{I}[\mathsf{Tech.Prox} = 0]$). To do so, we run the following regression:

$$Y_{i,t-4} = \beta \mathbb{I}[\mathsf{Tech.Prox}_{i,t} > 0] + \alpha_i + \phi_s + \delta_t + \varepsilon_{i,t}$$

where $Y_{i,t}$ can be either: Tobin's Q, leverage, log of market capitalization, return-on-equity (ROE), price-earning ratio (PE), internal cost of capital (R), size (log of assets), age, the number of new patents issued and the stock of patents (both normalized by the average value of fixed assets). α_i, ϕ_s, δ_t are respectively firm, NAICS3 industry and time fixed effects. As shown in Table D.1, one year before the release of the standard, firms with a positive proximity measure did not significantly differ from firms with zero proximity in these several dimensions. In particular, we want to highlight the fact that firms that have issued more patents just before the standard release are not more likely to receive a positive value for $\mathsf{Tech.Prox}$ (column 9). Similarly, firms that have accumulated over time a larger portfolio of patents are not more likely to register positive values for $\mathsf{Tech.Prox}$ (column 10). This suggests that innovation activity at the extensive margin cannot explain our measure of technological proximity and that the strategic release of patents just before the publication of the standard is not present in the data.

D.5 Main results controlling for patenting activity

Here, we want to check if our tecnological proximity is just a by-product of patenting activity in the sense that a large number of patents, independently of their quality, might lead to a higher value of technological proximity at the firm-level. As the variable $\mathsf{Tech.Prox}_{i,t}$ is constructed by matching standards published in t to the number

Table D.1: DIFFERENCES IN FIRM-LEVEL CHARACTERISTICS

	(1) Q	(2) Leverage	(3) Mkt Cap	(4) ROE	(5) P/E	(6) R	(7) Size	(8) Age	(9) New Patents	(10) Stock of Patents
$\mathbb{I}[\texttt{Shock} > 0]$	0.02	0.01	0.025	0.01	8.43	0.06	0.02	0.01	0.00	-0.02
	(0.96)	(1.27)	(1.03)	(1.40)	(1.61)	(0.33)	(1.45)	(1.34)	(1.57)	(-1.13)
Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NAICS3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Return-on-equity (ROE) is defined as the ratio of the firms' quarterly income over the value of equity. PE is the price-earning ratio. R is the internal cost of capital. Size is the logarithm of the value of the assets of the firm. The variable *Stock of Patent* corresponds to the number of patents accumulated until that quarter and normalized by the average value of fixed assets. The variable *New Patent* corresponds to the number of patents released during the quarter and normalized by the average value of fixed assets. All other dependent variables (Age, Q, Leverage, Market Cap.) and dummy variable $\mathbb{I}[\texttt{Tech.Prox} > 0]$ are define in Section 2.3. t-statistics are reported in parenthesis. Standard errors are clustered at the firm-level. "*", "**" and "***" designate significance at the 1%, 5% and 10% level.

of patents accumulated up to $t-4$, we augment our baseline model as follows:

$$Y_{i,t} = \alpha_i + \phi_{s(i),t} + \sum_{n=-16}^{N=12} [\beta_n \texttt{Tech.Prox}_{i,t+n} + \gamma_n P_{i,t+n}] + X'_{i,t-1}\eta + \varepsilon_{i,t}. \tag{D.2}$$

where the variable $P_{i,t}$ is the log of (one plus) the number of patents issued between $t-8$ and $t-4$. We use this model to conduct the same analysis as in Sections 4.2–4.4. As shown in Figure D.5, results overall do not change: the effect of standardization remains significant. Results also hold when defining $P_{i,t}$ as the log of (one plus) the total number of patents accumulated until $t-4$. Therefore, we conclude that our measure is indeed capturing the extent to which firms are closer to the new frontier. This dimension matters to explain the dynamics of each dependent variable under consideration.

D.6 Main results controlling for patent similarity

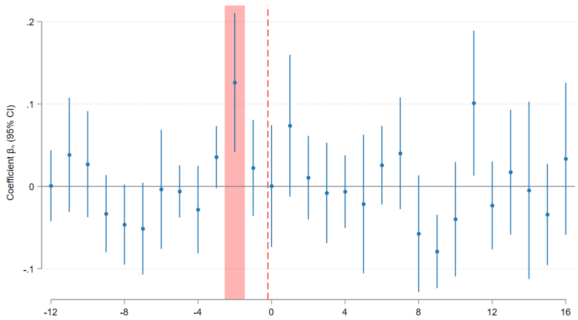
Here, we want to check if our proximity measure still has a significant impact also when controlling for firms that are similar in terms of their patent portfolio. We would therefore like to control for a "technological" category on top of the sector fixed effect that is already included in all models. This captures the idea that patenting intensity might differ across technological classes and more generally that our score can be naturally larger for some technologies than for others. To construct this category, we look at the initial portfolio of patents for each firm, and build a network based on the relative weights of each IPC 3-digit category. We then run a k-mean clustering algorithm using 100 categories which we denote by ts. We augment our baseline model as follows:

$$Y_{i,t} = \alpha_i + \phi_{s(i),t} + \omega_{s(i),t} + \sum_{n=-16}^{N=12} \beta_n \texttt{Tech.Prox}_{i,t+n} + X'_{i,t-1}\eta + \varepsilon_{i,t} \tag{D.3}$$

