WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

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

Antonin Bergeaud[1], Julia Schmidt[2] & Riccardo Zago[3]

June 2022, WP #876

ABSTRACT

When a technology becomes the new standard, the firms that are leaders in producing this technology have a competitive advantage. Matching the semantic content of patents to standards and exploiting the exogenous timing of standardization, we show that firms closer to the new technological frontier increase their market share and sales. In addition, if they operate in a very competitive market, these firms also increase their R&D expenses and investment. 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.

Keywords: Standardization, Patents, Competition, Innovation, Text Mining.

JEL classification: L15, O31, O33.

[1] Banque de France. antonin.bergeaud@banque-france.fr.
[2] Banque de France. julia.schmidt@banque-france.fr.
[3] Banque de France. riccardo.zago@banque-france.fr.

The authors want to thank without implicating Philippe Aghion, Simon Bunel, Maarten De Ridder, Michele Fioretti, Stèphane Guibaud, Jorge Lemu, Dimitris Papanikolaous, Gianluca Violante and seminar participants at Northwestern University, ESSIM 2022 and Collège de France, INSEAD for helpful comments.

NON-TECHNICAL SUMMARY

The development and production of goods and services is often subject to a myriad of technical standards. From payments systems to specifications for door frames 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 also entails the selection of one technology among competing ones as it aims at assuring the widespread proliferation of the best technologies and practices within each industry. In this sense, the process of standardization goes hand in hand with technological progress: when new technologies emerge, new standards are defined in order to facilitate their large-scale adoption.

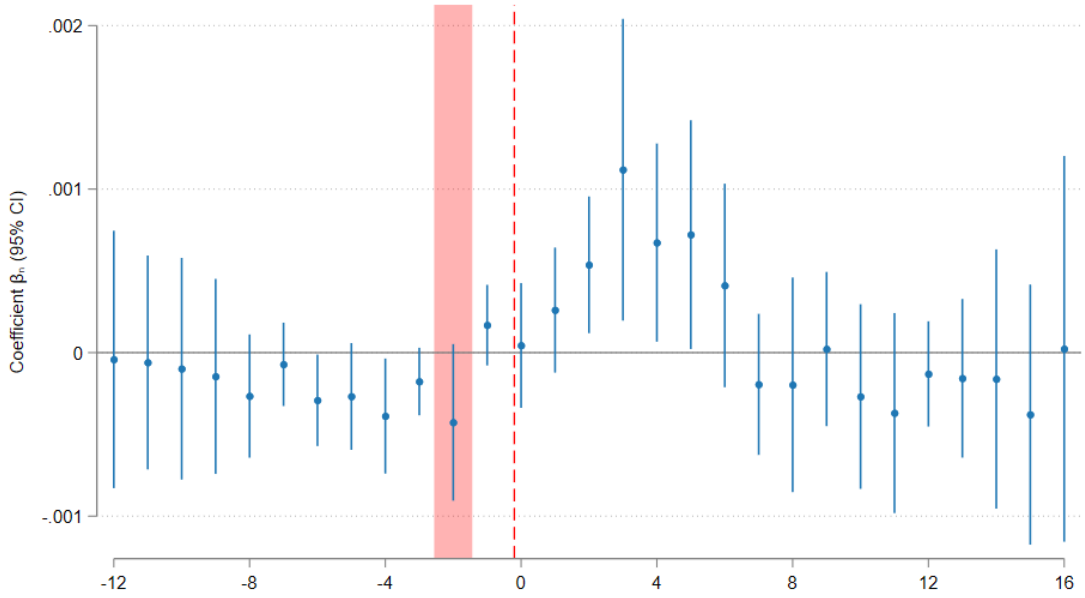
Yet, the ability of firms to adapt to the new standard -which we refer to as the new *technology frontier*- depends on their past technological choices. Indeed, some firms -given their innovation history- could be technologically better prepared to deploy the technologies described in the new standard. As such, firms close to the new frontier may have an immediate competitive advantage and benefit from a shift in market power in their favor. This raises a well-known trade-off between rewarding successful innovations and avoiding the creation of monopolies. This paper contributes to the debate. By introducing a new measure of proximity of firms to the technological frontier, we show how the selection of one technology among competing ones through standardization affects competition, innovation and growth.

Our new measure of proximity to the frontier uses text analysis to study the extent to which the semantic content of firms' patents overlaps the content of a newly issued standard. Hence, we use this measure to study the impact of standardization at the firm and sectoral level.

We show that, when a new standard is released, firms closer to the new technological frontier gain immediately in terms of sales and market shares. We also find that, if the market is competitive, frontier firms invest more in R&D and capital formation while this is not the case if the level of competition is too low. These results are consistent with the interpretation of standardization as a shock that reduces the level of competition, benefiting technological leaders. Yet, these effects are only temporary. In fact, standardization aims at creating a common ground which allows laggards to catch up in the long-run through spillovers. We show that this mechanism is in place and that the catching-up of followers ultimately drives higher long-term sectoral growth.

Hence, we conclude that standards overall do not create permanent monopolies or lead to rent-seeking behavior. On the contrary, they encourage innovation in the whole industry and contribute to economic growth in the long-run.

Figure 1 : Market Shares and Proximity to the Technological Frontier



Note : Figure 1 shows the dynamic of firm-level market shares when a firm owns patents that are close to the new frontier (defined by a technical standard published in time zeo).

Brevets et normes : implications pour la concurrence et la croissance

RÉSUMÉ

Lorsqu'une technologie devient la nouvelle norme, les entreprises leaders dans la production de cette technologie ont un avantage concurrentiel. En faisant correspondre le contenu sémantique des brevets aux normes et en exploitant l'exogénéité de la date de publication de ces normes, nous montrons que les entreprises plus proches de la nouvelle frontière technologique augmentent leur part de marché et leurs ventes. De plus, si elles opèrent sur un marché très concurrentiel, ces firmes augmentent également leurs dépenses et leurs investissements en R&D. Toutefois, ces effets sont temporaires puisque la standardisation crée une base technologique commune à tous qui permet aux suiveurs de rattraper leur retard et à l'économie de croître.

Mots-clés : standardisation, brevets, concurrence, innovation, text mining.

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 door frames 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 also entails the selection of one technology among competing ones as it aims at assuring the widespread proliferation of the best technologies and practices within each industry. In this sense, the process of standardization goes hand in hand with technological progress: when *new* technologies emerge, new standards are defined in order to facilitate their large-scale adoption.

Yet, the ability of firms to adapt to the new standard –which we refer to as the new *technology frontier*– depends on their past technological choices. Indeed, some firms –given their innovation history– could be technologically better prepared to deploy the technologies described in the new standard. As such, firms close to the new frontier may have an immediate competitive advantage and benefit from a shift in market power in their favor. This raises a well-known trade-off between rewarding successful innovations and avoiding the creation of monopolies. This paper contributes to the debate. By introducing a new measure of proximity of firms to the technological frontier, we show how the selection of one technology among competing ones through standardization affects competition, innovation and growth.

Addressing the above question empirically is challenging as it requires (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 demands 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 innovative activity at the firm-level (see Hall et al., 2005). To combine these two different sources, 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 that are closer to the newly defined frontier, gain in terms of sales and market shares. In fact, if a firm already possesses the capacity to develop products based

upon the new standard, it has an immediate competitive advantage that translates into higher market shares. We also find that, if the market is competitive, frontier firms invest more in R&D and capital formation while this is not the case if the level of competition is too low. These results are consistent with the interpretation of standardization as a shock that reduces the level of competition, benefiting technological leaders. However, this advantage is only temporary. In fact, standardization aims at creating a common ground which allows laggards to catch up in the long-run through spillovers. We show that this mechanism is in place and that the catching-up of followers ultimately drives higher long-term sectoral growth. Hence, we conclude that standards overall do not create permanent monopolies or lead to rent-seeking behavior. On the contrary, they encourage innovation in the whole industry and contribute to economic growth in the long-run.

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 technological adoption at the industry-level (e.g., see Baron and Schmidt, 2014). We show that this measure of *actual adoption* is meaningful as it 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).

In the second part of our investigation, we use the data from Kogan et al. (2017)[1] to match firm-level quarterly data from Compustat, Crisp and Ibes to patent data and our new measure of technological proximity (now aggregated at firm-quarter level). Then, we study whether standardization can actually be considered as an exogenous shock to the firm by looking at stock market reactions. We show that financial markets respond only at the time the content of the standard is made public. In fact, in that very moment, firms closer to the new technological frontier experience higher abnormal returns while professional forecasters review their expectations on firms' future earnings-per-share upwards. Therefore, since frontier and non-frontier firms are also similar in several dimensions, we conclude that the timing of release and content of the standard can be interpreted as exogenous.

Then, we investigate the implications of this shock for the real economy. For this purpose, we use a dispersed lead-lag model, which allows to capture the entire response dynamic to a standardization shock while mitigating the potential bias due to subse-

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

quent and previous standards releases. Under this identification strategy, we first show that firms closer to the new frontier gain both in terms of sales and market shares for roughly five quarters after the publication of the standard. In particular, we estimate that –for frontier firms– this translates into an (average) increase of sales and market share respectively by 6.0%. and 5.6% in the the first year following the standard release.

Thereafter, we consider the responses of investment in R&D and in capital following standardization, which depends 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 and capital is the prevailing effect. We estimate that frontier firms (on average) increase their investment in innovation and capital respectively by 4.4% and 7.2% in the first year following the standard release.

Finally, we move to the macroeconomic implications of standardization for growth, and we interpret our results under the lens of Schumpeterian theory. For example, a step-by-step growth model (see e.g. Aghion et al., 1997, 2001) would predict that standardization gives an advantage for the leading firm in the short-run, but this advantage only matters for innovation if the level of competition is high enough. Yet, in the long-run, standardization increases knowledge spillovers and leads to greater technology diffusion. Ultimately, this mechanism allows followers to catch-up and the economy to grow. Our empirical analysis finds evidence for this Schumpeterian effects: after four years following the release of the standard, sectoral growth increases by 0.11 percentage points, and this is driven by the catch-up process of laggards.

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 by 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 favouring one technology over another, 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). 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., 2020).[2]

The second strand of literature this paper speaks to is on the link between innovation and competition. In standard endogenous growth models (in particular Romer, 1990;

[2] Brachtendorf et al. (2020) also consider the link between standards and patents. Specifically, they use SEP declarations for one specific SSO, namely, the European Telecommunications Standards Institute (ETSI) 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 on the firm and macroeconomic level.

