The effect of standardization on innovation
A machine learning approach

Maddalena Agnoli * Petyo Bonev †

May 10, 2019

Abstract

In this study, we estimate the effect of standardization on innovation. A major difficulty arises because innovation itself potentially impacts standardization, which leads to reverse causality. To deal with the resulting endogeneity, we apply machine learning methods to predict counterfactual innovation paths. Our identification strategy exploits unpredictable standards, i.e. standards that could not be foreseen by the market. For the corresponding technologies, we use innovation history to predict what the amount of patents would have been in the case of no standards. We use these predictions as counterfactual (no-treatment) outcomes to estimate the effect of the standards. We find a positive effect of standardization on subsequent patenting activity, but no effect on patent quality.

*MINES ParisTech (CERNA) and PSL Research University, maddalena.agnoli@mines-paristech.fr
†Swiss Institute for Empirical Economic Research petyo.bonev@unisg.ch

1 Introduction

We evaluate the effect of technology standards on innovation. Technology standards consist of a set of voluntary rules and requirements for a technological system and have the objective to ensure interoperability across products. Through their development process, and through the choice of one of many competing technologies, standards potentially impact future innovation activities. Identifying their causal effect on innovation, however, is a non-trivial task. The major problem is the reverse causality relationship between standardization and innovation. In particular, standards might arise precisely due to already existing innovation in a certain technology area.

We establish a novel identification approach to the deal with the endogeneity of standards. Our approach consists of three steps. In a first step, we build a prediction of whether a standard will be established in a given period, a given country, and a given technological class. The objective of this first step is to mimic the expectation formation of the firms using market information on the uncertain future events of establishing a standard.

In a second step, we construct predictions of counterfactual post-treatment (i.e. post standard release) innovation paths. An innovation path in a certain technology class is defined in this paper as the number of patent applications within the technology class followed over time. We restrict our attention to those technology categories, for which a standard is established in a given period despite a high predicted probability from our first step of no standard establishment. The intuition for this choice is the following. Consider a technology group, for which firms anticipate a high probability that no standard will be released. Then their innovation activities just prior to the event of the standard will correspond to those that firms would have exerted in the counterfactual no-standard scenario. The event of establishing a standard in that technology class can be viewed as a shock to the market. Thus, for those technology classes, we can use pre-treatment (i.e. prior to the standard) information to predict the future, post-treatment innovation path for the counterfactual no-treatment case. In particular, the pre-treatment information does not contain anticipation effects.

Steps 1 and 2 are generic in the sense that they can be constructed with prediction approach. In our empirical evaluation, we use machine learning methods - neural networks and random forests, respectively. These methods have been shown to deal well with large number of covariates and nonlinear model functions.

In a third step, we compare the actual innovation paths in the "shocked" technology categories to the predicted innovation paths. This comparison allows us to estimate the treatment effect of the technology shocks (the standards). The treatment effect is local in the sense that this is a treatment effect on a particular group of treated units: those, where no anticipation effects took place.

Our three-step approach complements existing econometric techniques. It relies on the assumption is that the information contained in our dataset correctly accounts for anticipation of standardization events. This assumption underlies the validity of the first step. The assumption is related to CIA- and conditioning-on-the-propensity-score-type assumptions (e.g. as in matching estimation). The major advantage of our approach is that we do not require common support in the covariates. Each (technology) unit serves as its own counterfactual match. The second step of our approach draws on the paper of Burlig et al. [2017] who also use past histories to construct counterfactuals. Our method accounts for the complex nature in which innovation activities are planned and implemented at the level of the firm. In particular, anticipation effects are likely to shift the paths of innovation already prior to treatment, invalidating the approach of Burlig et al. [2017]. Our initial prediction step accounts for the anticipation of the firms and ensures an unbiased prediction in the second step.

We contribute to the evidence on the effect of standardization on innovation in particular to DTI [2005], Swann [2010], Blind et al. [2017] and Layne-Farrar [2013].

Our results indicate a positive effect of standardization on subsequent patenting activity, but no effect on patent quality. Furthermore, our results do not support the technology selection theory of standards at the level of the economy.

The paper is structured as follows. Section 2 describes the institutional backgroud of standardization and the data, and section 3 our empirical strategy. Section 4 shows the results and is divided in two subsections. In section 4.1, we present the results of the first step of our methodology, the prediction of standard events, while in section 4.2, we create the counterfactual innovation path for unpredicted standards and estimate the effect of standardization on innovation. Section 5 concludes.

2 Institutional background and data description

2.1 A brief overview of the institutional setup

The ecosystem of standard setting organizations

Standards may be defined as "documents that provide requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose."[1] Another function of standards is the reduction in variety which leads to a selection among competing technologies (Tassey [2000], Blind and Jungmittag [2008]). Unlike regulations, standards are rules with no compulsory character, whose success depends solely on whether companies voluntarily decide to adopt them or not. Standards, therefore, can be understood as self-regulatory actions of an industry (Rysman and Simcoe [2008]).

Since standards are voluntary rules, they can be established by any company or group of companies. In practice, one can distinguish between four types of agents that establish standards. First, single companies might develop a standard on their own. Standards developed by single companies are referred to as "proprietary specifications" (Bekkers et al. [2014b]). The firm retains full control over the specification and its evolution, and the specification typically serves the particular interests of the firm. When the specification gains market success, it is referred to as a "de facto standard". An example here is the Video Home Standard (VHS) developed by the JVC.

More commonly, standards are established by formal standard developing organizations (SDOs). Depending on the scope of their standards, formal SDOs can be national or international. National SDOs are entities that are formally recognized by the regulators as standard developing organizations (Bekkers et al. [2014a]). They are "membership-driven bodies that bring together standardization experts - often from competing companies and from governments, academia and civil society - to develop standards in response to priorities determined by public- or private-sector members" (Bekkers et al. [2014b]). An example for a formal national SDO is the German Institute for Standardization (DIN). Standards can also be established by quasi-formal SDOs, which are very similar in terms of structure and status to the formal SDOs but do not have a formal recognition by the regulators. An example here is the the World Wide Web Consortium (W3C).

[1] International Organization for Standardization (ISO), What is a standard?, http://www.iso.org/iso/home/standards.htm

Finally, standards can be developed also by informal industry organizations called consortia (also fora or Special Interest Groups (SIGs)). A consortium consists of private sector members that share a common interest. It may limit the number of participants in order to achieve a more efficient and quick standard development process. Consortia may be formed for developing a single standard or for a broader scope.

The process of developing a standard

Developing and adopting a standard is a complex process that may take up to several years. Although the characteristics for this process exhibit a substantial heterogeneity across different SDOs, there are also common features. As an example, a consensus among the members of the SDO on the standard's scope and context must be reached before a standard is released. Thus, typically, a standard must pass some sort of ballot. In addition, most of the SDOs have (formal or informal) rules that concern intellectual property rights (IPR) on technologies necessary for the adoption of the standard (the most prominent example being here standard essential patents (SEPs)). These policies aim at ensuring fair conditions for SEP holders and applicants once the standard is adopted. Examples for such policies are ensuring transparency about and licences for SEPs, preventing patent hold ups, preventing too high cumulative fees, and many more (see Bekkers et al. [2014a] for a detailed discussion).

Uncertainty related to the process of developing a standard

In many cases there is a substantial uncertainty in the process of developing a standard. Although often a working group is setup by the SDOs to elaborate a draft proposal, the final outcome might not be foreseeable to the participants until the very end. One rather amusing example is the development of the Computer Graphics Reference Model by the International Standardization Organization (ISO), (ISO [1992]). An ad hoc group was set up by ISO to investigate the feasibility of creating a standard, and a year later, two competing approaches were established (see Rada et al. [1994] for a detailed description of this case). It took the working group three further years to realise that the first approach was a process-oriented view and the second a data-oriented one. The two approaches were subsequently merged into one.

Even after a draft is established, a substantial uncertainty still resides in the subsequent standard setting phase. One part of it is related to the negotiation process that reflects the complex interplay of (often conflicting) interests. An unsatisfactory ballot can result in a subsequent refinement of the standard before a final consensus is achieved. A second part of the uncertainty is due to the disclosure of SEPs by the patent holders. In particular, it may be the case, that even the SEP holders are not aware of all of their standard-relevant patents. Such a patent might however be discovered ex post. In addition, participants might not be aware of patents owned by third parties that have been disclosed. Finally, a participant might simply realize that a known SEP has a much higher value for the firm. All of these cases might lead to a revision of the standard (e.g. by using an alternative technology) or to a withdrawal of the standard altogether if the former is not feasible (Bekkers et al. [2014a]).

All these aspects of the standard setting procedure can make its outcome uncertain even for involved participants. Our empirical strategy relies on identifying this uncertainty.

2.2 Data sources

Data used for prediction of standards

To predict the timing of a release of a standard, as described in section 3.3, we use data from several sources.

First, we retrieve data on standards from the database Perinom. It contains information on national standards from 27 countries, as well as on European and global standards. We dispose of information on the publication date, the issuing SDO, the country, information on the content of the standards (such as the title, the abstract, the language), as well as the technological classification according to the International Classification for Standards (ICS) (a given standard can be categorised with a combination of several ICS classes). We can also track international relationships between standards, i.e. to which extent standards from different SDO's are related to each other and how similar they are. In particular, when releasing a standard, a SDO should publicly disclose if the standard is equivalent to some already existing one (or a modified version of it), a process which, ironically, is itself standardized by an ISO standard (ISO/IEC Guide 21). We use this information to determine which countries have implemented international and European standards, and to distinguish whether a country has developed or adopted a standard. We only use newly developed standards and exclude adopted ones in our analysis since we are interested in the effect of technological shocks to the market. Standard adoption can also affect innovation, however, the effect might not be comparable to the implementation of a new and unexpected standard. Furthermore, expectations about standard development and adoption might follow different patterns.

Second, for each of the 27 countries in the Perinom database, we extract GDP per capita, total population, R&D expenditure as a percentage of GDP, the mean tariff rate, the natural resource rent as a percentage of GDP, and the categorization of countries in high vs. low income countries (time varying) for each year from the World Bank's World Development Indicators (WDI) database. In addition, we also obtain trade data on country and product level from the United Nation's UNCTAD database.

The choice of these variables is motivated in section 2.5.

Patent data

We extract patent data from Patstat, a database of the European Patent Office (EPO) that collects and structures information on patents from 38 patent authorities worldwide. We observe detailed characteristics for each patent such as technology category, owner, inventor, filing date, and all kinds of changes in the patent's life time (renewal, withdrawal, etc.).[2] Our main dependent variable is the number of patent applications in a given period, country and technology area. This variable can be interpreted as a proxy for innovation. To account for patent quality, we also calculate the average number of forward citations to patents of a technology category.[3]

2.3 Linking data sources and sample definition

Standards and patents are classified with different technology classification systems. In particular, patents are typically classified according to the International Patent Classification (IPC) system and/or according to the Cooperative Patent Classification (CPC), whereas standards are classified according to their own system (the ICS). As of today, there exists no publicly available concordance table that links the IPC (or CPC) to the ICS categories. In order to identify the relevant patent technology classes for a standard, we use data on declared SEPs from the Searle Center database Database on Technology Standards and Standard Setting Organizations. This database combines data on standards similar to Perinorm with information on the SDOs themselves. A major ad-

[2] For a detailed description of the database see European Patent Office [2018].