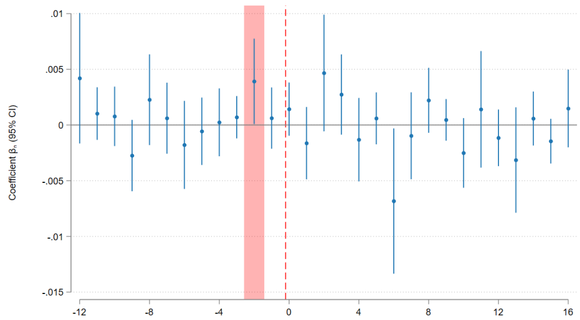
where $\omega_{s(i),t}$ is a technological category fixed effect (interacted with a time dummy) capturing whether a firm belongs to a pool of firms similar in terms of patent portfolio. This control allows us to check if –also within these groups– the effects of standardization on the dependent variables still hold. As Figure D.6 shows, this is indeed the case.

Figure D.1: MAIN RESULTS CONTROLLING FOR SSO MEMBERSHIP

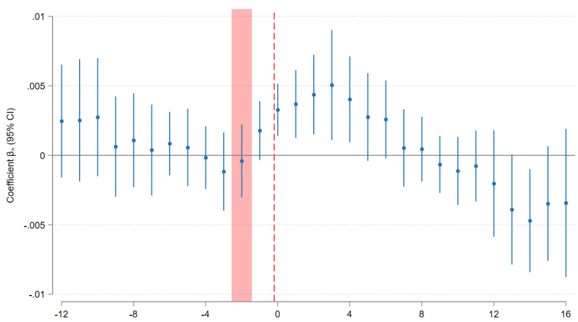
(a) ar^{NAICS3}



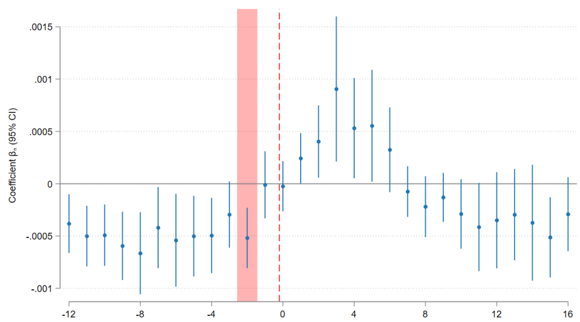
(b) $\Delta\mathbb{E}(EPS)$



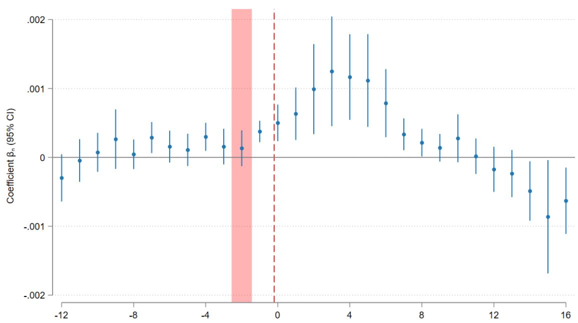
(c) Sales



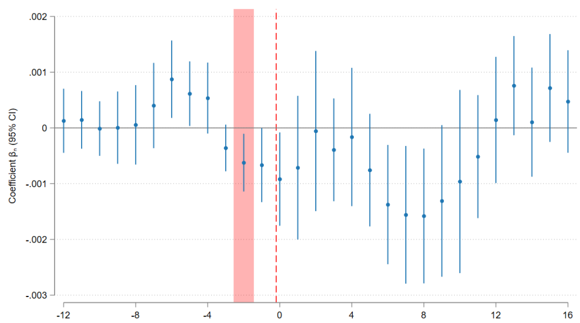
(d) Market Share



(e) R&D (Competitive Ind)

