LEM
WORKING PAPER SERIES

When authors become inventors: an empirical analysis on patent-paper pairs in medical research

Arianna Martinelli [a]
Elena Romito [a]

[a] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy.

When authors become inventors: an empirical analysis on patent-paper pairs in medical research

Arianna Martinelli[1] and Elena Romito[1]

[1]Scuola Superiore Sant'Anna - Istituto di economia

July 2019

Abstract

This paper investigates the effect of patenting on follow-on knowledge in cancer research. Using a difference-in-difference approach on an original dataset of patent-paper-pairs we are able to estimate the causal effect of the granting of a patent on scientific development in the same domain. Furthermore, we disentangle between private companies and universities in order to assess whether patenting impacts differently on the two groups. In addition, we study to which extend the degree of applicability of an innovation is further affects the relation. To address these issues we build a novel dataset matching patent data (retrieved from USPTO) and publication data (retrieved from Thompson-Web of Science). Results show that patenting reduces the rate of citations of the paired publication indicating a decrease of related scientific activity only in case the citing agent belongs to a public institution. In addition, the the more invention is applied, the weaker is the negative effect. This paper makes a contribution to the debate on IPR and economics of science.

1 Introduction

The importance of basic research for the development of medical innovation is in no doubt. One might wonder what would have happened if Vincenzo Tiberi, who discovered the chemotherapeutic power of some kinds of molds 35 years before Fleming officially invented Penicillin, patented his discovery? How would the pace of subsequent scientific research have changed? Would Fleming have focused on a different research area? These types of questions are even more relevant relevant today in light of evidence that the patenting of fundamental medical discoveries is more widespread than in the past (Thursby and Thursby, 2011; Merz and Cho, 2005). To address this topic we develop a thought experiment to analyze the counterfactual case. Patent and publications data provide the context for this experiment.

Since the influential work of Vannevar Bush (1945) on the linear model of innovation, several studies have theoretically addressed and empirically tested (Mansfield, 1998; Nicholson, 2012) the hypothesis that basic research is a critical ingredient in closer-to-the-market research (e.g. applied research). However, few works study the opposite direction - from downstream to upstream research. Notable exceptions are studies of patent-paper pairs (PPP) (Huang and Murray, 2009; Murray, 2002; Murray and Stern, 2007) which provide a means to precisely identify pieces of knowledge disclosed first in publications and later protected by patents.

Traditionally, basic and applied research had different motivations, and were associated to different means of appropriation. Universities and publicly funded research institutions usually involved in upstream research, were less likely to protect their inventions using intellectual property rights. Instead, the practice of patenting became widespread at the industry level Partha and David (1994); Murray (2010).

Currently, due to changes in both the legal environment and routines, the distinction

between *public* and *proprietary* knowledge is less pronounced (Eisenberg, 1996) and more research on the potential effect that the practice of granting patents has on the follow-on rate of basic knowledge becomes of crucial importance.

This paper contributes by determining how patenting affects follow-on scientific research in the same area. Our interest is in understanding whether the practice of using intellectual property rights (IPRs) to protect an innovation disclosed initially on a public platform is helpful or detrimental to future cumulative research activity. We are interested also in identifying whether patenting has a greater effect on private companies or on universities.

While industry has for long been characterized by patenting and licensing activity, universities and public research centers tend to be more open and known for their scientific disclosure (Partha and David, 1994). However, there seems to be trend towards universities developing dedicated structures to facilitate relations and interactions with the industry (i.e. technological transfer offices) and to manage their increasing patent portfolios (Baldini et al., 2014; Wu et al., 2015). Nevertheless, lack of organizational models related to dealing with the increasing privatization of research output, research material and tools results in higher transaction costs associated to accessing science for universities rather than commercial firms. In this paper, we try to assess the effect of these possible differences disentangling the effect of patents on follow-on scientific developments carried out by private and public actors. Also, we study how different actors might be differently affected by the granting of a patent depending on the degree of applicability (i.e. closeness-to-the market) of the innovation.

The paper focuses on a rather specific field of medical innovation - research on the detection of cancer. There are two reasons for this choice. First, research on medical conditions and therapeutics are relevant at both the political and economic levels and also have ethical implications. Second, medical research processes are most likely to be

3

protected by patenting of pieces of knowledge related to basic information rather than applicable and easily marketable knowledge (Holman and Munzer, 2000).

We use a novel dataset of PPP (i.e. patents and publications with common content) as suggested first by Murray (2002). Patents granted by the USPTO (United States Patent and Trademark Office) between 2004 and 2011 in the technological class related to Detecting cancer (class 435/6.14) were manually matched to publications included in the WoS (Web of Science) on the basis of correspondence between inventors and authors, patent application dates and publication period, patent and paper abstracts. Additional patent information was retrieved from the European Patent Office (EPO)-PATSTAT Database (October 2014). The starting patent sample includes 1,652 patents, of which 373 are associated to a publication. This latter sample is the set of *paired-patents* used for the empirical analysis in this paper.

Following the identification strategy in Huang and Murray (2009) we can treat the time of patent granting as an exogenous shock which allows us to measure the effect of patent granting on the production of subsequent new knowledge (measured as citations to the paired publication) and to isolate the effect of making *proprietary* an innovation that previously was *public*. A reaction in the number of annual citations received by the publication after the granting of the corresponding patents indicates a change in the rate of follow-on public knowledge.

This work contributes to the growing literature on the overall desirability of a patent system. While patents generally are considered necessary to reward R&D investment to incentivize innovation activity, their potential adverse effects such as the possible stifling of scientific research need also be evaluated.

The paper is organized as follows. Section 3.2 reviews the literature and formulates some hypotheses; section 3.3 describes the empirical strategy and the data; section 3.4

presents the results; and section 3.5 concludes the paper.

2   Theory, contribution and hypotheses

2.1   Scholars at risk of infringement

The legal research exemption[1] allows researchers to use proprietary inventions without infringing the monopolistic rights of the patent holder. While United States patent law provides this possibility, its real applicability and scope are quite limited (Dreyfuss, 2004).