[3] For patent counts and citations as proxies for innovative activities see for example Acs and Audretsch [1989], Hagedoorn and Cloodt [2003].

vantage of this database is that it contains information on SEP declarations and reports the patent identification numbers of the SEPs. These declarations make it possible to link the ICS classes of the standards to the International Patent Classification (IPC) numbers of the patents. SEP declarations do not exist for all standards, either because the standard does not include patented technologies or because the relevant patents are not declared publicly as SEPs. Furthermore, some argue that firms tend to over-declare, i.e. declare patents that are not really essential to the standard (Stitzing et al. [2017]). This limits our sample of standards to those where information on SEPs is available.

We only consider newly developed standards and exclude adopted standards. We use international relationships in order to link international or European standards to countries. The developing country is then the country that first adopted an international or European standard.

In order to link trade to technology classes, we use the concordance table provided by Lybbert and Zolas [2014]. This table links product categories from the Standard International Trade Classification (SITC) (used also in the UNCTAD database) to IPC classes. The matching is done through keyword searches in the patent documents and allocates probability weights to each SITC-IPC pair. We obtain trade data on the technology level by weighting exports and imports from the UCTAD database with these probability weights and linking them to standards through the ICS-IPC matches.

2.4 Descriptive statistics

Our final dataset contains 143,451 observations and 253 technologies (i.e. ICS combinations). Every observation is defined as a combination of a technology, a country and a year. We count 13,244 standardization events, i.e. about 9% of the observations. The scarcity of standard events makes prediction particularly challenging. A naive prediction of no standardization event for all years would already result in a prediction accuracy of 91%, i.e. 91% of cases are predicted accurately. Our data represents only a fraction of all available standards in Perinorm. This is mainly due to the ICS-IPC matching using SEP data which was not possible for all technologies. We also loose some standardization events by considering the period of observation from 1995 to 2015 which excludes older and some more recent standards. Figures 1 and 2 show the distribution of standardization events over the available years and countries. Figures 8 and 9 in the appendix show the same distributions for the whole set of standards in Perinorm for the same time period. These exclude international and European standards which have bee allocated to countries as described in section 2.3.

Figure 1: Number of standards over years
13,244 standards

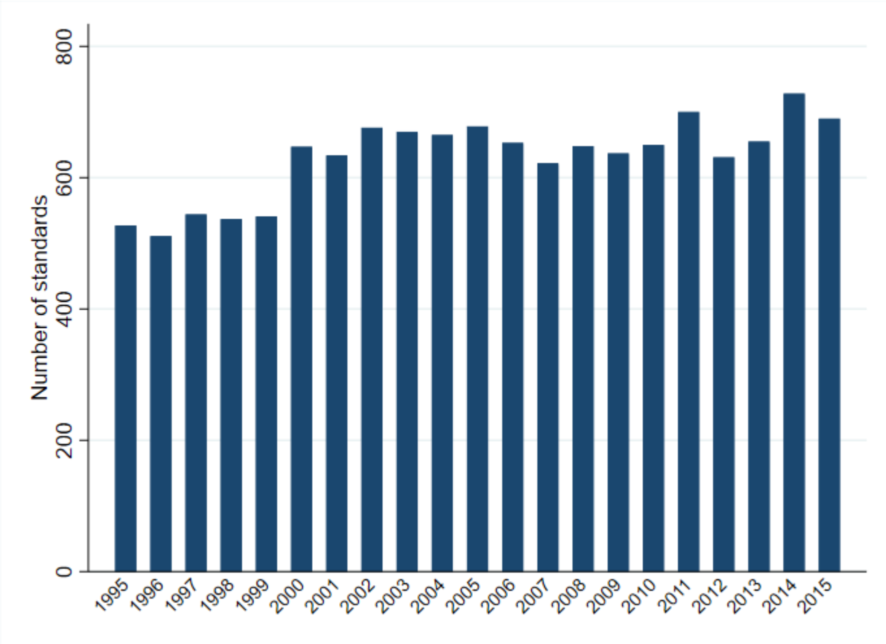


Figure 2: Number of standards over countries
13,244 standards

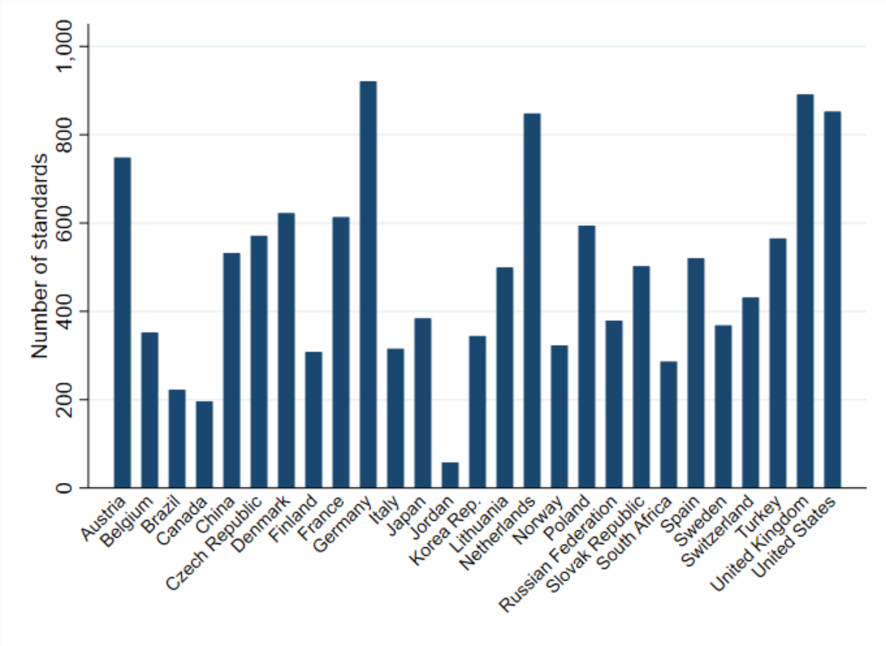


Table 1 presents descriptive statistics of all input variables for standard prediction as well as of the standardization event variable itself and our variables of interest, the number of patent applications and the average number of 5-years forward citations, which serve as our proxy for innovation. Furthermore, we show average values of all vairables by country in tables 15 and 16 in the appendix.

Table 1: Descriptive statistics

	Mean	Standard dev.	Min.	Max.	No. of observations
Standard event (0/1)	.0923	.2895	0	1	143451
Exports in tech. (mio. USD)	253	575	.0003	8784	143451
Imports in tech. (mio. USD)	234	500	.0025	6631	143451
Patent stock in tech. (thous.)	38	111	0	1049	143451
Total patent stock (mio.)	11	41	0	621	143451
GDP per capita (thous. USD)	32	21	1	92	143451
Total population (mio.)	99	244	3	1371	143451
Standard stock in tech. (thous.)	.6527	3	0	43	143451
R&D expenditure (% GDP)	.0179	.0088	-.0026	.0429	143451
Mean tariff rate	.0438	.0399	0	.2382	143451
Natural resource rent (% GDP)	.0179	.034	.0001	.2175	143451
Number of patents in tech. (thous.)	8	21	0	186	143451
Total number of patents (mio.)	2	8	0	108	143451
Number of patent citations	7	71	0	2880	132825
Cumulative number of patent citations	35	323	0	11776	132825
Average number of 5 years citations	91	499	0	7101	132825

Note: The unit of observation is on the country - technology category - year level.

Table 2 compares descriptive statistics of all input variables of years where a standardization event occurs in the following years with years where without standardization. Years preceding a standardization event are characterized on average by less exports and imports, smaller total and technology related patent stocks, a lower mean tariff rate and a lower natural resource rent. GDP per capita, population size, standard stocks within the technology area and national R&D expenditure are higher for these years. T-statistics for the mean comparisons are reported in the table. All differences are significant at the 95% confidence level.

Table 2: Descriptive statistics by standard and no-standard years

	Standard		No standard				
	Mean	Standard deviation	Mean	Standard deviation	Difference in means	t-statistic	No. of observations
Exports in tech. (mio. USD)	171	475	254	561	-82	-16	143451
Imports in tech. (mio. USD)	165	445	235	489	-69	-15	143451
Patent stock in tech. (thous.)	30	99	38	111	-9	-9	143451
Total patent stock (mio.)	8	35	11	41	-3	-9	143451
GDP per capita (thous. USD)	34	19	32	21	2	10	143451
Total population (mio.)	109	255	98	242	11	5	143451
Standard stock in tech. (thous.)	4	6	.2524	2	4	181	143451
R&D expenditure (% GDP)	.0182	.0084	.0177	.0089	.0005	6	143451
Mean tariff rate	.0358	.032	.0452	.041	-.0095	-25	143451
Natural resource rent (% GDP)	.0145	.0301	.0186	.0348	-.0041	-13	143451
Number of patent citations	6	60	7	74	-.6669	-.9661	143451
Cumulative number of patent citations	25	220	30	282	-5	-2	143451
Average number of 5 years citations	77	412	92	508	-16	-3	143451

In table 3, we compare our output variables by years with and without standardization event. More precisely, the table reports the average increase of patent applications and citations one to five years after the year of observation compared to the average number of patent applications and citations of the preceding five years. The number of patent applications increases for both, years with and without standard. However, the increase is on average higher if no standardization event happened. This naive comparison between treated and untreated years would lead to the conclusion that standardization reduces innovation. No difference can be found for patent citation. The differences are small and not significant at the 95% confidence interval. As we will see in section 4, our estimation method, which accounts for predictability of standards, suggests a positive relationship between standardization and patent applications. The simple comparison of means between treated and untreated is therefore misleading.

Table 3: Patent applications and citations - Before-after and Diff-diff

	Standard		No standard				
Years after treatment	Mean	Standard deviation	Mean	Standard deviation	Difference in means	t-statistic	No. of observations
Patent applications							
1 year	242	4938	414	5986	-171	3	129789
2 years	356	5612	614	6701	-258	4	122958
3 years	463	6219	803	7460	-341	5	116127
4 years	562	6858	1001	8271	-439	5	109296
5 years	678	7616	1181	9010	-503	5	102465
Patent citations							
1-year lead	3	61	2	67	.6945	1	132825
2-years lead	4	64	4	71	.3941	.587	126500
3-years lead	6	69	5	73	.5482	.7668	120175
4-years lead	7	70	6	74	.3908	.5209	113850
5-years lead	7	71	7	76	.0525	.0663	107525

Note: The table reports the difference between the average patent count 1-5 years after the treatment period and the lagged patent counts averaged over the 5 years preceding the treatment period.

2.5 Variable selection for standard prediction

To our knowledge, the likelihood of standardization has not been studied empirically yet. In order to discuss the predictability of standards, it is important to understand how standards are created. The first step of standardization emerges from the idea of one or more market participants. The initiators then have to find enough support for their idea and a standardization body to sponsor it. Standards are built on the beliefs and understanding of its authors about the market and are either created in anticipation of market changes or based on current practice. Anticipatory standards are especially implemented in sectors with short product life cycles such as the ICT sector (Cargill [2011]). Chiao et al. [2007] describe standard development as a process which often takes place at an early stage of the technology development. Firms can seek to obtain a comparative advantage by initiating the standardization process at this early stage. Cargill [2011] examines sources of standardization failure at different stages of standardization and argues that in the very early stage standardization can fail due to a lack of interest of market participants to standardize or to bear the costs of standardization. Another early source of failure is disagreements between different parties, notably about intellectual property rights. Furthermore, standardization is influenced by the innovative activity within the technology area. Chiao et al. [2007] indicate, for example, that firms devote a lot of effort to the standardization process by making important investments in R&D. Loyka and Powers [2003] discuss factors influencing global product standards and relate it to market, industry and company factors. Market factors describe country specific aspects such as consumer characteristics, economic development and infrastructure. Industry factors include market structure, product and production particularities, competition and technological aspects. Company factors relate rather to the adoption of product standards within the companies. He also argues that standardization becomes necessary as an economy develops due to the increasing complexity of the society and the industrialization of the economy. Moreover, the international trade literature has identified standards as potential barriers or promoters of international trade (Hallikas et al. [2008], Biddle et al. [2012], Chiao et al. [2007], Swann [2010], World Trade Organization). Standards are often implemented in response to the countries' position in the international market space and frequently create tensions between the developed and the developing world due to differences in adoption costs and an unequal distribution of intellectual property rights (Gibson [2007], Ernst [2011]). It is also worth noting that standardization is a costly process and requires a certain institutional structure. Only very few low income countries dispose of a standardization body.

