Is this novel technology going to hit?
Genetic markers predicting technological novelty diffusion

Michele Pezzoni[1], Reinhilde Veugelers[2], Fabiana Visentin[3]

Abstract: Despite the high interest of scholars in identifying breakthrough inventions, little attention has been devoted to follow how the technological content of those breakthrough inventions is re-used over time. We overcome this limitation by focusing on the dynamics of diffusion of the novel technologies that inventions incorporate. Specifically, we consider the factors affecting the time needed for a technology to be legitimated as well as its technological potential. We find that the dynamics of diffusion of a novel technology are affected by the characteristics of its building blocks, i.e. the technological components that combined together for the first time generate a novel technology. Combining similar technological components, components familiar to the inventors' community and with a high level of appropriability generates a technology that requires a short time to be legitimated but with a low technological potential. Combining technological components with a science-based nature generate technologies with a longer time to be legitimated but a high technological potential.

Keywords: technological novelty, diffusion, initial characteristics, patent data

[1] Université Côte d'Azur, CNRS, GREDEG, France; ICRIOS, Bocconi University, Milan, Italy; BRICK, Collegio Carlo Alberto, Torino, Italy; email: michele.pezzoni@unice.fr
[2] KULeuven; email: reinhilde.veugelers@kuleuven.be
[3] Chair in Economics and Management of Innovation – École Polytechnique Fédérale de Lausanne; BRICK, Collegio Carlo Alberto, Torino, Italy; email: fabiana.visentin@epfl.ch

1. Introduction

While there is a rising interest in identifying individual breakthrough inventions, little work has investigated how the novel technologies embodied in these inventions diffuse generating follow-up inventions (Arts et al., 2013). We consider novel technologies as the result of a combination of existing technological components (Schumpeter 1939; Nelson & Winter 1982). Then, we analyze the diffusion curve of a novel technology by looking at all the patented inventions that build-up on that technology. Looking at the diffusion curve of a novel technology, we are particularly interested in identifying the time that a novel technology needs to be legitimated within the inventors' community and its technological potential defined as the number of follow-on inventions that the novel technology can generate. We assume that the characteristics of the combined components are transmitted to the novel technology. Specifically, we investigate how the characteristics of the recombined components influence both novel technology legitimation and technological potential. Using a biological analogy, we identify novelty as the emergence of a mutation in the technological DNA of an invention, and we trace the diffusion of this mutation by following its genetic markers, i.e. the combined technological components, in subsequent inventions. Our study shows how the genetic marker characteristics impact on the mutation diffusion.

The so-called "onco-mouse'', a mouse largely used in laboratory cancer research, is an example of a well-recognized breakthrough invention embodying a novel technology. In 1985, Philip Lader and Timothy Steward at the Harvard laboratories had the revolutionary idea of isolating a gene and injecting it into a female mouse. The novel technology, embodied in the onco-mouse invention, was the result of the combination of two existing technological components. "Gene isolation" and "injection of material into animals" technological components appeared together

for the first time. The technology generated a variety of follow-up inventions represented by transgenic animals used in labs to treat a variety of diseases. For instance, the "diabetic mouse" created in 1996 by Seo Jeon Sun, a professor from Seoul National University is an example of patented invention building up on the novel onco-mouse technology. Looking at the "diabetic-mouse" patent documentation, the onco-mouse patent does not appear in its backward citation list, since citations refer to the specific prior art used and not necessarily to the technology content of an invention. In this work, we go beyond citations. Instead, we look at the patents' technological content and trace the reuse of the novel combination in follow-on patents. With this approach, the "diabetic-mouse" is included in the diffusion of the onco-mouse novel technology. The two combined components at the base of the onco-mouse technology have each their own characteristics and so has the resulting combination. The "gene isolation" was rather new within the technological community, as well as the "injection of material into animals". Both components were similar to each other. They were furthermore both strongly science-based. By investigating how the combined components impact on legitimation and technological potential, we aim to reply to a series of *what-if* questions such as, what would have been the diffusion of the novel onco-mouse technology if the combined components were more established? What would have been the diffusion of the novel onco-mouse technology if the two components were more science-based? Or, more generally, what are the "genetic markers" predicting a "hit" technology?

In this paper, we conduct an empirical study that traces the diffusion of more than 10,000 novel technologies over the period 1985-2000. In our study, the existing technological components correspond to the technological classes used in the European Patent Office classification, and novel technologies correspond to an unprecedented combination of technological classes in the history of patenting (Strumsky and Lobo, 2015).

The contribution of our study is twofold. First, we provide insights to the economics of innovation literature that is interested in identifying the key drivers of the evolution of a "technological trajectory" (Dosi, 1985)[4]. Second, we provide relevant managerial implications by predicting the legitimation time and the potential of a technology since its early appearance.

The remaining of the paper is organized as follow. Section 2 develops the hypothesis on how the characteristics of the technological components recombined affect the diffusion of the resulting technology. Section 3 illustrates the method used in the analysis. Section 4 presents the results that are discussed in Section 5. Section 6 concludes.

2. Technological novelty and its diffusion

A large number of studies has investigated the diffusion of an invention looking at its use by the relevant population of potential adopters (Hall, 2004). A limited number of studies has investigated technological innovation dynamics tracing the diffusion path of a novel technology in the subsequent inventions that incorporate the technology. Extant studies rely on case studies developed for specific technologies. For instance, Achilladelis (1993) follows the diffusion of technological innovations in the antibacterial medicines sector over 70 years tracing all the patents appeared in this temporal window. Differently from the extant literature, our study adopts a systematic quantitative approach that allows us to identify the introduction of a novel technology, to trace its diffusion, and to assess which factors that impact on its diffusion.

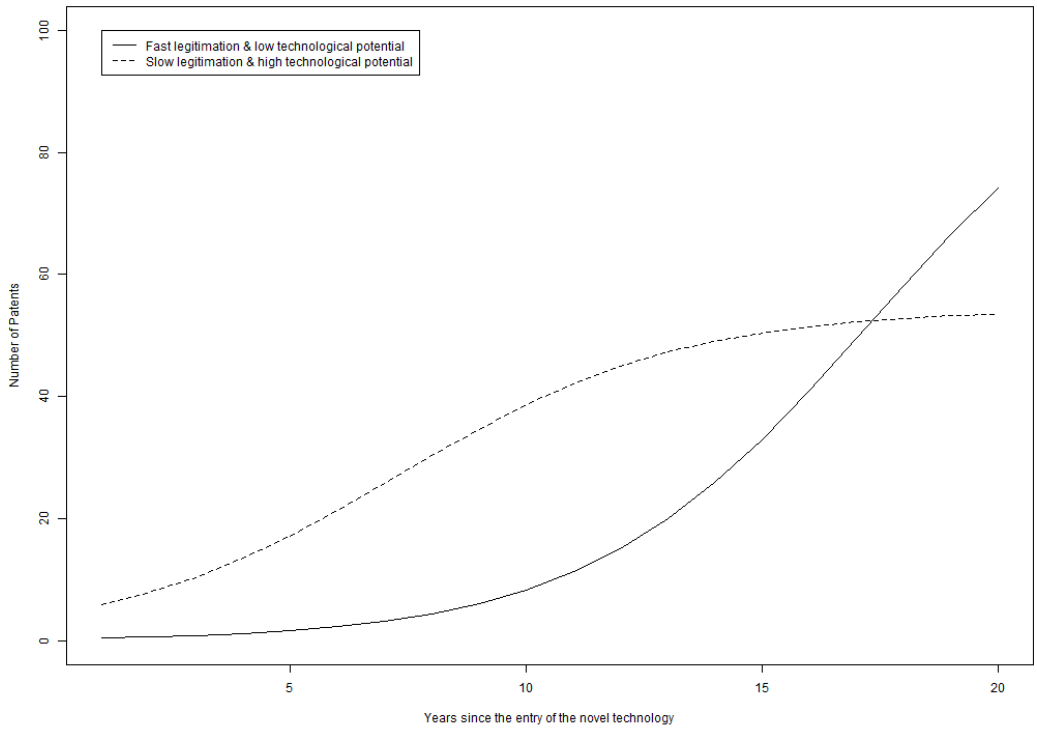
[4] A neo-Schumpeterian scholar would define a novel technology as a new technological paradigm and the diffusion of a novel technology as a technological trajectory. Over time, radical inventions introduce new technological paradigms. Technology advancement is cumulative and new paradigms open new venues for further technological developments shaping new technological trajectories. Technological trajectories can be traced over time by following the patent paper trail related to the technological paradigm (Achilladelis, 1993).