These limitations have been revealed in certain high profile decisions such as the *Roche Products Inc v. Bolar* and the *Madey v. Duke University*. In the first case, the court narrowed the scope of the exemption hugely indicating that it was limited to experiments "for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry" and did not extend to use for business reasons. In the second case, the court specified further that "the profit or non-profit status of the user is not determinative". The consequence of these decisions is that scientists and universities could be sued for patent infringement if they use proprietary technologies in their research. Therefore, they need to negotiate formal access via licensing agreements.

Other landmark court cases have further defined the boundaries between scientific research and patent holders. One of the most prominent examples is the *Myriad vs. University of Pennsylvania* case where the commercial company Myriad sued the University of Pennsylvania for patent infringement[2]. This case was brought to stop the universitys research and testing of patient samples for BRCA1 (Breast Related Cancer Antigen 1) patented by Myriad.

Other universities have faced similar scenarios in cases brought against them by DuPont.

---

[1]See Dent et al. (2006) for a review on research exemptions in different OECD legislations.
[2]See Williams-Jones (2002) for an account.

Harvard University patented the called OncoMouse and licensed it exclusively to DuPont in 1988. Initially, DuPont did not aggressively enforce patent protection in the case of academic institutions. However, over time, more and more universities have been involved in legal cases with DuPont for OncoMouse patent infringements.

In addition, there have also been cases where informal settlement agreements have *de facto* contributed to establishing the principle that scientists are obliged to respect the conditions related to the patenting of a piece of knowledge which is necessary for their academic research. An example is the case of Miami Childrens Hospital (MCH) and Pennsylvania University. In this instance, the source of the discord was a genetic test for the detection of Canavan disease (a disorder which damages the ability of nerve cells in the brain to send and receive messages). Specifically, the MCH allowed other institutions who agreed to sign a license contract, to perform this test. Although, initially Pennsylvania University refused to pay for the MCH license, following a settlement agreement drafted by a MCH advisor, Pennsylvania University was forced to sign a license contract and to pay royalties ($12.50 each) for tests they had performed in the past (Colaianni et al., 2010).

These cases show that the scientific exception does not protect scholars performing research, and their perceived fears of infringement are justified (Reitzig et al., 2007). In practice, this means that scientists also need to gain access to the proprietary technologies (e.g. research tools, materials) they need to use. Some studies show that bargaining for access represents a burden especially in the case of biomedical research, and delays research projects (Eisenberg et al., 2001). This problem is exacerbated in technological domains (pharmaceuticals and biotech) where patents are used increasingly.

The empirical evidence suggests that lower costs for accessing existing research have a positive effect on subsequent research and increase exploratory research (Murray et al., 2016). A similar positive effect occurs if specific institutions such as biological resource

centers provide easier access to biological material (Furman and Stern, 2011). Following the seminal work by Murray (2002) who introduced the notion of PPP related to tissue engineering, Huang and Murray (2009) use PPP and find that the patenting of basic innovations related to genetic research reduces the degree of follow-up public innovation in the same field by about 17%. In a study with a similar setting but employing also a control group of publications from the same journal, Murray and Stern (2007) find that citation rate after the patent grant declines by approximately 10 to 20 percent.

In a very recent paper, Sampat and Williams (2019) present novel evidence that on average gene patents do not impose an economically meaningful reduction on follow-on scientific and commercial investments. They interpret their result as a sign that the market for cross-licensing operates "at least somewhat efficiently" (Sampat and Williams, 2019, p.227). While they do not have sufficient heterogeneity of the "for-profit" and "not-for-profit" status of the contributors to follow-on innovations to test their proposition, they suggest that there might be some differences. These differences could be related to the transaction costs faced by these two types of actors and their different capabilities in engaging in the complex licensing process (e.g. different budget, less capable university technology transfer offices). Based on this review, we expect that:

HB1: The granting of a patent negatively affects further scientific research carried out by public institutions

HB2: The granting of a patent does not affect further scientific research carried out by private companies

2.2 The role of applicability

Willingness to prosecute infringers might differ depending on the type of invention and its degree of applicability. Stokes (1997) observes that knowledge has both basic and applied

7

value and that can be produce with and without any consideration of use. Fundamental scientific research might not have any commercial applications (i.e. being in the so-called "Bohr's Quadrant" in Stoke's model) or might be use-inspired basic research (i.e. being in the so-called "Pasteur's Quadrant" in Stoke's model). This differentiation is reflected in the choice of different appropriation means and also in different reactions to increase of formal appropriation through patenting Murray and Stern (2007).

According to Walsh and Arora (2003) patenting in biotech does not affect scientific research in the same area. The large majority of the industry respondents to the authors survey declared they would be unwilling to sue university researchers conducting noncommercial research activity. Biotech firms benefit from the biotech research conducted by academics, and legal procedures against universities may be associated to bad publicity for the firm. The survey revealed also that patents are less likely to be enforced by industry if the scientist is involved in research that is far from commercialization, and the patent owner is interested in maintaining its position in the "research business".

The three court cases discussed in the previous subsection are in line with this reasoning. The OncoMouse patent dispute marked a relatively new trend in being the first case of relatively aggressive patent enforcement by a private firm against a broad set of universities[3]. According to Walsh and Arora (2003), DuPonts change of attitude coincided with the decision to abandon its "molecular research business". This made the company more inclined to challenge the universities since its ambition to contribute to the stream of molecular research had dissipated.

The disputes over the BRCA1 and Canavan disease were related to two rather similar innovations which are genetic tests. While these diagnostic tools could be used also to perform scientific research, they were already a marketable product, and therefore more

---

[3]For a complete overview on the timeline of events that characterize the OncoMouse incident see Murray (2010).

likely to be protected by a commercial firm. Although the field of analysis for our study is quite narrow, our patents cover different types of inventions such as the discovery of a new protein sequence and the introduction of a new molecular entity. In line with Walsh and Arora (2003), we expect that the higher the level of applicability, the stronger the impact on the stream of follow-up scientific knowledge. Therefore our additional hypotheses are:

H1: The more applied the scientific discovery, the more negative the relationship between patent award and further scientific research carried out by a public institution

H2: The more applied the scientific discovery, the more negative the relationship between patent award and further scientific research carried out by a private company

3 Methodology and Data

3.1 Building *patent-paper pairs*

To test our hypotheses, first, identify the PPP in the field of cancer detection. Following Huang and Murray (2009), we start with the sample of patents[4]. We use the current U.S. patent classification[5] and the USPTO Patent Full-Text and Image Database (PatFT) to retrieve all USPTO granted patents in the class Detection of cancer (i.e. 435/6.14 class). This class includes 3,618 patents and we focus the analysis on the 1,652 patents granted between 2004 and 2011.