Our input variables for standard prediction include technology related exports and imports in order to capture international trade effects on standardization. Macroeconomic development is captured by GDP per capita, population and a high income dummy.

Technology related and total patent stocks as well as R&D expenditure as a percentage of GDP capture the importance of intellectual property rights and innovation for standardization. We furthermore include the age of the technology and the number of existing standards worldwide within the technology category in order to control for the anticipatory character of the standards. Finally, we include country and year dummies in order to capture trends in time and space.

3 Empirical strategy

3.1 Microeconomic foundations

We use a standard on the (ethical) use of Artificial Intelligence (AI) in industrial application to motivate the first approach of our empirical strategy. A proposal for such a standard is currently under development by an expert committee of ISO (the ISO/IEC JTC1 committee). Consider a firm that can spend in period $t = 0$ a total of 1 on R&D activities. The firm can adopt a technology that optimally prepares the firm for a standardization event in $t = 1$. As an example, the firm could hire researchers trained in developing certain types of algorithms. Or it can change those of its current models that use a certain type of information (e.g. race) in order to make them "more ethical". Adopting this technology is costly and the cost equals $C \in [0, 1]$. When a standardization event occurs in $t = 1$ and the firm is properly prepared, the R&D activities of the firm yield a return of ρ_1. When there is no standardization event in $t = 1$ and the firm prepares in $t = 0$, the R&D activities yield a return of ρ_2. Finally, when the firm does not prepare and there is an event, the R&D activities yield a return ρ_3. We assume that $\rho_1 > \rho_2 > \rho_3$ and set $\eta_i := 1 + \rho_i$ for $i = 1, 2, 3$. The motivation behind this assumption is that if the firm correctly predicts a standardization event, its preparation might give it an early-adopter advantage, e.g through developing patents on standard-based algorithms.[4] Let the firm-specific discount rate be τ and the (possibly subjective) probability for a standardization event in $t = 1$ (as seen from $t = 0$) be p. When the firm adopts a technology as a preparation for a future standard, its expected profit in $t = 0$ is

$$\Pi_{0,S} = -C + \frac{p(1 - C)\eta_1 + (1 - p)(1 - C)\eta_2}{1 + \tau}. \tag{1}$$

[4] In Europe, an algorithm can be patented only if it is a part of mixed-type invention, which also solves a technical problem in an innovative way (IAM [2018]).

In the case of no preparation, the expected profit of the firm in $t = 0$ is

$$\Pi_{0,N} = \frac{p\eta_3 + (1-p)\eta_2}{1+\tau}. \tag{2}$$

The firms adopts a standardization technology iff

$$\Pi_{0,S} > \Pi_{0,N}, \tag{3}$$

which is equivalent to

$$p > \frac{C(1+\tau+\eta_2)}{(1-C)\eta_1 + C\eta_2 - \eta_3} =: \bar{p} \tag{4}$$

Thus, if the probability for a standard is lower than a threshold $\bar{p}$, the firm behaves in $t = 0$ as if there will be no standard in $t = 1$ (namely, it does not invest in a future standard). We refer to this case as "No anticipation".

The idea of our identification strategy is to exploit this decision rule in the following way. Suppose that we can estimate the probability p. If we knew $\bar{p}$, then we could isolate all the cases in which there was a surprise for the firm, i.e. it decided not to invest in standardization technology and there was a standard or vice versa. Using these surprises yields a source of identifying variation. In particular, one can use the non-prepared trajectory of patents until $t = 0$ in order to predict how many patents there would have been in $t = 1$ had there been no standardization event. This prediction can be interpreted as a counterfactual post-treatment patent trajectory.

There are two pitfalls related to this strategy. First, in our paper we consider innovation on a national level, and not on a firm level. This problem could be solved by considering all firms in the given technological area of the national market. In particular, if we knew all thresholds of the firms operating on this market, we could either aggregate the procedure firm by firm, or simply pick the highest threshold in a given period (and use considerations identical to those in the previous paragraph applied to this highest threshold).

Second, $\bar{p}$ is not known to the researcher. Estimating it would involve substantial assumptions on the future profits of the firm, which are hard to be elicited from the data particularly in the case of standardization. We therefore choose the threshold $\bar{p}$ with the highest prediction accuracy in our main results and analyze the sensitivity of our results with respect to changes in $\bar{p}$.

3.2 Econometric framework

We cast our econometric problem in the Rubin Causal Model framework (Rubin [1974]). Denote by D_{it} the random standardization indicator for time t and technology category i, $i = 1, \ldots, n$, $t = 1, \ldots .T$, where $D_{it} = 1$ denotes the event "A standard is introduced" (we omit the country index for simplicity). Define $Y_{i,t}(d)$ to be the potential outcome of interest in period t and technology i when the treatment is equal to $d \in \{0, 1\}$. For simplicity of exposition, assume that the standards are introduced at the beginning of a period and the outcome is realized at the end of the same period. The notation can be generalized to a multi-period gap between treatment and outcome in a straightforward way.

We are interested in estimating the average causal effect

$$ATE = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)]. \tag{5}$$

However, for each t and i, only one of $Y_{i,t}(1), Y_{i,t}(0)$ is observed. This problem is referred to as the Fundamental Problem of Causal Inference (Holland [1986]).

Our approach identifies a conditional version of equation (5) in three steps. The first step aims at isolating those standards that surprised the market. We pursue this step by using a rich dataset to predict whether a standard will be released or not. In particular, let $\mathbb{F}_l$ be the information at some point in time l that agents can use to estimate the propensity score $p_{i,t} = P\{D_{i,t} = 1\}$ for technology i at time t. Denote the estimate with $\hat{p}_{i,t}$. We assume that market participants form a prediction $\hat{D}_{i,t}$ for $D_{i,t}$ using a simple Bayes classifier:

$$\text{Set} \quad \hat{D}_{i,t} = 0 \quad \text{if} \quad \hat{p}_{i,t} \leq \bar{p} \quad \text{and} \quad \hat{D}_{i,t} = 1 \quad \text{if} \quad \hat{p}_{i,t} > \bar{p}, \tag{6}$$

where $\bar{p}$ is a threshold probability.[5] We define the set of standards with $\hat{D}_{i,t} = 0, D_{i,t} = 1$ to be the Non-Anticipated standards (NA).

This definition of NA-standards has two advantages. First, the actual implementation is straightforward. The researcher can use either standard econometric classification approaches such as logit or Machine Learning techniques. We discuss the empirical implementation in subsection 3.3 below. Second, this definition of a missclassified standard

[5] The standard Bayes classifier uses $\bar{p} = 0.5$.

is closely related to the microeconomic discussion from the previous section. Market participants are "surprised" by the standard, in the sense, that prior to the standard they behave as if no standard will be released.

In a second step, we predict the outcome variable for all standards in the NA group using only pre-treatment (i.e. pre-standard) characteristics, including pre-treatment outcomes. This step and the following step are borrowed from the paper by Burlig et al. [2017]. Denote the predicted outcome for technology i and period t by $\hat{Y}_{i,t}(0)$. The motivation for this step is that for the NA-set of standards, the history $(Y_{i,l}, X_{i,l})_{l\leq t-}$ does not contain anticipatory effects and can be used to construct an unbiased prediction for the counterfactual non-treatment outcome. The notation of the prediction reflects this assumption by including an indicator for the potential outcome.

In a final third step, each outcome in the NA-group is compared to its predicted counterfactual. The resulting estimator is defined as

$$\hat{\beta}^T = \mathbb{E}[Y_{i,t}(1) \widehat{- Y_{i,t}}(0) \mid NA] =\mid NA \mid^{-1} \sum_{i,t}(Y_{i,t} - \hat{Y}_{i,t}(0)), \tag{7}$$

where $\mid NA \mid$ is the number of observations in the NA group.

Before we discuss the actual implementation of steps 1-3, we briefly discuss the main underlying assumptions through a comparison of the estimator to the standard matching on the propensity score. Both, equation (**??**) and the matching estimator rely on estimating the propensity score. However, the matching estimator crucially relies on a common support assumption, which ensures finding similar treated and non-treated units. Our approach, on the contrary, builds for each unit in the NA group its own counterfactual prediction. The two crucial assumptions behind our estimation approach are (i) that the information available to the econometrician is sufficient to identify the NA group and (ii) the counterfactual predictions for this group are unbiased. Assumption (i) is similar in spirit to the CIA assumption invoked by the matching estimator. It is a non-testable assumption. Assumption (ii) can be defended in a way similar to defending the parallel trend assumption used in a DID estimator: by predicting pre-treatment outcomes based on their histories

Predictions might not follow the exact same pre-treatment path as actual outcomes, but follow a parallel trend. In order to take this into account, the prediction error on pre-treatment innovation can be subtracted:

$$\hat{\beta}^{TD} = \mathbb{E}[Y_{i,t}(1) \widehat{- Y_{i,t}}(0) \mid NA] - \mathbb{E}[Y_{i,t-\epsilon}(1) \widehat{- Y_{i,t-\epsilon}}(0) \mid NA] \; , \tag{8}$$

where $t - \epsilon$ denotes some pre-treatment period.

Although our estimation method does not depend on the selection of an untreated control group, i.e. a comparable sample without standardization, we follow Burlig et al. [2017] in randomly selecting untreated observations. Untreated means that no standard has been released, but also that no standard was predicted by the model. Yet, the timing of a standard event cannot be defined for untreated units. Burlig et al. [2017] propose a solution that consists of randomly assigning a treatment date to those units. We decided to repeat the random selection of untreated years 100 times and use average counterfactual outcomes in order to avoid that the estimated effects are due to the specific random sample. The comparison with this control group allows us to control for global trends and shocks that could lead to prediction errors in the whole sample. Just as in equations (7) and (8), we can calculate $\hat{\beta}^{U}$ and $\hat{\beta}^{UD}$ for this control group. We obtain two additional measures of the treatment effect that account for general trends across country-technology pairs, the difference-in-differences estimator

$$\hat{\beta}^{DD} = \hat{\beta}^{T} - \hat{\beta}^{U} \tag{9}$$

and the triple difference estimator

$$\hat{\beta}^{3D} = \hat{\beta}^{TD} - \hat{\beta}^{UD} \; . \tag{10}$$

3.3 Empirical implementation

For the first step, we use a neural network with one hidden layer in order to predict the occurrence of a standard in technology category i at time t in a given country. It has been shown that in many cases one hidden layer is sufficient for an accurate prediction due to the universal approximation theorem that states that any continuous function can be approximated using a feed-forward neural network with a single hidden layer and a finite number of neurons under mild assumptions on the activation function. Neural networks have been used increasingly for classification problems. One advantage of neural networks is that they allow for complex non-linear relationships between the input variables and the output without imposing a specific functional form ex-ante. The neural network implicitly selects the most useful variables for prediction among the input variables by allocating weights to each variable in the input layer. Furthermore, each neuron in the hidden layer obtains also a weight which can introduce non-linearity in the prediction model. There is a trade-off between the number of neurons in the hidden layer and the calculation time until convergence. We vary the number of neurons in the hidden layer in order to evaluate the sensitivity of our prediction results with respect to the number of neurons. The model issues a prediction value between zero and one which can be understood as the probability of a standardization event to occur in t. Prediction accuracy is defined as the ratio between correctly predicted realization to the number of realizations. Correctly predicted events include true positives, i.e. years with a standard event that have been predicted correctly, and true negatives, i.e. years without a standard event which have been predicted as zeros. False positives and negatives represent false predictions. In order to decide whether the prediction of a standard event is set to zero or one, we have to decide on the threshold $\bar{p}$. We choose the threshold that leads to the highest prediction accuracy as a baseline and conduct a sensitivity analysis relative to the choice of the threshold. In particular, we compare our results with a the ones of a very low threshold which we arbitrarily set to 10%. With a very low threshold NA-standards are considered extremely unlikely by the prediction model and can therefore be considered more confidently as actual shocks to the market.