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 et al. (2005) and 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 can be interpreted as a shock that reduces 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 level of competition at the firm and aggregate level.

Our empirical strategy relies on the exogenous timing of the standardization. To study the plausibility of this assumption, 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 show that markets positively reacts to news on patenting activity.[3] All these papers demonstrate that the market efficiency hypothesis (among the many, 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. 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, 2021; 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 con-

[3] 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 firms' obsolescence of their innovation portfolio.

struct measure of novelty based on the amount of textual dissimilarity with 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 aggregate implications of our results and Section 6 concludes.

2 Data construction and matching

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. It prevents or stops others from commercially exploiting the patented invention. 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 which 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

[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).

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.

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

[5]The inverse document frequency is based on a measure of how often a word shows up in a database 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. 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.

Selection. Based on the extraction of the first 100 million matches, we report in Table 1 descriptive statistics of our score. The first row reports the distribution of the score based on the first 100 million matches extracted from the matching procedure. We also compute 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 the second panel of Table 1. In Table 1, 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. Standardization may actually 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. 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).

However, for our analysis, we are interested in the firm-level effects of the standardization of a firm's patent portfolio and therefore exclude patent-standard matches occurring after the release of the standard. Restricting the sample to only those matches where the release of the standard occurs the same year or subsequent to the filing of

documents. See Appendix B for details.

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.

the patent application reduces the number of matched standards. The median time lag for this restricted sample is 8.0 years while the average is slightly higher, at 10.1 years.

In the final line of Table 1, 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.

In Section 3, we evaluate the meaningfulness of our score on the patent-level by investigating its relation with measures of economic and scientific patent value. As we will show and discuss in more detail later, we find that there is a clear, positive association of our aggregated score with other measures of economic importance. Another 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. 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.

2.3 Firm-level data

Aggregation of scores at the firm-level. Given the mapping between each patent of the firm and the corresponding standard, we aggregate patent-to-standard scores at the firm-quarter level by weighting the sum of patents' scores with the relative importance of each 3-digit IPC classes in the firm 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 –belonging to the IPC class j and issued up to $t-4$ – to a standard published at time t. Then,

the weighted aggregation of scores over IPC classes can be written as the following measure of proximity which we refer to as "Shock" throughout:

$$\text{Shock}_{i,t} = \sum_{j \in J} \omega_{j,t_0} \sum_{p \in j} \text{Score}_{i,p,j,t},$$

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.

In conclusion, the variable $\text{Shock}_{i,t}$ is a firm-quarter level information shock expressing the (IPC-weighted) proximity of the stock of patents (accumulated until $t-4$) of a firm to the standard released in quarter t. This shock 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 higher is the shock the closer is the stock of patents of the firm to the newly released standard.[6]

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 four: sales, capital investment, 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. Capital investment is the gross (flow) expenditure for new capital net of depreciation. Since it is usually under-reported, R&D expenditure is measured as a 4-quarter moving average. For comparability across firms, we normalize these three variables by the (mean) level of fixed assets (property, plant, equipment).[7] 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 (build 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, com-

[6] 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.

[7] As we show in Bergeaud et al. (2022), the value of assets is sensitive to the standardization shock. 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.

munication equipment, scientific instruments, medical instruments, and software) as defined in Chan et al. (1990). Finally, we follow De Loecker et al. (2020) to construct NAICS3 industry mean markups. 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 markup of its industry is above the 75^{th} percentile of the distribution. Hence, we construct a dummy variable accordingly.

Financial market data. As explained in Mitchell and Stafford (2000), abnormal returns are useful to study short-term market reactions to corporate events. Following this line, we want to evaluate how markets interpret the standardization shock. Since our analysis focuses on the real effects of the shock 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:

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

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

Table 2: DESCRIPTIVE STATISTICS

	Mean	SD	p1	p5	p25	p50	p75	p95	p99	N
(A) Standardization Shock										
Shock	0.34	2.02	0.00	0.00	0.00	0.00	0.10	1.27	6.24	24,162
$\mathbb{I}[\text{Shock} > 0]$	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
CapX	0.02	0.03	0.00	0.00	0.01	0.02	0.03	0.07	0.14	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 Shock measures the proximity of the stock of patent of the firm to the standard. $\mathbb{I}[\text{Shock} > 0]$ is a dummy that takes value one for positive values of the variable Shock. 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 and CapX are respectively the level of R&D expenditure and capital investment 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 U.S dollars. The dummy variable $\mathbb{I}(\text{Tech} - \text{firm})$ 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.

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 data and our standardization shocks. 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 standardization shock on 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 shock. As from panel (B), the mean level of sales is 62% of the value of fixed assets. Mean (flow) investments

in research and development (R&D) and capital (CapX) are respectively equal to 4% and 2% 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 reduces. 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 verify the validity and quality of our matching procedure by looking 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 value.

3.1 Economic value of a patent à la Kogan et al. (2017)

Kogan et al. (2017, hereafter KPSS (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 the KPSS (2017) 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. It is therefore particularly meaningful to study questions of market share and competition. The economic value à la KPSS (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 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 as calculated by KPSS (2017), we sum the score across all associated standards on the patent-level, essentially weighing each patent-standard association by their individual score (unmatched patents have a zero score). We then merge these data with the KPSS (2017) dataset and run the following patent-level regression:

$$\log(\text{value}_i) = c + \alpha \log(1 + \text{score}_i) + \beta \log(1 + \text{cit}_i) + \gamma f_i + \varepsilon_i \quad (1)$$

Table 3: Regression Results for Financial, Scientific and Private Patent Value

	(1)	(2)	(3)	(4)	(5)
	KPSS (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 KPSS (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 grant year/quarter fixed effect as in KPSS (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 KPSS (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 value_i is the economic value of patent i (in millions USD) from KPSS (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, taken from KPSS (2017), as a control variable as well as fixed effects f_i, namely the interaction of the year and quarter of the grant date with the 3-digit IPC class.

Columns 1 and 2 of Table 3 summarize the results. 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 where we match our database to the KPSS (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 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 the economic value of a patent not in terms of its external valuation by financial markets or other patenting firms, but how patent owners themselves value their patents. Patent holders have to pay maintenance or renewal fees to keep a patent in force.[8] 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. 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 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

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

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.

4 Standardization as a competition shock: firm-level results

In this section, we move from patent- to firm-level data and show that the release of a standard generates the same firm-level response as a temporary negative competition shock, i.e. a shock that expands the market share of some firms at the detriment of others. In particular, we provide evidence that the variable $Shock_{i,t}$, i.e. the firm-level aggregation of patent-to-standard scores, measures well the proximity of a firm to the newly set technological frontier and consequently, it captures the technological advantage of that firm with respect to others. Our empirical strategy relies on the exogeneity of the timing and magnitude of this variable, which we explore by considering the market reaction and, namely, its absence of anticipation. Then, we use $Shock_{i,t}$ to assess the causal impact of standardization on various firm-level outcomes.

4.1 Empirical strategy

Our goal is to analyse the response of firms to standardization shocks. To do so and to better tailor our empirical strategy, it is important first to understand how the standardization procedure works in practice, what is the timing of events in the approval path, and why firms have the incentive to comply with the new standard.

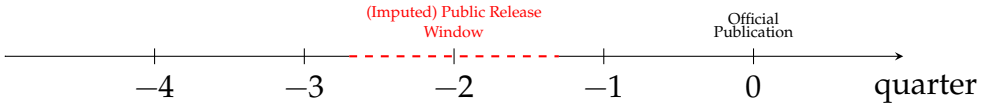
We can briefly summarize the publication path of a standard as follows.[9] Once the standard is proposed and drafted, it goes under the scrutiny of a committee. This first phase concludes with a vote. If the committee's vote is positive, then the draft of the standard is publicly released and circulated 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

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

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 publication occurs in 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[10] (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.[11] Actually, 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.[12] In a similar manner, value chains (both domestic and across borders) require that downstream and upstream producers agree on com-

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

[11] 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.

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

mon 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 a standardization shock on firm dynamics. Yet, it is important to stress that different standards can be released in subsequent periods such that firms can receive multiple shocks throughout time. Therefore, in order to better isolate the effect of a specific shock, 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 dynamic of the response. In particular, in our setting, we know that a static model would be biased since the firm's response could be affected also by subsequent and previous shocks. Our generic model is described in equation (2):

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

where $Y_{i,t}$ is the firm-level dependent variable under consideration. α_i is a firm fixed effect, $\phi_{s,t}$ 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.). $Shock_{i,t}$ expresses the proximity of the stock of patents of firm i at time $t-4$ to the standard publicly released at t. We include 12 lags and 16 leads of the shock (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 a shock happening at $t+n$ on the value of Y measured at t, controlling for the effect of all previous and future shocks. Our identification strategy relies on the assumption that the variable $Shock_{i,t}$ is not correlated with previous realizations of Y. We will check that the response of the firm to future shocks 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 Exogeneity of the standardization shock

The potential to innovate is heterogeneous across firms (see Baumol, 2002 and Griliches, 2007) and this certainly matters for standardization. In fact, in the long-run, firms that innovate more and better are more likely to see their patents become the basis of future standards. In this sense, we can think of standardization as a long-run endogenous process. However, in the short-run, the timing of standardization, its detailed

content and potential impact can be considered exogenous to the firm. In fact, firms do not know ex-ante when the standard will be released and to which extent their stock of patents match the frontier defined by the standard itself. This fact is key for our identification. We dedicate this section to the demonstration that the standardization shock (i.e. the magnitude and timing of the variable $Shock_{i,t}$) is indeed unexpected and exogenous.