The publications matching these patents are retrieved from Clarivate AnalyticsWoS database. Despite some potential bias towards English-language titles, lack of citation counts for books, and differences in coverage among research fields (especially arts and humanities), most publications data based studies rely on the WoS database (Meho and Yang, 2007). Also, the use of WoS to extract the articles ensures inclusion of research published

---

[4]Note that the pair construction could have started with the sample of papers (Murray, 2002).

[5]We use the field "Current U.S. classification" on `http://patft.uspto.gov/netahtml/PTO/search-adv.htm`

in the most prestigious and influential journals. Note that, although the USPTO and the WoS databases include very detailed information on respectively patents and papers, these are independent of one another. Therefore, construction of the pairs was done manual by matching each of the 1,652 patents in the initial sample.[6]

The manual matching involved three steps. The first is related to the people involved in the inventive process. For each patent inventor, we searched for whether they were listed as an author on any of the publications included in the WoS database. We used a rather restrictive definition according to which all patent inventors must also be an author[7]. The second step relates to the lag between patent application and paper publication. For those cases where the author matching condition was satisfied, we proceeded to compare patent application and paper publication dates; we require that they must not be more than two years[8] apart. This temporal correspondence is fundamental for the empirical strategy presented below. The third step in the matching procedure relates to topic. To ensure correspondence of topics, we checked that patent abstracts and paper abstracts matched. We used keyword search to confirm also that patent and paper contents were the same. From the initial set of 1,652 patents, we identified 373 patents related to knowledge disclosed previously in a scientific journal publication. In what follows, we consider only this set of PPP.

---

[6]Magerman et al. (2015) detect PPP using text-mining algorithms. Their techniques identified a rather small number of pairs: from an initial 948,432 scientific publications in biotechnology they identified 584 PPP.

[7]Note that the reverse relation does not hold: a papers authors may include more individuals than the number of the patents inventors.

[8]Wherever applicable we compare patent priority date to publication date. We identified some cases where despite a considerable lag between USPTO application date and paper publication, the patent reflected the content of the paper. When we investigated further, we found a correspondence between priority date (rather than application date) and paper publication date. This applied particularly to non-American inventors who applied for a patent in another country before applying for a USPTO patent. The USPTO database provides filing dates but refers also to applications filed in different legislations where the USPTO application claims priority. Therefore, we also consider suitable matched patents with a priority data no more than two years different from the data on the publication.

3.2 Empirical strategy and econometric model

To test our hypotheses, we developed an econometric model to isolate the effect of being granted a patent on the rate of follow-on public knowledge. Following the literature (de Solla Price, 1965; Posner, 2000), we use number of yearly citations received by a (paired) publication to measure follow-on public knowledge. To understand the rationale behind our identification strategy, we would highlight two points. First, the so-called grant lag identifies the time elapsed between date of filing the patent application and date of patent award. Since there is correspondence between patent application and paper publication, the grant lag identifies the period when the new knowledge is freely available and usable. After a patent has been granted, it becomes risky (in terms of infringement) for scientists to continue to exploit the knowledge without obtaining a license. Second, the average grant lag in our sample is 4.5 years with considerable heterogeneity ($\sigma = 755.23$). Figure 1 is a graphical representation of the grant lag distribution in our sample.

The high variation in grant lags allows us to consider award of a patent as an exogenous shock to the corresponding publication during "its life". In this setting we can implement a difference-in-difference model where the annual citations received by papers associated to patents with longer grant-lags constitute the "control" for citations to a paper associated to a patent with a shorter patent lag. Figure 2 provides a graphical representation of the identification strategy.
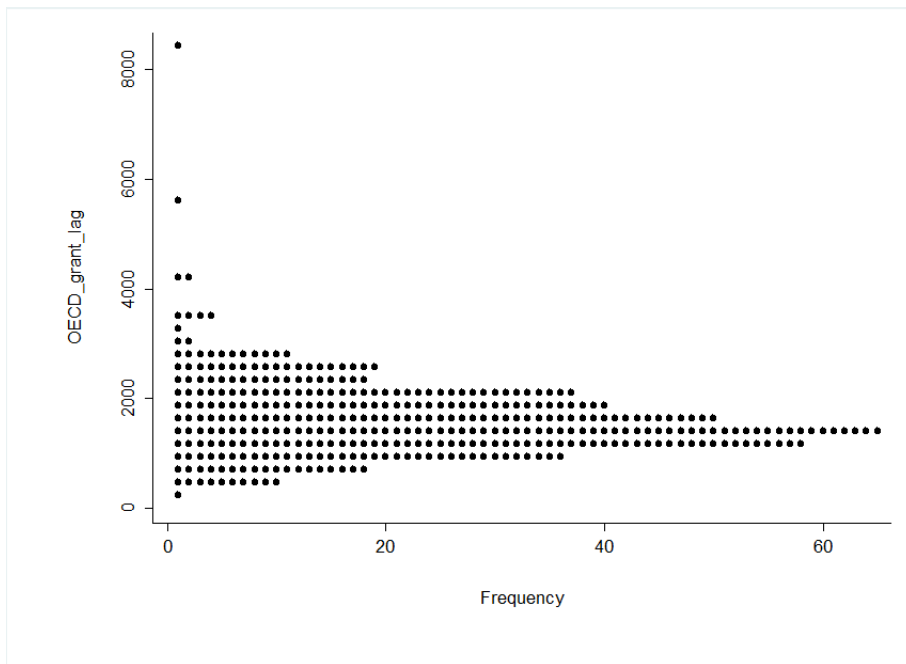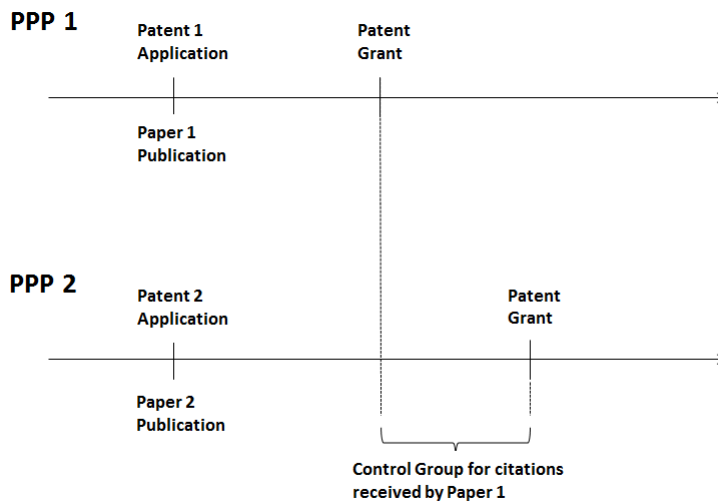
11