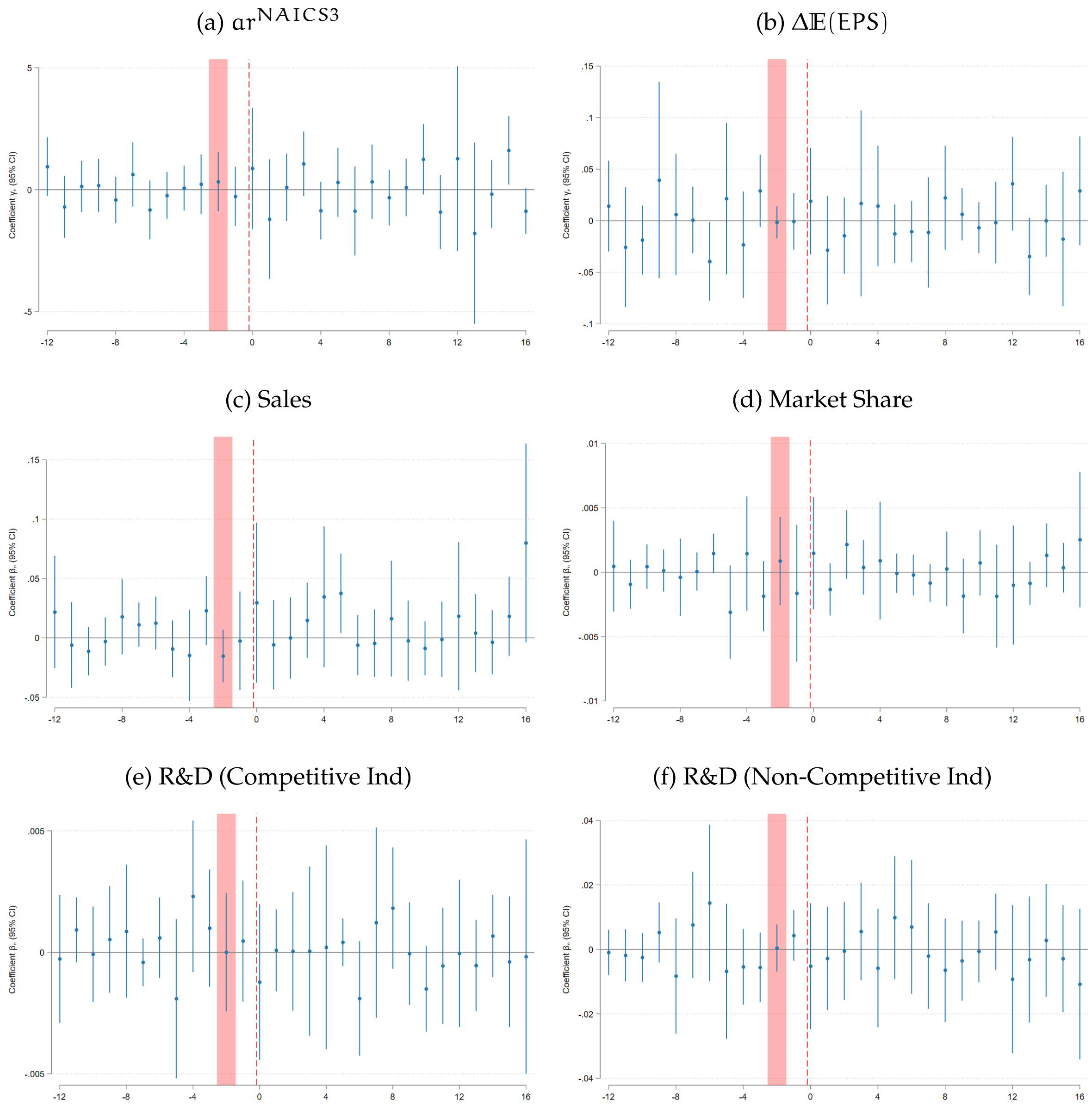


(f) R&D (Non-Competitive Ind)



