Bergé, Laurent R.; Doherr, Thorsten; Hussinger, Katrin

Working Paper

How patent rights affect university science

ZEW Discussion Papers, No. 22-034

Provided in Cooperation with:
ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Bergé, Laurent R.; Doherr, Thorsten; Hussinger, Katrin (2022) : How patent rights affect university science, ZEW Discussion Papers, No. 22-034, ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim

This Version is available at:
https://hdl.handle.net/10419/264391

// NO.22-034 | 08/2022

DISCUSSION PAPER

// LAURENT R. BERGÉ, THORSTEN DOHERR,
AND KATRIN HUSSINGER

How Patent Rights Affect University Science

Leibniz Association

ZEW

How Patent Rights Affect University Science*†

Laurent R. Bergé,‡ Thorsten Doherr§ and Katrin Hussinger¶

July 12, 2022

Abstract

How do intellectual property rights influence academic science? We investigate the consequences of the introduction of software patents in the U.S. on the publications of university researchers in the field of computer science. Difference-in-difference estimations reveal that software scientists at U.S. universities produced fewer publications (both in terms of quantity and quality) than their European counterparts after patent rights for software inventions were introduced. We then introduce a theoretical model that accounts for substitution and complementarity between patenting and publishing as well as for the direction of research. In line with the model's prediction, further results show that the decrease in publications is largest for scientists at the bottom of the ability distribution. Further, we evidence a change in the direction of research following the reform towards more applied research.

Key words: patent rights; publications; economics of science; difference-in-difference estimation; model of science production

JEL Codes: I23; O31; O34; O38; L38

*This paper is forthcoming in Industrial and Corporate Change.

†We would like to thank Konstantinos Arkolakis, Stefano Bianchini, Yann Bramoullé, Nicolas Carayol, Robin Cowan, Lee Fleming, Dan Gross, Karin Hoisl, Francesco Lissoni, Henry Sauermann, Ivan Savin, Dan Spulber, Russell Thomson, Martin Watzinger and Emmanuel Lorenzon for insightful discussions. We thank four anonymous referees for constructive comments. For helpful comments we also thank seminar participants at the universities of Bordeaux, EPFL Lausanne, Luxembourg, Strasbourg and the Copenhagen Business School, at the 2017 ZEW conference in Manheim, the 2017 Academy of Management conference in Atlanta, the 2017 and 2018 EPIP conferences in Bordeaux and Berlin, R&D Management Conference 2017, the CIE Symposium at Maastricht University, the 2018 Luxembourg IP and Innovation workshop and the 2019 Northwestern/USPTO Conference on Innovation Economics in Chicago. We are grateful to Dimitris Kafteranis for research assistance. Katrin Hussinger is grateful for financial support from a Marie Curie Career Integration Grant.

‡BxSE, UMR CNRS 6060, University of Bordeaux; email: laurent.berge@u-bordeaux.fr; mail: Maison de l'économie – BxSE, University of Bordeaux, 6 avenue Léon Duguit, 33600 Pessac, France.

§Centre for European Economic Research (ZEW); email: doherr@zew.de; mail: ZEW – Leibniz Centre for European Economic Research, L 7, 1, 68161 Mannheim, Germany

¶University of Luxembourg, ZEW and KU Leuven (email: katrin.hussinger@uni.lu, mail: Campus Kirchberg, Université du Luxembourg, 6, rue Richard Coudenhove-Kalergi L-1359 Luxembourg, Luxembourg.

1 Introduction

Knowledge created at universities is seen as a major source of technological opportunities and progress (Griliches, 1979, 1992; Klevorick et al., 1995). Traditionally, university science was characterized by an open regime that facilitates disclosure and diffusion of inventions and discoveries (Dasgupta and David, 1994). University knowledge was non-excludable and non-rival in use which contributed to its impact on growth in income per capita and increasing returns to scale within an economy (Aghion and Howitt, 2005; Jones, 2005). Benefits spread to the private sector in form of accelerated corporate innovation (Jaffe, 1989; Toole, 2012) and productivity growth (Adams, 1990).

Since the 1980s, a series of intellectual property (IP) policies that aimed at spurring innovation and entrepreneurship based on university discoveries has shaped IP regulations for U.S. universities (Mowery and Sampat, 2001). The most prominent change was the U.S. Bayh-Dole Act of 1980 that strengthened universities' institutional ownership rights for discoveries made by federally funded scientists (Henderson, Jaffe and Trajtenberg, 1998; Mowery et al., 2001; Mowery and Ziedonis, 2002; Sampat, Mowery and Ziedonis, 2003; Mowery and Sampat, 2005).

Another drastic IP law change was the introduction of software patents in the U.S. (Graham and Mowery, 2003; Bessen and Hunt, 2004, 2007). While not aimed at universities, this law change had strong side effects on the academic sector as we show in this paper. The introduction of software patents was a response to the belief that the changing nature of technology should be reflected in IP legislation which became visible in a general trend towards stronger and wider patent protection (Merges and Nelson, 1994; Scotchmer, 1991). The implication for the private sector was a rise in software patents (Kortum and Lerner, 1999; Graham and Mowery, 2003) which has been attributed to strategic considerations rather than to an increase in innovation (Bessen and Hunt, 2007; Noel and Schankerman, 2013).[1] As patents for general purpose technologies being used in complex product industries, software patents are strongly associated with legal disputes (Bessen and Meurer, 2005) and market entry barriers (Cockburn and MacGarvie, 2011).

[1]Scholars have disputed the breadth of software patent claims (Burk and Lemley, 2003; Rai, 2003) and the allegedly poor quality of prior art documentation (Lunney Jr, 2000) questioning the validity of software patents per se.

While the value and implications of software patents remain disputed for the private sector (see Gallini, 2002, for a discussion of the arguments), the effect of the introduction of IP rights for software for university research has not yet been investigated.[2]

In this paper, we investigate the effects of the introduction of software patent rights on the scientific publications of U.S. university scientists. Universities present an environment in which individual-level incentives to publish are essential. We argue that the introduction of patent rights for software affects U.S. university researchers working on software related issues by increasing their incentives to produce patents. Our empirical results of a difference-in-difference analysis show that the reform toward the patentability of software inventions led to an important decrease (both in terms of quantity and quality) in U.S. computer scientists' production of scientific publications compared to their European peers. We evidence a 20% reduction in U.S. publication numbers.

To investigate the mechanisms, we introduce a theoretical model of science production which integrates substitution and complementarity between patenting and publishing, and which also accounts for the direction of research. In light of our model, the overall negative effect that we observe in our empirical results implies that the complementarity between patenting and publishing does not compensate the substitution effect. The model further predicts that the decrease in publications is dependent on scientists' ability, the ones facing the largest decrease being the ones with the lowest ability. The second prediction of the theoretical model holds that the direction of research changes towards topics more complementary to patents (i.e. more applied), and the importance of this change also depends on scientists' ability.

In line with the theoretical model, further results show that scientists with the *ex ante* lowest number of citation-weighted publications, which is our proxy for ability, are the ones that suffer the biggest drop in publications. They produced 31% less publications following the introduction of patent rights for software inventions. The magnitude of the drop in publications decreases along the productivity distribution. At the other end of

[2]There are a few studies that descriptively look into the topic. Graham and Mowery (2003) provide descriptive evidence for selected universities concluding that the surge in software patents observed in the private sector was not accompanied by an equally strong increase of university software patents. Rai et al. (2009) suggest that universities participate in the market for software patents, potentially because of economies of scale realized through technology transfer offices. Love (2014) contributes a survey among university scientists about the usefulness of patents in software.

the scientist publication ability distribution, top U.S. academic computer scientists were only slightly affected by the reform. Using a measure approximating the appliedness of publications we find that: i) the direction of research changed towards more applied topics, ii) the largest change towards more applied research topics is observed for scientists with the lowest publication ability.

The causality of the results is supported by a comparison of the publication pattern of software- and hardware-focused computer scientists at U.S. universities. Hardware computer scientists, whose scientific work is patentable since Bayh-Dole, did not change their publication output around the time of the introduction of patent rights for software inventions. Overall, our results cannot rule out concerns about negative implications for science of strengthened IP ownership rights for universities.

This paper is organized as follows. Section 2 details the institutional setting in the U.S. and Europe respectively. Section 3 introduces the data and Section 4 provides some descriptive evidence. The identification strategy and main empirical results are presented in Section 5. Section 6 describes a theoretical model of science production from which we derive testable implications. Section 7 presents empirical evidence for the mechanisms, Section 8 discusses the results and Section 9 concludes the paper.

2 Background

This section provides an overview of the patentability of software in the U.S. and Europe. For further details on the legal background we refer to Graham and Mowery (2003) for the U.S. case, Bakels et al. (2008) for the European case and Guntersdorfer (2003) for a comparison.

2.1 The U.S. case

In the 1970s, the predominant method to protect software in the U.S. was copyright. Algorithms were deemed not to be patentable at the United States Patent and Trademark Office (USPTO) which was confirmed by a number of Supreme Court decisions in the 1970s (Graham and Mowery, 2003). The case of Gottschalk versus Benson (409 U.S. 63 [1972]) explicitly rejected software as patentable subject matter and the 1976 Copyright Act

explicitly endorsed copyright as an appropriate protection regime for software. As noted by Hall and MacGarvie (2010), however, patents and copyrights protect very different aspects of software. While copyright is awarded to creators of original works and protects a specific computer code as an "original expression", it does not protect the functions performed by the code. The function of a software program may be protected by a patent.