In the second step, we use random forests in order to construct counterfactual innovation paths. Random forests are increasingly popular methods for both, classification and regression problems. They combine a multitude of decision trees in order to improve out-of-sample predictions. Decision trees are prone to overfitting, i.e. to match training samples to closely and can therefore lead to poor out-of-sample predictions. In other words, they lead to low-bias, but high-variance predictions. Random forests reduce variance by randomly selecting a subset of input variables for each tree (therefore reducing the risk of growing too strongly correlated decision trees) and averaging the predictions of the different trees. In our model, we set the number of decision trees to 100. Our random forest uses a bagging algorithm for model averaging, i.e. the algorithm generates a number random subsamples with replacement and averages prediction output over these samples. Decision trees are grown deep, i.e. potentially fit the data in the respective subsample very well. The reduction in variance is achieved through bagging and the random selection of input variables in each decision tree.

4 Results

4.1 Standard prediction

We randomly select 20 percent of our data as the training sample and 80 percent as the validation sample. Standardization events are predicted one to five years ahead. The dataset consists of a panel of 253 technology groups and 27 countries between 1995 and 2015 and contains 143,451 observations.

Figure 3 shows the prediction accuracy for the different number of neurons in the hidden layer of the neural network (predictions from period -1). A higher number of neurons in the hidden layer leads to a better prediction accuracy. This is the case for all prediction leads. This reflects the complex non-linear relationship between the input variables and the output. In figure 5 we show the prediction accuracy by prediction lead, i.e. for how many years ahead the standardization event has been predicted. The maximum accuracy is 94.4% (the values of figure 5 are presented in table 17 in the appendix). Here, we use predictions made with 15 neurons in the hidden layer, since they lead to the highest prediction accuracy. The prediction accuracy is very similar for the different leads, but becomes slightly better closer to the standardization event. The prediction accuracy is generally the highest at a prediction threshold of 0.5-0.55. The predictions are better than a naive predictor of setting predictions to 0 for all periods which would lead to a prediction accuracy of 90.8% because of the scarcity of standardization events. Using a prediction lead of 1 year, 15 neurons in the hidden layer and a threshold of 0.5, we are able to correctly predict 98.1% of all non-standardization events (true negatives), and 57.9% of standardization events (true positives). Compared to the naive predictor, we loose 1.5% of possible true negative predictions, but are therefore able to predict more than half of the standardization events correctly.

Figure 3: Accuracy of standard prediction by number of neurons

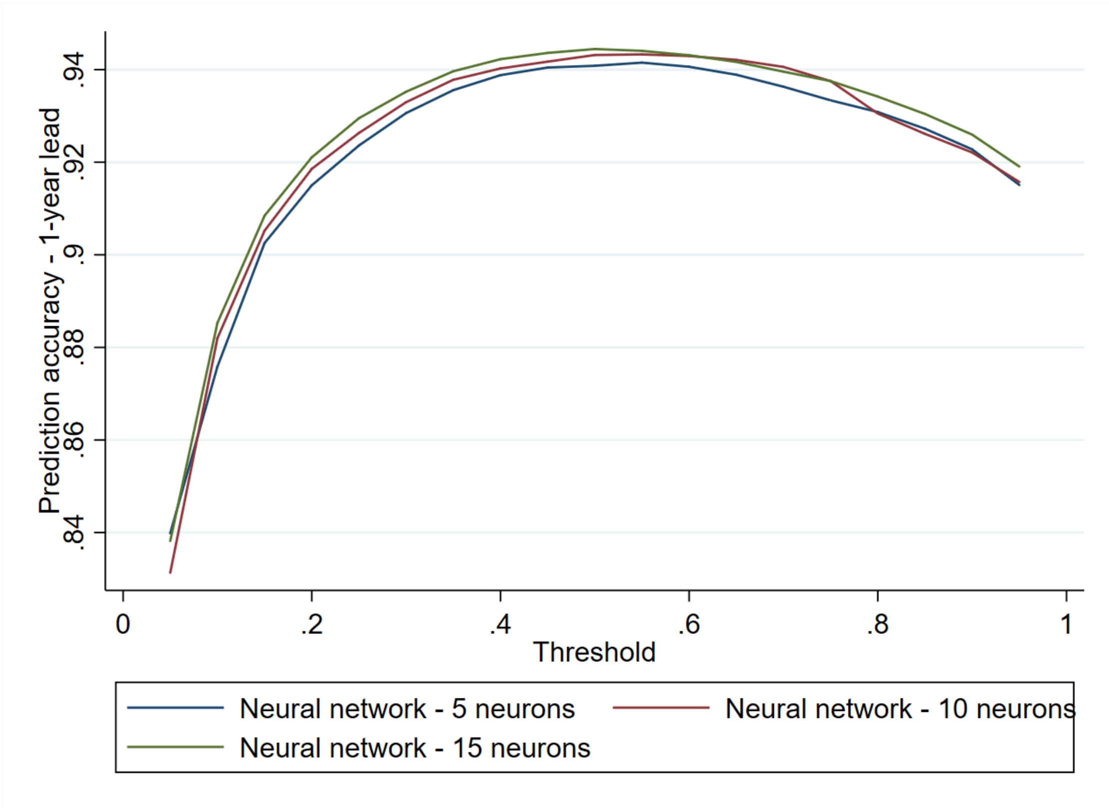
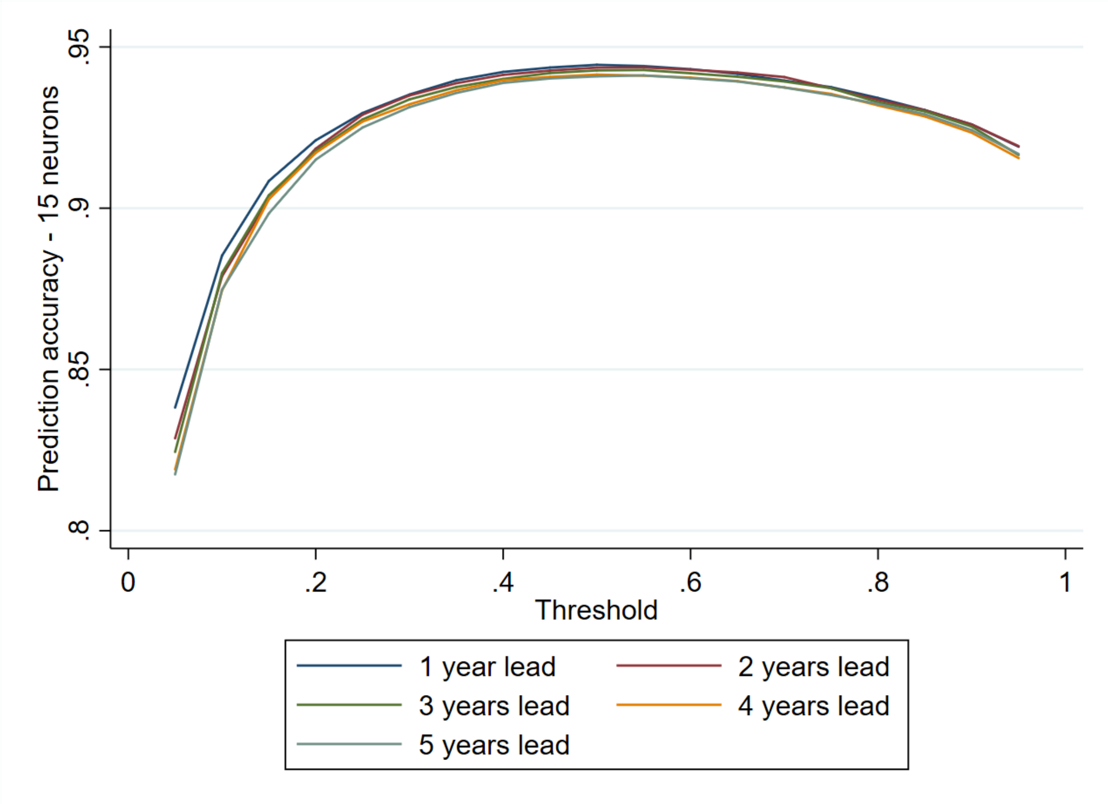


Figure 4: Full sample

Note: Neural network with 1 hidden layer for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions 1 year ahead, i.e. using inputs from $t-1$. Accuracy = true predictions/number of observations.

Figure 5: Accuracy of standard prediction by prediction lead



Note: Neural network with 1 hidden layer with 15 neurons for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions x=1,...,5 years ahead, i.e. using inputs from $t - x$. Accuracy = true predictions/number of observations.

Table 4 reports the prediction accuracy as well as the share of true positives and negatives with a decision threshold of 0.5. More than half of all standards are predicted accurately. As shown in table 18 in the appendix, the prediction accuracy is very similar for the training and test samples. Our neural network outperforms predictions obtained by simple regression, probit or logit models. The results are presented in tables 19 to 21 in the appendix. Figure 10 plots the prediction accuracy of the logit model for the different prediction leads. The prediction accuracy and the number of correctly predicted negatives are only slightly lower, but all models do worse in predicting standard events (true positives) then the neural network. The share of correctly predicted positives ranges from 29 to 43 percent depending on the model, while it ranges from 55 to 58 percent for the neural network.