Figure 1: Distribution of grant lag

Figure 2: Scheme of the patent-paper pair identification strategy

Our econometric models to test the baseline hypothesis (HB1 and HB2) are:

$$CIT\_PUBLIC_{i,t} = \alpha + \beta_0 PATENT\_WINDOW_{i,t}+$$
$$+\beta_1 PATENT\_IN\_FORCE_{i,t} + \gamma_i + \delta_t + \epsilon_{it} \tag{1}$$

$$CIT\_PRIVATE_{i,t} = \alpha + \beta_0 PATENT\_WINDOW_{i,t}+$$
$$+\beta_1 PATENT\_IN\_FORCE_{i,t} + \gamma_i + \delta_t + \epsilon_{it} \tag{2}$$

where $CIT\_PUBLIC_{i,t}$ is the number of citations received by the paired publication i in year t from publications whose authors affiliation is a public research center, and $CIT\_PRIVATE_{i,t}$ is the number of citations received by paired publication i in year t from publications whose authors' affiliation is a firm. The variable $PATENT\_IN\_FORCE_{i,t}$ is a dummy variable that is equal to 1 for all the years t when the patent associated to the publication i is valid (i.e. in force). $PATENT\_WINDOW_{i,t}$ is a dummy variable that is equal to 1 for the year t in which the patent associated to publication i was granted. Finally, (γ_i) is the set of paired publication fixed effects, (δ_t) is a set of year fixed effects, and (ϵ_{it}) is the error term.

To test our hypotheses H1 and H2 on the role of knowledge applicability, we expand equations 1 and 2 by interacting our primary variable of interest $PATENT\_IN\_FORCE_{i,t}$ with the variable $APPLIED\_KNOWLEDGE_i$. We therefore estimates:

$$CIT\_PUBLIC_{i,t} = \alpha + \beta_0 PATENT\_WINDOW_{i,t}+$$

$$+\beta_1 PATENT\_IN\_FORCE_{i,t} + \beta_2 APPLIED\_KNOWLEDGE_i+$$

$$+\beta_3 APPLIED\_KNOWLEDGE_i * PATENT\_IN\_FORCE_{i,t}+ \qquad (3)$$

$$+\gamma_i + \delta_t + \epsilon_{it}$$

$$CIT\_PRIVATE_{i,t} = \alpha + \beta_0 PATENT\_WINDOW_{i,t}+$$

$$+\beta_1 PATENT\_IN\_FORCE_{i,t} + \beta_2 APPLIED\_KNOWLEDGE_i+$$

$$+\beta_3 APPLIED\_KNOWLEDGE_i * PATENT\_IN\_FORCE_{i,t}+ \qquad (4)$$

$$+\gamma_i + \delta_t + \epsilon_{it}$$

While β_1 represents the effect of an enforceable patent; β_3 represents the additional effect related to the degree of knowledge applicability.

All the models are estimated using ordinary least square (OLS) estimations with high dimensional fixed effects (Guimares and Portugal, 2010). While our dependent variables are count variables, we chose OLS rather than a non-linear specification of the model since it causes no convergence problems when the entire set of fixed effects is included into the analysis. Given the importance of fixed effects in our econometric design, we therefore rely on the OLS model. Furthermore, linear estimations provides easier interpretations for interaction terms than non linear model (Karaca-Mandic et al., 2012). As a robustness check, we performed a Poisson quasi maximum likelihood (PQML) analysis. All 373 paired publications i in our set are observed from year of publication up to 2015. Since their year of publication differs, they are observed for different numbers of year, and the resulting panel dataset is unbalanced. The final dataset includes 4,455 observations, and average

time length of 7.17 years.

3.3 Variables

To test our hypotheses, we included the following variables.

Dependent variables

$CIT\_PUBLIC_{i,t}$ is the number of citations received by the paired publication i in year t from publications whose all author affiliations are a public research center. While citations to scientific articles is a well-established measure for follow-on public knowledge, this variable distinguishes subsequent scientific research developed in a public center.

$CIT\_PRIVATE_{i,t}$ is the number of citations received by the paired publication i in year t from publications whose all author affiliations are a private firm. Similar to the previous variable, this allows us to measure the level of follow on scientific research conducted by private entities.

All the dependent variables were computed based on information retrieved from Clarivate Analytics WoS database.[9]

Main variables of interest

$PATENT\_IN\_FORCE_{i,t}$ is a dummy variable that is equal to 1 for all the years t in which the patent associated to the publication i is in force. This variable corresponds to the treatment in our empirical setting. Therefore, this variable coefficient is especially important.

$PATENT\_WINDOW_{i,t}$ is a dummy variable that is equal to 1 for the year t in which the patent associated to publication i was granted. This variable captures the immediate

---

[9]Note that this information is not readily available in the Clarivate Analytics WoS database. Therefore, we reconstructed the citing publications using the Python PyAutoGUI package.

effect of the award of a patent on a piece of knowledge previously disclosed only on a public platform.