In the 1980s, patent law started to slowly change in favor of software patents with the ruling that software can be patented if tied to physical or mechanical processes. The change was initiated by the Diamond versus Diehr case decision (450 U.S. 175 [1981]) in which the Supreme Court decided on the patentability of a rubber-curing process that used software to calculate the cure time. The physical transformation of rubber "into a different state or thing" took the invention out of the realm of abstraction. The subject matter was declared patentable even though the software implementation represented the only novel feature of the invention. During the first half of the 1990s, a number of court decisions, which are described in more detail by, e.g., Graham and Mowery (2003) and Hall and MacGarvie (2010), spurred a discussion about the broadening of the patentable subject matter with important implications for software. An important step toward new legislation was taken in 1994 by the Court of Appeals of the Federal Circuit (CAFC), which distinguished between patentable software as "rather a specific machine to produce a useful, concrete, and tangible result" and unpatentable software as a disembodied mathematical concept such as a law of nature, natural phenomenon, or an abstract idea (Sterne and Bugaisky, 2004). After a series of further cases in 1994[3] – with the last one being that the CAFC ruled that the rejection of a software patent application at the USPTO by IBM was erroneous in 1995[4] – the U.S. Commissioner of Patents issued new patentability guidelines in 1996 which allowed inventors to patent any software embodied in physical media (Sterne and Bugaisky, 2004).[5]

Although their announcement was perceived negatively by the stock market, the new

[3]In re Alappat (5440676), In re Warmerdam (6089742), In re Lowry, In re Trovato.

[4]In re Beauregard (5710578)

[5]The guidelines specified that a distinction should be made between (a) "a computer or a programmable apparatus controlled by software as a statutory 'machine'", (b) computer-readable memory used to direct a computer such as a memory device, a compact disc or a floppy disk as a statutory 'article of manufacture' and (c) a series of steps to be performed on or with the aid of a computer as a statutory's process" (USPTO guidelines, 1996, https://www.uspto.gov/web/offices/com/sol/og/con/files/cons093.htm).

patentability guidelines were followed by a surge in software patenting in the private sector during the period 1996-1999 (Hall and MacGarvie, 2010), which has been largely attributed to strategic considerations rather than to an increase in innovation (Bessen and Hunt, 2007; Noel and Schankerman, 2013).

With a focus on the university landscape, software was already one of the fields in which universities had licensing agreements before the introduction of software patent rights (Mowery et al., 2001). After 1996, the number of university-held software patents decupled over the period 1982-2002 from 37 patents in 1982 to 396 patents in 2002, which corresponded to a 4% increase in the share of software patents among university patents (Rai et al., 2009). The disproportionate increase in university software patents has been attributed to economies of scale realized through technology transfer offices (TTOs) (Rai et al., 2009; Graham and Mowery, 2003). TTOs started to play a more active role in the field of computer science after the introduction of patent rights (Rai et al., 2009). TTOs took a "one size fits all" approach in the sense that the propensity to apply for patent protection for a software invention was predominantly determined by the TTO's tendency to seek patent protection in other disciplines (Rai et al., 2009). This implied that computer scientists faced closer scrutiny of their inventions by the TTO. Moreover, patents became important for tenure, promotion and annual salary raises across the U.S. (Love, 2014). Due to the significant changes for computer scientists with regard to the role of the TTO, output expectations and career requirements, we refer to the introduction of software patents as a regime shift.

After the introduction of software patents, computer scientists might face some institutional and career pressure to patent. We expect that scientists nevertheless show a strong ambition to publish and that – if demanded to produce patentable results in addition – they reallocate their time from projects with low scientific prospects to patenting, keeping the efforts put into projects with expected high scientific quality constant or reduce it least. Prior evidence mostly suggests that patenting and publication activities are complementary at the researcher level (Stephan et al., 2007; Fabrizio and Di Minin, 2008; Czarnitzki, Glänzel and Hussinger, 2009).[6] Patents can occur as byproducts of scientific

[6]Note that prior evidence suggests that patenting and scientific quality of publications are negatively correlated (Murray and Stern, 2007; Fabrizio and Di Minin, 2008; Czarnitzki, Glänzel and Hussinger, 2009).

research projects that spawn both, patents and scientific publications. Projects conducted by university researchers are designed to consist of modules that differ with regard to the extent to which they are patentable or publishable in scientific journals. Some researchers employ inter-personal economies of scope and follow a dual knowledge disclosure strategy, setting set up their projects ex ante to address both an industry and academic audience (Murray, 2002; Murray and Stern, 2007; Magerman, Van Looy and Debackere, 2015).

2.2 The European case

Whereas the U.S. Patent Act of 1952 laid the foundation for the expansion of the patentable subject matter (Sterne and Bugaisky, 2004), Article 52(2) of the European Patent Convention (EPC) explicitly excludes specific categories of inventions such as business methods and software. These inventions do not fulfill the technical contribution requirement. Article 52(2) specifies that software is not patentable "as such". Further guidelines are provided by the case decisions of the Technical Board of Appeal of the EPO.[7] According to those, software may, for instance, be patented if tied to physical or mechanical processes. A proposal for a Directive on the Patentability of Computer-Implemented Inventions (knows as the CII Directive) which was intended to improve clarity on the treatment of software inventions under European patent law was rejected in 2005 (González, 2006). Hence, the legal situation in Europe corresponds to the legal situation in the U.S. before the introduction of the new patent guidelines and after Diamond versus Diehr (Gunterstorfer, 2003).

[7]The "as such" clause leaves some room for interpretation. A decision of the Technical Board of Appeal of the EPO from 1988 holds, for instance, that "even if the basic idea underlying an invention may be considered to reside in a computer program a claim directed to its use in the solution of a technical problem cannot be regarded as seeking protection for the program as such within the meaning of Article 52(2)(c) and (3) EPC" (T 0115/85, 1988). However, "a computer program product is not excluded from patentability under Article 52(2) and (3) EPC if, when it is run on a computer, it produces a further technical effect which goes beyond the "normal" physical interactions between program (software) and computer (hardware)" (T 1173/97, 1998). In other word, if the invention covers a "trick" of how to do something rather than a program code, it is patentable (Bakels et al., 2008).

3 Data, variables and descriptive statistics

3.1 Data

The population of interest includes all scientists working on software-related topics at U.S. and European universities. The study period is defined as 1989-2004 to enable observations over a sufficient number of years before and after the policy change. We focus on the year 1996 as the year in which our treatment starts because the new patentability guidelines were introduced in this year accompanied by a steep increase in software patents (Hall and MacGarvie, 2010).

Based on publication records retrieved from the Web of Science (WoS), we constructed a panel dataset of academic scientists in the field of computer science following a multistep procedure, which is detailed in Appendix A. Since we are interested in academic scientists, we restrict the panel to scientists who are solely affiliated to universities and discard the ones affiliated to companies. Computer science is defined by the WoS subfields Artificial Intelligence, Information Systems, Theory & Methods, Software Engineering, Interdisciplinary Applications, and Cybernetics. Since the field of computer science is broad, scientists working in other university departments frequently publish in this field. Therefore, only scientists with a minimum of three computer science publications in three different years qualified for our sample. Two of those publications must be from before 1996. Using other thresholds or no threshold at all does not alter the findings (see Table B1 in Appendix). In Appendix B.3, we replicate the main estimations with an alternative data set: DBLP, which is a publication data set curated for computer science known for its good coverage of conferences. The results from all estimations that could be replicated hold.

Scientists working on hardware-related topics have been excluded from the treatment group,[8] as well as scientists who switched affiliations between the U.S. and Europe. For each computer scientist, our data contain the scientist's history of publishing during 1989-2004, supplemented by the scientist's patent record for the same period. Computer science publications include articles published in scientific journals and in conference proceedings.

[8]The results are not altered by using different definitions of software scientists, such as, for instance, including scientists with publications on hardware-related topics in the treatment group (see Table B2 in Appendix).

Conference proceedings are an important outlet for computer scientists, enabling the speedy dissemination of results (Patterson, Snyder and Ullman, 1999; Visser and Moed, 2005; Bar-Ilan, 2008). However, the most impactful research of computer scientists is published in journals (Bar-Ilan, 2010; Franceschet, 2010). Our final sample consists of 8,133 computer scientists working on software-related topics at universities, including 4,437 based in Europe. In total, these scientists produced 86,756 unique publications.

3.2 Measuring science production

We investigate the effects of the introduction of patent rights for software inventions on the quantity and quality of academic researchers' scientific output.

The quantity measure depicts the number of publications produced in a given year by a scientist. This measure includes all publications contained in the WoS data base in the field "computer science", including both journal articles and conference proceedings. We excluded publications in non-peer-reviewed journals.[9]

We create several measures reflecting publication quality. The first measure is the citation-weighted number of publications, measured as the sum of the citations to the publications of a scientist in a given year. The number of citations is defined as the number of citations received from other WoS publications published up to 2016, the year in which we extracted the data. The fact that articles with an earlier publication date are cited more frequently than those published recently should not bias our results as: *i*) we use a control group with a similar distribution of publications over time; and *ii*) we include year fixed-effects. Citation data are highly skewed, however: in our sample, the 3% most-cited publications receive 50% of all citations. This means that a few scientists have an excessive weight on the estimates so that we complement the citation measure with alternative quality measures.

The second quality measure is the number of top-cited publications, $TOP\,5\%$. To construct this variable, we consider all worldwide publications in computer science in a given year to define the publications in the 95th percentile of the citation distribution. For each scientist, we count the number of these top 5% publications per year. Since

[9]The excluded journals are Datamation, Byte, Dr Dobb's Journal, Computer Design, Sharp Technical Journal and Hewlett-Packard Journal.

citation patterns can be field-specific and may reflect the evolution of the field instead of quality, in Appendix B we complement these two measures with subfield-corrected citation measures. The results, in Appendix B4, are qualitatively similar using these alternative measures.