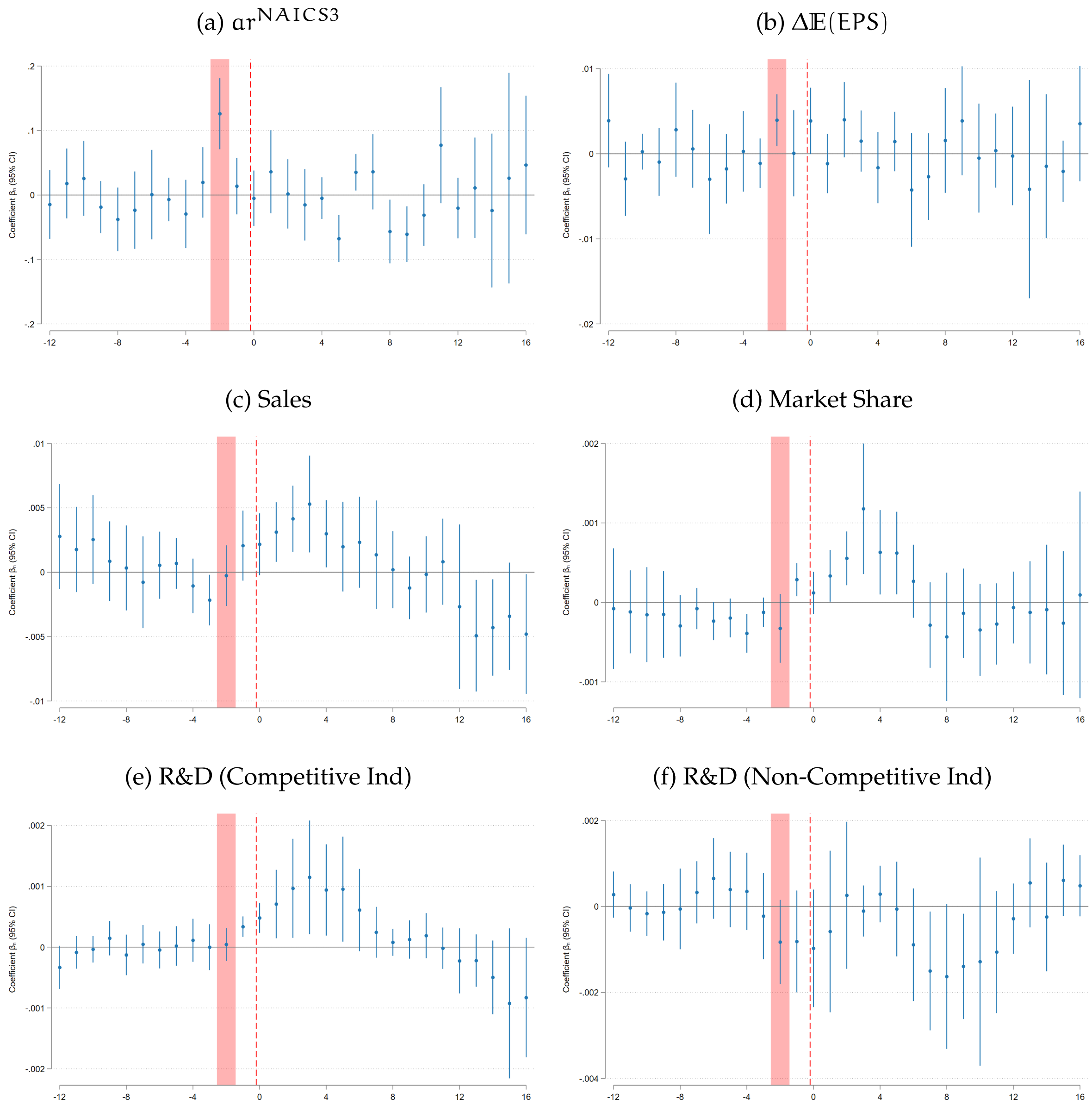
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They include an additional control for SSO membership as explained in Appendix D.1. The coefficient plotted correspond to β_n (see equation (D.1)).

Figure D.2: Main Results: SSO Membership Effects



Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They include an additional control for SSO membership as explained in Appendix D.1. The coefficients plotted correspond to γ_n (see equation (D.1)).

Figure D.3: Main Results Excluding Standards from American SSOs



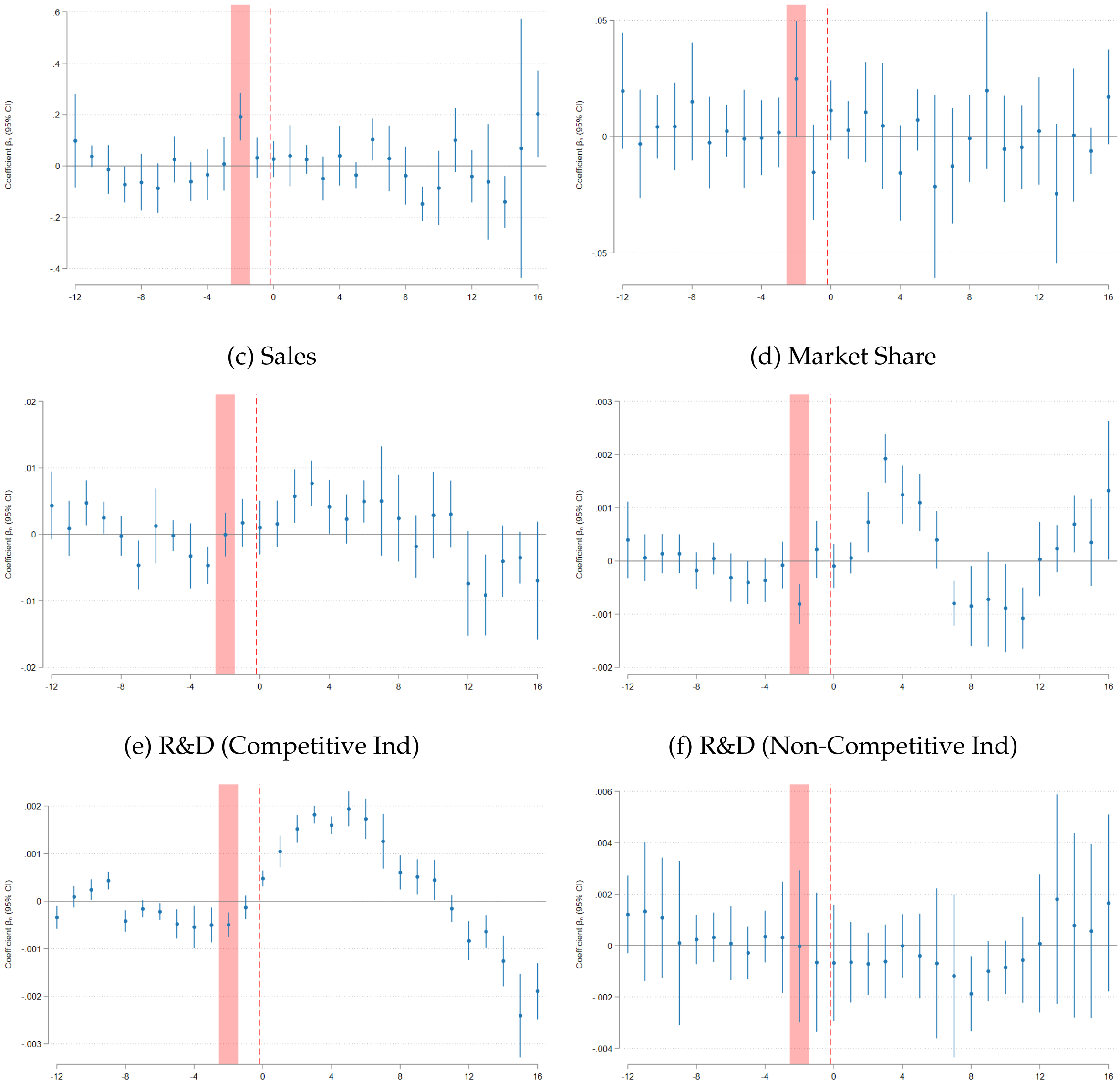
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They exclude standards from American SSOs in the construction of the proximity measure at the firm-level as explained in Appendix D.2.