$APPLIED\_KNOWLEDGE_i$ is a dummy variable which takes the value 1 if the paired paper i is published in an applied journal according to the CHI-classification (Hamilton, 2003). The CHI classification provides four levels of applicability from clinical observation (level 1) to basic biomedical information (level 4); we set our variable equal to 1 if the corresponding journal belongs to "clinical mix" journals (level 2) and equal to zero if the journal belongs to the "clinical investigation" (level 3 or level 4).

Controls

$PAPER\_AGE_{i,t}$ is the age of the paired publication i in year t, that is the number of years since the paper was published. Since we expect a nonlinear relation between citations received and age due to the normal progress of science, we add the fourth order polynomial expansion.

Tables 1 and 2 present the summary statistics and the correlations, respectively.

Table 1: Summary statistics

	count	mean	sd	min	max
CIT_PUBLIC	4455	9.657	24.22	0	306
$CIT\_PRIVATE$	4455	0.535	1.666	0	28
PAT_WINDOW	4455	0.0837	0.277	0	1
$PATENT\_IN\_FORCE$	4455	0.513	0.500	0	1
$PAPER\_AGE$	4455	6.170	4.493	0	29
$APPLIED\_KNOWLEDGE$	4176	0.459	0.498	0	1
Observations	4455				

16

Table 2: Cross-correlation table

Variables	CIT_PUBLIC	CIT_PRIVATE	PAT_WINDOW	PATENT_IN_FORCE	PAPER_AGE	APPLIED_KNOWLEDGE
CIT_PUBLIC	1.000					
CIT_PRIVATE	0.763	1.000				
PAT_WINDOW	0.023	-0.002	1.000			
PATENT_IN_FORCE	-0.022	-0.084	-0.310	1.000		
PAPER_AGE	-0.024	-0.090	-0.091	0.712	1.000	
APPLIED_KNOWLEDGE	-0.043	-0.050	0.024	-0.055	-0.102	1.000

4 Results

Table 3: Estimation for the baseline hypothesis

DEPENDENT VARIABLES:	CIT_PUBLIC		$CIT\_PRIVATE$	
	(OLS)	(PQML)	(OLS)	(PQML)
	(1)	(2)	(3)	(4)
$PATENT\_IN\_FORCE$	-2.269***	-0.239***	-0.006	-0.117
	(0.660)	(0.058)	(0.075)	(0.121)
PAT_WINDOW	-1.228*	-0.159***	-0.121	-0.290**
	(0.619)	(0.034)	(0.070)	(0.102)
$PAPER\_AGE$		0.460***		0.377***
		(0.027)		(0.048)
$PAPER\_AGE^2$		-0.051***		-0.060***
		(0.004)		(0.008)
$PAPER\_AGE^3$		0.001***		0.002***
		(0.000)		(0.000)
$Constant$	10.960***		0.550***	
	(0.394)		(0.045)	
Observations	4439	4375	4439	3266
R^2	0.869		0.644	
ll		-9484.1		-2206.5

Note: Column 1 and 3 are estimated using a linear model with multiple levels of fixed effects. Column 2 and 4 are estimated using the Poisson Quasi-Maximum Likelihood model. Standard errors are reported in parenthesis.
Legend:* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 3 presents the results related to our baseline hypotheses (HB1 and HB2). Columns 1 and 3 present the estimates of the models described in equations 1 and 2, respectively. The coefficient of the variable $PATENT\_IN\_FORCE$ reported in column 1 is negative and significant, indicating that award of the patent has a negative effect on the rate of follow-on

scientific research by a public institution. In particular, we find a decline in the expected number of annual citations of about 2.27. Since the average number of citations is 9.657, this corresponds to a reduction of 23.5% (significant at the 0.01% level)[10]. This supports hypothesis HB1 confirming that the "privatization" of formerly freely available knowledge stifles further scientific contribution from public actors. The coefficient of the variable *PATENT_IN_FORCE* reported in column 3 is negative but not statistically significant. In line with hypothesis HB2, we find that scientific contribution by private comapanies is not affected by the patent granting. These results confirm Sampat and Williams (2019) intuition about potential differences in licensing efficiency between non-profit and commercial firms. In particular, while the sudden introduction of "fencing" against scientific knowledge does not affect its use by commercial firms it is detrimental to research conducted in universities and public research centers.

Columns 2 and 4 show that these results are robust to a different estimation model (i.e. PQML). Since our dependent variables are count variables and are highly skewed, the PQML model is a suitable choice (Gourieroux et al., 1984; Cameron and Trivedi, 2013).[11] Since not all the fixed effects can be included in the PQML estimation, we add the polynomic terms of the variable *PAPER_AGE*. Together with the paper fixed effects, these variables control for temporal specifications. The results of the PQML estimations are in line with those obtained using OLS in terms of both significance and magnitude.

Table 4 shows the results related to hypotheses H1 and H2 obtained by estimating equations 3 and 4, respectively. The interaction term *PATENT_IN_FORCE* *AP-PLIED_KNOWLEDGE* measures the additional effect that the patenting of more applied

---

[10]Note that the magnitude of the coefficient is in line with Huang and Murray (2009) who estimates an impact of 17%

[11]As explained in Section 3.2 we focus on OLS because for a valid empirical strategy we need to include the whole set of fixed effects reported in equations 1 and 2. However, the PQML does not allow this because of non-convergence problems.