The first step in our analysis is to define a novel technology. In his seminal work on the origins of innovation, Schumpeter claimed that "innovation combines components in a new way, or that it consists in carrying out New combination" (1939, pp. 88, Schumpeter 1939). Fleming (2001) elaborated on the concept of invention as a process of "recombinant search" (pp. 118) by arguing that inventors search among existing technological components and recombine them to realize something new. According to this "recombinant search" approach, we consider a novel technology as the result of pre-existing technological components that are combined for the first time (Verhoeven et al., 2016).

In reconstructing the diffusion of a novel technology, we follow the mainstream literature on diffusion assuming that a diffusion process follows a sigmoid-curve (Griliches, 1957). A S-curve results from two distinct phases: a first phase of slow growth followed by a second phase of asymptotical convergence to a ceiling level. In our study, we assume that the first phase is driven by a social process of legitimation. Initially a technology diffuses slowly since it needs time to gain legitimation within the inventors' community and with consumers. Inventors using the novel technology for follow-up inventions need time to learn and familiarize with the novel technology and abandon the established competing technologies. Once legitimated, the diffusion of the novel technology accelerates since it becomes the dominant technology. The second phase of asymptotical convergence to a ceiling level is determined by the potential of a technology. Two different trends can lead to the diffusion ceiling. First, all the possible applications of the technology have been implemented exhausting the inventive opportunities. Second, the technology loses its appeal in favor of emerging new technologies. Figure 1 compares the diffusion curves of a technology having a long legitimation phase and a high technological potential with another

technology having a short legitimation phase but a low technological potential. The time span considered in Figure 1 is twenty years since the introduction of the novel technology.

Figure 1: Comparison between diffusion curves



An insightful example of technology experiencing a first phase of legitimation followed by a second phase of convergence to its technological potential, is represented by the Sulpholamide drugs. The Sulpholamide is a class of synthetic antibacterial drugs introduced in 1935. The first Sulpholamide-based drug, Prontosil, was developed and launched by Bayer. In the next 10 years following the Prontosil introduction, other companies and public research laboratories gradually developed and marketed more than 5,000 innovative Sulpholamide-based drugs. The success of Sulpholamide-based drugs was exhausted by the mid-‘50s for two reasons. First, additional

improvements of the therapeutic benefits of this class of drugs become more difficult. Second, new antibiotics appeared (for details on the case, see Achilladelis, 1993).

There is a general consensus among innovation scholars that novel technologies spur from unprecedented combinations of existing components. “At any point in the technological evolution, any component is at risk of being recombined with any other component” (Fleming, 2001; pp.119). Nonetheless, previous studies did not consider the characteristics of the combined components as a predictor of diffusion of the novel technology. In this paper we assume that the characteristics of the combined components are transmitted to the novel technology influencing its diffusion curve. We identify technological, cognitive and institutional characteristics of the combined components as relevant for the diffusion of the novel technology. Technological characteristics are those related exclusively to the technological aspects of the combined components. Cognitive characteristics are those related to the understanding of the combined components by the inventors’ community. Finally, institutional characteristics are those related to the level of protection granted to the inventions by the extant intellectual property rights.

As technological characteristics we consider both the technological similarity and science-based nature of the combined components. Novel technologies can result from the combination of components that differ substantially or that are similar between each other. We expect that a novel technology resulting from the combination of two similar components is legitimated earlier than a novel technology resulting from the combination of two dissimilar components. Combining similar components requires less inventive effort, and a larger number of inventors might have the required competences to use the resulting technology. On the contrary, combining dissimilar components requires the contribution of specialists with different technological competences, often organized in teams of inventors, slowing down the legitimation. At the same time, we expect

that the combination of components that are based on similar technological principles generates a novel technology that only marginally differs from the existing components, harming its technological potential. On the contrary, combining dissimilar components generates a novel technology with a greater original content, enhancing its technological potential.

We expect that:

> *Hypothesis 1 (combining similar components): A novel technology resulting from the combination of two similar components (a) requires a shorter time to be legitimated and (b) has a lower technological potential.*

Novel technologies can result from the combination of science-based or applied components. Science-based components are those components where the knowledge used has general and abstract content, while applied components are those components where specialized knowledge is used for specific applications (Breschi et al., 2000). Novel technologies resulting from science-based components are expected to be characterized by a broad set of potential technological applications. At the same time, the generality of the scientific knowledge used is expected to be far from applied purposes. On the contrary, novel technologies relying on applied components use less abstract knowledge and are closer to a limited set of possible technological applications. We expected that a novel technology resulting from the combination of science-based components, being far from possible applied purposes, takes more time to be legitimated in the inventors' community. We expect also that a novel technology resulting from the combination of science-based components, having a broad set of applications, has a higher technological potential while the combination of applied components has lower technological potential.

Hypothesis 2 (combining science-based components): A novel technology resulting from the combination of science-based components (a) requires a longer time to be legitimated and (b) has a higher technological potential.

Cognitive characteristics are relevant since components are recombined by individuals who need to receive and interpret information before being able to use it in a creative way (Cohen and Levinthal, 1990). As cognitive characteristics, we consider the familiarity of the inventors' community with the use of the components. A novel technology can result from the combination of components that have been frequently or rarely used within the inventors' community. A component frequently used is well-known and can be exploited, while a component rarely used is unknown and has to be explored (March, 2001). Following Fleming's argument (2001), we assume that there is a greater chance that inventors build-up their inventions on components they are familiar with. Exploiting familiar components reduces the risks for unexpected results and facilitates the legitimation of the resulting novel technology. However, familiar components have been already largely exploited in the past so the novel technology resulting from their combination is expected to have a lower technological potential.

We formulate the following hypothesis:

Hypothesis 3 (combining familiar components): *A novel technology resulting from the combination of familiar components (a) requires a shorter time to be legitimated and (b) has a lower technological potential.*

Finally, as institutional aspects we consider the level of appropriability of the results of the inventive effort characterizing the components. We define the level of appropriability as the possibility to protect inventions from imitation. Appropriability largely depends on the scope of

the protection that the extant intellectual property legislation guarantees over inventions. A broad scope of protection provides an incentive to inventors and firmsto start new research activities but, at the same time, it prevents outsider potential inventors to benefit from the latest technological advancements (Breschi et al., 2000). Specifically, when the scope of the protection is board, outsiders are deterred to start new research activities by the risk of infringing the intellectual property rights granted to the incumbent inventors (Merges and Nelson, 1994). As a result of the barriers to the entry of new incumbent, the community of inventors remains small. The limited number of inventors facilitates the diffusion of information and leads to a faster legitimation of the technology. The trade-off of having a small community of inventors is a limitation on the aggregated inventive potential. As a consequence, we expect that the combination of components with high appropriability leads to a novel technology with lower technological potential.