Figure D.4: Main Results with Most Innovative Firms Excluded

(a) $ar^{\mathtt{NAICS3}}$ (b) $\Delta\mathbb{E}(\mathtt{EPS})$

(c) Sales (d) Market Share

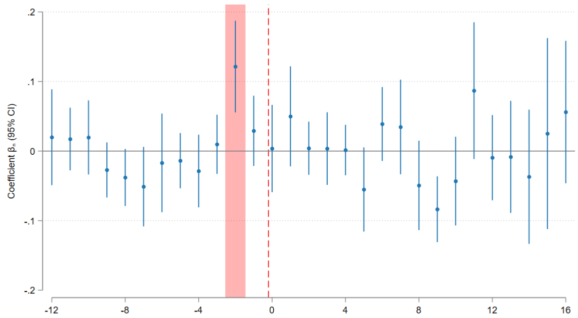
(e) R&D (Competitive Ind) (f) R&D (Non-Competitive Ind)



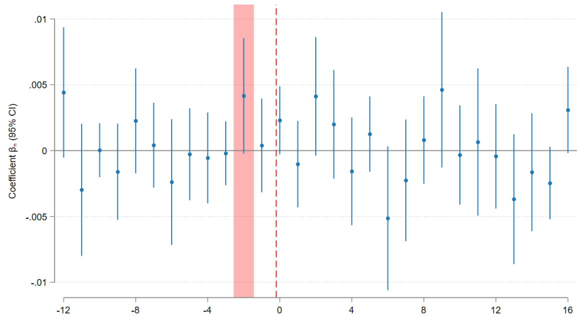
Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They exclude the top 25% firms with the largest value of $\mathtt{Tech.Prox}$ at the sector level as explained in Appendix D.3.