The third quality measure, $JIF\,10\%$, is the number of publications in highly recognized journals and conference proceedings, such as the ACM Computing Surveys or the IEEE Conference on Computer Vision and Pattern Recognition. To identify top publication outlets, we use the journal impact factors (JIF) from Scopus for the year 2015.[10] In total, 13% of the publications appear to be published in a top 10% JIF outlet.

3.3 Descriptive statistics

Table 1 reports some descriptive statistics for U.S. and European academic software computer scientists. Disregarding any temporal effect, these figures depict two samples that appear to be very similar in terms of scientific output. The yearly output of the average scientist is 0.58 articles and 0.38 conference proceedings. Taking publication quality into account, a gap between the U.S. and European scientists becomes apparent. U.S. scientists have almost twice the number of citation-weighted publications as European scientists, but with a higher variance as well. The pattern is similar with regard to the two alternative quality measures, with U.S. scientists producing an average of 0.25 publications in top journals per year, as opposed to 0.16 for European academic computer scientists. These numbers are respectively 0.11 and 0.068 for the yearly production of top-cited publications. For both groups, patenting is a rare event, as their 90th percentile in patents per year is 0. However, the patenting rate is almost four times higher for U.S. academic scientists.

4 Descriptive evidence

Before describing the multivariate analysis, we provide some descriptive evidence for the evolution of the number of publications over time for the treatment and control group.

[10] About 53% of WoS publications were not found in the Scopus database of JIFs. Hence, we can safely conclude that these are not among the top 10% of JIF journals.

Table 1: Descriptive statistics of the production of U.S. and European university software scientists – across all years (1989–2004).

	U.S. scientists							European scientists						
Variables	Min	Median	Q3	90%	Max	Mean	SD	Min	Median	Q3	90%	Max	Mean	SD
Yearly # Publications	0	0	1	2	62	0.93	1.6	0	1	1	3	42	0.96	1.5
Yearly # Articles	0	0	1	2	38	0.55	1	0	0	1	2	15	0.58	1
Yearly # Proceedings	0	0	0	1	28	0.38	1	0	0	0	1	33	0.38	0.96
Yearly # of citation-weighted publications	0	0	8	41	8400	20	116.9	0	0	5	27	4852	12.9	78
Yearly # Publications in top 10% JIF journal/conference	0	0	0	1	12	0.25	0.65	0	0	0	1	12	0.16	0.51
Yearly # top 5% cited publication	0	0	0	0	7	0.11	0.39	0	0	0	0	8	0.068	0.31
Yearly # Patents	0	0	0	0	13	0.029	0.26	0	0	0	0	7	0.0077	0.11
# Scientist-year				50,191							58,838			
# Scientists				3696							4437			

Notes: The data consists of active computer scientists working in software. An active computer scientist is defined by having at least two publication before 1996 (with one before 1994), and at least one publication in the period 1997-2004. Observations correspond to *scientist* × *year*.
Source: Authors' own calculations based on publication data from Web of Science and patent data from the USPTO.

Panel A in Figure 1 compares the scientific output of U.S. and European university software scientists. Before 1996, the trends for both European and U.S. scientists appear to be similar. After the introduction of software patents, however, the trends differ. European scientists experienced an upwards trend, while the U.S. scientists faced a downwards trend. This simple graphical analysis suggests that the IP reform had a negative effect on the publication volume of U.S. university computer scientists.

In order to ensure that the pattern illustrated in Panel A is software specific, Panel B of Figure 1 shows the same figure for university scientists working on hardware-related topics. Hardware scientists work in a subfield within computer science, but were not affected by the reform, as they have been able to patent their inventions since the Bayh-Dole Act of 1980. We identified a total of 4,158 hardware computer scientists based on the WoS subfield "Hardware & Architecture", 2,347 of which are affiliated with U.S. universities and 1,811 with European universities.

Panel B shows that for hardware scientists, the publication outcome curves of U.S. and European university scientists are roughly parallel across the entire period. This evidence shows that the drop in publications is specific to academic software scientists, which speaks in favor of a causal relationship between the introduction of software patent rights and the decrease in scientific publications.

Further, Panel A may suggest that software scientists reallocate part of their time from scientific to commercialization activities. This is in line with Figure 2, which depicts the patent applications of U.S. university software scientists in our sample over time. The patenting rate was around 0.02 patents per year before 1996, while it greatly increased afterwards, hovering around 0.035.

5 Effect on scientific output: Econometric approach and empirical results

5.1 Econometric approach and identification

Our analysis considers difference-in-difference (DiD) regressions, in which the introduction of the new patentability guidelines granting patent rights for software inventions defines

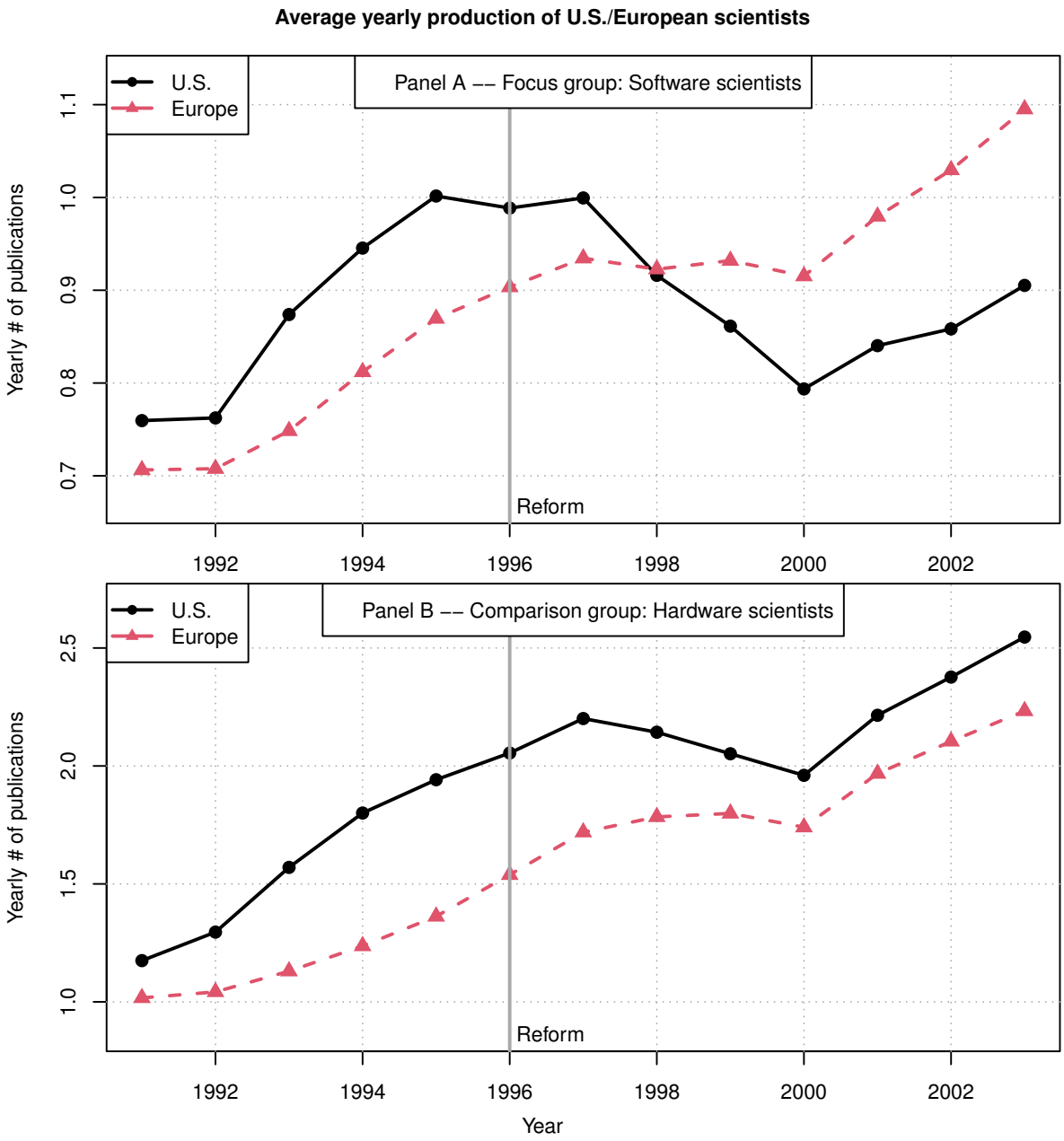

Figure 1: Publication trends of U.S. and European computer scientists.
Notes: Sample of U.S. and European university computer scientists. Each point represent the average yearly production which was smoothed using a 3 years window ($y_t^{3yw} = (y_{t-1} + y_t + y_{t+1})/3$).
Source: Authors' own calculations based on Web of Science data.