Hypothesis 4 (combining components with high appropriability): A novel technology resulting from the combination of components with high appropriability (a) requires less time to be legitimated and (b) reaches lower technological potential.

3. Methods

Our analysis relies on a sample including all the patents filed to the European Patent Office (EPO) in the period 1978-2015. Following Fleming and his coauthors (2007), we consider the technology classes in which a patent is classified (International Patent Classification -IPC- codes) as the technological components on which the patent build-up. We mark the appearance of a novel technology as the first time ever[5] appearance of a combination of IPC codes in a patent. Limiting our definition of novel technology to the first time ever appearance of a combination, might generate trivial pairs of IPC codes resulting from the mechanical merge of all IPC codes included in all the EPO patent applications. Therefore, we restrict novel technology definition only to those novel combinations re-used in at least 20 patents in the 20 years following the novel technology appearance. We end up with a study sample that includes 11,009 novel technologies which generated 505,135 patents.

To represent the diffusion curves of each novel technology we proceed in two steps. First, we count the yearly number of patents that embodied the novel technology (y_t). Doing so, we construct the actual patent cumulated distribution over a window-period of 20 years ($Y_t = \sum_{p=1}^{t} y_p$) following the appearance of the novel technology[6]. Second, we use the actual cumulated distribution of each novel technology to fit the corresponding trend function. The trend function represents an algebraic approximation of the diffusion curve of the technology. Following the arguments detailed in Section 2, we opt for an S-curve (or sigmoid curve) characterized by a slow

[5] To flag a technology as novel, we would need the complete history of its component to be able to evaluate when a pair of component appear together for the first time. We use the first available data period at EPO, 1978-1984, as a buffer period to capture the history of our components and we track novel combinations starting from 1985.

[6] We consider all the novel technologies until the patent cohort of 1996 to leave a 20-year window forward to this last cohort.

initial growth and by an asymptotic convergence to a ceiling level. The use of an estimated diffusion curve, instead of actual data, allows us to measure the technology potentials and the legitimation period also for those technologies that do not exhaust their innovative potentials during the observation period of 20 years. For instance, technologies with particularly slow legitimation might not show any asymptotic convergence to a ceiling level after the 20 years covered by the observed data.

The technological diffusion S-curve is identified by three parameters, i.e, *midpoint*, *slope* and *ceiling* and can be expressed by the following equation:

$$\hat{Y}_t = \frac{ceiling}{1 + e^{\left(-\frac{(t-midpoint)}{slope}\right)}}$$

(Equation 1)

where $\hat{Y}_t$ is the cumulated number of patents predicted at time *t*, *t* is the number of years elapsed since the appearance of the novel technology, the parameter *ceiling* is defined as the upper asymptote of the S-curve, *midpoint* is the required time to reach the fifty percent of the ceiling, and *slope* is the diffusion speed at the midpoint (Griliches, 1957). We use the three parameters describing the S-curve as proxies of the concept of *legitimation* and *technological potential* illustrated in Section 2. We assume that a technology is legitimated when it is accepted by the community and we fix the "acceptance" point as the moment when the technology reaches the ten percent of its ceiling. Using the formula in Equation 1, legitimation is calculated as a linear combination of *midpoint* and *slope*, *midpoint-ln(9)*slope* (See Appendix A for the details concerning the mathematical formulation) and is expressed in number of years since the

appearance of the novel technology. The potential of a technology equals to ceiling and indicates the maximum number of inventions that the novel technology is expected to generate.

In this paper, we conduct a set of regression exercises aiming to estimate the impact of the technological, cognitive and institutional characteristics of the components combined on the legitimation and technological potential of a novel technology. Specifically, we estimate with an Ordinary Least Squares the following two equations:

$$\textit{Legitimation} = \beta_0 + \textbf{Component characteristics} * \beta_1 + \textbf{Inventors' characteristics} * \beta_2 + \textbf{Applicants' characteristics} * \beta_3 + \textbf{Other controls} * \beta_4 + \varepsilon$$

(Equation 2)

$$\textit{Technological potential} = \beta_0 + \textbf{Component characteristics} * \beta_1 + \textbf{Inventors' characteristics} * \beta_2 + \textbf{Applicants' characteristics} * \beta_3 + \textbf{Other controls} * \beta_4 + \varepsilon$$

(Equation 3)

where *Component characteristics* is a vector of variables that includes the characteristics of the components that combined to generate the novel technology namely, *similarity, science-based content*, *familiarity*, and *appropriability*. The other three vectors of variables represent our controls. *Inventors' characteristics* is a vector including the inventors' characteristics. *Applicants' characteristics* is a vector accounting for applicants' characteristics. As *other controls* we include a set of time-dummies representing the calendar year when the novel technology appeared (*Technology entry year*).

To measure *similarity* between the components combined, we exploit the hierarchical structure of the IPC code classification where each additional digit denotes a higher degree of refinement of

the technological classification. Precisely, we define two components being similar when they have the first three digits of their IPC codes in common. To construct the variable *Science-based content*, we consider the patent applications including the two combined components over a rolling time window from t-1 to t-4. For each component we compute the average number of references to the non-patent literature per patent application (Meyer-Krahmer & Schmoch, 1998). Finally, we calculate the *Science-based content* variable as the average number of references per patent between the two components. To measure *familiarity*, we refer to Fleming's measure (2001). To this end, we count the number of patent applications in a four-year rolling window for each of the two combined components. Then, we calculate the average number of patents between the two components. Similarly, to construct the variable *Appropriability*, we calculate the average number of claims of the patents in a four-year rolling window for each of the two combined components (Merges and Nelson, 1994; Tong and Frame, 1994). Then, we calculate the average between the two components. In our regression model we consider the logarithm transformation of the variables *similarity, Science-based content, familiarity, and Appropriability* to interpret the estimated effects as semi-elasticities.

Several time-invariant unobserved characteristics of the components might bias our estimations of the technological, cognitive and institutional characteristics such as, the unmeasurable intrinsic propensity of the component to generate additional inventions and its technological complexity. Therefore, we include as explanatory variables a set of 118 three-digit technological class dummies referring to the combined components. A technology class dummy equals one if the first three digits of the IPC codes of at least one of the two combined components equal to the three-digit identifying the technological class dummy, zero otherwise. A complete list with the detailed description of the 118 technological class dummies is reported in Appendix B.

As inventors' characteristics we consider *inventors' experience* and *inventors' team size*. To build the *inventors' experience* variable, we proceed in three steps. First, we extract the inventors listed in the patents filed during the first year when the novel technology appears. Second, we reconstruct the patent history of each inventor and count the number of patents she filed. Finally, in case the patent introducing the novel technology has multiple inventors, we sum the stock of patents of each inventor. The *team size* is the number of inventors appearing in the patents filed during the first year when the novel technology appears.