Table 4: Estimation results on the role of knowledge applicability

DEPENDENT VARIABLES:	CIT_PUBLIC (OLS) (1)	CIT_PRIVATE (OLS) (2)
PATENT_IN_FORCE	-3.175*** (0.775)	-0.183* (0.0877)
PATENT_WINDOW	-1.248 (0.660)	-0.154* (0.0747)
PATENT_IN_FORCE*APPLIED_KNOWLEDGE	1.519* (0.634)	0.311*** (0.0718)
Constant	11.46*** (0.421)	0.596*** (0.0476)
Number of observations	4160	4160
R^2	0.871	0.649

Note: Estimations are performed using a linear model with multiple levels of fixed effects. Standard errors are reported in parenthesis.
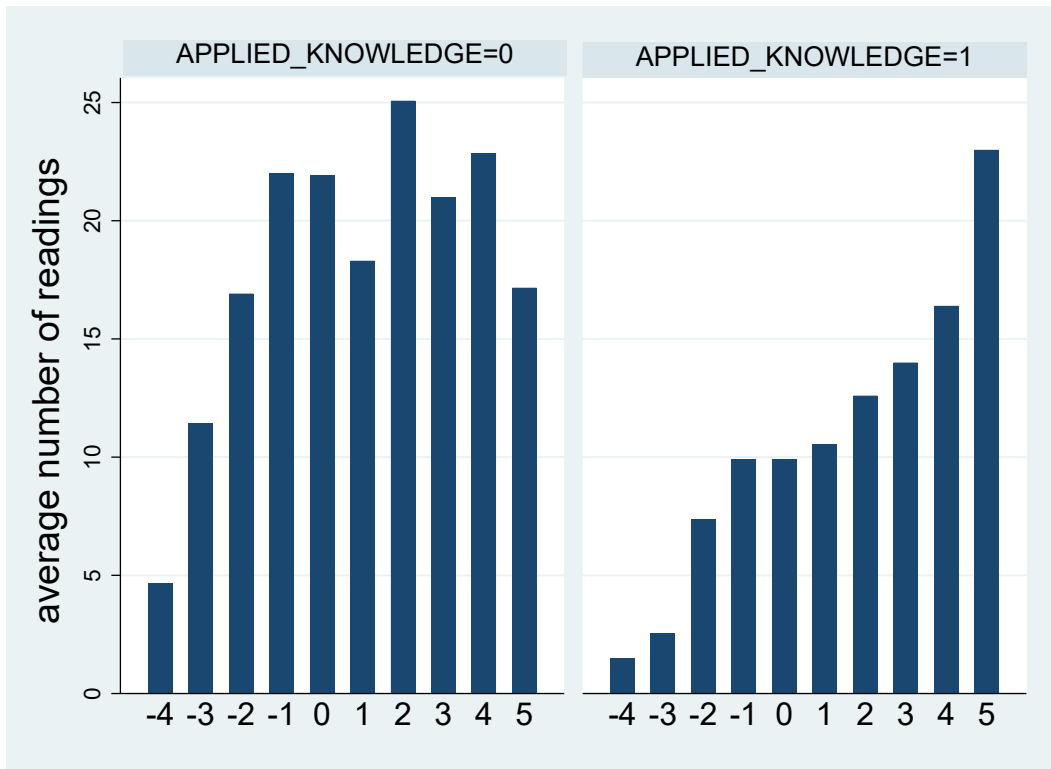Legend:* $p<0.05$, ** $p<0.01$, *** $p<0.001$

knowledge might have on subsequent scientific developments. The coefficient of the interaction term reported in table 4 column 1 is positive and significant. This means that if the scientific knowledge of a publication is more applied the negative effect of patenting on further scientific knowledge developed by public research centers is mitigated. Since in table 4 column 2 the interaction term is also positive and significant this result is confirmed for subsequent scientific research developed by private firms. These results reject hypotheses H1 and H2 and call for further explanation. One of the foundations to patent economic theory is that patents encourage disclosure, and more generally, generate rapid and wide diffusion of technical information on the most recent inventions (Machlup, 1958). In line with this theory, Mazzoleni and Nelson (1998) suggest that universities may also publicize potentially commercial research through their patents. In the case of PPP, patent visibility might spill over to publication generating an "advertising effect". In our setting, more applied publications might benefit from this "advertising effect" of patents and result in an increased number of citations. To test this, we use Scopus data provided by PlumX Metrics[12] to retrieve the number of annual accesses to papers to proxy for the evolution of publication visibility over time. Data availability constrains to this check to the sub-sample of publications paired to a patent granted in 2011 or 2012[13].

Figure 3 depicts the average number of reads received by a scientific publication before and after the patent grant data in the respective cases of applied and non-applied research. The X-axis indicates the number of years since patent grant data both backward and forwards, and zero is the moment of the patent award. Figure 3 shows that the visibility of non-applied papers does not change significantly after a patent has been granted; however, the reverse is true for applied publications. This confirms Mazzoleni and Nelsons (1998)

---

[12]See https://blog.scopus.com/topics/plumx-metrics

[13] This corresponds to 81 publications (25% of the sample)Data availability constraints is due to missing values for older periods.

Figure 3: Visibility differences

intuition about the role played by patents as advertising mechanisms for the paired publication. Even more interesting is that we observe an evident difference in visibility between the two groups even before the patent is granted. This could be interpreted as a "freedom to advertise effect", rather than a simple advertising effect. Although the scientific publication should play an important role in disclosure, it would seem that the difference in visibility across the two groups could be driven by the tendency to hide pieces of applied knowledge before the patent is awarded. The need of hiding vanishes as knowledge becomes protected by a formal IPR and the inventor becomes free to advertise his invention with no-risk of copying. Further checks on this interpretation would require more systematic data on visibility, including presentations at scientific conferences before and after patent award. While am interesting direction for further research, this is beyond the scope of the present paper.

4.1 Robustness checks

In this section we propose two robustness checks to strengthen our analysis. Table 5 columns 1 and 2 report the OLS and PQML estimations of our baseline with, as dependent variable, the number of citations received by public-private collaborations. The variable $CIT\_COLLABORATIVE_{i,t}$ is the number of citations received by the paired publication i in year t from publications with authors from both private and public organizations. Since this can be considered an intermediate case between the two main dependent variables ($CIT\_PUBLIC_{i,t}$ and $CIT\_PRIVATE_{i,t}$) we would expect an intermediate result. Table 5 column 1 shows a negative effect, although significant only at 10% of confidence level. The size of the coefficient is smaller and considered the average (2.55) results in a 9.3% decrease in the number of received citations. The same result is obtained with the PQML estimation.