Figure D.5: Main Results Controlling for Patenting Activity

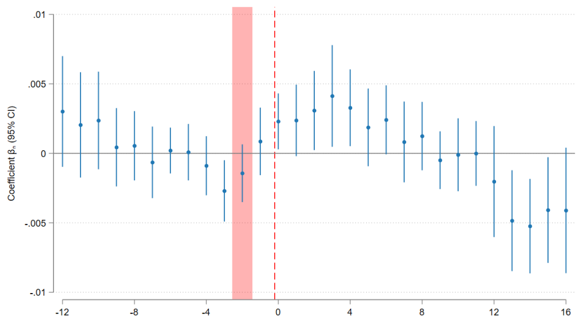
(a) ar^{NAICS3}



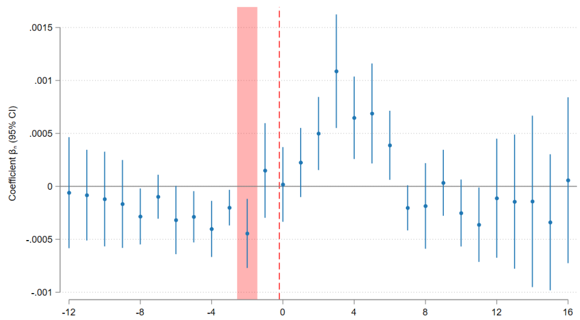
(b) $\Delta\mathbb{E}(EPS)$



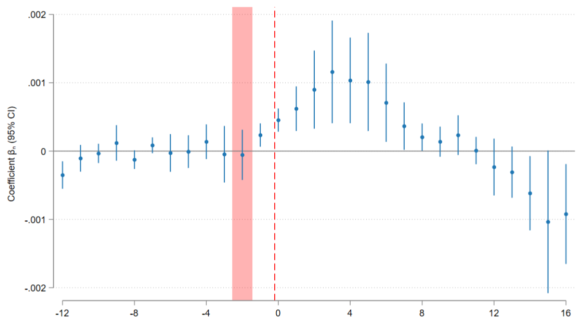
(c) Sales



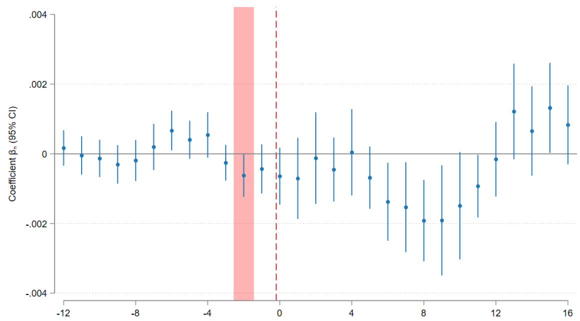
(d) Market Share



(e) R&D (Competitive Ind)

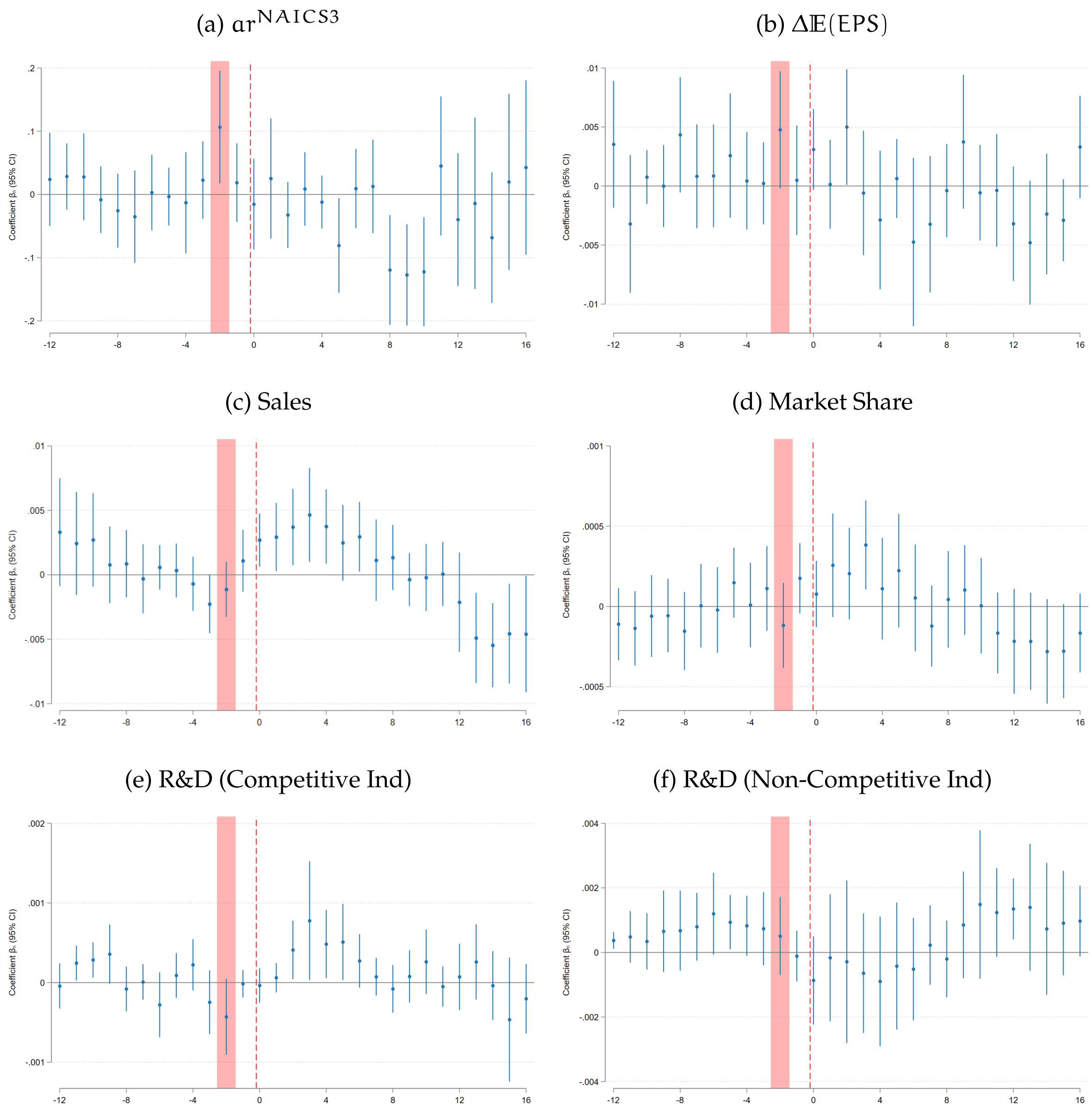


(f) R&D (Non-Competitive Ind)



Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They add an additional control for the number of newly filed patent as explained in Appendix D.5. The coefficients plotted correspond to β_n (see equation (D.2)).

Figure D.6: Main Results Controlling for Patent Similarity



Notes: These figures replicate Figures 2a, 2b, 3a, 3b, 4a and 4b respectively. They add an additional control, namely a dummy for each of the 100 categories of firm based on the similarity of their technological classes of their patent portfolio interacted with a time fixed effect as explained in Appendix D.6.

E Industry-level aggregation and results

Industry-level data. For the sample of firms described in Section 2.3, we aggregate data at NAICS3 industry-level as follows. First, we define as *leaders* those firm-quarter observations for which the variable $\mathsf{Tech.Prox}_{i,t}$ is strictly positive, and as *followers* all the others. By doing so, we take into account that a firm can be in the group of followers in one period, but in the group of leaders in the next one (or vice versa). Given this, we proxy the capability of a sector to adapt to the new standard by taking the cross-firm mean of positive values of technological proximity for each quarter and industry. Then, we aggregate and construct the total amount of sales, patents, the level of market capitalization, the Q-value of investment, the level of leverage for both groups within each industry. Moreover, for each group, we proxy the age of the representative firm with the mean age of firms in that group. Finally, for each industry and quarter we take the number of leaders and followers.

Thereafter, we move to industry-level aggregate figures by aggregating group-specific numbers. Hence for each industry, we build the quarterly growth rate of sales of the industry and its decomposition between leaders and followers, the aggregate Q, leverage, market capitalization and a dummy taking value one for tech-industries. For the mean age of firms at industry-level, we take the weighted average of the mean age of leaders and followers. The weight used is the share of leaders (followers) in each industry-quarter.

Table E.1 shows descriptive statistics of the industry-level data. As from panel A, the mean-industry has a value equal to 0.33. As from panel B, followers in the mean industry spend on aggregate 30% of total R&D expenditure at the industry-level, they issue 29% of all patents in the industry, they have an average market share equal to 36%. The average age of followers across industries is 47 quarters. The aggregate Q-value is on average 2.94 for followers at industry-level, while leverage is 22%. Followers total market capitalization is on average 64 billion dollars. The share of followers in each industry is on average 77%. As from panel C, the industry average growth rate (i.e. the average growth rate of sales) is 2%, with followers and leaders contributing by the same amount. 19% of industries are high-tech. The mean age of firms in the industry is 67 quarters, the mean Q-value is 1.73 and leverage is 22%. The mean market capitalization is 205 billion US dollars.

Results. In Section 5 we use this industry-level data to study how the process of standardization and the proximity of leaders in the industry to the new standard affect sales, investment in R&D, research output (patents), and growth. Do do so, we consider the lead-lag model introduced in Section 4.1, but now defined for an industry-level panel dataset that aggregates firm-level variables. In practice, the model is now:

$$Y_{s,t} = \phi_s + \delta_t + \sum_{n=-16}^{N=12} \beta_n \mathsf{Tech.Prox}_{s,t+n} + X'_{s,t-1}\eta + \varepsilon_{s,t}, \tag{E.1}$$

where $Y_{s,t}$ is the dependent variable for NAICS3 industry s at quarter t. $\mathsf{Tech.Prox}_{s,t}$ is the mean value for leaders in the industry. $X_{s,t-1}$ is the usual set of controls now