As applicants' characteristics we consider their experience, type (university or public research center versus private company), being a single applicant, and their country. The *applicants' experience* is computed in the same way as for inventors. *Type of applicants* is a dummy that equals one if at least one of the applicant is a university or a research center, zero otherwise. *More than one applicant* is a dummy that equals one if there is more than one applicant, zero otherwise. *Applicant's country* is a set of dummies, one for each applicant's country reported on the patent documents, that identifies the geographical location of the applicant(s). *Applicants' concentration* is measured as the Herfindahl–Hirschman Index (HHI). For the stock of patents of each component, we calculate the share of patents owned by each applicant. Based on these shares we calculate the HHI of each component and we use in the regression exercise the log-transformed average of the HHI of the two components. As for the inventors' characteristics, the set of variables that refer to applicants (excluding *Applicants' concentration*) is computed on the patents filed during the first year when the novel technology appears.

Table 1 reports the descriptive statistics of all our dependent and independent variables.

Table 1: Descriptive statistics.

	Obs.	Mean	Sd	Min	Max
Dependent variables					
Technological potential [# patents]	11009	65.97	93.84	16.32	998.29
Legitimation (10%) [# years]	11009	5.86	3.13	0.00	16.78
Other sigmoid parameters					
Midpoint (50%) [# years]	11009	12.37	3.32	1.53	24.89
Slope at midpoint	11009	2.96	1.16	0.21	7.47
Component characteristics					
Similarity 3 digits (dummy)	11009	0.33	0.47	0.00	1.00
Science-based content	11009	3.50	7.41	0.02	283.99
Familiarity	11009	346.24	427.84	0.50	5490.50
Appropriability	11009	10.91	2.09	1.50	28.75
Inventor's characterisics					
Inventors' experience (dummy)	11009	0.47	0.50	0.00	1.00
Inventors' team size	11009	3.36	2.89	1.00	55.00
Applicant characteristics					
Applicants' experience	11009	637.19	1503.72	0.00	17279.00
University applicant (dummy)	11009	0.05	0.22	0.00	1.00
More than one applicant (dummy)	11009	0.25	0.43	0.00	1.00
Applicants' Concentration (HHI)	11009	0.03	0.05	0.00	0.89

The onco-mouse case, provides an illustrative example of how we identify the diffusion curve of a novel technology and of how we calculate the characteristics of the combined components. We mark as onco-mouse novel technology the combination of IPC codes "Introducing […] material into […] the body of animals" (IPC code A01K67) and "[…] DNA or RNA concerning genetic engineering […]" (IPC code C07H21) that appeared for the first time in 1985 patent. We observe the novel onco-mouse technology for a 20-year window, i.e., until 2004. Figure 2 shows the actual cumulated distribution (Y_t) of the user patents embodying the combinations of the IPC codes A01K67-C07H21 (dotted line). Referring to the patent distribution, we estimate the three parameters of the corresponding S-curve by using a maximum likelihood estimation methodology. The solid line in Figure 2 represents the fitted S-curve curve ($\hat{Y}_t$). This curve is identified by an

estimated ceiling equal to 267.38 patents, a midpoint of 15.49 years, and a slope of 3.45. In terms of legitimation and technological potential, it results that the legitimation of the technology is reached after 7.91 years since the onco-mouse appearance while 267.38 patents correspond to the technological potential reachable by the technology. The use of a specific functional form allows us to predict the diffusion of a novel technology at any point in time by means of the three estimated parameters. In our example, by substituting the estimated parameters in Equation 1, we can predict that, after 10 years since the onco-mouse technology appearance ($\hat{Y}_{10}$), the cumulated number of user patents embodying the onco-mouse technology equals 45.24 ($\frac{267.38}{1+e^{(-\frac{(10-15.49)}{3.45})}}$=45.24). In the same way, we can predict that after 30 years from its appearance ($\hat{Y}_{30}$), the cumulated number of user patents embodying the onco-mouse technology equals to 263.45 ($\frac{267.38}{1+e^{(-\frac{(30-15.49)}{3.45})}}$=263.45). Figure 3 plots all the estimated diffusion curves of the novel technologies in our sample. The black line highlights the onco-mouse technology diffusion curve.

Concerning the characteristics of the combined components generating the onco-mouse, we find that they are not similar (*Similarity* = 0 [sample average = 0.33]), they are science based (*Science-based content* = 9.51 [sample average = 3.50]), the inventors' community is not familiar with the components (*Familiarity=160.5* [sample average = 346.24]), and finally, that their *appropriability* is slightly higher than the average of the other components (*Appropriability=13.11* [sample average = 10.91]).

Figure 2: Diffusion curve of the novel onco-mouse technology

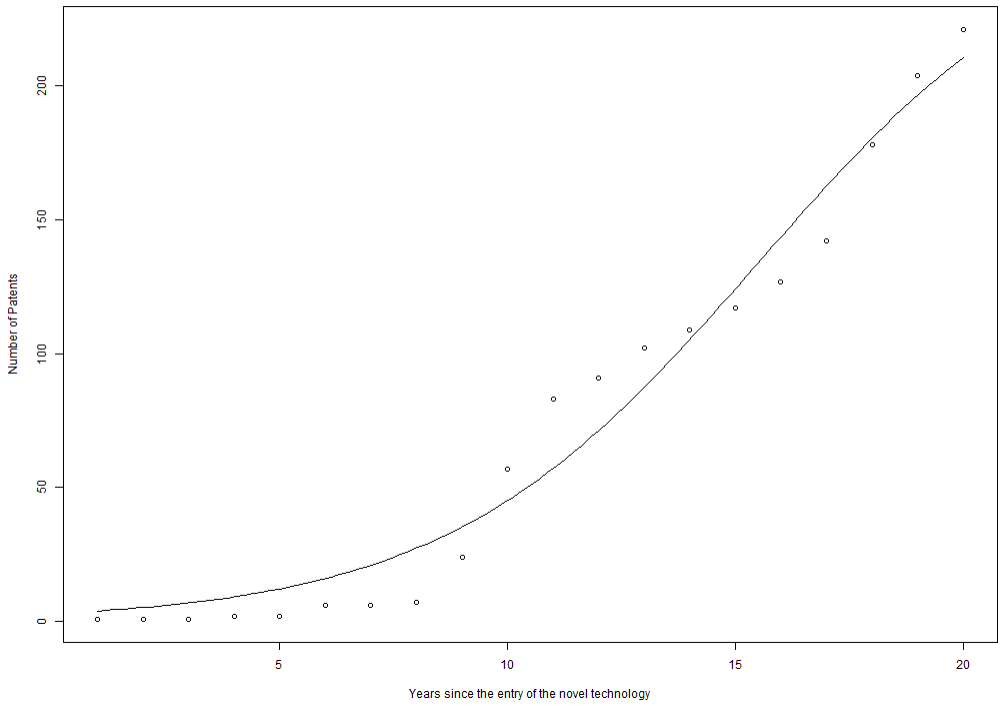
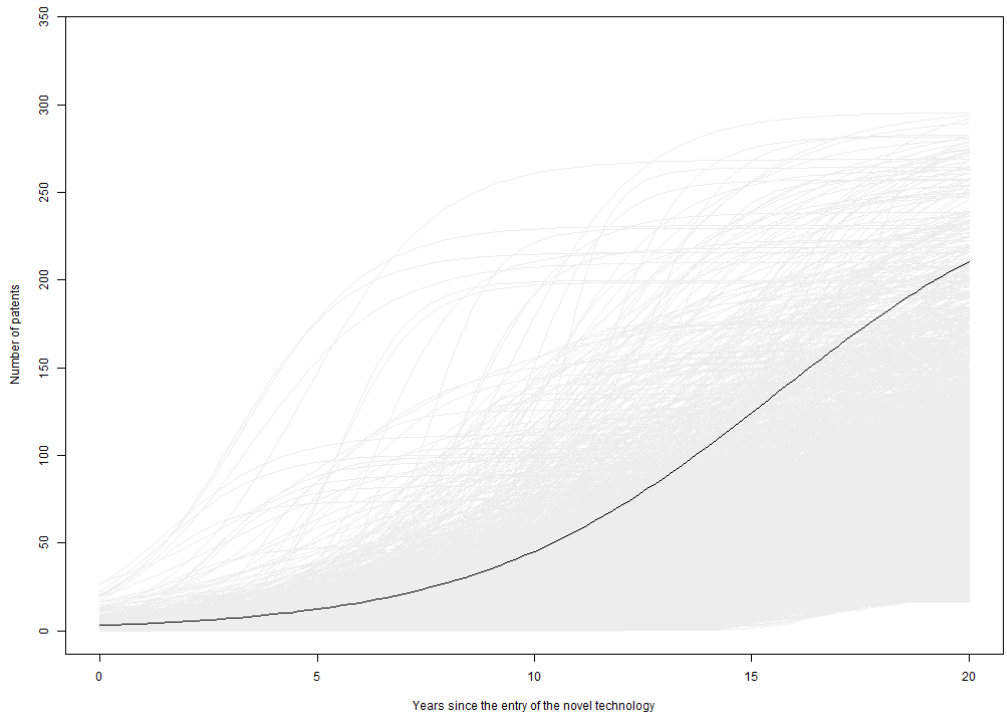