Table 4: Accuracy of standard prediction by treatment outcome

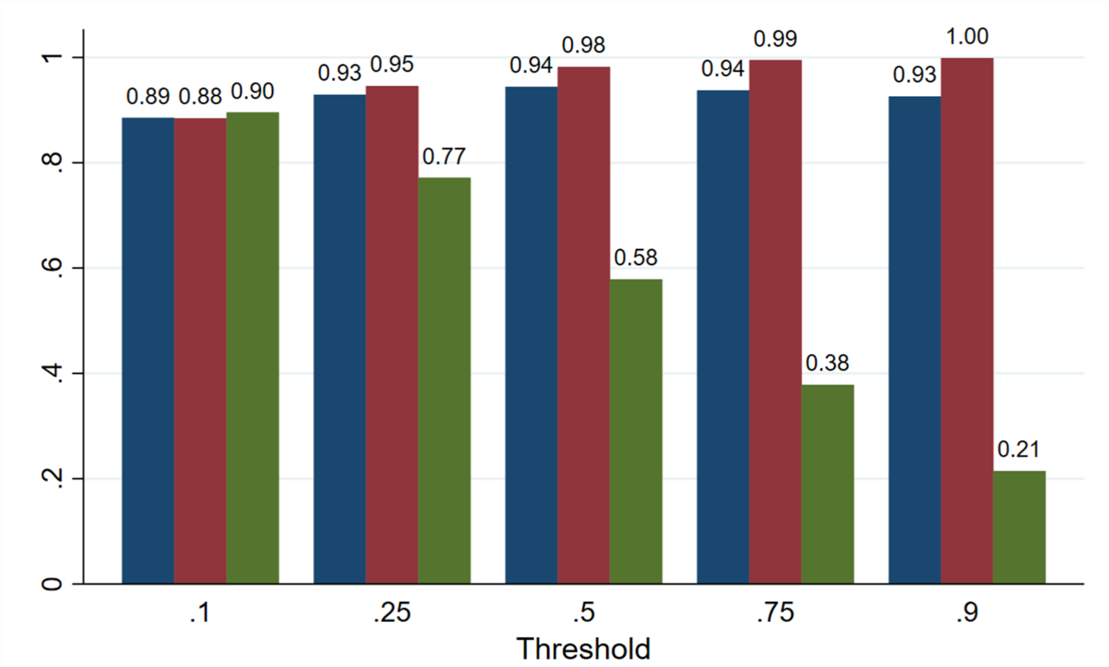
		Full sample		
Prediction lead	Accuracy	True positives	True negatives	Number of obs.
1 year	.944	.579	.982	136620
2 years	.944	.569	.982	129789
3 years	.943	.562	.983	122958
4 years	.941	.562	.982	116127
5 years	.941	.548	.983	109296

Note: Neural network with 1 hidden layer with 15 neurons for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions x=1,...,5 years ahead, i.e. using inputs from $t - x$. Accuracy = true predictions/number of observations.

Prediction accuracy is maximal at a threshold of 0.5. However, the choice of the threshold is arbitrary and implies an assumption about how market participants make predictions about standard events. Prediction accuracy measures the total share of correct predictions, i.e. gives the same weight to correct predictions of the occurrence and the absence of standardization events. Risk-averse market participants who want to avoid investing in the wrong technology might fear false positive predictions more than false negatives and might, therefore, choose a higher threshold. On the other hand, our analysis is based on the assumption that the selected standards are truly unpredictable by the market. This might especially be the case for very unlikely standardization events, i.e. where the predicted probability of a standard to occur is very low.

For these reasons, we conduct a sensitivity analysis when constructing the innovation counterfactual with respect to the decision threshold above which a standardization event is assumed. Figure 6 illustrates the prediction accuracy (share of correctly predicted events), as well as the percentage of correctly predicted positives and negatives by decision threshold. The percentage of correct positive (negative) predictions decreases (increases) mechanically with the threshold.

Figure 6: Standard prediction by decision threshold



Note: Neural network with 1 hidden layer with 15 neurons for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions 1 year ahead, i.e. using inputs from $t-1$. Accuracy = true predictions/number of observations. Well classified positives = correctly predicted years *with* standardization event. Well classified negatives = correctly predicted years *without* standardization event.

For the following analysis of the causal effect of standardization on innovation we use standard predictions from the neural network with 15 neurons in the hidden layer which led to the best prediction results.

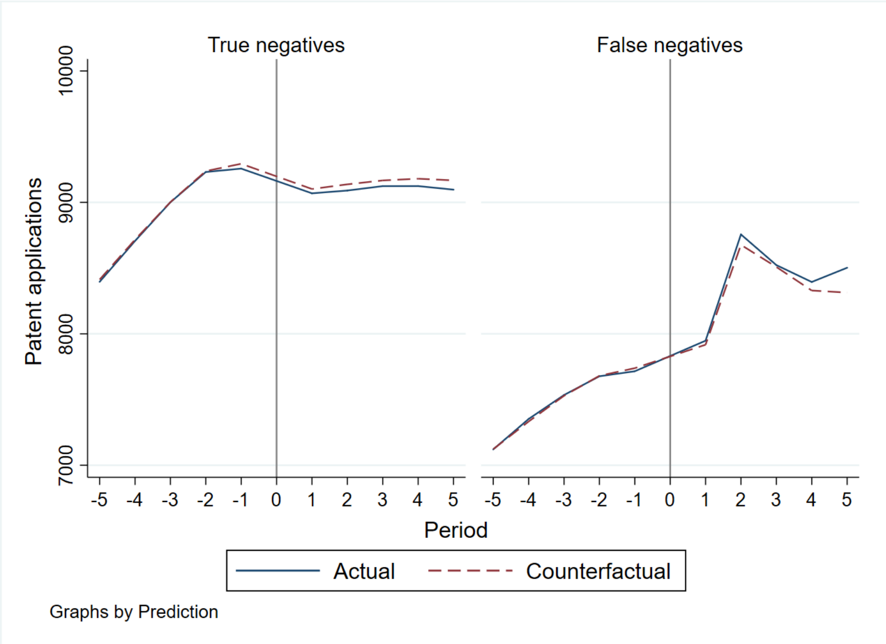
4.2 Counterfactual innovation and the causal effect of standardization on innovation

In this section, we present our findings on the effect of standardization on innovation. We use patent application counts to proxy innovative activity and compare our findings with the standards' effect on forward citations to patents of a technology and the share of a technology's patent applications in the country's total applications in order to measure different aspects of innovation. Patent counts measure the overall patenting activity within a technology class. Forward citations are often used to measure patent quality, since patents can be of very different quality and importance to an industry. The technology's share of patent applications refers to its importance in the national market and shows whether innovation efforts are shifted towards or away from a standardized technology.

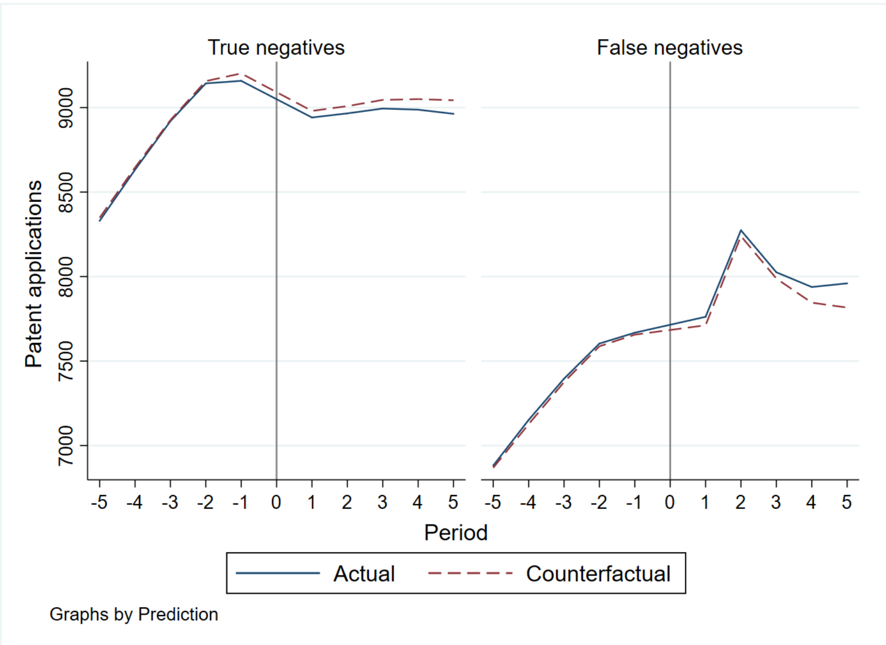
We predict counterfactual innovation paths for five years before and after the treatment period and compare them with actual innovation paths. We also test for differences and parallel trends in pre-treatment counterfactual and actual innovation paths. Predictions are made for false negatives and a control group that consists of average predictions of 100 random untreated samples, i.e. true negatives (see section 3.2). We identify all years with an unexpected standard, i.e. all years where no standard has been predicted, but a standardization event occurred (false negatives), as our treatment group. We use the same pre-treatment variables to create the counterfactual innovation path as have been used for standard prediction. Since our prediction model was not able to predict these standards, those variables do not contain information on the standard itself and can therefore be used to create a counterfactual situation of innovation without standardization. Only pre-treatment (i.e. pre-standard) information is used, i.e. prediction inputs from periods -1 to -5. Period 0 represents the year of the standardization event for false negatives and the randomly selected pseudo-treatment period for the control group. The results below use standard predictions one year ahead (i.e. standard predictions made with inputs from period -1). Results are similar for other prediction leads.

Figure 7 shows average actual and counterfactual patent application counts for the thresholds 0.5 and 0.1. The latter represents a very low threshold, i.e. false negatives include only very unlikely standards. The lower the threshold, the more unlikely have been the unexpected standards, i.e. the bigger the surprise of a standardization event to happen. The figure shows that pre-treatment patent applications follow a very similar path as the counterfactual predictions, while post-treatment counterfactuals deviate from actual outcomes.

Figure 7: Actual vs. counterfactual patent application counts



(a) Threshold = 0.1



(b) Threshold = 0.5

Note: Estimation of counterfactual patent counts using random forest with 100 decision trees. Period 0 = treatment or pseudo-treatment period. For false negatives, predictions are averaged over 100 random control group draws.

The figures suggest that patent applications are higher after an unexpected standardization event than the would have been without the standard. For true negatives the counterfactual patent application path seems to be slightly higher than the actual path after the pseudo-treatment period. In tables 5 and 6 we calculate the different treatment effects discussed in section 3.2. The treatment group (T) refers to all years with false negative predictions, i.e. years with a NA-standard. The control group consists of the randomly selected false negatives, i.e. years without standardization where no standard has been predicted by the model (U). β^T and β^U are the simple differences between actual and counterfactual patent application counts. β^{TD} and β^{UD} represent the DID estimators within each group where pre-treatment outcomes are averaged over the five years preceding the treatment period 0. β^{DD} and β^{3D} calculate the difference in the post-treatment differences as well as the triple-differences estimator between false and true negatives. The columns refer to the years after the treatment period. The results show that patent applications are higher than they would have been without standardization. However, for true negatives the treatment effects are negative, which suggests that our prediction model does not perfectly predicted patent applications for the control group. The difference in post-treatment differences and the triple-differences estimators show that ignoring the prediction error in the control group would lead to an underestimation of the positive effect of standardization on patent applications.

Table 5: Treatment effects of standardization on patent applications (threshold 0.1)

	Post-treatment period in years					
	1	2	3	4	5	Average
β^{T}	30.74	79.29	14.49	65.18	188.99	75.74
β^{TD}	31.55	80.1	15.3	65.99	189.8	76.55
β^{U}	-32.83	-47.14	-43.21	-56.19	-70.47	-49.97
β^{UD}	-18.5	-32.8	-28.88	-41.86	-56.13	-35.63
β^{DD}	63.57	126.43	57.71	121.37	259.46	125.71
β^{3D}	50.05	112.91	44.18	107.85	245.93	112.18

Note: Estimation of counterfactual patent application counts using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period.

Table 6: Treatment effects of standardization on patent applications (threshold 0.5)

	Post-treatment period in years					
	1	2	3	4	5	Average
β^{T}	49.96	34.39	36.96	92.82	143.45	71.52
β^{TD}	32.79	17.22	19.8	75.66	126.29	54.35
β^{U}	-38.79	-42.82	-51.16	-62.43	-80.19	-55.08
β^{UD}	-19.79	-23.81	-32.15	-43.42	-61.18	-36.07
β^{DD}	88.75	77.2	88.12	155.25	223.64	126.59
β^{3D}	52.58	41.03	51.95	119.08	187.47	90.42

Note: Estimation of counterfactual patent application counts using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period.