To show this, we look at how financial markets and operators react when the content of a standard becomes public. In fact, if markets are efficient (e.g., see Eberhart et al., 2004, Daniel et al., 1998, Mitchell and Stafford, 2000) and the release of the first version of the standard –along with its content– is unexpected, we should observe movements in stock market returns and changes in market expectations around that date. 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.[13]

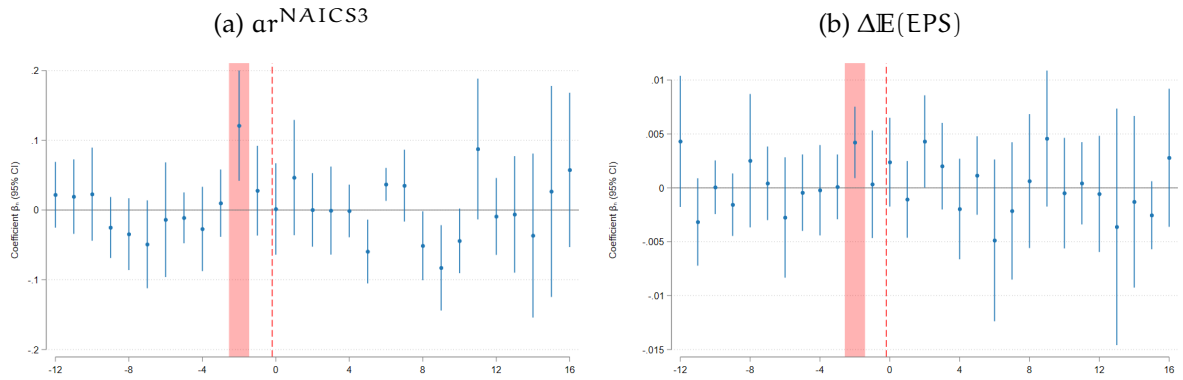
The vector of controls $X_{i,t-1}$ includes age, q-value of investment, leverage and market capitalization of firm i along with a dummy variable taking value one if the firm is operating in a high-tech industry. We consider these variables to take into account respectively for how long a firm has been 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 of the standard, as published by the gazette of 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 $t =$

[13] 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: STANDARDIZATION SHOCK AND FINANCIAL MARKETS' REACTION



Notes: Figure 2a plots the estimated coefficients of equation (2) 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 in the gazette of the standards' organization.

-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 markets efficiently internalize the proximity of the firm to the technological content of the standard only at the moment of the information release. 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 public release of the standard. In words, once the information is public, firms whose stock of patents is closer to the standard are now expected to have a higher EPS in one year. In Appendix C.1 we show that these results hold also 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.[14]

Despite the absence of a pre-trend, some endogeneity concerns still remain. In particular, there might be self-selection into the standardization shock. As mentioned, there might be some (time-varying) firm-specific characteristics that could explain why a firm is always shocked more than others. To address this point, in Appendix C.2 we show that the realization of the shock is not driven by ex-ante firm-level characteristics as firms receiving a positive shock ($\mathbb{I}[Shock > 0]$) do not significantly differ from others in several dimensions. Another similar concern is that some firms are always

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

more innovative and successful than others, or have the network and lobbying influence to push their patents to become a standard.[15] Although these (supposedly) time-invariant characteristics are captured by the firm-level fixed effect in equation (2), in Appendix C.3 we show that the results of this section hold also when removing the top 25% of most innovative firms (i.e. those firms that we believe might have ex-ante pushed or lobbied for their patents to become a standard as they always receive larger and more frequent shocks).

In light of this evidence, we characterize our standardization shock as an information shock and conclude that its timing and content is orthogonal to firms' characteristics and exogenous to investors and operators, who internalize it and react to it only at the moment of the information disclosure.

4.3 Implications for sales and market shares

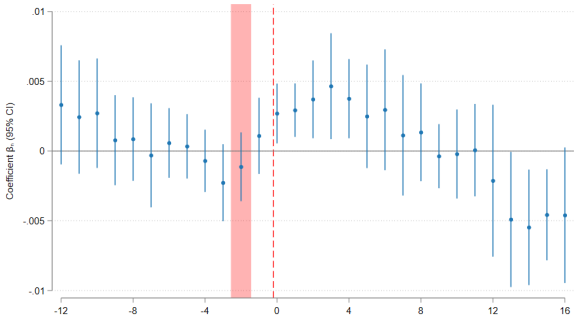
Why do markets value more those firms that are closer to the new technological frontier? What does the standardization shock stand for? Typically, in the context of a corporate event, markets react accordingly by discounting today future cash-flows that are expected to follow the event that revises their forecasts. In this section, we investigate whether a standardization shock indeed changes future cash-flows. In particular, we study what are the real effects of the shock 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 stock 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 standardization shock (*demand effect*) or whether the shock leads also 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. The shock can therefore affect competition and market concentration for roughly one year and a half.

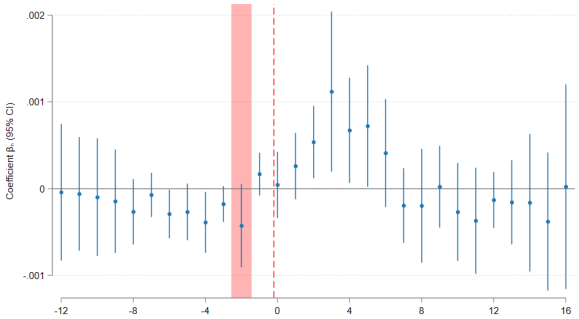
[15] It is important to stress that the role of SSOs is to ensure that the new standard reflects the current consensus in the industry and not the influence of few players. Spulber (2019) shows in a theoretical model that the voting process in SSOs assures that standards are defined efficiently, as a sufficiently large number of industry participants share its economic benefits. This therefore outweighs the detrimental impact of conveying too much market power to firms that might profit from the chosen standard.

Figure 3: STANDARDIZATION SHOCK, 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 in the gazette of the standards' organization.

Can we better quantify the effect of standardization? Given the way we build the shock, it is hard to interpret the estimated coefficients of Figure 3a and 3b. For this reason, we re-estimate equation (2) but include in the sample only firms with either a zero-shock or a shock above the 75^{th} percentile of the distribution of positive shocks. Moreover, instead of the continuous variable $Shock_{i,t}$, we use the dummy $\mathbb{I}[Shock_{i,t} > 0]$ as explanatory variable in the regression. Thus, we can measure the (average) effect of standardization on sales (now in logs) and market share 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 shock, we find that frontier firms increase sales and market share respectively by 6.0% and 5.6% by the end of the first year after the publication of the standard.

We run a number of robustness checks in the appendix. In particular, in Appendix C.4-C.8, we show that these results hold also when clustering errors at the firm-level, when including non-listed firms in the sample, when excluding the top 25% of most innovative firms in each industry, when considering the standardization shock at the intensive margin (which demonstrates that proximity to the new standard really matters) and when using alternative measures of scores for the computation of the shock.

To conclude, the above evidence suggests that the publication of a standard (which proxies technology adoption at the industry-level) attributes a comparative advantage to those firms with a stock of patents closer to the new technological frontier. This advantage translates into higher sales and higher market shares. For this reason, we claim that our standardization shock operates on the market as a (negative) temporary competition shock.

4.4 Implications for R&D and capital expenditure

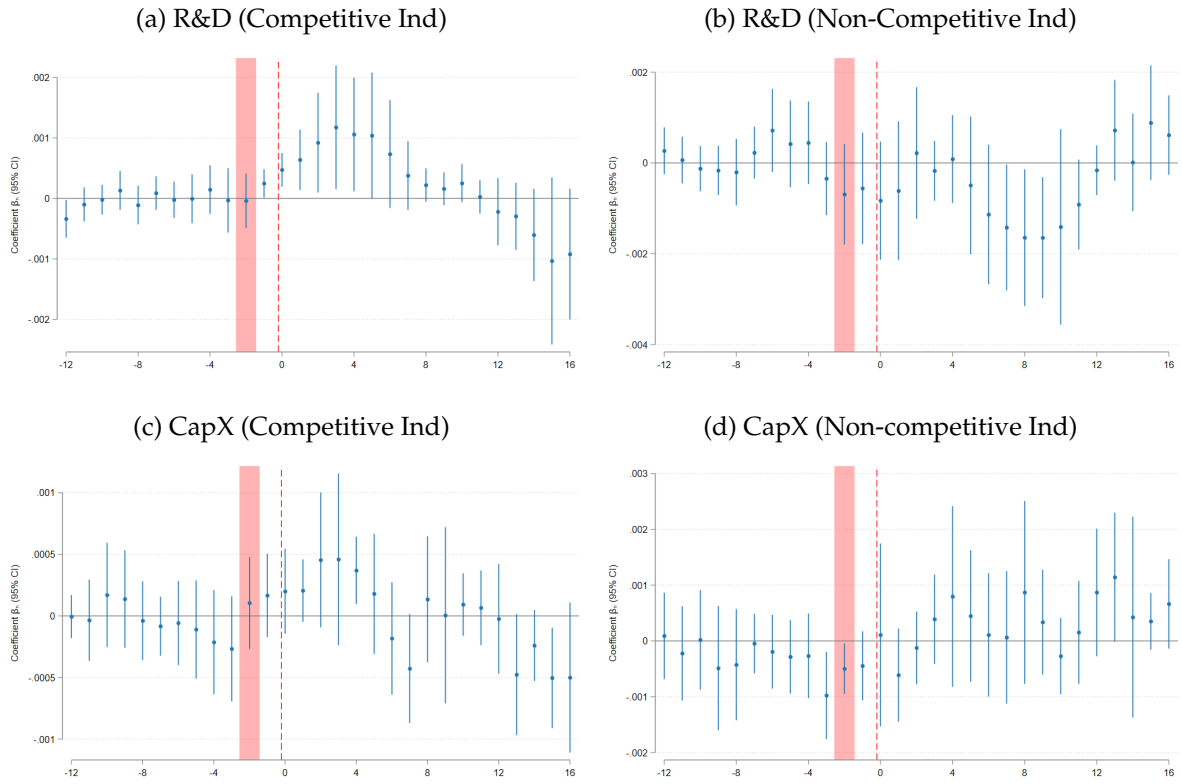
If the shock leads to higher sales and market shares, it may affect also firm-level incentives to invest and innovate in the future. However, the incentives to do so should depend on competition. In fact, if firms are operating in a competitive market prior to the shock, then the short-run advantage that standardization gives to frontier firms is very large relative to others. Consequently, given the highly competitive structure of the market, frontier firms will strive to keep the lead in the future by investing more into R&D and new capital. On the contrary, if the level of competition is too low, i.e. if leading firms were already far ahead of laggards, then the standardization shock gives lower incentives to innovate and invest 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^{th} percentile (non-competitive industries) and those below (competitive industries). We then use our lead-lag model to study the impact of the standardization shock on R&D and CapX investments in competitive and non-competitive industries. If the standardization shock is really a negative competition shock, 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 when the standardization shock realizes. 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. Now, we repeat the same analysis with CapX as dependent variable. As shown in Figure 4c, 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 4d, we find that the standardization shock leads to a decline in capital investment already around the imputed date of release of the first version of the standard. As shown in Appendix C.4-C.8, these results hold to the same robustness checks previously listed.