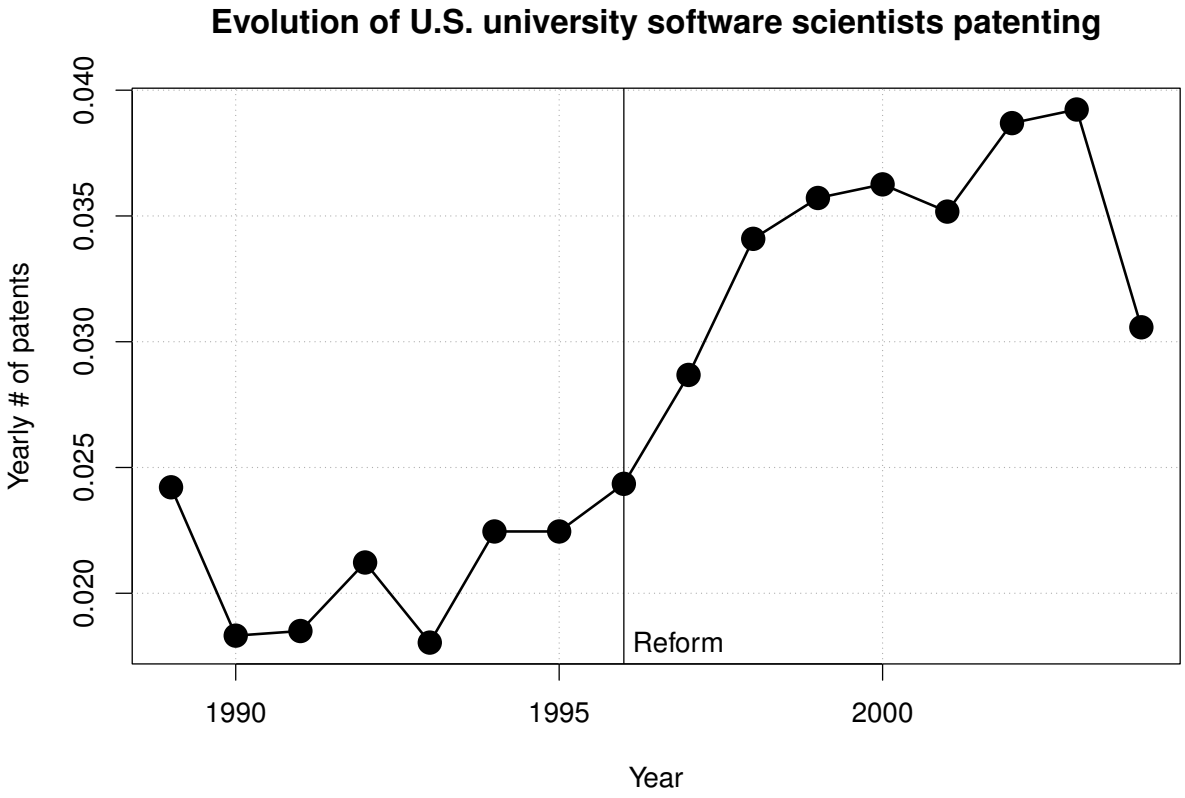

Figure 2: Patenting trends of U.S. software scientists.

Source: Authors' own calculations based on USPTO data.

the pre- and post-change period. The effect of the legal change can be obtained by the following regression:

$$E\left(y_{it}\right) = \exp\left(\alpha_i + \gamma_t + a \times Treat_i \times Post_t\right), \tag{1}$$

where y_{it} represents the (quality-weighted) outcome of scientist i in year t, α_i are scientist fixed-effects, γ_t are time fixed-effects, $Treat_i$ is a dummy variable, given a value of 1 if i is a U.S. scientist and $Post_i \equiv 1\left\{t \geq 1996\right\}$ is a dummy variable identifying the post-change period. The coefficient of interest is a. It represents the total impact of the legal change on scientists' productivity over the post-change period.

Since the dependent variables are count variables, we estimate Equation (1) using Poisson models (Santos Silva and Tenreyro, 2006). Poisson models provide consistent estimates as long as the conditional mean is properly specified. Further, we cluster the standard errors at the scientist level in order to avoid biases due to possible serial correlation (Bertrand, Duflo and Mullainathan, 2004). Since the citation-weighted number of publications is over-dispersed, with a standard-deviation to mean ratio of about 6 (see Table 1), we employ Negative Binomial estimations for this variable to ensure more efficient estimates.[11]

In addition, we estimate the following equation to obtain yearly treatment effects:

$$E\left(y_{it}\right) = \exp\left(\alpha_i + \gamma_t + \sum_{t'=1989,\, t' \neq 1995}^{2004} a_{t'} \times Treat_i \times 1_{t'=t}\right). \tag{2}$$

In this set-up, we estimate the effect of the treatment for each of the sample years, i.e., each a_t. We use the year before the reform, 1995, as a reference category. If the policy implies a change in the science production of U.S. scientists, there should be a shift in the coefficients a_t for the period after 1995. This model enables a clear visual representation of the policy change over time and tests for parallel trends before the policy change.[12]

[11] The Poisson and Negative Binomial estimates are almost identical for the other dependent variables.

[12] Note that we do not focus on the direct involvement of scientists in patenting as has been done in previous studies (e.g., Azoulay, Ding and Stuart, 2009; Czarnitzki, Glänzel and Hussinger, 2009). We do so because the number of patenting software scientists is a highly selective group of scientists while the regime change affects *all* scientists. Focusing only on patenting scientists would lead to a very incomplete picture of the consequence of the change. We would miss all those scientists who invest time in commercialization efforts, but fail to obtain a patent. See the discussion in Section 8.3.

Table 2: Effect of the legal change on scientists' production.

Dependent Variables:	Yearly # Publications		
Column:	(1)	(2)	(3)
Variables			
Constant	-0.0795***		
	(0.0121)		
Treat	0.0780***	0.0747***	
	(0.0186)	(0.0186)	
Post	0.0610***		
	(0.0172)		
Treat×Post	-0.1655***	-0.1621***	-0.1645***
	(0.0262)	(0.0262)	(0.0272)
Fixed-Effects			
Year		✓	✓
Scientist			✓
Fit statistics			
Observations	109,029	109,029	109,029
Adj-pseudo R^2	0.00057	0.00356	0.21675
Log-Likelihood	-163,676.3	-163,188.1	-128,272.4

*Clustered (scientist) standard-errors in parentheses. Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The coefficients correspond to maximum likelihood Poisson estimates. The sample consists of active computer scientists working in software. An active computer scientist is defined by having at least two publication before 1996 (with one before 1994), and at least one publication in the period 1997-2004. Observations correspond to *scientist* × *year*.
Source: Authors' own calculations based on publication data from Web of Science.

5.2 Empirical results

Table 2 reports the regression results. Column 1 displays the estimation results of Equation (1), in which the overall effect of the regime shift is estimated without time dummies and scientist fixed-effects. We find a significant negative effect of -0.16, meaning that the IP reform reduced the production of journal publications and publications in conference proceedings of U.S. scientists by 15% (i.e. $100 \times [1 - \exp(-0.16)]$), compared to European scientists. In column 2, we include time fixed-effects and, in column 3, scientist fixed-effects. These inclusions do not change the magnitude of the estimate.

Turning to publication quality, Table 3 reports fixed-effects DiD estimates for the three quality variables. The estimated coefficients for the treatment effect are equally negative, with a higher order of magnitude. The estimated coefficients are -0.47 for the citation-weighted number of publications, -0.26 for the number of articles in the top JIF journals and -0.27 for the number of top-cited publications.

Table 3: Effect of the legal change on scientists' quality-weighted production.

Dep. Variables:	Yearly # of Citations-Weighted Pub.	Yearly # of Top 10% Ranked Articles (JIF)	Yearly # of Top 5% Cited Articles (Worldwide)
	Neg. Bin.	Poisson	Poisson
Column:	(1)	(2)	(3)
Variables			
Treat×Post	-0.4743***	-0.2665***	-0.2760***
	(0.0481)	(0.0429)	(0.0565)
Fixed-Effects			
Year	✓	✓	✓
Scientist	✓	✓	✓
Fit statistics			
Observations	106,907	71,384	46,649
# Scientist	7,959	5,255	3,406
Adj-pseudo R^2	0.06218	0.18248	0.14863
Log-Likelihood	-239,093.7	-43,598.3	-22,094.8
Over-dispersion	0.16391		

*Clustered (scientist) standard-errors in parentheses. Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The coefficients correspond to maximum likelihood Negative Binomial (column 1) or Poisson estimates (columns 2 and 3). The sample consists of active computer scientists working in software. An active computer scientist is defined by having at least two publication before 1996 (with one before 1994), and at least one publication in the period 1997-2004. Observations correspond to *scientist* × *year*. Although the same sample is used across all regressions, the number of observations vary because, due to the Negative Binomial/Poisson fixed-effects setup, all scientists whose dependent variable is equal to 0 across all periods are dropped.
Source: Authors' own calculations based on publication data from Web of Science.

To provide insight into the temporal dynamics, yearly treatment effects for the quantity and the three quality variables are reported in Figure 3. The estimation follows Equation (2) and includes both year and scientist fixed-effects. The upper left panel represents the number of publications for which the pattern is clearest. We observe that before the reform, the estimates fluctuate around 0. After 1996, the coefficients become strongly negative. As could be expected, a few years elapsed between the introduction of software patent rights in 1996 and a sharp drop in publications. From 1999 onward, the estimated effect fluctuates around -0.3, which hints at a long-term decrease of publications of 25%.

Regarding the different citation-weighted publication measures, we find qualitatively very similar results. Before the reform, the coefficients hover around 0.3 with a large standard-error, while a few years after the reform, the estimated coefficients become strongly negative, with a magnitude of about -0.5 (i.e. about 40% less in terms of yearly number of citation-weighted publications). This shows a decrease of the U.S. scientists' citation advantage due to the IP reform.[13] For the number of publications in the top JIF journals, there is no significant advantage for U.S. scientists before the IP reform, while the treatment effect becomes significantly negative afterwards. For the number of top cited publications, the pattern is less clear: the yearly estimates also decrease after the reform but are not significantly different from 0 in the post-reform period. Overall, the yearly effects clearly display a drop in publication volume and quality after the reform, at least for three of the four dependent variables.

6 A model of science production

The decrease in publications raises many questions. First, what can explain this decrease? Second, does scientist heterogeneity matter? Third, is there an impact on the direction of research? To find some answers, we first introduce a model of science production whose outcome will be interpreted in light of our previous empirical result. We then extend it

[13]Note that when using subfield-corrected citations, as in Figure B2 in Appendix, the pre-reform estimates are even closer to 0, while the post reform estimates are higher in magnitude.