Figure 3: Novel technologies' diffusion curves. The back curve identifies the onco-mouse technology



4. Results

Table 2 shows the results of our econometric exercise. Columns 1 and 2 report the estimation of a baseline model including only the component characteristics and technology entry year dummies. Columns 3 and 4 add as controls the inventors' and applicants' characteristics. Columns 5 and 6 add the technological class dummies. According to our most complete model specification, represented in Equation 2 and 3 and estimated in columns 5 and 6, we find that combining two similar components reduces significantly the technology legitimation time. Specifically, if the novel technology results from the combination of two similar components the legitimation time is 15.12 months[7] shorter than a technology resulting from the combination of two dissimilar components. Component similarity significantly affects also the technological potential. Having similar component decreases the technological potential by 14.8 patents. Then, hypothesis one, *combining similar components*, is confirmed.

Combing components with a higher science-based content generates a novel technology that requires more time to be legitimated but has a higher technological potential. Precisely 50% more of science-based content of the technology increases the legitimation time by 1.38 months, while it increases the technological potential by 9.1 patents. Therefore, hypothesis two, *combining science-based components*, is confirmed.

A novel technology resulting from the combination of two familiar components requires a shorter time to be legitimated. Increasing the level of familiarity by 50% decreases the time needed for a novel technology to be legitimated by 1 month. The same technology has lower technological

[7] 15.12 is obtained multiplying the coefficient of Similarity in the regression with Legitimation as dependent variable (-1.26) by the number of months in a year (12).

potential by 3.5 patents. These results are in line with hypothesis three, *combining familiar components*.

Combining components with a high appropriability generates a novel technology that requires shorter time to be legitimated and has lower technological potential. Precisely 50% more of appropriability decreases the legitimation time by 5.94 months and decreases the technological potential by 13.65 patents. Therefore, hypothesis four, *combining components with high appropriability*, is confirmed.

Concerning the controls, we find that the inventors' experience reduces the time needed for a technology to be legitimated while the inventors' team size increases the technological potential. The experience of the applicants, having multiple applicants, and the presence of a university among the applicants promoting the novel technology lead to a technology with a higher potential. The legitimation time is negatively affected by the presence of multiple applicants. The applicant concentration in the sector, foster the technological potential.

In Appendix C we estimate two models with the same specification as those reported in Equation 2 and 3 but replacing the dependent variables with the other two parameters of the S-curve, respectively *Midpoint* and *Slope*. Moreover, we show that the linear combination of the coefficients estimated for *Midpoint* and *Slope* regressions allows us to calculate the coefficients estimated in the legitimation regression (Column 5, Table 2).

In Appendix D we check the robustness of our main results reported in Table 2. Specifically, we change the functional form by using the log transformation of our dependent variables legitimation and technology potentials. In this regression exercise, the estimated coefficients for the component characteristics can be interpreted as elasticities.

Table 2: Regression results. OLS estimations for Equation 1 and 2.

VARIABLES	(1) Legitimation (10%)	(2) Technological potentials	(3) Legitimation (10%)	(4) Technological potentials	(5) Legitimation (10%)	(6) Technological potentials
Component characteristics						
Similarity 3 digits (dummy)	-0.0047	17.3***	0.077	18.1***	-1.26***	-14.8***
log(Science-based component)	0.49***	25.0***	0.49***	22.3***	0.23***	18.2***
log(Familiarity)	-0.16***	-6.87***	-0.092***	-4.96***	-0.17***	-6.91***
log(Appropriability)	-3.05***	-64.8***	-2.51***	-54.4***	-0.99***	-27.3***
Inventor's characterisics						
Inventors' experience (dummy)			-0.30***	-3.08	-0.20***	-1.18
log(Inventors' team size)			-0.17***	4.03***	-0.063	5.55***
Applicant characteristics						
log(Applicants' experience)			-0.027**	0.14	-0.010	0.60*
University applicant (dummy)			0.10	7.62*	0.17	10.3***
More than one applicant (dummy)			-0.77***	7.76***	-0.86***	5.71**
log(Concentration)			0.26***	9.81***	0.042	11.0***
Dummy Technology class (3 digits)	No	no	no	no	yes	yes
Dummy Applicant's country	no	no	yes	yes	yes	yes
Dummy technology entry year	yes	yes	yes	yes	yes	yes
Constant	14.1***	254***	14.3***	238***	12.9***	248***
Observations	11,009	11,009	11,009	11,009	11,009	11,009
R-squared	0.090	0.074	0.128	0.096	0.243	0.155

5. Discussion

Ideally an attractive technology has a short legitimation period and a high technological potential. To visualize the impact of the component characteristics on the novel technology diffusion, we simulate four scenarios where the novel technology is obtained by combining (1) similar components (compared to dissimilar components), (2) components with a science-based content increased by 50%, (3) components with familiarity increased by 50% and (4) components with appropriability increased by 50%. The simulations are based on the estimations of the econometric models presented in Table 2, columns 5 and 6. Figure 4 shows the increase/reduction of legitimation time obtained simulating the four different scenarios, while Figure 5 reports the diffusion curve for each simulation up to twenty years after the introduction of the novel technology. Although our model allows to predict technological potential for any time span, we selected twenty years because it seems a reasonable period over which both managers and policy makers might be requested to predict the technological potential of a novel technology.

Our simulations show that the characteristics considered, i.e., similarity, science-based content, familiarity, and appropriability always imply a trade-off between the legitimation and technological potential. Looking at Figures 4 and 5, we observe that by augmenting the value of the variables science-based component, the corresponding technology has higher technological potential. However, it requires a longer time to be legitimated. On the contrary, augmenting the value of familiarity, the novel technology decreases the time needed to be legitimated and, as a drawback, reduces its technological potential. Higher Similarity as well as higher Appropriability decrease the legitimation time and the technological potential.