All in all, these asymmetric responses corroborate the idea that our standardization shock is a temporary (negative) competition shock that gives a comparative advantage to frontier firms. Since their stock of patents better complies with the standard, they are able to expand their market share and –if the market was very competitive before

Figure 4: STANDARDIZATION SHOCK, R&D AND CAPX



Notes: Figure 4a and 4b plot the estimated coefficients 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. Figure 4c and 4d plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

the shock– they invest more in R&D and CapX in order to reinforce and protect their position.

Yet, it is important to mention that –when considering all firms in the sample– the increase in CapX and R&D 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 4.3, i.e. we compare frontier firms to firms not directly affected by the standardization shock. In this case, we find that frontier firms increase R&D and CapX respectively by 4.4% and 7.2% by the end of the first year following the publication of the standard.

5 Aggregate effects and implications

The step-by-step model developed in Aghion et al. (1997, 2001) provides a useful framework to think about the growth effect of standardization. In this model, sectors are characterized by a leader and a follower. The leader has a better process efficiency

and can produce at a lower cost thanks to its past innovation choices. Leaders and followers can increase their productivity through successful innovation, driven by different motives: followers aim at reducing the productivity gap with the leader and –potentially– surpass them (*catching-up effect*), while leaders strive to keep their dominant position in the market (*escape competition effect*).

The effect of standardization can be naturally introduced in this model. First standardization gives an advantage to the leader, the firm who masters the technology that is impacted by the standardization shock. We view this as an increase in the size of the productivity gap between the two firms, which should positively impact the profit and sales of the leader. But a second effect of standardization is to increase knowledge spillovers from the leader to the follower in order to facilitate the catching-up process of the latter.[16] This mechanism of technology diffusion is reminiscent of the results by Bloom et al. (2013): R&D efforts create rents for innovating firms, but overall technology spillovers to other firms dominate in the long-run. As shown in Rysman and Simcoe (2008), standardization aims at improving and incentivizing this diffusion process by creating a common ground and knowledge base for industry participants to build upon. Therefore, standardization could be an important driver of economic growth as knowledge diffusion not only allows for catching-up, but also fosters new innovation (Hegde et al., 2022; Furman et al., 2021).

In light of this, the aggregate effect of standardization should be thought as a combination of a short-term effect of increasing the advantage of leaders, and a long-term effect where followers benefit from technological spillovers and increase their research effort. Which one of these two competing effects on aggregate growth dominates is an empirical question that our data allow to tackle.

To do so, we first study if standardization leads to higher growth at industry-level both in the short- and long-run. In particular, we study how much of the change in growth due to standardization can be explained by leaders and followers. Second, we analyse the dynamics of followers to the standardization shock.

Sectoral growth through standards. 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 positive shocks ($\text{Shock}_{i,t} > 0$), 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.

[16] This is similar to the effect of a relaxation of intellectual property right policy which is discussed in Acemoglu and Akcigit (2012) in the same types of model.

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.[17] We estimate this model with the dependent variable being the industry growth rate as well as the growth rate of leaders and followers. The explanatory variable is the average value of shocks 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,[18] and sum the coefficients over the first year and first-to-forth year after the publication of the standard. 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 receiving a higher aggregate shock, i.e. where leaders are already very close to the new technology frontier. The percentage increase of leaders' growth is 0.08pp for the average shock. This effect is counterbalanced by the growth rate of followers. Since by definition followers are far away from the frontier, the more leaders in the same industry are on top of the new technology, the less followers grow in the short-run. For the average shock 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 the shock, the more sectoral sales increase (by 0.11pp for the average shock 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.

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 and capital 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

[17] Since we are now dealing with a panel where dependent variables and covariates are defined at 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 D explains in detail the construction of the industry-level data and the estimating model used in this section.

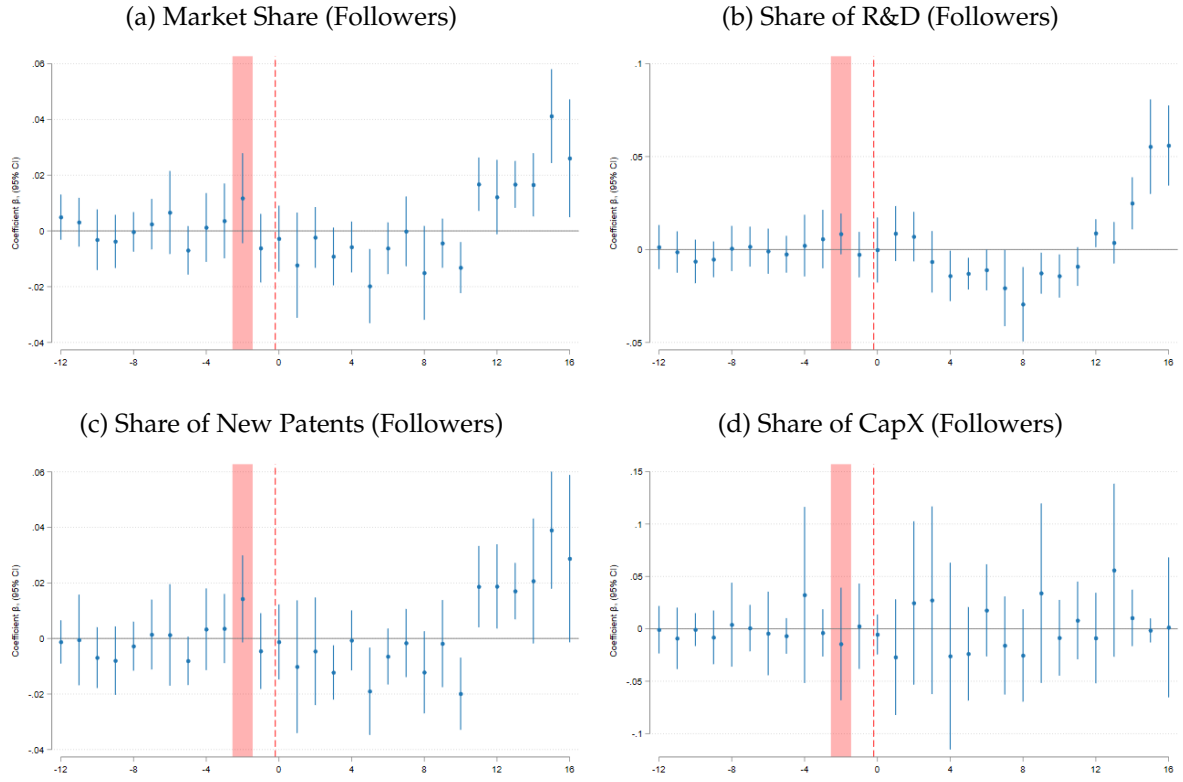
[18] See Figure D.1 in Appendix D.

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 D for details on data and estimation.

Figure 5: THE MACRO-ECONOMIC IMPLICATIONS OF STANDARDIZATION



Notes: Figures 5a-5d 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, the followers' share of total capital expenditure in the NAICS3 industry. See Appendix D 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 in the gazette of the standards' organization.

potentially generate stronger spillovers. To check this, we use our industry-level panel and relate leaders to laggards 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, (iv) the followers' share of total capital expenditure. Then, we use (i)-to-(iv) as dependent variables in model (2). Hence, by doing so, it is possible to understand which group drives sales, innovation and investment 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, laggards 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 laggards 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. On the other hand, as shown in Figure 5d, there is no significant effect on followers' capital investment share.

Overall, these results are compatible with the effects of a competition shock (to a highly competitive market) in the theoretical framework of Aghion et al. (2005). In other words, if the market is competitive, standardization gives temporary premia (in terms of sales and market shares) to frontier firms. In the short-run, these firms innovate more in order to escape competition. Yet, due to spillovers and research efforts, laggards catch up in the long-run and outperform frontier firms. This mechanism leads to higher long-run sectoral (endogenous) growth as the effort of laggards to catch up allows the industry to expand.

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 the timing of release and content of a new standard are exogenous to the firm, and (ii) how firm dynamics change depending on the proximity of the firm's stock 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 can be considered as a competition shock since it gives a temporary comparative 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 invest more and do more R&D after the release of the standard to escape future competition.

In the final part of the paper, we investigate the aggregate implications of standardization at the industry-level. We find that sectors in which leaders are closer to the newly adopted technology exhibit higher growth in the long-run. Yet, 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 industries where leaders are generally closer to the frontier and the potential for spillovers is larger, followers invest more in R&D and their research output is higher. This allows followers 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 laggards 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.

Acemoglu, Daron and Ufuk Akcigit, "Intellectual property rights policy, competition and innovation," *Journal of the European Economic Association*, 2012, *10* (1), 1–42.

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.

_ **,** _ **, Peter Howitt, and John Vickers**, "Competition, imitation and growth with step-by-step innovation," *The Review of Economic Studies*, 2001, *68* (3), 467–492.

_ **, 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.

Baumol, William J, "Entrepreneurship, innovation and growth: The David-Goliath symbiosis," *Journal of Entrepreneurial Finance, JEF*, 2002, *7* (2), 1–10.

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.

Bergeaud, Antonin and Cyril Verluise, "The International Diffusion of Technology: A New Approach and Some Facts," 2021. Mimeo Banque de France.

_ **, 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, "Truly Standard-Essential Patents? A Semantics-Based Analysis," Discussion Paper 14726, Centre for Economic Policy Research 2020.

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.

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.

_ , *13. Patent Statistics as Economic Indicators: A Survey*, University of Chicago Press, 2007.

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.

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.

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.[19]

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

[19] 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.[20]

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:

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

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.

[20] One might wonder why we do not consider groups of words as they appear in the standard's keywords list. The reason is that we still believe that matching part of a k-grams still bring 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 then compute a score $S(i,j)$ for each pair of patent and 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 match the same keyword several time.[21]. 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).