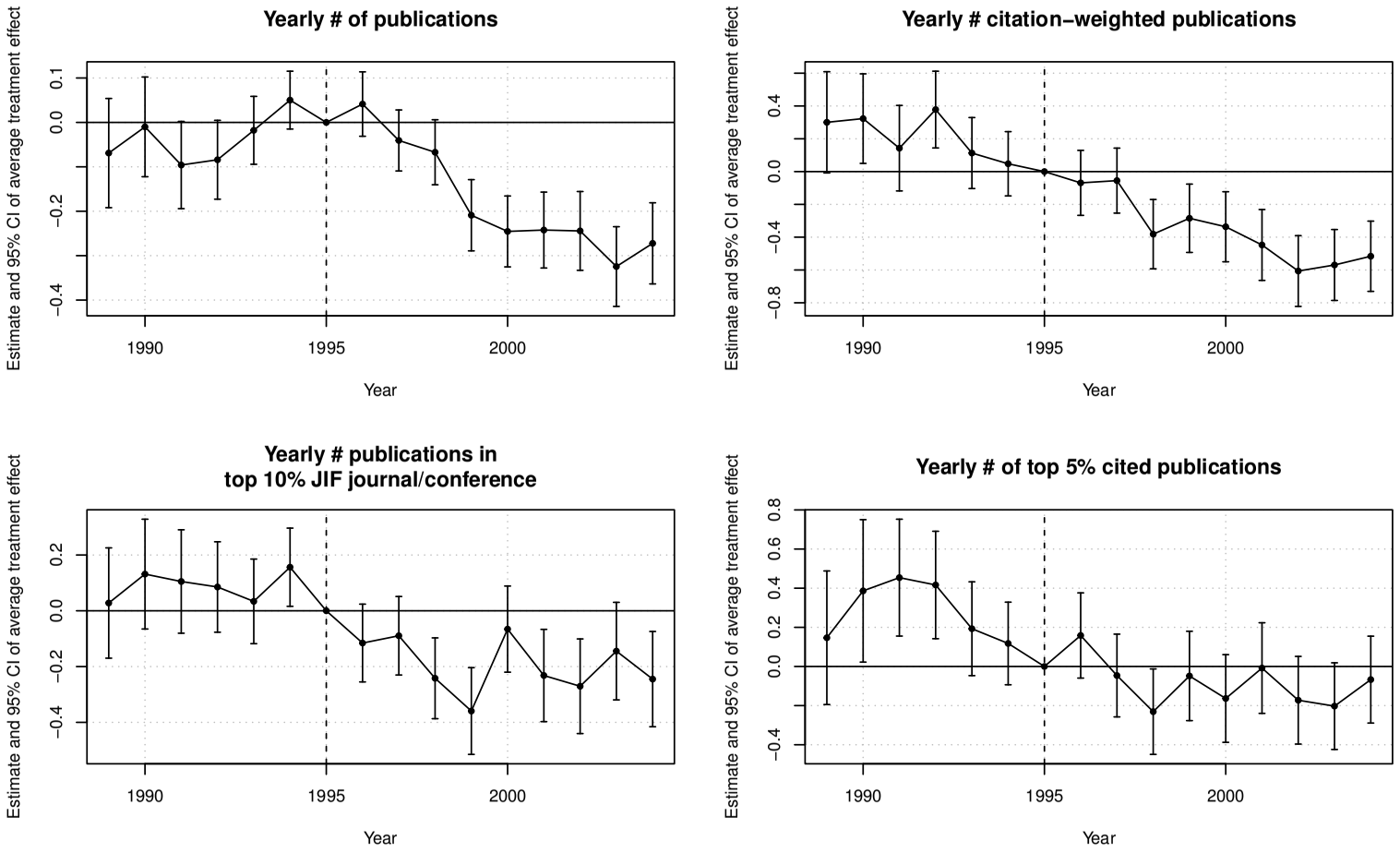

Figure 3: Yearly treatment effects.

Notes: The figures represent estimates of the yearly treatment effects with their 95% confidence intervals for different dependent variables, the reference year being 1995 (the year preceding the shock). Each panel shows the result of a pseudo Poisson, or Negative Binomial (for the citations variable),maximum likelihood estimation with scientist and year fixed-effects. The standard-errors are clustered at the scientist level.

to shed light on the impact on the direction of research.[14]

6.1 Model

Following Galasso, Mitchell and Virag (2018), we assume that scientists engage in both science and commercialization activities which lead to publications and patents, the two outcomes of our model.[15] Further, we assume complementarity between publishing and patenting. In line with the idea that the same knowledge can serve both the production of publications and patents (dual knowledge à la Murray and Stern, 2007), the complementarity in our model assumes that some efforts of working on publications spills over to the production of patents.

Let e_{it}^{pub} and e_{it}^{pat} be the respective efforts scientists exert to produce scientific work resulting in publications (publishing), and to perform commercialization activities resulting in patents (patenting) in period t. Similarly, let γ_{it}^{pub} and γ_{it}^{pat} be scientist i's ability for publishing and patenting, respectively. We suppose that in a given period of time the production of publications and patents follows a Poisson process which depends on a scientist's ability and the respective effort exerted in each activity. Formally, the production of publications and patents is respectively given by the following processes:

$$Pub_{it} \sim Poisson\left(\gamma_{it}^{pub} e_{it}^{pub}\right), \quad Pat_{it} \sim Poisson\left(\gamma_{it}^{pat}\left(e_{it}^{pat} + \lambda e_{it}^{pub}\right)\right),$$

with Pub_{it} (Pat_{it}) representing the number of publications (patents) produced by scientist i in period t. We suppose that the complementarity is exogenous and captured by the parameter $\lambda \in [0, 1]$, with $\lambda = 0$ leading to no complementarity.

Scientist i derives utility U_{it} from exerting efforts e_{it}^{pat} and e_{it}^{pub}. We assume a quasi-linear utility function in which both patents and publications confer the same utility to the scientist, normalized to unity.[16] This can be thought of as a scientist striving for tenure or a promotion at a university that gives credit for both publications and patents.

[14] Note that we only consider the idea of quantity of publication but the model also applies to the idea of quality, see Appendix C.1.

[15] Love (2014) shows that patents matter for the career of U.S. computer scientists. More than half of the computer scientists who patent do so because it is formally or informally considered for their tenure evaluation or a promotion decision.

[16] Note that assuming different utilities for the two activities would lead qualitatively to the same results.

Scientist i's utility is given by the following relation:

$$U_{it}\left(e_{it}^{pub}, e_{it}^{pat}\right) = Pub_{it} + Pat_{it} - \left[e_{it}^{pub}\right]^2 - e_{it}^{pub} \times e_{it}^{pat} - \left[e_{it}^{pat}\right]^2,$$

where the two activities are defined as substitutes due to the structure of effort costs, following Bénabou and Tirole (2016).

Assuming that scientists are risk neutral, the problem of scientist i in period t is to find the efforts $e_{it}^{pub^*}$ and $e_{it}^{pat^*}$ maximizing their expected utility:

$$\begin{aligned} \left\{e_{it}^{pub^*}, e_{it}^{pat^*}\right\} &= \arg \max_{e_{it}^{pub}, e_{it}^{pat}} E\left[U_{it}\left(e_{it}^{pub}, e_{it}^{pat}\right)\right] \\ s.t. \qquad & e_{it}^{pub} \geq 0 \text{ and } e_{it}^{pub} \geq 0 \end{aligned}$$

which yields the following optimal level of efforts:

$$e_{it}^{pub^*} = \frac{2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat}}{3}.$$

Henceforth we assume that $2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat} > 0$ and $(2-\lambda)\,\gamma_{it}^{pat} - \gamma_{it}^{pub} > 0$, so that the optimal efforts in both activities are positive. Outside this range, scientists optimally allocate the totality of their time either to working on publications, or working on commercialization.

Hence, given the optimal level of effort implemented by scientist i, the expected number of publications at time t is given by

$$\begin{aligned} \widehat{Pub_{it}} &\equiv E\left(Pub_{it}\right) \\ &= \gamma_{it}^{pub} e_{it}^{pub^*} \\ &= \frac{\gamma_{it}^{pub}}{3}\left(2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat}\right). \end{aligned}$$

Let us now consider the implications of the reform. Suppose that the introduction of patent rights for software inventions translates into an increase in γ^{pat} of Δ, and consider a scientist i with her counterfactual k. The counterfactual is a scientist who holds the same characteristics of scientist i but is not affected by the reform. After some rewriting

and a first order approximation, the log of the new production of publications of scientist i is equal to:

$$\begin{aligned}
\log\left(\widehat{Pub_{it}}\right) &= \log\left(\frac{\gamma_{it}^{pub}}{3}\right) + \log\left(2\gamma_{it}^{pub} - (1-2\lambda)\left(\gamma_{it}^{pat} + \Delta\right)\right) \\
&= \log\left(\widehat{Pub_{kt}}\right) + \log\left(1 - \frac{(1-2\lambda)\,\Delta}{2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat}}\right) \\
&\approx \underbrace{\log\left(\widehat{Pub_{kt}}\right)}_{\substack{\text{publication outcome} \\ \text{without reform}}} - \underbrace{\frac{(1-2\lambda)\,\Delta}{2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat}}}_{\substack{\text{change in publication outcome} \\ \text{due to the reform}}} . \qquad (3)
\end{aligned}$$

We define the treatment effect, TE_{it}, of the reform on publications as:

$$TE_{it} \equiv -\frac{(1-2\lambda)\,\Delta}{2\gamma_{it}^{pub} - (1-2\lambda)\,\gamma_{it}^{pat}}. \qquad (4)$$

6.2 Insights from the model

First of all, we can see that the model maps the empirical specification exactly. The term on the left of Equation (3) is equivalent to the fixed-effect estimate of scientists' production of Equation (1). This means that the average treatment effect from the empirical results is equivalent to TE, the treatment effect in the model.