Although each component characteristic implies a trade-off, the extent of these trade-offs differ significantly according to the characteristic considered. Augmenting the science-based content

increases to a considerable extent the technology potential, augmenting only slightly the legitimation time. Familiarity exhibits limited effects both on legitimation and technological potential. Appropriability and similarity are characterized by the largest trade-offs, namely shortening the legitimation time is obtained at the cost of significantly lower technological potential and *vice versa*. We conclude that Familiarity has a limited effect, science-based content increases significantly the technology potentials with a limited drawback in terms of legitimation time, appropriability and similarity show relevant trade-offs between legitimation and technological potentials.

Our general results can be used to reply to a series of *what-if* questions concerning specific novel technologies. One might ask, what would have been the diffusion of the novel "onco-mouse" technology, if its components had different characteristics? For instance the characteristics of the components of the "Global Positioning System on vehicles" (GPS) novel technology? This kind of question is relevant when managers and policy makers are asked to choose between two novel technologies based on their predicted diffusion. Table 3 reports the characteristics of the combined components that generated the two novel technologies respectively in 1985 (onco-mouse) and in 1986 (GPS).

Table 3: Legitimation and technological potentials of the onco-mouse vs. GPS

	Oncomouse	GPS	Overall (average)	GPS-Oncomouse	ΔLegitimation [years]	ΔTechnological Potential [patents]
Similarity	0	0	0.33	0	0	0
log(Familiarity)	5.08	3.26	5.26	-1.82	+0.31	+12.58
log(Science-Based)	2.25	0.2	0.74	-2.05	-0.47	-37.36
log(Appropriability)	2.57	2.19	2.37	-0.39	+0.38	+10.62
				Total:	+0.22	-14.16

The components that generated the onco-mouse technology were more science based, more familiar to the inventors' community, and more appropriable than those that generated the GPS technology. If the onco-mouse had the same characteristics of the GPS, it would have been characterized by a longer legitimation period (+0.22 years) and by lower technological potential (-14.16 patents). Precisely, the shortening of the legitimation period induced by the lower science-based content (-0.47 years), would have had more than compensated by the increase in the legitimation period due to a lower appropriability (+0.38 years) and familiarity (+0.31 years) characterizing the GPS technology. Similarly, the technological potential would have been reduced by the lower science-based content (-37.36 patents) and this reduction would have been partially compensated by the increase of the technological potential due to the reduction of familiarity (+12.58 patents) and appropriability (+10.62 patents).

Figure 4: Novel technologies' legitimation time increase/reduction obtained by changing the values of the technological component characteristics. The average legitimation time in the study sample equals 5.86 years

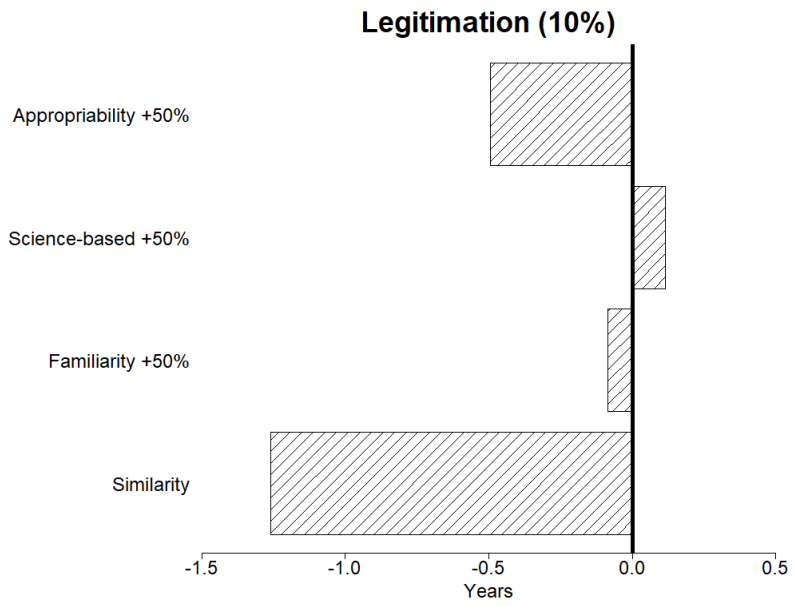
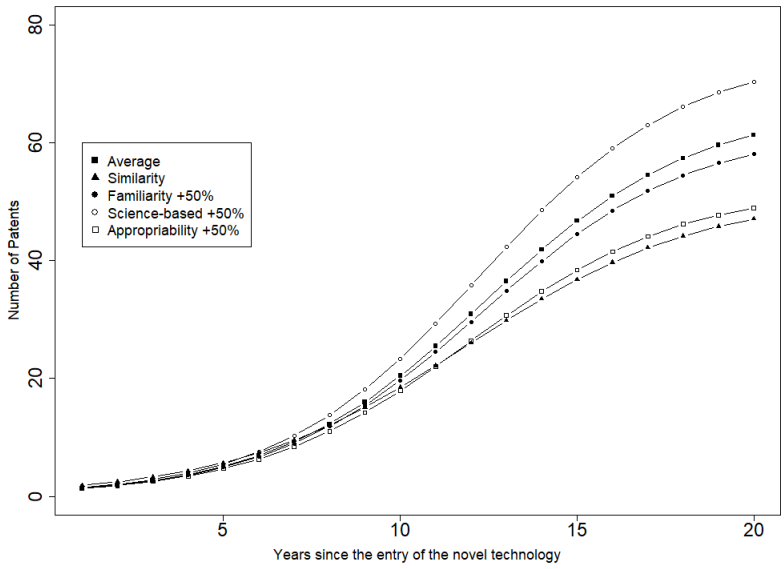


Figure 5: Novel technologies' diffusion curves obtained by changing the values of the technological component characteristics



6. Conclusion

In this paper we investigate if the conditions at the time of the appearance of a novel technology impact on its subsequent diffusion curve. To do that, we conduct an empirical study that traces the diffusion of 11,009 novel technologies over the period 1985-2000. Using patent data, we identify a novel technology as an unprecedented combination of existing technological components, as represented by the technological classes where a patent is classified. Then, we trace all the patents incorporating that novel technology and we estimate the corresponding diffusion curve. Under the assumption that a diffusion curve has an S-shape, we look at the technological, cognitive and institutional characteristics of the combined components affecting the duration of the legitimation phase (first part of the curve) and the potential of a technology (ceiling of the curve).

We find that a high degree of similarity between the combined components, familiarity of the inventors with the components, and high level of appropriability shorten the legitimation time and decrease the technological potentials. The science-based nature of the combined components plays an opposite role and increases the legitimation time, while positively affects the technological potential of a novel technology. By looking at the economic impact of our results, we find that increasing the science-based nature of the combined components significantly augments the technological potential with a limited increase in the legitimation time. Familiarity, has a limited effect both on technological potentials and legitimation. Finally, similarity and appropriability show a trade-off of relevant extent: decreasing the legitimation period has a significant cost in terms of technology potentials (and *vice versa*).

Our results have strategical implications in predicting how the legitimation period and the technological potential vary according to a set of characteristics of the novel technology observed at its appearance, its genetic markers. These predictions might be used in the private sector, when

managers face the choice between two emerging technologies for their businesses, as well as in the public sector, when policy makers are asked to choose, for instance, the technological orientation of a region.

7. References

Achilladelis, B. 1993. The dynamics of technological innovation: The sector of antibacterial medicines. ***Research Policy***, *22*: 279–308.

Arrow, K. 1962. Economic welfare and the allocation of resources for invention. In ***The rate and direction of inventive activity: Economic and social factors***. New Jersey: Princeton University Press.

Arts, S., Appio, F. P., & Van Looy, B. 2013. Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? ***Scientometrics***, *97*: 397–419.