The above calculated treatment effects do not tell us anything about the significance of the effects. Furthermore, the DID estimators rely on a crucial assumption, the assumption of pre-treatment parallel trends. In order to test this, we estimate a DID regression model where we regress patent applications on the period, the treatment and their interactions (see Pischke [2005]). The periods refer to the years around the standardization event or the pseudo-treatment period, where 0 represents the treatment period. We use the year before treatment (period -1) as our reference period. The treatment indicator is 1 for actual realizations of patent applications and 0 for their predicted counterfactuals. This setting allows us to test for two things. First, we are able to test whether pre-treatment counterfactual paths are parallel to actual paths. If pre-treatment trends are parallel, the interaction terms between the treatment indicator and the pre-treatment periods should be insignificant, i.e. the difference between actual and counterfactual outcomes does not vary over time before treatment, or in other words, is not significantly different from the reference period. Second, we can test whether there is a significant difference in trends after treatment, i.e. whether the DID estimator is significant. Since we have several post-treatment periods, we are also able to evaluate how the effect of standardization on patent applications evolves over time. Note that the coefficients of the interaction terms correspond to $\hat{\beta}^{TD}$ and $\hat{\beta}^{UD}$ (see section 3.2).

Tables 7 and 8 present the regression results for true and false negatives. Pre-treatment trends are parallel for false negatives, but not for true negatives. For false negatives, the DID estimator is significant and positive in period 5. This means that patent applications are higher five years after the unexpected standardization event than they would have been had no standard occurred. Depending on the threshold, they exceed the counterfactual by 213 or 132 patent applications. The DID estimator for false negatives is valid since pre-treatment trends are parallel. However, it cannot account for eventual global prediction errors that affect our whole data sample, i.e. also true negative predictions. A prediction model that only excludes information of the standardization event itself should show parallel trends before and after treatment for true negatives. Since counterfactual predictions deviate from actual patent applications for true negatives, we have to account for this when calculating the effect of standardization on patent applications. Since pre-treatment trends are not parallel for true negatives, the DID estimator is not correct for this sample. However, for our triple-difference estimator it is important that the trends in the difference between actual and counterfactual patent applications of false and true negatives are parallel before treatment.

Table 7: DID estimation for patent applications within groups (threshold of 0.1)

	False negatives		True negatives	
T	-24	-24	-37***	-37***
Period -5	-616***	870	-881***	53***
Period -4	-406***	612	-577***	30***
Period -3	-210*	323	-292***	10
Period -2	-58	50	-56***	3
Period 1	178	335	-192***	-63***
Period 2	938**	9	-157***	-68***
Period 3	770	-173	-127***	-80***
Period 4	591	-303	-114***	-77***
Period 5	575	-355	-126***	-78***
T × Period -5	22	22	19***	19***
T × Period -4	43	43	27***	27***
T × Period -3	29	29	36***	36***
T × Period -2	20	20	30***	30***
T × Period 1	54	54	4	4
T × Period 2	103	103	-10*	-10*
T × Period 3	38	38	-6	-6
T × Period 4	89	89	-19***	-19***
T × Period 5	213**	213**	-34***	-34***
Year dummies	No	Yes	No	Yes
Country dummies	No	Yes	No	Yes
Technology dummies	No	Yes	No	Yes
Observations	9906	9906	1328326	1328326

Standard errors are clustered at the country - technology level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Linear least-squares regression with patent applications (flow) as depend variable. T = 1 if actual, 0 if counterfactual outcome. Period 0 = treatment or pseudo-treatment period. False negatives = unpredicted standardization events. True negatives = correctly predicted years without standardization. For false negatives, predictions are averaged over 100 random control group draws. Standard prediction 1 year ahead.

Table 8: DID estimation for patent applications within groups (threshold of 0.5)

	False negatives		True negatives	
T	12	12	-44***	-44***
Period -5	-787***	756	-854***	41***
Period -4	-530***	504	-557***	21***
Period -3	-281***	287	-276***	4
Period -2	-69	73	-45***	0
Period 1	56	230	-222***	-61***
Period 2	583*	-324	-193***	-68***
Period 3	332	-506	-157***	-68***
Period 4	189	-470	-152***	-68***
Period 5	160	-605	-159***	-61***
T × Period -5	1	1	25***	25***
T × Period -4	14	14	31***	31***
T × Period -3	8	8	38***	38***
T × Period -2	5	5	31***	31***
T × Period 1	38	38	5	5
T × Period 2	23	23	1	1
T × Period 3	25	25	-7	-7
T × Period 4	81	81	-19***	-19***
T × Period 5	132**	132**	-36***	-36***
Year dummies	No	Yes	No	Yes
Country dummies	No	Yes	No	Yes
Technology dummies	No	Yes	No	Yes
Observations	15042	15042	1410334	1410334

Standard errors are clustered at the country - technology level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Linear least-squares regression with patent applications (flow) as depend variable. T = 1 if actual, 0 if counterfactual outcome. Period 0 = treatment or pseudo-treatment period. False negatives = unpredicted standardization events. True negatives = correctly predicted years without standardization. For false negatives, predictions are averaged over 100 random control group draws. Standard prediction 1 year ahead.

In tables 9 and 10 we regress the difference between actual and counterfactual patent applications on the period, an indicator variable which is 1 for false negatives and 0 for true negatives, and their interactions. The differences are parallel for pre-treatment periods since their interaction terms with the false negatives indicator (FN) are insignificant. The assumption of pre-treatment parallel trends between false and true negatives holds. We find a positive and significant difference in the fifth post-treatment period, i.e. the difference between actual and counterfactual patent applications in period 5 experienced a significantly higher increase with respect to the pre-treatment reference period for false negatives than for true negatives. Depending on the threshold, this difference amounts to 158 or 113 more patent applications when controlling for year, country and technology fixed effects. This confirms our finding using only within-group treatment effects for false negatives.

Table 9: DID estimation for treatment effects on patent applications across groups (threshold of 0.1)

	(1)	(2)
FN	13	76
Period -5	19***	178***
Period -4	27***	142***
Period -3	36***	111***
Period -2	30***	69***
Period 1	4	-21***
Period 2	-10*	-13**
Period 3	-6	9*
Period 4	-19***	16***
Period 5	-34***	20***
FN × Period -5	2	-39
FN × Period -4	16	-16
FN × Period -3	-7	-39
FN × Period -2	-10	-38
FN × Period 1	50	44
FN × Period 2	113	45
FN × Period 3	44	-12
FN × Period 4	108	49
FN × Period 5	246***	158*
Year dummies	No	Yes
Country dummies	No	Yes
Technology dummies	No	Yes
Observations	669116	669116

Standard errors are clustered at the country - technology level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Linear least-squares regression with the difference between actual and counterfactual patent applications as depend variable. FN = 1 if false negatives, 0 if true negatives. Period 0 = treatment or pseudo-treatment period. False negatives = unpredicted standardization events. True negatives = correctly predicted years without standardization. For false negatives, predictions are averaged over 100 random control group draws. Standard prediction 1 year ahead.

Table 10: DID estimation for treatment effects on patent applications across groups (threshold of 0.5)

	(1)	(2)
FN	56*	62*
Period -5	25***	170***
Period -4	31***	135***
Period -3	38***	105***
Period -2	31***	66***
Period 1	5	-18***
Period 2	1	-4
Period 3	-7	5
Period 4	-19***	11**
Period 5	-36***	9
FN × Period -5	-24	-59**
FN × Period -4	-17	-43
FN × Period -3	-30	-56**
FN × Period -2	-26	-46*
FN × Period 1	33	34
FN × Period 2	22	-23
FN × Period 3	33	-4
FN × Period 4	100	62
FN × Period 5	168***	113*
Year dummies	No	Yes
Country dummies	No	Yes
Technology dummies	No	Yes
Observations	712688	712688

Standard errors are clustered at the country - technology level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Linear least-squares regression with the difference between actual and counterfactual patent applications as depend variable. FN = 1 if false negatives, 0 if true negatives. Period 0 = treatment or pseudo-treatment period. False negatives = unpredicted standardization events. True negatives = correctly predicted years without standardization. For false negatives, predictions are averaged over 100 random control group draws. Standard prediction 1 year ahead.

In the following, we calculate treatment effects for different innovation measures and different country groups.

Patent quality

Patents can be of very different quality and importance to the industry. Therefore, patent counts are often adjusted for quality. The literature has identified different measures for patent quality, each with their advantages and inconveniences. One of the most used ones is the number citations towards the patents, so-called forward citations. This measure captures the use of a patent for other inventions and can, therefore, proxy the patent's technological importance. Since patents are of different age, it is common to use only forward citations within the first years of the patent's life. Here, we use the average number of forward citations a patent application of a given technology has received in the first 5 years after the application filing year.

Table 11: Treatment effects of standardization on average 5-years citation counts

	Post-treatment period in years					
	1	2	3	4	5	Average
β^T	-1.56	-1.38	-2	-2.02	-1.64	-1.72
β^U	-.11	.21	-.13	-.12	-.39	-.11
β^{DD}	-1.45	-1.59	-1.87	-1.9	-1.25	-1.61
β^{TD}	0	.18	-.43	-.46	-.08	-.16
β^{UD}	0	.32	-.02	-.01	-.29	0
β^{3D}	0	-.14	-.41	-.44	.21	-.16

Note: Estimation of the counterfactual using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period.

Table 11 shows that the effect of standardization on early patent life citations is very small, almost zero. This suggests that standardization leads to an increase in patent applications within the technology area of the standard, but not to an increase in the average number of citations to the technology class, thus, the average quality of the patents. One possible explanation could be that standard implementation and diffusion takes time and related patents might gain importance later in their life.

Technology selection

In section 2.1 we mentioned the technology selection function of standards. Here, we make a first attempt to test this hypothesis. To do so, we use the share of a technology's patent applications in the total number of patent applications in a given country and year as output variable of our random forest model. This measures the degree of patenting within a technology compared to all other technologies and proxies its relative R&D intensity.

Table 12: Treatment effects of standardization on the share of standard related patent counts in total patent counts

	Post-treatment period in years					
	1	2	3	4	5	Average
β^T	-.0065	-.0114	-.0093	-.0111	-.0159	-.0108
β^U	.0022	.0028	.0013	.0009	.0028	.002
β^{DD}	-.0088	-.0142	-.0106	-.012	-.0187	-.0129
β^{TD}	.0044	-.0004	.0016	-.0002	-.0049	.0001
β^{UD}	.0004	.001	-.0005	-.0009	.001	.0002
β^{3D}	.004	-.0015	.0021	.0007	-.0059	-.0001

Note: Estimation of the counterfactual using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period.

The variable ranges from 0 to 100, thus, represents percentage points. Table 12 shows an effect very close to zero, hence, does not suggest that standardization leads to technology selection in terms of patenting concentration within the standard's technology area. However, our measure of technology selection refers only to the economy as a whole and cannot reveal selection mechanisms within the technology fields. It also does not take into account standardization patterns in other technology fields. Hence, the technology selection theory of standards demands further research.

Country income groups

Tables 13 and 14 show the treatment effects on patent applications for high income and low income countries separately. The average positive effects found above seem to be driven by the high income countries, while treatment effects are generally negative for low income countries. Standards frequently create tensions between the developed and the developing world due to differences in adoption costs and an unequal distribution of intellectual property rights (Gibson [2007], Ernst [2011]). Low income countries might benefit less from standardization due to their second mover disadvantage on global markets and intellectual property right distribution. It is therefore possible that firms in low income countries try to innovate around standardized technologies.