Observation 1. The effect of the reform on publications can go in two directions depending on the level of complementarity. If complementarity is high enough ($\lambda > 1/2$) stronger patent rights increase the total returns of effort spent on publications since part of this effort can be translated into more patents. This leads to an increase in publication efforts and in the end an increase in publications. However, if complementarity is not as high ($\lambda < 1/2$) then the substitution effect prevails: some effort is shifted from publications to commercialization, leading to a decrease in publications. There is a tension between complementarity and substitution whose total effect is captured by the sign and the magnitude of TE.

The fact that our empirical estimates show a negative effect implies that $TE < 0$ and hence that the level of complementarity is not high enough to compensate the substitution

effect.[17]

Observation 2. The magnitude of the treatment effect TE depends negatively on the value of γ^{pub}.

This implies that scientists with the highest publication ability should be the least affected by the reform. In contrast, the ones with the lowest ability should be affected the most since they have the highest incentives to shift to patenting.

6.3 Complementarity and the direction of research

So far the value of λ, the complementarity, was taken as fixed when in reality it is tied to the topic the scientist works on. Research in a given topic may be much more difficult to patent than research in another topic. Similarly, the facility to publish may differ across topics. To account for these effects, we extend the previous model by including two variables varying according to the topic. Let $r_i(\tau) \in [0,1]$ represent the ease to publish in the topic τ for scientist i and $\lambda(\tau)$ the level of complementarity for the topic τ. Further, we assume that the research topic can be represented by a real value ranging from 0 to 1. The main properties of $r_i(\tau)$ and $\lambda(\tau)$ are[18]

$$\begin{aligned} r_i(0) &= 1 \qquad & \lambda(0) &= 0 \\ r_i(1) &= 0 \qquad & \lambda(1) &= 1 \\ r_i' &< 0 \qquad & \lambda' &> 0 \\ r_i'' &< 0 \qquad & \lambda'' &< 0. \end{aligned}$$

[17] Section 8.3. discusses the implications of complementarity more broadly.

[18] The form of r_i, the ease to publish, seems restrictive but is in fact very general. Indeed, it is specific to each researcher and hence includes any kind of scientist-specific characteristics making some topics more easy to publish in: one's own taste for a given topic, skills, demand for the topic, etc. Note that we use a general notation for the topics with τ, but best is to see it as a remapping of existing topics into each scientist's own ordering of topics: $\tau \equiv m_i(\tilde{\tau})$ with $\tilde{\tau}$ the topics common to all scientists and with m_i the mapping defined as making the properties of r and λ hold. This layer of complexity is not needed for the exposition and is then omitted since the simplification can be made without loss of generality.

In other words, there is a trade-off[19] between the ease to publish and the complementarity, and both functions are concave in the direction of their maximal value.[20] The new production functions are:

$$Pub_{it} \sim Poisson\left(r_i\left(\tau_{it}\right)\gamma_{it}^{pub}e_{it}^{pub}\right), \quad Pat_{it} \sim Poisson\left(\gamma_{it}^{pat}\left(e_{it}^{pat}+\lambda\left(\tau_{it}\right)e_{it}^{pub}\right)\right).$$

Proposition 1. The reform induces all scientists to switch to topics which are more complementary to patents. The scientists with the lowest publication ability, γ^{pub}, are the ones making the most important topic changes.

For the proof see Appendix C.2 where we establish that the topic changes by an amount ϵ approximately equal to:

$$\epsilon = \frac{\lambda'\left(\tau_0\right)\Delta}{-r_i''\left(\tau_0\right)\gamma_{it}^{pub}-\lambda''\left(\tau_0\right)\left(\gamma_{it}^{pat}+\Delta\right)},$$

with τ_0 the topic in which the scientist was working before the reform. Since the second derivatives of r and λ are negative, the change is always positive. Further, as γ^{pub} or γ^{pat} increases, *ceteris paribus*, the magnitude of the change decreases. This means that a) scientist move to topics which are more difficult to publish in but are more complementary to patents and b) scientists with the highest ability change the least.

Accounting for the change in topic, the total effect of the reform can be written as:

$$TE_{it}^{topic} = \epsilon\frac{r_i'\left(\tau_0\right)}{r_i\left(\tau_0\right)} + 2\epsilon\frac{r_i'\left(\tau_0\right)\gamma_{it}^{pub}+\lambda'\left(\tau_0\right)\left(\gamma^{pat}+\Delta\right)}{\Omega} - \frac{\left(1-2\lambda\left(\tau_0\right)\right)\Delta}{\Omega},$$

with $\Omega \equiv 2r_i\left(\tau_0\right)\gamma_{it}^{pub}-\left(1-2\lambda\left(\tau_0\right)\right)\gamma_{it}^{pat}$. There are three terms. The first is a flat decrease in publications, since $r'\left(\tau_0\right)<0$, depends only on the ease to publish. The second term depends on the curvature of the ease to publish and the complementarity functions, the sign depending on the relative importance of the two: for instance if complementarity increases much faster than the ease to publish decreases, then this leads to a positive

[19]The trade-off is needed to exclude the unrealistic possibility that there exists a topic which is at the same time easier to publish in, and more complementary, than any other topic, so that by construction scientists would never change from this topic.

[20]Let $\tilde{r}_i\left(x\right) \equiv r_i\left(1-x\right)$ be a remapping of r into an increasing funtion from the minimal to the maximal value of r. Then $\tilde{r}'\left(x\right) = -r_i'\left(1-x\right) > 0$ and $\tilde{r}''\left(x\right) = r_i''\left(1-x\right) < 0$, proving that $\tilde{r}$ is indeed concave.

effect. Finally, the third term is similar to the effect without topic change: the substitution effect, mediated by complementarity. Linking these insights to our empirical results, this means that the decrease in publications can come both from a substitution effect and a topic-change effect. While the former effect only influences the number (and quality) of publications, the latter effect can influence the direction of research as a whole. It is hence interesting to investigate empirically if, in line with the model, scientists did change their topics towards research with greater complementarity to patents, that is, more applied.

7 Mechanisms

7.1 Scientist heterogeneity

According to the model, scientists with the lowest publication ability before the IP reform have the highest incentives to reallocate their efforts towards commercialization activities. To test this mechanism, we approximate the publication ability of the scientists by citation-weighted publications. The citation-weighted variable speaks to the theoretical model presented in Appendix C.1. We split the sample into five groups of scientists, based on their positions in the ex ante distribution of the average number of citations per publication. Reflecting the skewness of the citation distribution, the first three groups are the first three quartiles, the fourth group comprises scientists in the $[75; 90]$ percentile, while the last group constitutes the most-cited scientists $]90; 100]$ percentile). The distribution for the U.S. and Europe is given in Table 4. It appears that, except for the top group, there are stark differences between European and U.S. computer scientists, the latter having almost twice the number of citations along the distribution. Hence we categorize U.S. and European computer scientists according to the distribution of their respective region.

We estimate DiD models with scientist and time fixed-effects for the five groups. The results (coefficient estimates and standard errors) are given in Figure 4, in which the dependent variable is the number of publications. In line with the model predictions, we observe that the effect is largest for the first quartile, followed by the second quartile. The associated coefficient estimates are -0.25 and -0.24 respectively. We note that the

Table 4: Distribution of average citations per article before 1996, separated between European and U.S. scientists.

	Ex ante average # of citations per publication							
Percentile:	0%	10%	25%	50%	75%	90%	99%	100%
Europe	0	0	1	3.6	10.8	26.4	110.8	1230.7
U.S.	0	0.25	2	7	19.9	44.3	214.9	1269.6

Source: Authors' own calculations based on Web of Science data.

effect size of the reform decreases along the ability distribution, reaching a non-significant coefficient estimate of -0.007 for the ex ante highly cited scientists. This pattern is corroborated when using quality weighted measures: the decrease is most important for the lowest quartiles and becomes gradually less important as we go up the quartiles.

Finally, in line with the model, Figure 5 also shows that U.S. low publication ability scientists are the group with the highest increase in patenting rates. The average number of patents per year of that group rises from 0.012 before 1996 to 0.029 after the change in the law. This is the highest increase across all citation categories, both in absolute value and relative terms (+139%).[21]

7.2 Do scientists change the direction of their research?

One key question of interest is whether this change in output is also linked to a change in content. Tracking changes in content is a difficult endeavor since content is not a quantitative characteristic but a qualitative one. In this section, we proxy the idea of content with an appliedness measure based on the keywords from publications and abstracts from patents.

Our appliedness measure captures the following idea: how much do publications look like software patents? After extracting software patents from the USPTO, we categorize publication keywords into three categories: 1) applied (frequently appearing in patents abstracts), 2) neutral and 3) basic (almost never appearing in patent abstracts). After normalization, we end up with a measure ranging from -1 to 1. The value -1 represents a publication with only basic keywords while a value of 1 represents a publication with

[21] The graph further illustrates a positive correlation (especially *ex ante*) between the scientists' position in the citation distribution and the number of patents produced.