We thus considered five scores that more or less reflect those criteria. Starting from the simplest possible one:

$$S_1(i,j) = \sum_{l \in \mathcal{A}(j)} \sum_{k \in \mathcal{B}(i)} \mathbb{1}(l = k) \text{IDF}(l) \tag{B.1}$$

$$S_2(i,j) = \sum_{l \in \mathcal{A}(j)} \frac{n(k,i)}{|\mathcal{B}(i)|} \text{IDF}(l) \tag{B.2}$$

$$S_3(i,j) = \sum_{l \in \mathcal{A}(j)} \frac{n(k,i)}{|\mathcal{B}(i)|} \text{IDF}(l) \left(|\mathcal{A}(j) \cap \mathcal{B}(i)|\right) \tag{B.3}$$

$$S_4(i,j) = \sum_{l \in \mathcal{A}(j)} \left(\frac{n(k,i)}{|\mathcal{B}(i)|}\right)^{s(l)} \text{IDF}(l) \left(|\mathcal{A}(j) \cap \mathcal{B}(i)|\right) \tag{B.4}$$

$$S_5(i,j) = \sum_{l \in \mathcal{A}(j)} \sqrt{\left(\frac{n(k,i)}{|\mathcal{B}(i)|}\right)^{s(l)}} \text{IDF}(l) \left(|\mathcal{A}(j) \cap \mathcal{B}(i)|\right) \tag{B.5}$$

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

where we have denoted

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

the number of times k-gram l appears in $\mathcal{B}(i)$. The first score S_1 in (B.1) 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. The second score S_2 in (B.2) standardizes this score by the length of patent i's abstract $|\mathcal{B}(i)|$, and score S_3 in (B.3) adds a multiplicative term for the number of common k-grams between $\mathcal{A}(j)$ and $\mathcal{B}(i)$. Score S_4 in (B.4) 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. Finally, score S_5 in (B.5) 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 main part of the paper, we will consider score S_5 for all of our empirical exercises to measure proximity between patents and standards. In Appendix C.8, we report results using alternative shocks 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

In this Section we provide further evidence of the exogeneity of the standardization shock by using other measures for cumulative abnormal returns.

To construct abnormal returns, now we consider two alternative statistical models. First, we consider the baseline CAPM model, with the SP500 as market portfolio. Second, we use the the French-Fama 3-factor model[22], 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 estimating windows 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) $ar_{i,t}^{French-Fama}$, 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 Figure C.1a and C.1b, firms whose stock 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. Hence, 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, in this case we do not find any effect. In words, professional forecasters do not significantly change their expectations when considering how the EPS will be two fiscal years from now. This view is consistent with the dynamic of sales observed after the publication of the standard: as explained in Section 4.3, sales increase only for five consecutive quarters.

C.2 Differences across "treated" and "untreated" firms

In this Section, we study whether there are significant differences across firms that do receive a positive shock ($\mathbb{I}[Shock > 0]$) and those that do not ($\mathbb{I}[Shock = 0]$). To do so we, we run the following regression:

$$Y_{i,t} = \beta\mathbb{I}[Shock_{i,t} > 0] + \alpha_i + \phi_s + \delta_t + \varepsilon_{i,t}$$

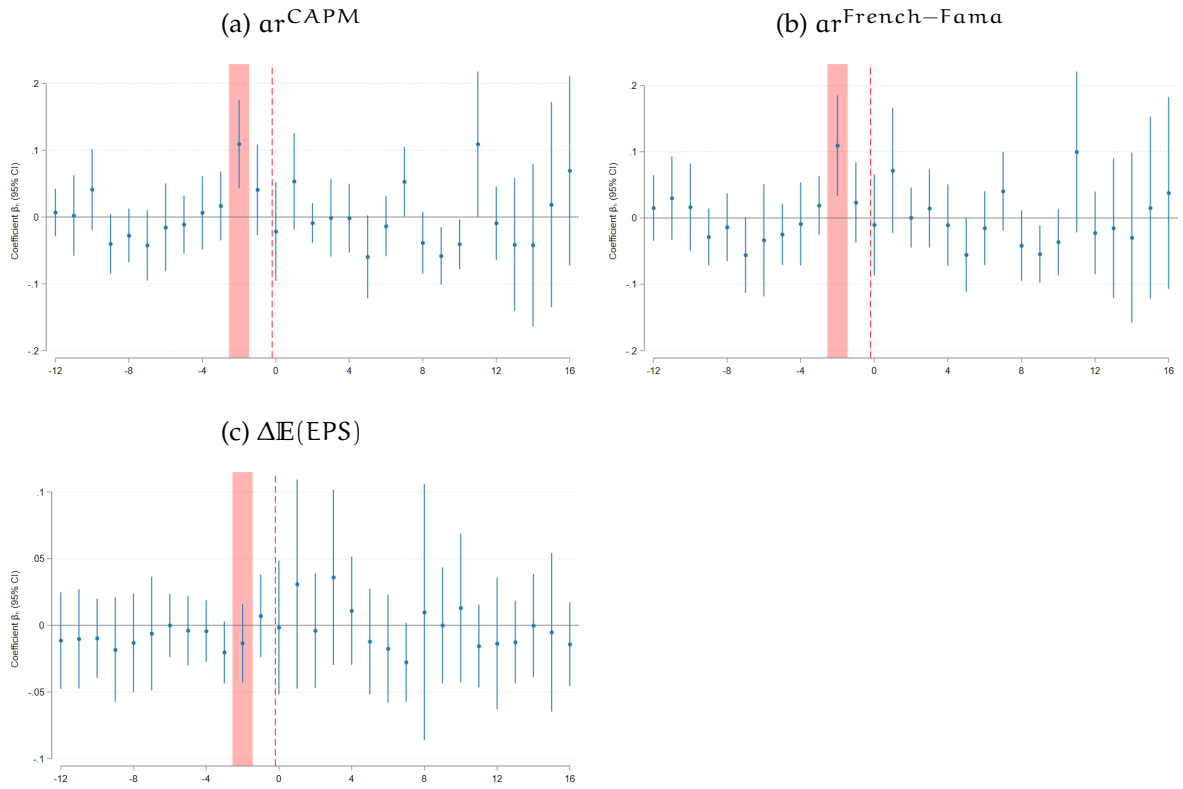
where $Y_{i,t}$ can be either: the age of the firm, Tobin's Q, leverage, log of market capitalization, return-on-equity (ROE), return-on-assets (ROA), price-earning ratio (PE), internal cost of capital (R), size (log of assets). α_i, ϕ_s, δ_t are respectively firm, NAICS3 industry and time fixed-effects. As shown in Table C.1, firms with a positive shock do not significantly differ from firms with a zero shock in these several dimensions.

C.3 Financial markets' reaction when excluding most innovative firms

It can be that it is always the same few firms that experience a positive shock ($Shock_{i,t} > 0$) in a specific industry. In this section, we control that our results are not driven only

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

Figure C.1: Standardization Shock and Financial Markets' Reaction



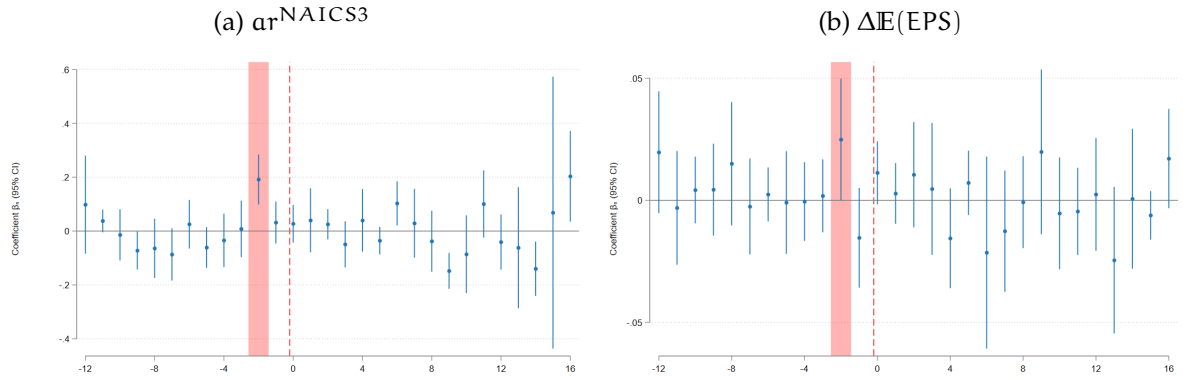
Notes: Figure C.1a and C.1b plots the estimated coefficients of equation (2) when the dependent variable is the firm-level abnormal return computed through the CAPM model and French-Fama 3-factor model. Figure 2b plots the estimated coefficients when the dependent variable is the change in the 2-year 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 in the gazette of the standards' organization.

Table C.1: Differences in firm-level characteristics

	(1) Age	(2) Q	(3) Leverage	(4) Mkt Cap	(5) ROE	(6) ROA	(7) PE	(8) R	(9) Size
$\mathbb{I}[Shock > 0]$	0.01	0.02	0.00	0.02	0.00	-0.00	-0.82	0.81	0.03
	(1.34)	(1.12)	(0.97)	(0.92)	(0.14)	(-0.43)	(-0.14)	(1.15)	(1.71)
Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Naics3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Return-on-assets (ROA) is defined as the ratio of the firms' quarterly income over the value of assets. 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. All other dependent variables (Age, Q, Leverage, Market Cap.) and dummy variable $\mathbb{I}[Shock > 0]$ 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.

Figure C.2: STANDARDIZATION SHOCK AND FINANCIAL MARKETS' REACTION WITH MOST INNOVATIVE FIRMS EXCLUDED



Notes: Figure C.2a plots the estimated coefficients of equation (2) 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 C.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 in the gazette of the standards' organization.

by these group of firms. This is important as we are aware that some firms are capable to lobby for their patents to become a standard and that, in some industries, some firms are dis-proportionally more innovative than others.

To do this check, first we study how much the shocks of a single firm explain the sum of shocks received by the entire NAICS3 industry. Formally, for a firm i belonging to NAICS3 industry s, we define:

$$[\textit{Shock Concentration}]_{i,s} = \frac{\sum_t \text{Shock}_{i,t}}{\sum_{i \in S} \sum_t \text{Shock}_{i,t}}$$

as a concentration measure capturing by how much a single firm explains the total amount of shocks received by it industry across time. This variable has mean 0.9% (median equal to 0%) and standard deviation equal to 6%, which means that the average firm explain alone only 0.9% of the shocks realized in its corresponding industry. Then, within each NAICS3 industry we drop the top 25^{th} percentile of firms that explain the most of the shocks received at sectoral level. Finally, we re-estimate the results of Section 4.2.

Figure C.2 shows results. Also under this sample selection, the essence of the results do not change: results are not driven by firms that consistently score more in their industry.

C.4 Main results under other clustering procedure

Since standards have an impact at industry level, in Section 4.2-4.4 we chose to double-cluster errors at the (NAICS3) industry and date level in order to account for correlation of the error term for firms belonging to the same industry and "shocked" by the standard release in the same period. Here instead, we assume the shock to have purely

a firm-level impact. Therefore, we cluster errors at the firm-level thus taking into account how residuals auto-correlate within each firm and over time. As shown in Figure C.3, results do not change.