Breschi, S., Malerba, F., & Orsenigo, L. (2000). Technological Regimes and Schumpeterian Patterns of Innovation. **The Economic Journal**, 110(463), 388–410.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive capacity: a new perspective on learning and innovation. ***Administrative Science Quarterly***, 128–152.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. ***Research Policy***, *11*:147–162.

Fleming, L. (2001). Recombinant uncertainty in technological search. ***Management Science***, *47*: 117–132.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. ***Administrative Science Quarterly***, *52*: 443–475.

Geroski, P. A. 2000. Models of technology diffusion. ***Research Policy***, *29*: 603–625.

Griliches, Z. 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. ***Econometrica***, *25*: 501.

Hall, B. H. 2004. *Innovation and diffusion*. Cambridge, Mass.: National Bureau of Economic Research.

Meyer-Krahmer, F., & Schmoch, U. 1998. Science-based technologies: university–industry interactions in four fields. ***Research Policy***, 27 835–851.

March, J. 1991. Exploration and exploitation in organizational learning. ***Organization Science***, 2 71-87.

March, J. G. 2013. ***Handbook of Organizations***. U.K.: Routledge.

Merges, R. P., & Nelson, R. R. 1994. On limiting or encouraging rivalry in technical progress: The effect of patent scope decisions. ***Journal of Economic Behavior & Organization***, *25*: 1–24.

Nelson, R. R., & Winter, S. G. 2004. ***An evolutionary theory of economic change***. Cambridge, Mass.: Harvard Univ. Press.

Schumpeter, J. A. 1939. ***Business cycles***. New York: McGraw-Hill.

Strumsky, D., & Lobo, J. 2015. Identifying the sources of technological novelty in the process of invention. ***Research Policy***, *44*: 1445–1461.

Verhoeven, D., Bakker, J., & Veugelers, R. 2016. Measuring technological novelty with patent-based indicators. ***Research Policy***, *45*: 707–723.

Appendix A: Legitimation as a linear combination of midpoint and slope of an S-curve technology diffusion curve

In this appendix details the mathematical formulation relating legitimation (defined as the time needed for diffusion curve to reach the 10% of its ceiling point) to the midpoint and slope. Starting from Equation 1, we extract *t* as follow:

$$\hat{Y}_t = \frac{ceiling}{1 + e^{(-\frac{(t-midpoint)}{slope})}}$$

$$1 + e^{(-\frac{(t-midpoint)}{slope})} = \frac{ceiling}{\hat{Y}_t}$$

$$-\frac{(t - midpoint)}{slope} = \log(\frac{ceiling}{\hat{Y}_t} - 1)$$

$$-t + midpoint = slope * \log(\frac{ceiling}{\hat{Y}_t} - 1)$$

$$t = midpoint - slope * \log\left(\frac{ceiling}{\hat{Y}_t} - 1\right)$$

Then we set the 10% of the ceiling point, $\frac{\hat{Y}_t}{ceiling} = \frac{1}{10}$ and compute the corresponding t

$$t_{10\%} = midpoint - slope * \log(10 - 1)$$

$$t_{10\%} = midpoint - slope * 2.2$$

Appendix B: List of the technological class dummies of our sample (three digits)

Tech class	Number of patents (whole Patstat)	Legitimation	Potential	Class name
C12	226231	5.40	56.74	biochemistry; beer; spirits; wine; vinegar; microbiology; enzymology; mutation or genetic engineering
C40	4502	6.28	74.89	combinatorial technology
B82	5427	6.19	62.41	nanotechnology
A61	972410	6.19	58.17	medical or veterinary science; hygiene
C07	593444	5.12	53.73	organic chemistry
B81	5593	5.44	59.87	microstructural technology
C13	1402	4.39	30.79	sugar industry
C30	11506	4.91	40.13	crystal growth
A01	100309	5.85	63.02	agriculture; forestry; animal husbandry; hunting; trapping; fishing
G06	333626	6.53	65.50	computing; calculating; counting
G01	344963	5.58	51.31	measuring; testing
A23	61675	5.48	46.07	foods or foodstuffs; their treatment, not covered by other classes
H04	568514	6.83	75.78	electric communication technique
G10	23284	5.76	56.73	musical instruments; acoustics
C01	41421	4.29	47.89	inorganic chemistry
G11	83155	5.11	58.21	information storage
G02	102565	5.86	48.71	optics
H01	424044	6.05	54.34	basic electric elements
C05	5273	4.57	30.47	fertilisers; manufacture thereof
H03	69422	5.77	44.52	basic electronic circuitry
C04	35183	4.63	46.52	cements; concrete; artificial stone; ceramics; refractories
B01	190188	5.30	46.96	physical or chemical processes or apparatus in general
A21	8172	4.82	42.78	baking; equipment for making or processing doughs; doughs for baking
C23	41382	5.25	40.21	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; inhibiting corrosion of metallic material or incrustation in general
C22	34422	4.76	34.53	metallurgy; ferrous or non-ferrous alloys; treatment of alloys or non-ferrous metals
G09	43751	6.19	56.42	educating; cryptography; display; advertising; seals
C08	337146	4.60	38.52	organic macromolecular compounds; their preparation or chemical working-up; compositions based thereon
E21	29316	5.24	47.15	earth or rock drilling; mining
C03	28360	4.56	35.42	glass; mineral or slag wool
C09	132566	5.30	42.99	dyes; paints; polishes; natural resins; adhesives; compositions not otherwise provided for; applications of materials not otherwise provided for
G07	33692	5.32	48.83	checking-devices
C02	23804	5.02	38.74	treatment of water, waste water, sewage, or sludge
G03	75366	5.70	43.36	photography; cinematography; analogous techniques using waves other than optical waves; electrography; holography
G05	42106	5.67	47.14	controlling; regulating
D01	17333	4.65	36.77	natural or man-made threads or fibres; spinning
G21	15183	5.76	36.89	nuclear physics; nuclear engineering
C11	39725	5.19	43.40	animal or vegetable oils, fats, fatty substances or waxes; fatty acids therefrom; detergents; candles
B09	4642	3.61	37.39	disposal of solid waste; reclamation of contaminated soil
A62	10026	3.53	30.05	life-saving; fire-fighting
C25	19093	5.50	42.70	electrolytic or electrophoretic processes; apparatus therefor