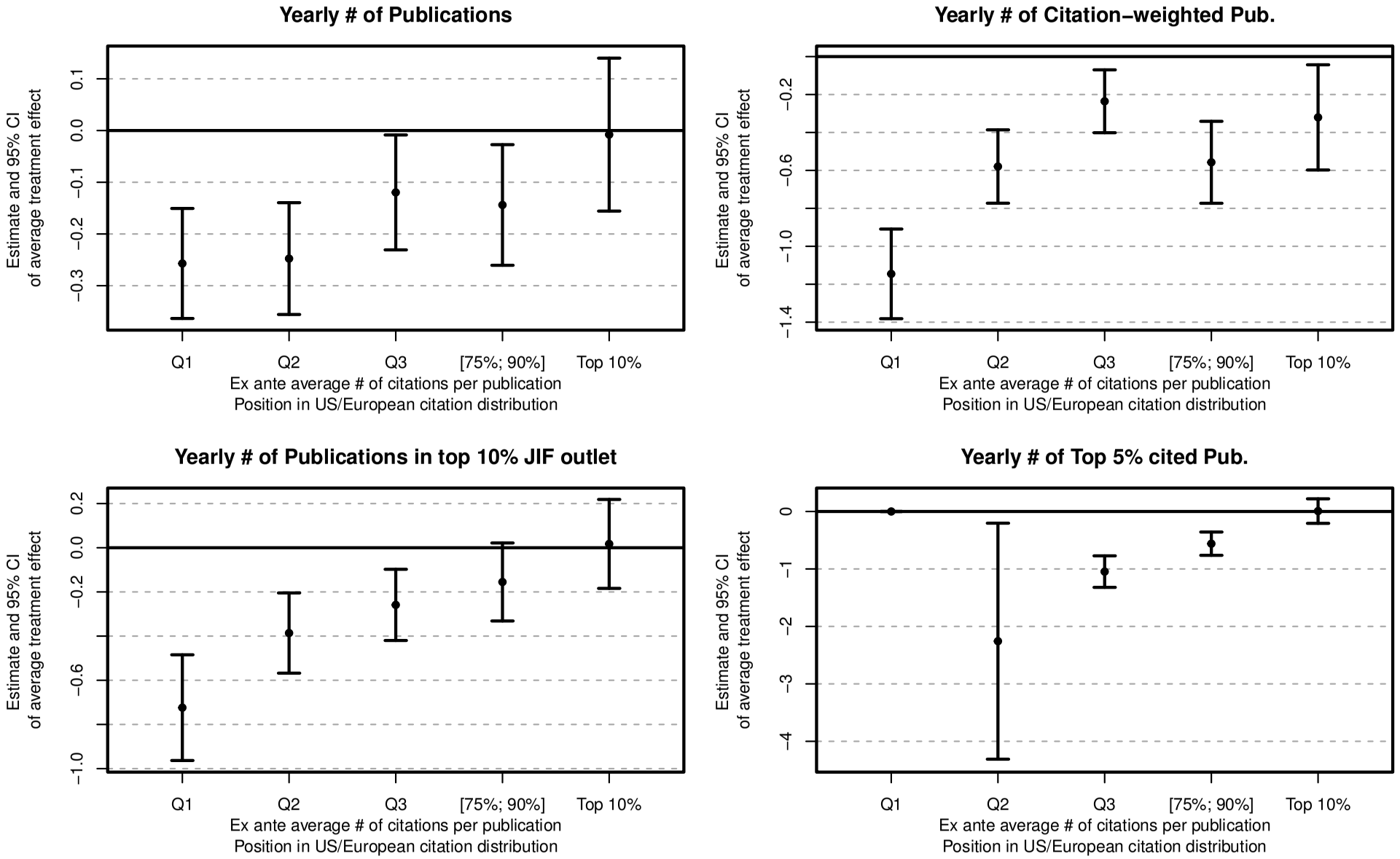

Figure 4: The effect of software patent introduction, mediated by an *ex ante* measure of ability.

Notes: The graph reports the estimates and 95% confidence intervals of the average treatment effect for 5 separate regressions, where the full sample is split according to the ex ante (1989-1996) average number of citations received per publication. Each regression is a Poisson, or Negative Binomial (for the citations variable), fixed-effects estimation with scientist and year fixed-effects. In the last pane, the first quartile had no scientist with a top 5% cited publication.
Source: Authors' own calculations based on Web of Science data.

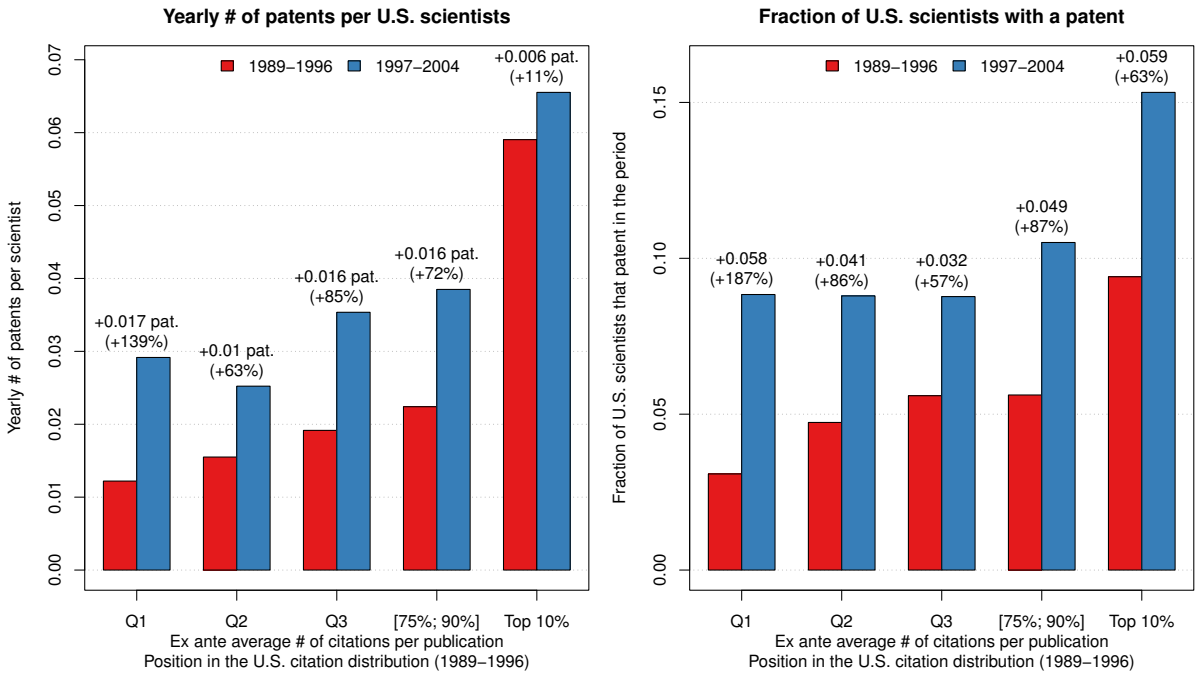

Figure 5: Evolution of patent production by U.S. university software scientists, split by *ex ante* citation level.

Source: Authors' own calculations based on Web of Science and USPTO patent data.

only applied keywords. The construction of this variable is detailed in Appendix D with examples of its validity. The key advantage of this measure is that it is based on information external to the publications (abstracts from software patents), so that the value of this measure is independent of the patenting status of the scientists.

We now use this measure to track whether the scientists change the direction of their research following the reform. The dependent variable, appliedness, is only defined when a scientist has one or more publications with non-missing keywords. Overall, about 60% of publications do not report a keyword, leading to a large number of missing values. This in turn creates many holes in the panel which can introduce artifacts in the estimates. To avoid this problem, we pool the data for each scientist into two periods: 1990-1996 and 1997-2004.[22] This means that the appliedness measure is equal to the average appliedness across all publications with non missing keywords in the period. We end up with a full panel of two periods with no missing values.

Table 5 reports the difference-in-difference estimations for the full sample and when the sample is split by the *ex ante* average number of citations. The first thing we no-

[22]Keywords in publications are only available from 1990 onwards in our data set.

tice from the means of the appliedness measure reported in the table, is that appliedness is decreasing with the average *ex ante* ability of the scientists. For the first group, the appliedness score is -0.028, and this value decreases steadily across the groups until reaching -0.136 for the top 10% scientists. According to this measure, the scientists with the highest level of citations in computer science are the ones doing the most basic research.

The estimation for the full sample finds an estimate of the average treatment effect of 0.023, meaning that U.S. scientists tend to shift the content of their research towards more applied work following the reform. When we break down the estimation by the *ex ante* average ability, we can see that this result is mainly driven by the first group (the scientists with the lowest number of citations) for which the estimates reach an increase in appliedness of 0.042, an important increase given a mean of -0.028. The second group with the highest increase represents scientists in the 75-90 percentiles of the citations distribution, this group experiences an increase of 0.039 in appliedness, although this number is imprecisely estimated.

Overall, we can conclude from these results that the reform increased the appliedness of the works of U.S. computer scientists. The change was especially important for scientists at the left of the ability distribution, in line with the predictions of the theoretical model.[23]

8 Discussion

8.1 Quantifying the consequences

To evaluate the counterfactual scientific output of U.S. computer scientists had patent rights for software inventions not been introduced, we compute the total number of publications of the software scientists of our sample in the period 1997-2004, $prod_{US}^{1997-2004}$. The counterfactual situation can be written as $prod_{US}^{1997-2004} / \exp(-0.20)$ where -0.20 is the estimate of Table 1. We identify a total loss of 5,246 publications (95% confidence interval: $[3743; 6830]$). This drop is significant: MIT, for instance, had 3,735 publications

[23]Note that the results of this section hold when we use an alternative measure of appliedness: conference proceedings. Conference proceedings, on average, can be considered as hosting publications with more applied content. In results displayed in Appendix D.3, we observe a decline in journal articles much more pronounced than in conference proceedings, the decline in journal articles being highest for lower ability scientists. This is another suggestion that scientists shift their research direction towards more applied topics.

Table 5: Effect of the reform on the direction of research, estimations split by the ex ante position in the citation distribution.