Table E.1: INDUSTRY-LEVEL DESCRIPTIVE STATISTICS

	Mean	SD	p1	p5	p25	p50	p75	p95	p99	N
(A) Industry-level Standardization Shock										
Followers Mean Shock	0.33	1.06	0.00	0.00	0.02	0.06	0.17	1.81	5.88	1,512
(B) Followers Characteristics										
Followers Share of Industry R&D	0.30	0.29	0.00	0.00	0.05	0.21	0.47	0.95	1.00	1,512
Followers Share of Industry Patents	0.29	0.28	0.00	0.00	0.04	0.22	0.44	0.91	1.00	1,512
Followers Market Share	0.36	0.24	0.02	0.06	0.18	0.32	0.50	0.86	0.94	1,512
Followers Mean Age (quarters)	47.93	17.79	17.80	25.23	33.71	44.47	60.11	80.67	98.65	1512
Followers Q	2.94	9.14	0.02	0.05	0.34	0.77	1.59	12.26	53.00	1,512
Followers Leverage	0.22	0.10	0.03	0.08	0.15	0.21	0.28	0.41	0.53	1,512
Followers Market Cap. (Billion$)	64.00	124.07	1.32	2.46	9.42	19.19	59.68	319.30	692.66	1,512
Share of Followers in the industry	0.77	0.15	0.33	0.50	0.67	0.80	0.89	0.97	0.99	1,512
(C) Industry Characteristics										
Industry Quarterly Growth Rate	0.02	0.02	-0.02	-0.01	0.00	0.01	0.02	0.06	0.10	1,512
Contribution of leaders	0.01	0.01	-0.02	-0.01	0.00	0.01	0.02	0.04	0.05	1,512
Contribution of followers	0.01	0.02	-0.01	-0.01	-0.00	0.00	0.01	0.04	0.08	1,512
$\mathbb{I}(\textit{Tech-industry})$	0.19	0.39	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1,512
Industry Mean Age (quarters)	67.73	24.40	26.36	34.24	46.71	68.15	83.96	110.55	124.57	1,512
Industry Q	1.73	0.63	1.00	1.09	1.30	1.54	1.97	3.03	3.96	1,512
Industry Leverage	0.22	0.09	0.06	0.10	0.16	0.20	0.27	0.39	0.45	1,512
Industry Market Cap. (Billion$)	205.26	327.85	4.23	9.47	29.47	82.60	217.26	1022.42	1644.88	1,512

Notes: see Appendix E for details on data construction.

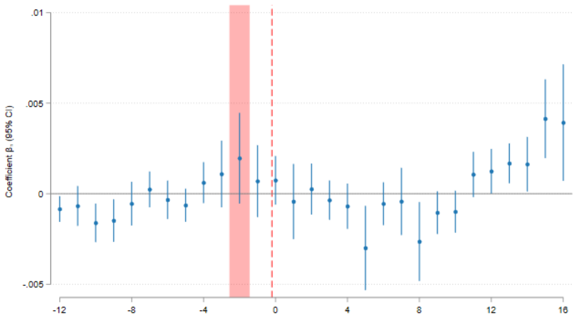
defined at the industry-level (age, Tobin's Q, market capitalization leverage, a dummy for high-tech industries) described in panel C of Table E.1.

We estimate this model where the dependent variable is sectoral growth and its components. Figure E.1 shows the estimated coefficients. Table 4 in Section 5 shows the cumulative effects of the mean value (0.33) when we aggregate estimates over the first four quarters after the publication of the standard, or over all periods after the publication.

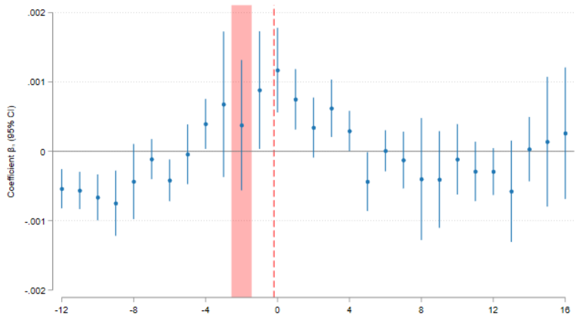
Figure 5 of Section 5 show results when we estimate model (E.1) with dependent variables being respectively the followers market share, their share of total expenditure in R&D, their share of the total research output (patents) in the industry. When considering this group-specific variables, the controls used are also defined at the group-level as described in panel B of Table E.1.

Figure E.1: SECTORAL GROWTH AND TECHNOLOGY ADOPTION

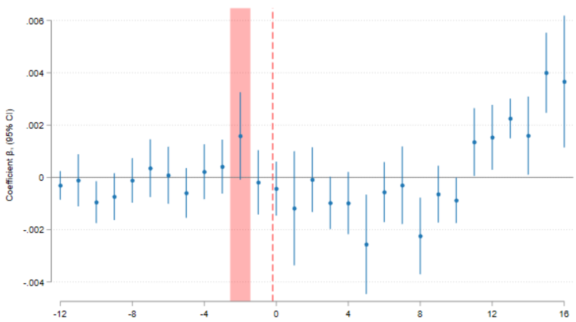
(a) Industry Growth



(b) Leaders' Growth



(c) Followers' Growth



Notes: Figure E.1 plots the estimated coefficients of equation (E.1) when the dependent variable is the quarterly growth rate of the NAICS3 industry and its decomposition between leaders and followers of the industry. See Appendix E for more details on data construction and estimation. In all figures, the 95% confidence intervals for each point-estimate is reported. Standard errors are double-clustered at (NAICS3) industry-level and date. The red area indicates the imputed time-window of public release of the standard's content, based on knowledge of the procedure of approval. The red-dashed line indicates the official publication of the standard, as reported by the standard-setting organization.