Table 5: Robustness check on baseline hypothesis

DEPENDENT VARIABLES:	$COLLABORATIVE\_CITATIONS$	
	(OLS)	(PQML)
	(1)	(2)
$PATENT\_IN\_FORCE$	-0.393	-0.198***
	(0.217)	(0.0596)
$PATENT\_WINDOW$	-0.185	-0.143***
	(0.203)	(0.0432)
$PAPER\_AGE$		0.477***
		(0.0313)
$PAPER\_AGE^2$		-0.0529***
		(0.00455)
$PAPER\_AGE^3$		0.00153***
		(0.000220)
$Constant$	2.783***	
	(0.130)	
Number of observations	4439	4187
R^2	0.835	
ll		-4993.9

Column 1 is estimated using a linear model with multiple
levels of fixed effects. Column 2 is estimated using the Poisson
Quasi-Maximum Likelihood model. Standard errors are
reported in parenthesis.
Legend:* $p<0.05$, ** $p<0.01$, *** $p<0.001$

The second robustness check uses an alternative measure of applicability. In the main analysis, we rely on the journal classification to distinguish between applied and non-applied knowledge. Table 6 repeats the testing of H1 and H2 using a different measure of applicability built on patent indicators rather than publication indicators. We construct the variable $LOW\_PRIVATE\_NPL_{i,t}$ that is a dummy variable equal to 1 if the granted patents assignee is a private firm, and the number of non-patent literature (NPL) references is below the median sample. This variable is used as an alternative measure of patent applicability and is used in the robustness checks. The source for this variable is the EPO-PATSTAT database (October 2014). Table 6 columns 1 and 2 confirm that subsequent scientific research developed at both public and private institutions stemming from more applied research are less affected by patents being awarded.

5 Conclusions

A large stream of empirical literature examines the channels through which firms use public knowledge. For instance, firms develop absorptive capacities not only through internal R&D (Cohen and Levinthal, 1990) but also exploiting public knowledge (Zucker et al., 1998). Furthermore, public knowledge provides new scientific techniques (Levin et al., 1987). Finally, the academic community provides non-monetary rewards (e.g., access to conferences and academic scholars) (Stern, 2004) to industrial scientists.

However, little is known about the "reverse" relation, that is the effects and mechanisms through which private knowledge affect public knowledge developments.

While some research in the field of sociology of science (Merton, 1973; Garfield, 1979) highlights the depth of differences in institutions, norms, and rules between the scientific community and industry. This research does not address the challenges posed by the convergence between public and private knowledge, and the likelihood of a clash between these

25

Table 6: Robustness check on the role of applicability

DEPENDENT VARIABLES	: CIT_PUBLIC (OLS) (1)	CIT_PRIVATE (OLS) (2)
PATENT_IN_FORCE	-2.511*** (0.666)	-0.0307 (0.0757)
PATENT_WINDOW	-1.211 (0.618)	-0.119 (0.0703)
PATENT_IN_FORCE*LOW_PRIVATE_NPL	2.366** (0.873)	0.237* (0.0992)
Constant	10.90*** (0.395)	0.544*** (0.0449)
Number of observations	4439	4439
R^2	0.870	0.644

Note: Estimations are performed using a linear model with multiple levels of fixed effects. Standard errors are reported in parenthesis.
Legend:* p<0.05, ** p<0.01, *** p<0.001

institutional spheres is high. For instance, disclosure and appropriation are established differently in science and technology, using publications and patents, respectively. Also the reward system works through recognition based on citations to science, and the granting of a temporary monopoly on the technology. The seminal work of Murray 2002 and her use of PPP allowed to identify points of convergence between public and private knowledge empirically. These pairs represent intersections between public and private spheres of knowledge where patents refer to the proprietary dimension and publication in a scientific journal contributes to the body of public knowledge.

This paper contributes in two ways. First, we estimated the effect of a patent grant on the rate of subsequent knowledge with a specific focus on the different citing actors. Second, we studied whether invention characteristics such as applicability mediate the relation between private knowledge appropriation and public knowledge generation. We addressed these issues empirically in the context of cancer research. PPPs were constructed from a set of patents granted between 2004 and 2011 and identified using current U.S. patent classification. A manual matching identified 373 matched patents (22% of the patents in our original set). The empirical strategy exploited heterogeneity in grant lags, allowing the patent award to be seen as an exogenous shock along the life of a publication. The effect of a patent on subsequent scientific development was estimated using a difference-in-difference approach.

Our results show that different type of institutions reacts differently to sudden "privatization" of knowledge. While scientists affiliated to public institutions stop citing scientific articles after a patent is granted; corporate scientists do not. We interpret this different behavior as the result of different transactions costs faced by these type of institutions while gaining access to the same piece of proprietary knowledge. In particular, public institutions face higher transaction costs because lacking the capabilities to engage in the complex and

lengthy licensing processes. We also find that weather the disclosed knowledge is applied or not differently affect the rate of follow on knowledge. In particular, contrary to our expectations, we found that applied scientific articles received more citations both from public and private institutions. With the support of data on scientific article visualization, we interpret this result as a *freedom to advertise effect*. As the granting of the patent marks the "privatization" of the knowledge, scientists are finally free to advertise their discovery triggering more scientific citations.

Our findings have important policy implications as regards the interplay between incentives to innovate and access to scientific knowledge. Our research contributes to investigating possible drawbacks related to patenting, indicating that different actors are differently affected.

Finally, we need to point to some limitations of this work. First, although we interpret the results in light of the propensity to litigate, we do not observe actual litigation rates. A possible extension to our work would be to explore litigation propensity among our assignees. Second, although our empirical strategy provides the possibility to suggest some causality links, the generalization of the results to other fields should be cautious.

References

BALDINI, N., R. FINI, R. GRIMALDI, AND M. SOBRERO (2014): "Organisational change and the institutionalisation of university patenting activity in Italy," *Minerva*, 52, 27–53.

BUSH, V. (1945): "Science: The endless frontier," *Transactions of the Kansas Academy of Science (1903)*, 231–264.

CAMERON, A. C. AND P. K. TRIVEDI (2013): *Regression Analysis of Count Data (Second Edition)*, Cambridge: Cambridge University Press.

COHEN, W. M. AND D. A. LEVINTHAL (1990): "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, 35, 128–152.