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

In Section 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 equity 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 of course only for listed firms). Finally, re-estimate our results. Figure C.4 shows results. Also under this augmented sample and different set of controls, the essence of the results do not change.

C.6 Main results excluding most innovative firms

Following the procedure explained in Appendix C.3, here we re-estimate the results of Section 4.3-4.4 by excluding the top 25^{th} percentile of firms that explain the most of the shocks received at sectoral level.

Figure C.5 shows results. Also under this sample selection, the essence of the results do not change: results are not driven by firms that consistently score more in their industry.

C.7 Intensive vs. extensive margin of the shock

As from Table 2, we know that 50% of firms receive a positive shock, i.e. they have patents whose content can be matched to a new released standard. Here, we exploit this fact to understand (i) if the intensive margin of the shock really matters or (ii) whether our results are explained by the extensive margin of the shock only.

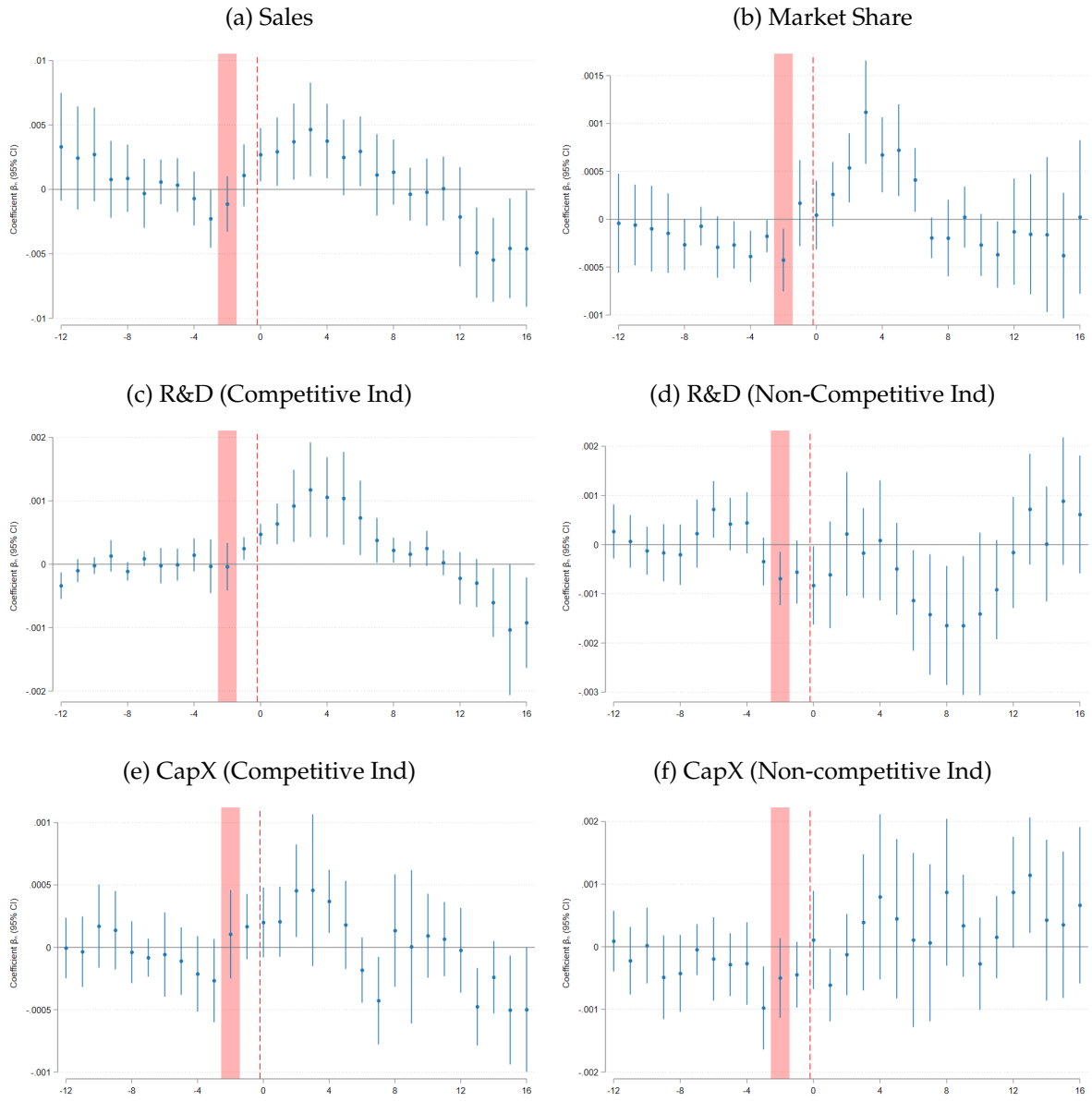
To answer the first question, we re-estimate the results of Section 4.2-4.4 when using only the sample of firms receiving a positive shock. As shown in Figure C.6, the intensive margin matters for our results to hold, with one exception: the effect of the shock on CapX for firms operating in a competitive industry (Figure C.6e) is significant only at 90% significance level. Overall, this evidence corroborates the idea that the size of the shock –i.e. the intensity of the shock– really matters.

To answer the second question, we consider the entire sample of firms and we modify our empirical model of equation (2) as follows:

$$Y_{i,t} = \alpha_i + \phi_{s,t} + \sum_{n=-12}^{N=16} \beta_n \mathbb{I}[Shock_{i,t+n} > 0] + X'_{i,t-1}\eta + \varepsilon_{i,t}$$

where $\mathbb{I}[Shock_{i,t} > 0]$ is a dummy variable taking value equal to one if firm i receive a positive shock at time t. We re-estimate the results of Section 4.2-4.4 under this specification. As shown in Figure C.7, the extensive margin clearly matters only for market

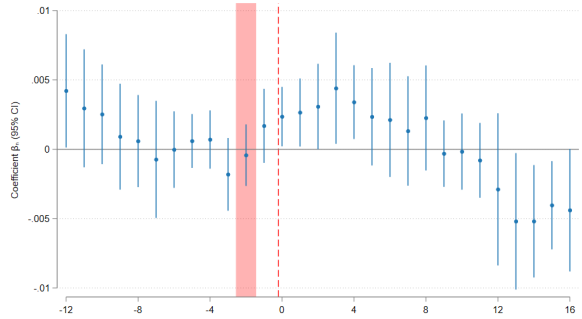
Figure C.3: Main Results under different Clustering



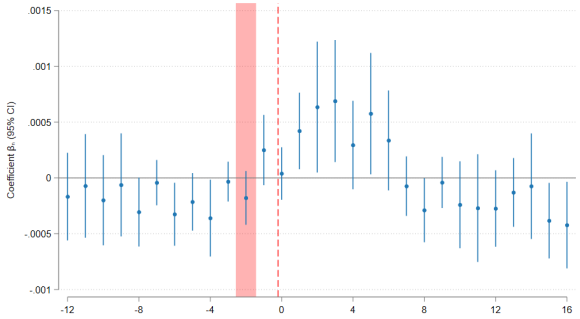
Notes: Figure C.3a and C.3b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is respectively the level of sales (normalized by the mean-level of fixed assets) and the firm-level market-share defined at NAICS3 industry level. Figure C.3c and C.3d plot the estimated coefficients 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. Figure C.3e and C.3f plot the estimated coefficients 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. 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 clustered at the firm-level. 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 in the gazette of the standards' organization.

Figure C.4: Main Results with Non-listed Firms Included

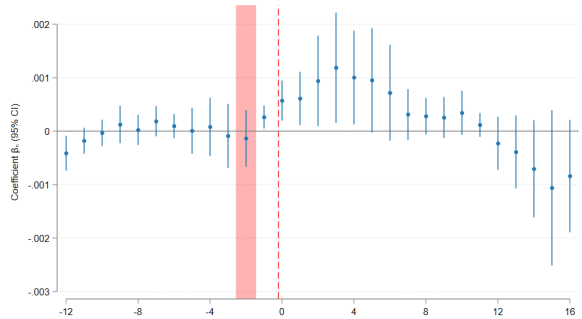
(a) Sales



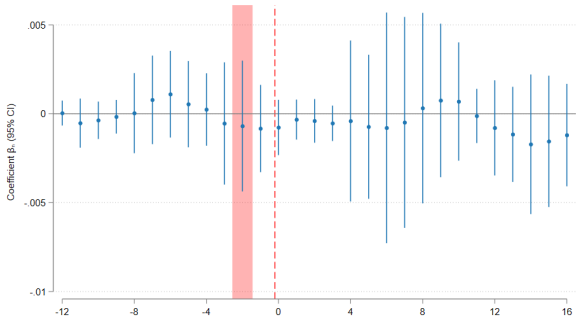
(b) Market Share



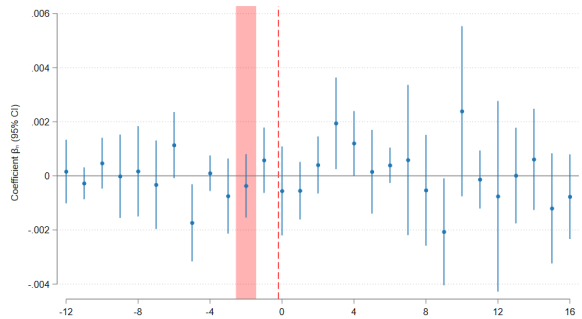
(c) R&D (Competitive Ind)



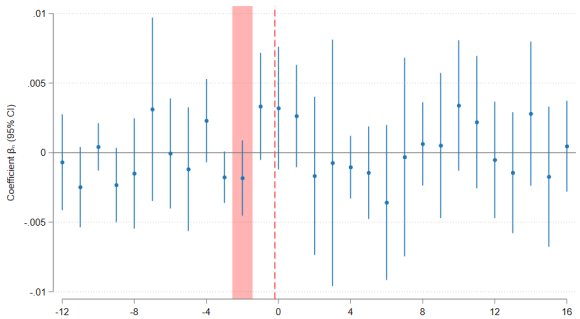
(d) R&D (Non-Competitive Ind)



(e) CapX (Competitive Ind)