B06	2548	6.36	42.86	generating or transmitting mechanical vibrations in general
A63	29203	5.37	51.46	sports; games; amusements
C10	60809	5.02	44.54	petroleum, gas or coke industries; technical gases containing carbon monoxide; fuels; lubricants; peat
B03	5096	6.76	45.44	separation of solid materials using liquids or using pneumatic tables or jigs; magnetic or electrostatic separation of solid materials from solid materials or fluids; separation by high-voltage electric fields
B22	25745	4.66	40.40	casting; powder metallurgy
B32	54028	5.47	44.85	layered products
G08	27261	7.00	60.63	signalling
C06	2871	4.76	44.14	explosives; matches
H05	74628	6.24	50.92	electric techniques not otherwise provided for
A24	5848	6.08	21.92	tobacco; cigars; cigarettes; smokers' requisites
B28	10629	3.89	32.62	working cement, clay, or stone
D21	30511	4.28	39.52	paper-making; production of cellulose
C21	18066	5.00	44.47	metallurgy of iron
B05	43991	5.20	38.17	spraying or atomising in general; applying liquids or other fluent materials to surfaces, in general
F21	43450	7.27	65.79	lighting
B41	71628	5.32	42.43	printing; lining machines; typewriters; stamps
D04	14911	5.36	35.52	braiding; lace-making; knitting; trimmings; non-woven fabrics
B29	118121	4.36	36.21	working of plastics; working of substances in a plastic state in general
H02	117217	5.71	42.32	generation, conversion, or distribution of electric power
B44	6913	5.72	33.29	decorative arts
D02	4015	3.63	32.48	yarns; mechanical finishing of yarns or ropes; warping or beaming
D06	40523	4.71	41.65	treatment of textiles or the like; laundering; flexible materials not otherwise provided for
B64	21484	5.76	38.66	aircraft; aviation; cosmonautics
B24	20637	5.42	50.81	grinding; polishing
B42	9003	5.36	36.31	bookbinding; albums; files; special printed matter
B31	6146	4.19	27.36	making articles of paper, cardboard or material worked in a manner analogous to paper; working paper, cardboard or material worked in a manner analogous to paper
F04	45443	5.94	41.67	positive-displacement machines for liquids; pumps for liquids or elastic fluids
B08	8716	5.09	43.42	cleaning
F17	6312	4.95	30.87	storing or distributing gases or liquids
A41	8219	5.05	37.20	wearing apparel
F41	9474	5.16	42.72	weapons
B27	9313	4.72	35.58	working or preserving wood or similar material; nailing or stapling machines in general
B25	26692	4.96	37.78	hand tools; portable power-driven tools; handles for hand implements; workshop equipment; manipulators
B04	4760	4.38	34.24	centrifugal apparatus or machines for carrying-out physical or chemical processes
C14	1182	1.81	29.61	skins; hides; pelts; leather
B23	73056	5.77	49.40	machine tools; metal-working not otherwise provided for
A43	10643	5.37	41.87	footwear
F01	64495	6.18	51.59	machines or engines in general; engine plants in general; steam engines
F03	14724	5.80	47.15	machines or engines for liquids; wind, spring, or weight motors; producing mechanical power or a reactive propulsive thrust, not otherwise provided for
G04	9336	6.27	35.19	horology
B21	33691	4.14	37.22	mechanical metal-working without essentially removing material; punching metal

A46	4839	3.86	31.61	brushware
B02	6196	3.75	33.50	crushing, pulverising, or disintegrating; preparatory treatment of grain for milling
F25	32049	6.53	43.86	refrigeration or cooling; combined heating and refrigeration systems; heat pump systems; manufacture or storage of ice; liquefaction or solidification of gases
F15	11656	4.17	37.40	fluid-pressure actuators; hydraulics or pneumatics in general
F23	26346	4.58	40.73	combustion apparatus; combustion processes
B67	9452	5.21	32.54	opening or closing bottles, jars or similar containers; liquid handling
F42	6421	5.56	38.51	ammunition; blasting
F27	11853	5.33	41.85	furnaces; kilns; ovens; retorts
F02	113956	6.27	50.83	combustion engines; hot-gas or combustion-product engine plants
B43	3051	5.02	39.64	writing or drawing implements; bureau accessories
B07	5771	3.38	32.54	separating solids from solids; sorting
F26	7296	3.87	26.95	drying
A45	12126	5.66	30.46	hand or travelling articles
E01	13889	4.20	34.12	construction of roads, railways, or bridges
B63	14208	4.62	29.88	ships or other waterborne vessels; related equipment
F28	23852	5.01	31.40	heat exchange in general
F22	3790	8.22	25.98	steam generation
D07	1360	2.72	24.97	ropes; cables other than electric
B65	165572	4.76	37.60	conveying; packing; storing; handling thin or filamentary material
E02	16012	5.45	50.15	hydraulic engineering; foundations; soil-shifting
A44	5202	5.14	48.71	haberdashery; jewellery
D05	2044	2.61	39.50	sewing; embroidering; tufting
B60	176856	6.27	58.58	vehicles in general
A47	59501	5.46	37.05	furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in general
E04	47926	4.74	37.33	building
D03	7892	3.08	47.50	weaving
B62	48784	7.01	53.75	land vehicles for travelling otherwise than on rails
F24	37270	6.76	37.86	heating; ranges; ventilating
F16	191793	5.81	45.69	engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
A22	3717	1.81	28.32	butchering; meat treatment; processing poultry or fish
B30	6065	6.81	31.84	presses
B61	9889	4.79	26.03	railways
B66	18098	5.22	41.57	hoisting; lifting; hauling
B26	14447	3.50	30.29	hand cutting tools; cutting; severing
E03	9308	4.55	42.68	water supply; sewerage
E06	15805	3.43	33.77	doors, windows, shutters, or roller blinds, in general; ladders
E05	37663	6.04	43.50	locks; keys; window or door fittings; safes

Appendix C: Set of regressions with midpoint and slope as dependent variables

In the main text we present a set of regression with legitimation and technological potential as dependent variables. While the later variable has a direct correspondence in one of the parameter of a sigmoid curve proxying a diffusion curve, legitimation is a linear combination of the remaining two parameters, midpoint and slope. Table C1 reports a separate set of regressions with midpoint and slope as dependent variables in order to disentangle the determinants of the two parameters linearly combined to obtain legitimation. Note that the coefficients estimated for each variable in Table 1, columns 5 and 6, can be calculated according to Equation 4 (see Appendix A for further details).

$$\hat{\beta}_{Legitimation} = \hat{\beta}_{midpoint} - \hat{\beta}_{Slope} * 2.2$$

(Equation 4)

Table C1: Regression results. OLS estimations using as dependent variables Midpoint and Slope and adopting the same specifications reported in Equation 1 and Equation 2.

VARIABLES	(1) Midpoint (50%)	(2) Slope at midpoint
Technological component characteristics		
Similarity 3 digits (dummy)	-0.48***	0.35***
log(Science-based component)	-0.20***	-0.20***
log(Familiarity)	-0.12***	0.022**
log(Appropriability)	-0.83***	0.074
Inventor's characterisics		
Inventors' experience (dummy)	-0.13**	0.031
log(Inventors' team size)	-0.078*	-0.0072
Applicant characteristics		
log(Applicants' experience)	-0.028***	-0.0081**
University applicant (dummy)	-0.30**	-0.21***
More than one applicant (dummy)	-0.27***	0.27***
log(Concentration)	-0.069	-0.051***
Dummy Applicant's country	yes	yes
Dummy technology entry year	yes	yes
Dummy Technology Class (3 digits)	yes	yes
Constant	18.1***	2.37***
Observations	11,009	11,009
R-squared	0.339	0.259

Appendix D: Elasticity estimation relying on an alternative functional form (log-log)

VARIABLES	(1) Log(Legitimation)	(2) Log(Technological potentials)
Technological component characteristics		
Similarity 3 digits (dummy)	-0.28***	-0.12***
log(Science-based component)	0.027**	0.12***
log(Familiarity)	-0.021***	-0.032***
log(Appropriability)	-0.089*	-0.17***
Inventor's characterisics		
Inventors' experience (dummy)	-0.047***	0.00077
log(Inventors' team size)	-0.024**	0.036***
Applicant characteristics		
log(Applicants' experience)	-0.0030	0.0040
University applicant (dummy)	0.042	0.071**
More than one applicant (dummy)	-0.21***	0.070***
log(Concentration)	-0.010	0.088***
Dummy Technology class (3 digits)	yes	yes
Dummy Applicant's country	yes	yes
Dummy technology entry year	yes	yes
Constant	2.71***	5.08***
Observations	11,009	11,009
R-squared	0.183	0.211