COLAIANNI, A., S. CHANDRASEKHARAN, AND R. COOK-DEEGAN (2010): "Impact of gene patents and licensing practices on access to genetic testing and carrier screening for Tay-Sachs and Canavan disease," *Genetics in Medicine*, 12, S5.

DE SOLLA PRICE, D. J. (1965): "The scientific foundations of science policy," *Nature*, 206, 233–238.

DENT, C., P. JENSEN, S. WALLER, AND B. WEBSTER (2006): "Research Use of Patented Knowledge," OECD Science, Technology and Industry Working Papers No. 2006/02, OECD Publishing, Paris.

DREYFUSS, R. (2004): "Protecting the public domain of science: Has the time for an experimental use defense arrived?" *Arizona Law Review*, 46, 457–472.

EISENBERG, R. S. (1996): "Public Research and Private Development: Patents and Technology Transfer in Government–Sponsored Research," *Va. L. Rev.*, 82, 1663.

EISENBERG, R. S. ET AL. (2001): "Bargaining over the transfer of proprietary research tools: is this market failing or emerging," *Expanding the boundaries of intellectual property: Innovation policy for the knowledge society*, 223, 225–231.

FURMAN, J. L. AND S. STERN (2011): "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research," *American Economic Review*, 101, 1933–63.

GARFIELD, E. (1979): *Citation Indexing: Its Theory and Application in Science, Technology and Humanities.*, New York: Wiley Interscience.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–720.

GUIMARES, P. AND P. PORTUGAL (2010): "A simple feasible procedure to fit models with high-dimensional fixed effects," *Stata Journal*, 10, 628–649.

HAMILTON, K. S. (2003): "Subfield and level classification of journals," *CHI Research.*

HOLMAN, M. A. AND S. R. MUNZER (2000): "Intellectual Property Rights in Genes and Gene Fragments: A Registration Solution for Expressed Sequence Tags," *Iowa Law Review*, 85, 735–750.

HUANG, K. G. AND F. E. MURRAY (2009): "Does patent strategy shape the long-run supply of public knowledge? Evidence from human genetics," *Academy of Management Journal*, 52, 1193–1221.

KARACA-MANDIC, P., E. C. NORTON, AND B. DOWD (2012): "Interaction terms in nonlinear models," *Health services research*, 47, 255–274.

LEVIN, R. C., A. K. KLEVORICK, R. R. NELSON, S. G. WINTER, R. GILBERT, AND Z. GRILICHES (1987): "Appropriating the returns from industrial research and development," *Brookings papers on economic activity*, 783–831.

MACHLUP, F. (1958): *An economic review of the patent system*, 15, US Government Printing Office.

MAGERMAN, T., B. V. LOOY, AND K. DEBACKERE (2015): "Does involvement in patenting jeopardize ones academic footprint? An analysis of patent-paper pairs in biotechnology," *Research Policy*, 44, 1702 – 1713, the New Data Frontier.

MANSFIELD, E. (1998): "Academic research and industrial innovation: An update of empirical findings," *Research policy*, 26, 773–776.

MAZZOLENI, R. AND R. R. NELSON (1998): "The benefits and costs of strong patent protection: a contribution to the current debate," *Research policy*, 27, 273–284.

MEHO, L. I. AND K. YANG (2007): "Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar," *Journal of the American Society for Information Science and Technology*, 58, 2105–2125.

MERTON, R. K. (1973): *The Sociology of Science*, University of Chicago Press.

MERZ, J. F. AND M. K. CHO (2005): "What are gene patents and why are people worried about them?" *Public Health Genomics*, 8, 203–208.

MURRAY, F. (2002): "Innovation as co-evolution of scientific and technological networks: exploring tissue engineering," *Research Policy*, 31, 1389–1403.

——— (2010): "The oncomouse that roared: Hybrid exchange strategies as a source of

distinction at the boundary of overlapping institutions," *American Journal of sociology*, 116, 341–388.

MURRAY, F., P. AGHION, M. DEWATRIPONT, J. KOLEV, AND S. STERN (2016): "Of Mice and Academics: Examining the Effect of Openness on Innovation," *American Economic Journal: Economic Policy*, 8, 212–52.

MURRAY, F. AND S. STERN (2007): "Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis," *Journal of Economic Behavior & Organization*, 63, 648–687.

NICHOLSON, S. (2012): "Financing research and development," *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, 47.

PARTHA, D. AND P. A. DAVID (1994): "Special Issue in Honor of Nathan Rosenberg Toward a new economics of science," *Research Policy*, 23, 487 – 521.

POSNER, R. A. (2000): "An economic analysis of the use of citations in the law," *American Law and Economics Review*, 2, 381–406.

REITZIG, M., J. HENKEL, AND C. HEATH (2007): "On sharks, trolls, and their patent preyUnrealistic damage awards and firms strategies of being infringed," *Research Policy*, 36, 134–154.

SAMPAT, B. AND H. L. WILLIAMS (2019): "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, 109, 203–36.

STERN, S. (2004): "Do scientists pay to be scientists?" *Management science*, 50, 835–853.

STOKES, D. (1997): *Pasteur's Quadrant: Basic Science and Technological Innovation*, Washington, D.C.: The Brookings Institution.

THURSBY, J. G. AND M. C. THURSBY (2011): "Has the Bayh-Dole act compromised basic research?" *Research Policy*, 40, 1077 – 1083, special Issue: 30 Years After Bayh-Dole: Reassessing Academic Entrepreneurship.

WALSH, J. AND A. ARORA (2003): "Research Tool Patenting and Licensing and Biomedical Innovation. IN COHEN, W. & MERILL., S.(Ed.) Patents in the Knowledge-Based Economy," .

WILLIAMS-JONES, B. (2002): "History of a gene patent: tracing the development and application of commercial BRCA testing," *Health LJ*, 10, 123.

WU, Y., E. W. WELCH, AND W.-L. HUANG (2015): "Commercialization of university inventions: Individual and institutional factors affecting licensing of university patents," *Technovation*, 36, 12–25.

ZUCKER, L. G., M. R. DARBY, AND M. B. BREWER (1998): "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *The American Economic Review*, 88, 290–306.