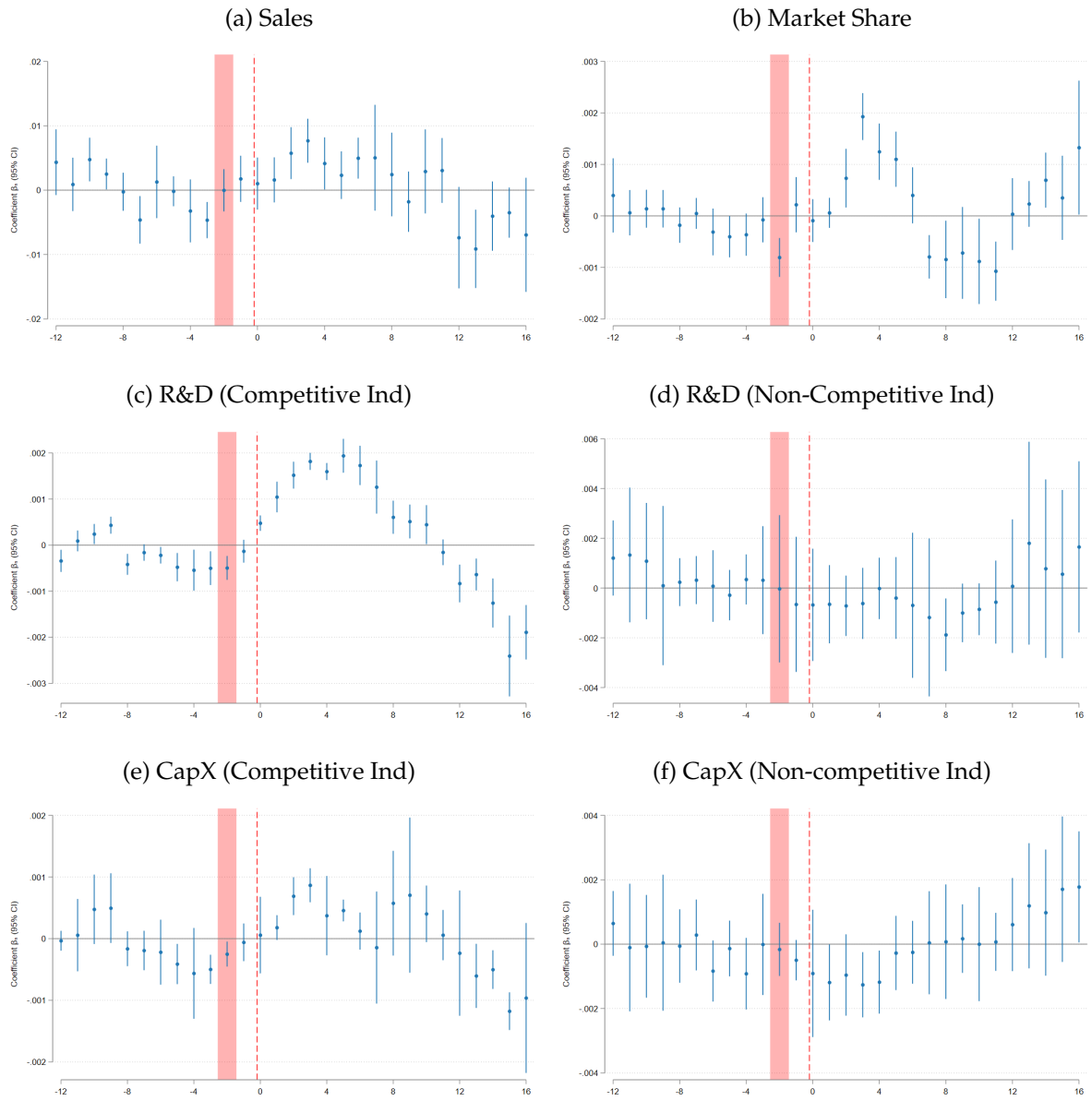


(f) CapX (Non-competitive Ind)



Notes: Figure C.4a and C.4b 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. Figure C.4c and C.4d plot the estimated coefficients 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. Figure C.4e and C.4f plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

Figure C.5: Main Results with Most Innovative Firms Excluded



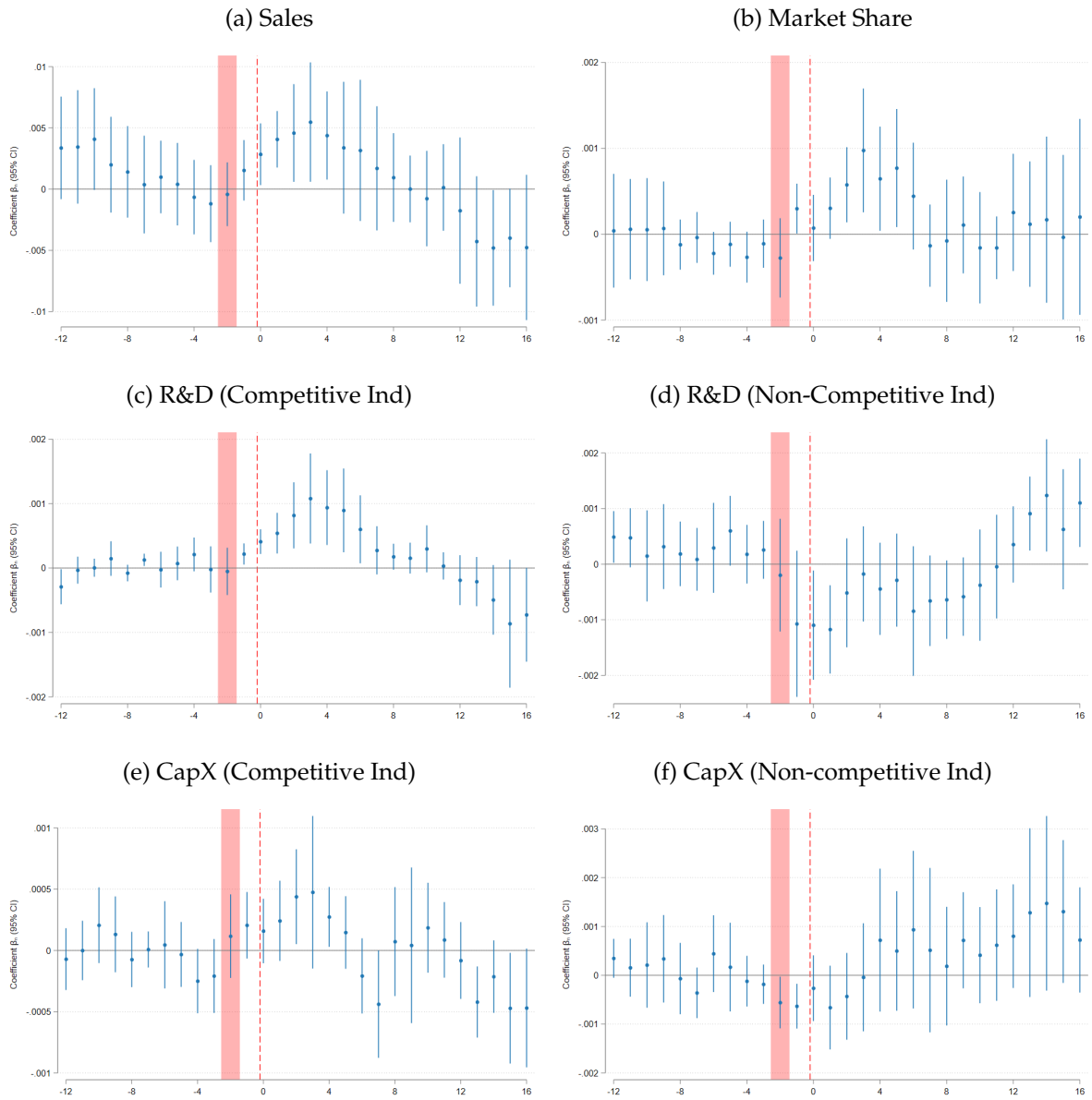
Notes: Figure C.5a and C.5b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is respectively the level of sales (normalized by the mean-level of fixed assets) and the firm-level market-share defined at NAICS3 industry level. Figure C.5c and C.5d plot the estimated coefficients 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. Figure C.5e and C.5f plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

shares: firms receiving a 0-shock immediately loose shares of sales to firms receiving a positive shock.

C.8 Main results under a different definition of the shock

Finally, we want to check whether our results differ much if we use another methodology to compute scores in the process of matching patents to standards. Here, we re-estimate the results of Section 4.2-4.4 when using score B.3 (see Appendix B.1.2) to build the firm-level standardization shock. As Figure C.8, results do not substantially change.

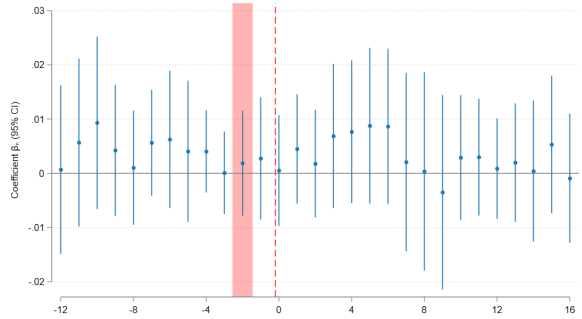
Figure C.6: MAIN RESULTS: THE INTENSIVE MARGIN OF THE SHOCK



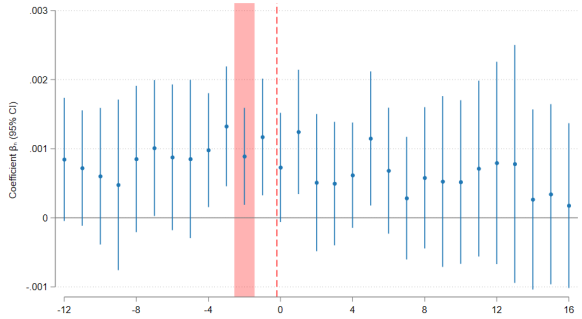
Notes: Figure C.6a and C.6b plot the estimated coefficients of equation (2) (see Section 4.1) when the dependent variable is respectively the level of sales (normalized by the mean-level of fixed assets) and the firm-level market-share defined at NAICS3 industry level. Figure C.6c and C.6d plot the estimated coefficients 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. Figure C.6e and C.6f plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