Table 13: Treatment effects of standardization on patent applications for high income countries (threshold 0.5)

	Post-treatment period in years					
	1	2	3	4	5	Average
β^{T}	66.4	53.41	60.24	120.38	174.54	94.99
β^{TD}	43.93	30.95	37.78	97.92	152.08	72.53
β^{U}	-45.87	-51.7	-65.64	-77.71	-98.17	-67.82
β^{UD}	-23.28	-29.11	-43.04	-55.11	-75.58	-45.22
β^{DD}	112.27	105.11	125.88	198.09	272.71	162.81
β^{3D}	67.21	60.06	80.82	153.03	227.65	117.76

Note: Estimation of the counterfactual using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period. Income groups according to the WDI database.

Table 14: Treatment effects of standardization on patent applications for low income countries (threshold 0.5)

	Post-treatment period in years					
	1	2	3	4	5	Average
β^{T}	-20.59	-48.44	-63.97	-25.07	2.23	-31.17
β^{TD}	-14.57	-42.43	-57.96	-19.05	8.24	-25.15
β^{U}	-16.85	-15.27	-6.27	-15.07	-24.45	-15.58
β^{UD}	-8.96	-7.39	1.61	-7.19	-16.57	-7.7
β^{DD}	-3.74	-33.16	-57.7	-10	26.68	-15.58
β^{3D}	-5.61	-35.03	-59.57	-11.86	24.81	-17.45

Note: Estimation of the counterfactual using random forest with 100 decision trees. Treatment period = year of standardization event for false negatives or pseudo-treatment period for true negatives. For false negatives, predictions are averaged over 100 random control group draws. Prediction errors for pre-treatment periods are averaged over the 5 years preceding the treatment period for β^{TD}, β^{UD} and β^{3D}. In column 6, post-treatment errors are averaged over 5 years following the treatment period. Income groups according to the WDI database.

5 Conclusion

In this paper, we investigated the causal effect of standardization on innovation. This is a non-trivial task due to the complex causal relationship between standardization and innovation. In order to solve this problem, we developed a novel methodology which accounts for anticipatory effects of standardization. First, we predict standards using a feed-forward neural network. Subsequently, we use pre-standard data in a random forest to create a counterfactual innovation path for non-anticipated standards. For this set of standards, we are able to estimate the causal effect of standardization on innovation, since pre-standard data do not contain information on the standard itself and are therefore not able to predict the standard.

We estimate the effect of standardization for our set of non-anticipated standards on the number of patent applications within a technology field, the average number of 5-years citations to patents from the technology and on the technology's share in total patent applications of a country. We find a positive effect of standardization on patent applications which is significant five years after standardization. We find no effect on patent citations and application shares.

This paper contributes to the literature by estimating the causal effect of standardization on innovation. Former studies have struggled to identify the causal effect properly due to the reverse causality relationship between standardization and innovation. We are able to identify this effect by excluding anticipation of standardization. We also provide a novel identification strategy which may be used in other settings. Finally, we contribute to the literature on technology shocks by predicting standardization events.

Further research is necessary on the effect of standardization on the quality and distribution of innovation. An additional topic is the effect of standard adoption rather than standard development on innovation. Our current work focuses on providing empirical evidence for the unpredictability of our non-anticipated standards which represents the crucial assumption of our identification strategy.

References

Acs, Z. J. and Audretsch, D. B. (1989). Patents as a measure of innovative activity. *Kyklos*, 42(2):171–180.

Bekkers, R., Birkman, L., Canoy, M., De Bas, P., Lemstra, W., Ménière, Y., Sainz, I., van Gorp, N., Voogt, B., Zeldenrust, R., et al. (2014a). Patents and standards: A modern framework for IPR-based standardisation. *European Commission*.

Bekkers, R., Dalais, M., Dore, A., and Volanis, N. (2014b). Understanding patents, competition & standardization in an interconnected world. *ITU*.

Biddle, B., Curci, F. X., Haslach, T. F., Marchant, G. E., Askland, A., and Gaudet, L. (2012). The expanding role and importance of standards in the information and communications technology industry. *Jurimetrics*, pages 177–208.

Blind, K. and Jungmittag, A. (2008). The impact of patents and standards on macroeconomic growth: a panel approach covering four countries and 12 sectors. *Journal of Productivity Analysis*, 29(1):51–60.

Blind, K., Petersen, S. S., and Riilloc, C. A. (2017). The impact of standards and regulation on innovation in uncertain markets. *Research policy*, 46(1):249–264.

Burlig, F., Knittel, C., Rapson, D., m. Reguant, and Wolfram, C. (2017). Machine learning from schools about energy efficiency. *NBER Working Paper No. 23908*.

Cargill, C. F. (2011). Why standardization efforts fail. *Journal of Electronic Publishing*, 14(1).

Chiao, B., Lerner, J., and Tirole, J. (2007). The rules of standard-setting organizations: an empirical analysis. *The RAND Journal of Economics*, 38(4):905–930.

DTI (2005). The empirical economics of standards. *DTI Economics Paper No. 12*.

Ernst, D. (2011). Indigenous innovation and globalization: The challenge for china's standardization strategy.

European Patent Office (2018). Data catalog. PATSTAT global.

Gibson, C. (2007). Globalization and the technology standards game: Balancing concerns of protectionism and intellectual property in international standards. *Berkeley Technology Law Journal*, 22(4):1403–1484.

Hagedoorn, J. and Cloodt, M. (2003). Measuring innovative performance: is there an advantage in using multiple indicators? *Research policy*, 32(8):1365–1379.

Hallikas, J., Varis, J., Sissonen, H., and Virolainen, V. M. (2008). The evolution of the network structure in the ict sector. *International Journal of Production Economics*, 115(2):296–304.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

IAM (2018). Can AI inventions be patented in Europe? https://www.iam-media.com/can-ai-inventions-be-patented-europe [Last accessed: 2019-04-06].

ISO (1992). Computer Graphics Reference Model. *ISO/IEC 11072:1992*.

Layne-Farrar, A. (2013). *Innovations in Organizational IT Specification and Standards Development*, chapter Innovative or indefensible? An empirical assessment of patenting within standard setting., pages 1–18.

Loyka, J. J. and Powers, T. L. (2003). A model of factors that influence global product standardization. *Journal of Leadership Organizational Studies*, 10(2):64–72.

Lybbert, T. J. and Zolas, N. J. (2014). Getting patents and economic data to speak to each other: an 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, 43(3):530–542.

Pischke, J.-S. (2005). Empirical methods in applied economics. http://econ.lse.ac.uk/staff/spischke/ec524/evaluation3.pdf [Last accessed: 2019-04-28].

Rada, R., Carson, S., Haynes, C., and Moore, J. (1994). IT standards development and consensus: Three case studies. *StandardView*, 2(1):50–54.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Rysman, M. and Simcoe, T. (2008). Patents and the performance of voluntary standard-setting organization. *Management Science*, 54(11):1920–1934.

Stitzing, R., Sääskilahti, P., Royer, J., and Audenrode, M. V. (2017). Over-declaration of standard essential patents and determinants of essentiality. *SSRN Working Paper*.

Swann, P. (2010). The economics of standardization: An update. *Report for the UK Department of Business, Innovation and Skills (BIS)*.

Tassey, G. (2000). Standardization in technology-based markets. *Research policyResearch policy*, 29(4-5):587–602.

World Trade Organization. Technical information on technical barriers to trade.

A Appendix

A.1 Figures

Figure 8: Total number of standards over years in Perinorm between 1995 and 2015
1,208,663 standards

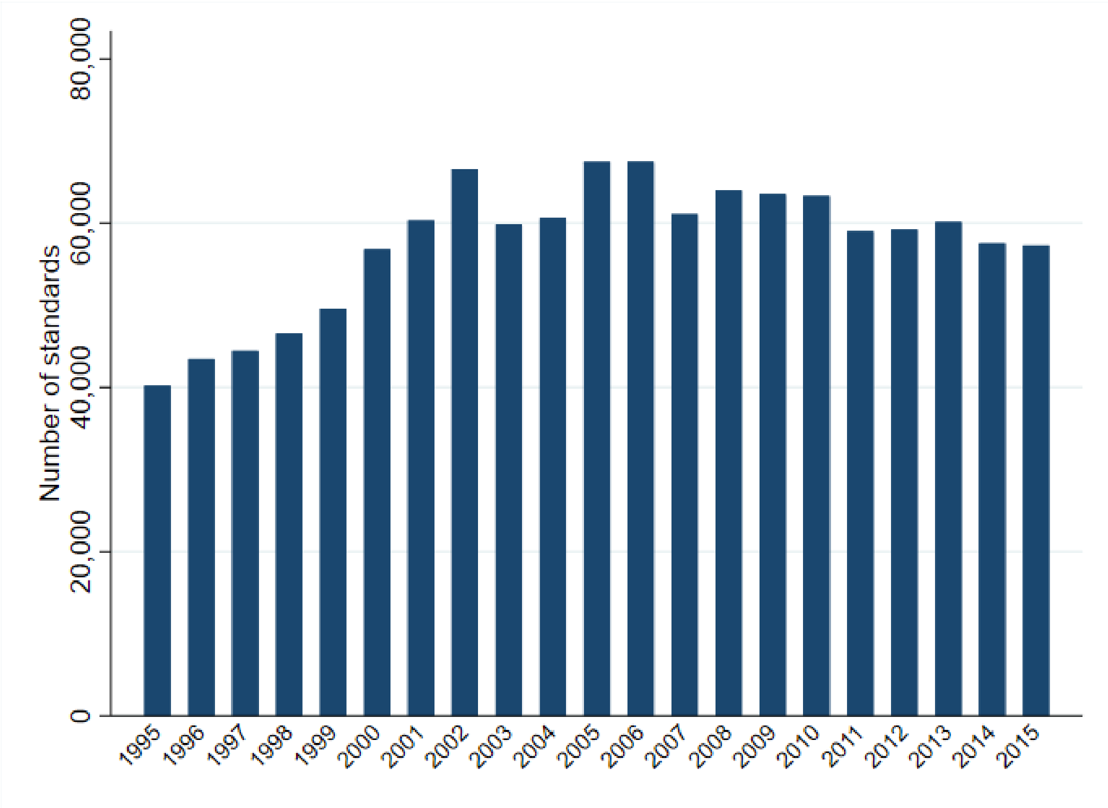
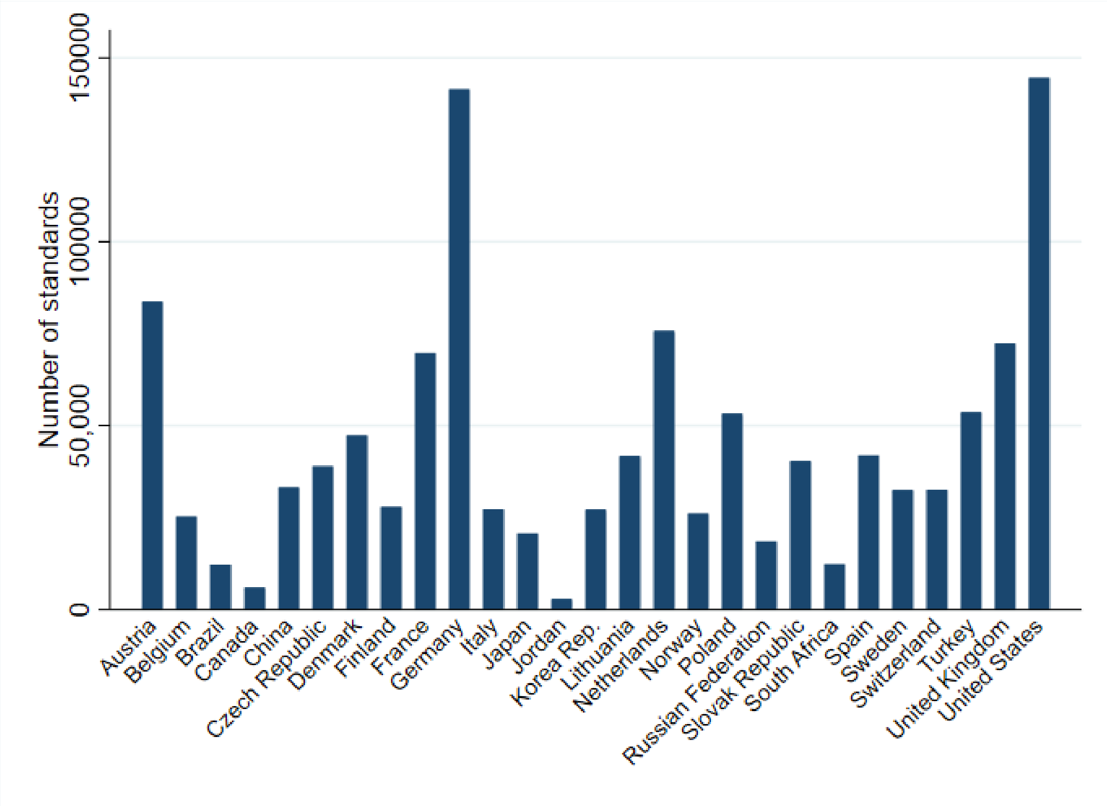


Figure 9: Total number of standards over countries in Perinorm between 1995 and 2015
1,208,663 standards



A.2 Tables

Table 15: Descriptive statistics by country

	Total number of standards	Mean										
		Exports in tech. (mio. USD)	Imports in tech. (mio. USD)	Patent stock in tech. (thous.)	Total patent stock (mio.)	GDP per capita (thous. USD)	Total population (mio.)	R&D expenditure (% GDP)	Mean tariff rate	Natural resource rent (% GDP)	No. of patent applications in tech. (thous.)	Total no. of patent applications (mio.)
Austria	748	104	96	6	2	44	8	.023	.024	.0017	1	.3644
Belgium	352	204	192	8	2	42	11	.02	.024	.0002	1	.3514
Brazil	222	65	106	1	.2067	10	187	.0124	.1407	.0342	.2373	.0443
Canada	196	220	272	19	5	45	32	.0187	.0467	.0332	4	1
China	532	1107	769	15	4	3	1298	.0128	.1166	.0388	4	.9048
Czech Republic	571	80	72	.5113	.0657	17	10	.0132	.024	.0066	.1095	.0144
Denmark	622	56	56	7	2	56	5	.0254	.024	.0114	1	.3274
Finland	308	74	50	11	3	43	5	.0315	.024	.0049	2	.4875
France	613	408	408	50	15	39	63	.0214	.024	.0005	10	3
Germany	921	870	657	116	39	40	82	.025	.024	.0012	22	8
Italy	315	300	239	17	6	36	58	.0121	.024	.001	3	1
Japan	384	604	359	251	74	44	127	.0313	.0283	.0002	50	15
Jordan	58	2	8	.0274	.0003	4	6	.0118	.1195	.0107	.0061	.0001
Korea Rep.	344	382	237	55	14	19	48	.028	.0851	.0002	12	3
Lithuania	499	9	9	.0254	.0005	10	3	.0079	.024	.007	.0066	.0002
Netherlands	848	307	294	27	7	47	16	.0188	.024	.0065	5	1
Norway	323	39	46	3	.723	85	5	.0188	.0134	.0827	.6101	.1419
Poland	594	84	95	.4497	.0602	10	38	.0072	.024	.0114	.1048	.0152
Russian Federation	379	58	120	1	.2181	9	145	.0109	.1004	.1451	.2728	.0467
Slovak Republic	502	30	33	.1218	.0053	14	5	.0073	.024	.0034	.0242	.0011
South Africa	286	23	47	.9698	.159	7	47	.0062	.0824	.0539	.1864	.0303
Spain	520	131	188	5	1	29	43	.0116	.024	.0006	.9834	.2586
Sweden	368	133	92	20	5	48	9	.0304	.024	.0053	4	1
Switzerland	431	142	112	28	9	70	8	.0273	.0114	.0002	6	2
Turkey	565	41	82	.4077	.0483	9	68	.0069	.043	.0038	.0982	.0122
United Kingdom	891	308	390	37	11	37	61	.0174	.024	.0078	7	2
United States	852	1039	1284	348	101	47	295	.0258	.0332	.0099	71	21

Table 16: Descriptive statistics by country

	Mean											
	Number of patent citations	C number of patent citations	A number of 5 years citations	Average operating revenue per firm in tech. (thous.)	Average number of employees per firm in tech. (thous.)	Number of firms in tech. (thous.)	Share of biggest firm's operating revenue	Share of biggest firm's number of employees	Country's share in global operating revenue in tech.	Country's share in global number of employees in tech.	Tech.'s share in total country's operating revenue	Tech.'s share in total country's number of employees
Austria	.2916	2	5	725	2	.0066	.2912	.2767	.0011	.001	.0003	.0003
Belgium	.7944	5	13	108	.066	2	.2465	.2178	.0115	.0058	.0003	.0003
Brazil	.0634	.2967	.8261	0	0	0	0	0	0	0	0	0
Canada	2	11	29	0	0	0	0	0	0	0	0	0
China	.767	4	13	293	2	.4902	.1496	.0951	.0283	.0563	.0002	.0002
Czech Republic	.04	.1496	.8394	14	.0903	2	.1708	.153	.0097	.0206	.0003	.0003
Denmark	.662	4	3	401	1	.0101	.196	.1958	.0009	.0009	.0002	.0002
Finland	.4171	3	6	103	.2311	.3999	.4497	.4022	.0111	.008	.0003	.0003
France	2	17	30	441	2	.8188	.2776	.2806	.0581	.0512	.0003	.0003
Germany	5	31	62	389	.5747	2	.2295	.2185	.1661	.117	.0003	.0003
Italy	1	10	17	41	.1098	6	.2034	.2089	.0598	.0351	.0003	.0003
Japan	5	49	67	2180	5	.3844	.4021	.3242	.1957	.1257	.0003	.0003
Jordan				76	.3705	.0012	.3649	.3404	0	0	.0002	.0002
Korea Rep.	1	7	18	19	.0556	2	.1846	.137	.0104	.0077	.0003	.0003
Lithuania				3	.0326	.6434	.2839	.2351	.0002	.0009	.0004	.0003
Netherlands	1	8	15	937	.2947	.8629	.4038	.3727	.0225	.0221	.0003	.0003
Norway	.1673	.7558	2	745	2	.01	.1959	.1993	.002	.0016	.0001	.0001
Poland	.0124	.0796	0	44	.2372	.0352	.4889	.2966	.0013	.0027	.0003	.0003
Russian Federation	.0582	.1942	1	9	.1683	6	.2259	.1852	.0133	.0601	.0003	.0003
Slovak Republic	.0116	.1124	.1357	49	.1237	.8693	.3771	.2435	.0055	.005	.0003	.0003
South Africa	.0951	.6504	1	822	4	.0034	.3536	.2758	.0006	.001	.0002	.0002
Spain	.386	2	6	15	.0387	10	.1292	.0828	.0699	.0481	.0003	.0003
Sweden	.8089	4	13	20	.054	3	.2452	.2198	.0153	.013	.0003	.0003
Switzerland	2	12	19	357	.8211	3	.2258	.2525	.0141	.0119	.0003	.0003
Turkey	.0302	.1592	.6347	1155	4	.0018	.1398	.1583	.0005	.0005	.0002	.0002
United Kingdom	4	21	45	229	.4708	1	.2663	.1754	.0704	.0466	.0003	.0003
United States	145	675	1900	1091	3	.1103	.2499	.211	.1267	.0804	.0003	.0003

Table 17: Prediction accuracy by lead

	Prediction lead				
Threshold	1 year	2 years	3 years	4 years	5 years
0.05	.838	.829	.824	.819	.817
0.1	.885	.879	.88	.875	.875
0.15	.908	.903	.904	.903	.898
0.2	.921	.918	.918	.917	.915
0.25	.93	.929	.928	.927	.925
0.3	.935	.935	.934	.932	.931
0.35	.94	.939	.938	.937	.936
0.4	.942	.941	.94	.94	.939
0.45	.944	.943	.942	.941	.94
0.5	.944	.944	.943	.941	.941
0.55	.944	.944	.943	.941	.941
0.6	.943	.943	.942	.94	.94
0.65	.942	.942	.941	.939	.939
0.7	.94	.941	.939	.937	.937
0.75	.938	.937	.937	.935	.935
0.8	.934	.934	.933	.932	.932
0.85	.93	.93	.93	.928	.929
0.9	.926	.926	.925	.923	.924
0.95	.919	.919	.917	.916	.917
Number of observations	136,620	129,789	122,958	116,127	109,296

Note: Neural network with 1 hidden layer and 15 nodes for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions are made 1 to 5 years ahead, i.e. inputs in period t are used for predictions in $t + x$, where $x \in \{1, ..., 5\}$. Accuracy = true predictions/number of observations.

Table 18: Accuracy of standard prediction for training and test samples

Prediction lead	Training sample	Full sample Number of obs.	Test sample	Number of obs.
1 year	.944	114761	.946	21859
2 years	.943	107930	.945	21859
3 years	.942	101099	.946	21859
4 years	.941	94268	.942	21859
5 years	.942	87437	.935	21859

Note: Neural network with 1 hidden layer with 15 neurons for prediction. Standard prediction is set to 1 if the prediction value of the neural network is larger than threshold. Predictions x=1,...,5 years ahead, i.e. using inputs from $t - x$. Accuracy = true predictions/number of observations. The restricted sample includes firm data from Orbis for prediction. Training/test sample contains 80/20 percent of the observations.

Table 19: Prediction accuracy of simple regression

Prediction lead	Accuracy	True positives	True negatives	Number of obs.
1 year	.9276	.2927	.9927	136620
2 years	.9271	.2963	.9926	129789
3 years	.9266	.2999	.9923	122958
4 years	.926	.3042	.9919	116127
5 years	.9256	.3104	.9915	109296

Note: Standard prediction using a linear prediction model. Prediction of a standard event for prediction values larger than 0.5.

Table 20: Prediction accuracy of probit

Prediction lead	Accuracy	True positives	True negatives	Number of obs.
1 year	.9305	.4237	.9825	136620
2 years	.9299	.4247	.9824	129789
3 years	.9292	.4251	.982	122958
4 years	.9282	.4251	.9815	116127
5 years	.9272	.4253	.981	109296

Note: Standard prediction using a probit model. Prediction of a standard event for prediction values larger than 0.5.

Table 21: Prediction accuracy of logit

Prediction lead	Accuracy	True positives	True negatives	Number of obs.
1 year	.931	.4344	.982	136620
2 years	.9301	.4328	.9817	129789
3 years	.9295	.4353	.9813	122958
4 years	.9284	.4339	.9808	116127
5 years	.9277	.4361	.9804	109296

** Prediction of a standard event for prediction values larger than 0.5

Note: Standard prediction using a logit model. Prediction of a standard event for prediction values larger than 0.5.

Figure 10: Prediction accuracy of logit