Figure C.7: Main Results: the Extensive Margin of the Shock

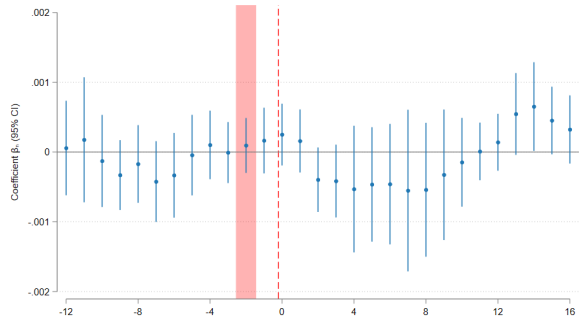
(a) Sales



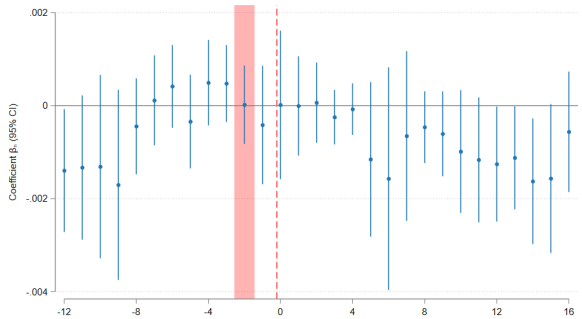
(b) Market Share



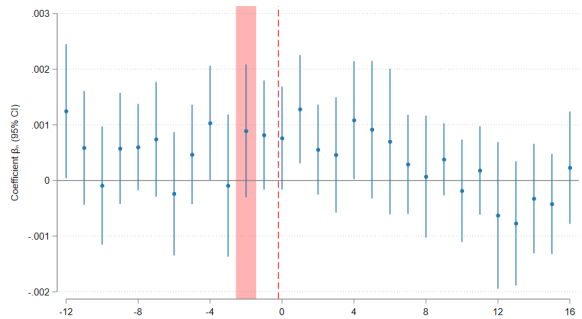
(c) R&D (Competitive Ind)



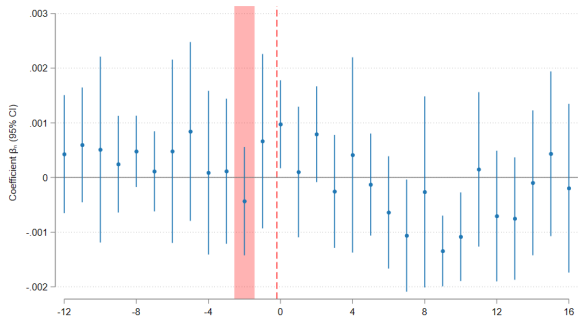
(d) R&D (Non-Competitive Ind)



(e) CapX (Competitive Ind)

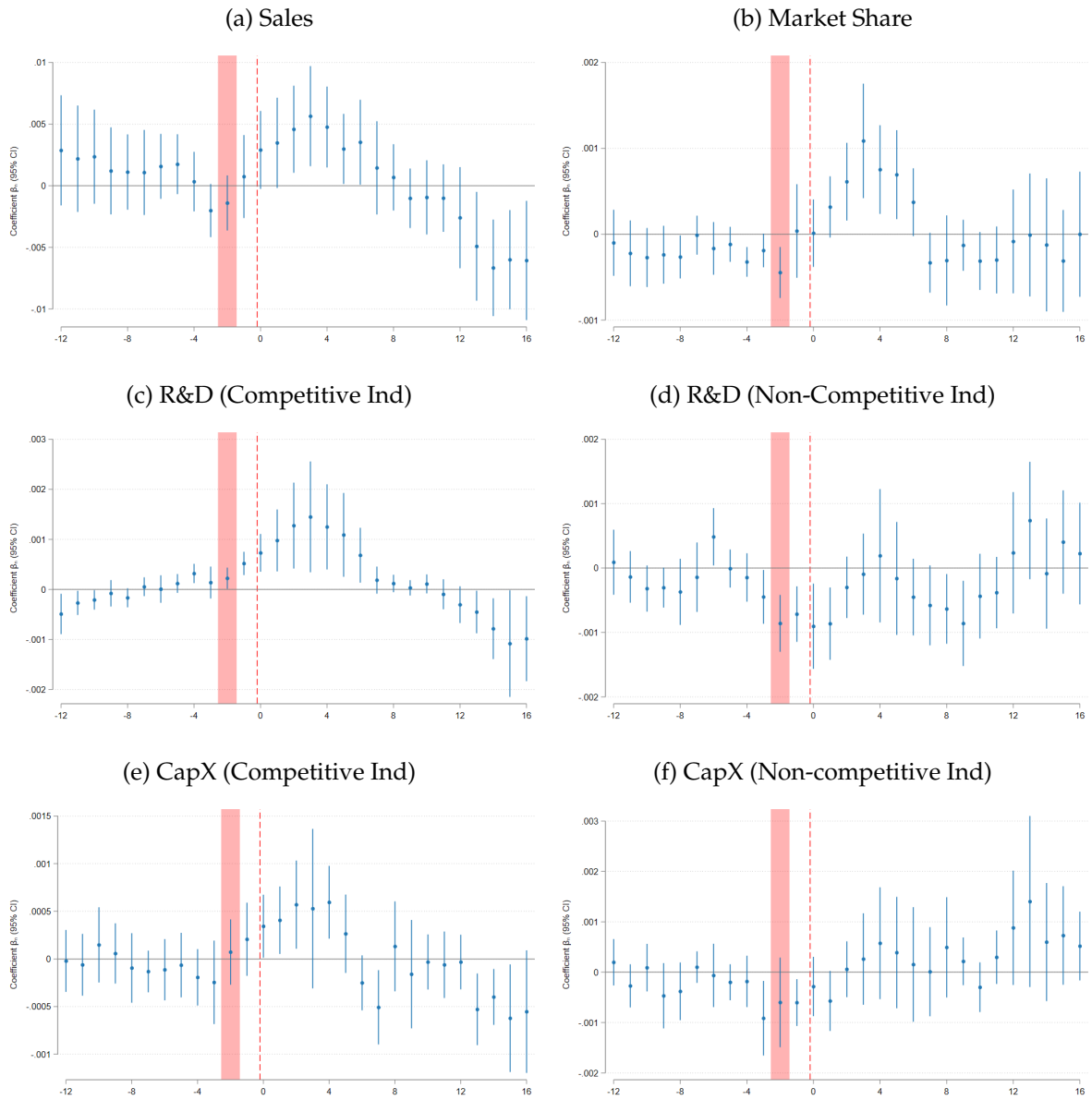


(f) CapX (Non-competitive Ind)



Notes: Figure C.7a and C.7b 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. Figure C.7c and C.7d plot the estimated coefficients 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. Figure C.7e and C.7f plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

Figure C.8: MAIN RESULTS UNDER OTHER DEFINITION OF THE SHOCK



Notes: Figure C.8a and C.8b 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. Figure C.8c and C.8d plot the estimated coefficients 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. Figure C.8e and C.8f plot the estimated coefficients 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. 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 in the gazette of the standards' organization.

D 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 $Shock_{i,t}$ is strictly positive, and as *followers* all the others. By doing so, we keep 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 viceversa). Given this, we proxy the capability of a sector to adapt to the new standard by taking the cross-firm mean of positive shocks for each quarter and industry. Then, we aggregate and construct the total amount of sales, patents, the total CapX expenditure, 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 D.1 shows descriptive statistics of the industry-level data. As from panel A, the mean-industry receives a shock 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 spend 51% of total CapX expenditure at the industry-level, 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 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 and CapX, research output (patents), and growth. Do do so, we consider the lead-lag model introduced in Section 4.1, but now defined for a industry-level panel dataset that aggregates firm-level variables. In practice, the model is now:

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

Table D.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 Share of Industry CapX	0.41	0.72	0.00	0.03	0.20	0.36	0.54	0.88	0.97	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
$\mathbb{I}$(*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 D for details on data construction.

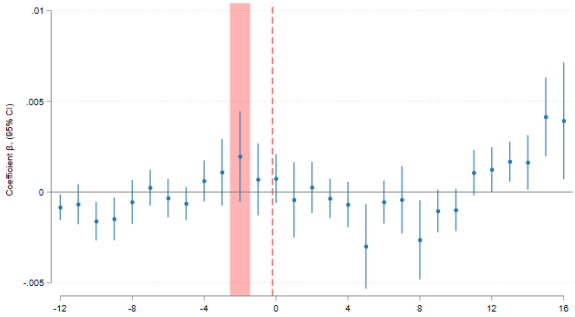
where $Y_{s,t}$ is the dependent variable for NAICS3 industry s at quarter t. $Shock_{s,t}$ is the mean shock of leaders in the industry. $X_{s,t-1}$ is the usual set of controls now defined at the industry-level (age, Tobin's Q, market capitalization leverage, a dummy for high-tech industries) described in panel C of Table D.1.

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

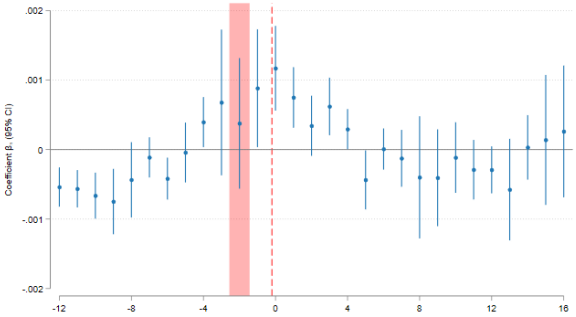
Figure 5 of Section 5 show results when we estimate model (D.1) with dependent variables being respectively the followers market share, their share of total expenditure in R&D and CapX, their share of the total research output (patents) in the industry. When considering this group-specific variables,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

Figure D.1: SECTORAL GROWTH AND TECHNOLOGY ADOPTION

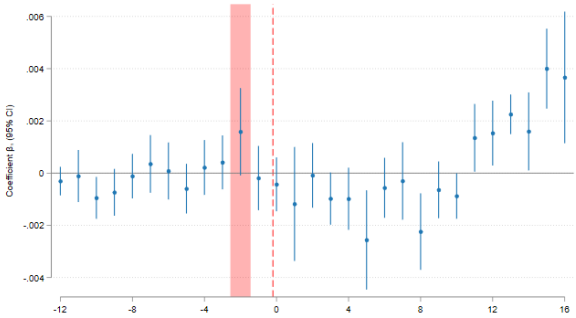
(a) Industry Growth



(b) Leaders Growth



(c) Followers Growth



Notes: Figure D.1 plots the estimated coefficients of equation (D.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 D 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 in the gazette of the standards' organization.