Dependent Variable:	Yearly Average Appliedness Score					
Ex ante position in the citations distribution	Full sample	Q1	Q2	Q3	[75%; 90%]	Top 10%
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Post	-0.0573***	-0.0498***	-0.0502***	-0.0610***	-0.0840***	-0.0447**
	(0.0055)	(0.0094)	(0.0122)	(0.0112)	(0.0143)	(0.0193)
Treat × Post	0.0229***	0.0424***	0.0183	0.0136	0.0397*	-0.0197
	(0.0081)	(0.0149)	(0.0171)	(0.0166)	(0.0212)	(0.0279)
Fixed-effects						
Scientist	✓	✓	✓	✓	✓	✓
Fit statistics						
Dependent variable mean	-0.07261	-0.02822	-0.05457	-0.08969	-0.11097	-0.13698
Observations	16,266	4,562	3,594	4,044	2,442	1,624
R^2	0.59890	0.55703	0.57598	0.59690	0.63557	0.62278
Within R^2	0.01710	0.01178	0.01343	0.02155	0.03357	0.01851

Clustered (Scientist) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The coefficients correspond to OLS estimates. The sample consists of active computer scientists working in software. An active computer scientist is defined by having at least two publication before 1996 (with one before 1994), and at least one publication in the period 1997-2004. Observations correspond to *scientist* × *period*, with the two periods being 1990-1996 (pre) and 1997-2004 (post).
Source: Authors' own calculations based on publication data from Web of Science and patents from the USPTO.

in computer science during 1997-2004. It represents about 1% of the worldwide WoS computer science publications in that period.[24] Using the same logic, we find a loss of 1,693 publications in the top 10% JIF journals, as well as a loss of 1,076 publications in the top 5% most-cited publications worldwide.

In order to investigate whether the loss in publications is outweighed by a gain in patents, we estimate a simple model assessing the change in patenting after the IP reform with scientist fixed-effects and a time trend:

$$E\left(y_{it}\right) = \exp\left(\alpha_i + \beta \times Post_t + \gamma \ln Trend_t\right),$$

where index i represents U.S. university software scientists. In the absence of a good counterfactual situation regarding patenting behavior, β provides an approximation of the gain in patents due to the reform. From the estimates reported in Table B6 in Appendix, we obtain a gain of 320 patents. By combining the two results, this leads to a total "price" of 16 publications per patent or 3.35 top-cited publications per patent.

Lastly, we approximate the importance of published software inventions for technology development before and after the reform, and attach a monetary value to the technology that the software publications inspire. We therefore track the scientific publications of our academic computer scientists in patent documents in the first five years after their publication date[25] (Ahmadpoor and Jones, 2017) and weight each matched patent by the monetary value measure of patents provided by Kogan et al. (2017) (henceforth KPSS).[26] Summing up the KPSS value for the patents inspired by each individual software publication provides a measure for the total monetary value generated by the published software invention.[27] Acknowledging the skewness of the patent value distribution, we also define two measures as the count of the number of patents above the 50th and 90th percentile

[24] This is a conservative estimate, since we applied the counterfactual only to the U.S. computer scientists who: i) do not work in hardware-related fields; and ii) have at least one publication before 1994. Therefore, in computing the counterfactual, we neglected the effect of the reform on scientists entering after 1994.

[25] We matched every patent from the USPTO applied in 1989-2009 to the publications of the U.S. computer scientists of our sample using the non-patent literature contained in patents. See Appendix A.5 for details on the matching procedure.

[26] Kogan et al. (2017) attribute monetary value to patents of public firms based on stock price variations around the grant announcement of patents.

[27] The monetary value is in 1982 constant dollars.

in KPSS values in their yearly patent application cohort that built on their software publication.

Table 6 presents a comparison of the pre- and post-reform citations in patents and the KPSS value of patents. The first row presents the number of patent citations per publication. U.S. publications have gained influence in the technology domain over time, since the average number of associated patent citations doubled between the pre- and post-reform period, rising from 0.16 to 0.36. Interestingly, about ten percent of the citations come from university patents. This is an important share given that university patents account for less than 2% of all patents at the USPTO.[28] Since the share of citations from university patents stay the same across the two periods, we can rule out that the surge in citations stem from the new university patents. Weighting the patents by their KPSS value leads to the opposite picture. In the pre-reform period, 1989-1996, U.S. publications were cited by patents worth a total of $3 million. This amount dropped to $1.8 million in the period 1997-2004.

The last two rows show the measures that weight patents by their position in the KPSS value distribution. It appears that the number of citations from patents with values above the median significantly increased, while the top 90% increased only marginally, indicating that the drastic KPSS value change is not only driven by a few highly valuable patents.

8.2 Comments on the dotcom bubble

A potential concern is that the dotcom bubble and its burst in 2000 might impact our results. The impact of the dotcom bubble on academic scientists might have been twofold. On the one hand, the financing of university research might have been increased due to the additional industry funding available during the dotcom period with the implication that the nature of university research shifted from rather basic to rather applied topics. On the other hand, university scientists might have faced increased incentives to leave the university to start their own company during the dotcom period.

A different financing pattern of university research would only be problematic for the

[28]The information on the type of applicant has been obtained from the EEE-PPAT data base (Callaert et al., 2011) which both harmonizes applicant names and categorizes applicants by type of institution (mostly whether the patent applicant is an individual, a private company or an university).

Table 6: Evolution of the influence of U.S. computer scientists' publications on innovation.

Unit of Observation: Publication	1989-1996 mean (s.e.)	1997-2004 mean (s.e.)	Diff.	t-stat
# Patent Citations	0.16 (1)	0.34 (1.8)	0.184	12.5
# Patent Citations from University Applicants	0.016 (0.16)	0.034 (0.3)	0.0182	7.7
Million $ weighted Patent Citations (KPSS)	3 (42.9)	1.8 (24.5)	-1.15	-3.05
# Patent Citations with KPSS value > p50	0.049 (0.46)	0.081 (0.69)	0.0317	5.41
# Patent Citations with KPSS value > p90	0.012 (0.17)	0.015 (0.19)	0.0023	1.31
Observations	15,925	22,338		

Notes: Each publication produced in year t is weighted by the number of patent citations it receives from patents applied between t and $t+5$, each patent is further weighted by their monetary value from KPSS. KPSS refers to the patent value created in Kogan et al. (2017) who use stock variations around patent grants to infer monetary value. The last two variables attribute a weight of 1 to patents whose KPSS value exceeds the according percentile across the set of all patents applied in the same year.
Source: Publication data from WoS, patent data from the USPTO, applicant type from EEE-PPAT (Callaert et al., 2011) and KPSS patent value from Kogan et al. (2017). Author's own calculations.

analysis at hand if a different trend for the U.S. and Europe is observed. More specifically, one might expect that the dotcom bubble had a stronger effect on U.S. universities, leading U.S. university researchers toward engagement in more applied research projects. Figure 6 shows that this is not the case. We see that - after the business share of higher education financing increased slightly in Europe and the U.S. - it dropped significantly in the U.S. after the dotcom bubble period, while there it remained on a high level for Europe. We therefore have no reason to expect that U.S. researchers would have shifted their efforts closer to commercialization than European universities due to different financing sources during the dotcom bubble period.

Second, thanks to the richness of financing opportunities prior to the burst of the dotcom bubble, computer scientists may have left academia to start their own businesses. In Table B7 in Appendix, we show the average treatment effect when only "stayers" are considered: scientists with at least one publication in the years 2003-2004 and who are still affiliated to an university. The coefficient estimates are similar to those of the full sample. This suggests that the main findings of the paper are not driven by researchers dropping out of academia to pursue commercialization activities.

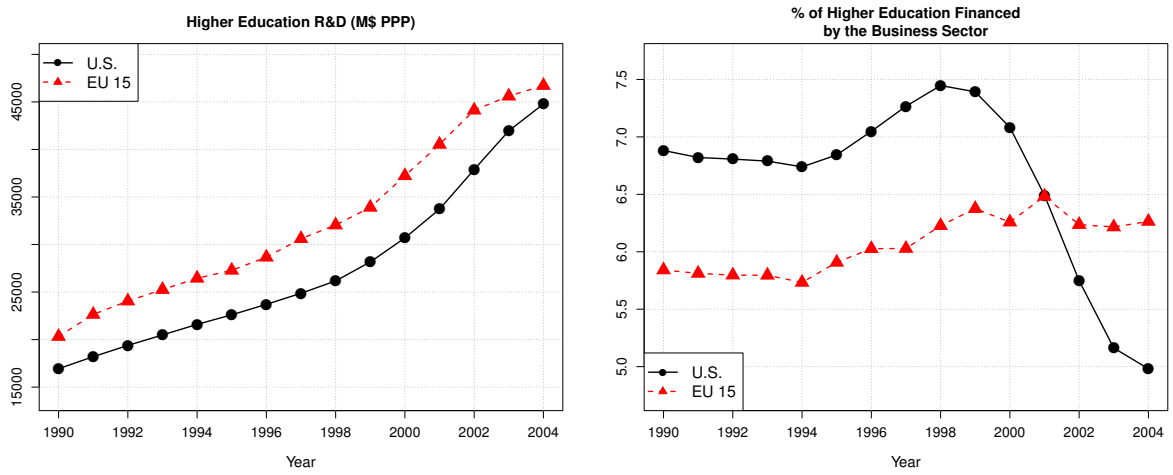

Figure 6: Evolution of the amount and type of financing of higher education R&D.
Source: OECD data.

8.3 Further discussion: Connections to the literature, generalization and avenues for future research

Our results clearly show that strengthening institutional IP ownership rights for universities has a negative effect on the research conducted at universities. It appears that citation-weighted publications decline in the aftermath, especially for researchers with low publication ability. Our results therefore speak to the literature that has long articulated concerns that the growing involvement of scientists in the commercialization of their research may have negative implications for the traditional research process (see, e.g., Blumenthal et al., 1996; Campbell et al., 2002; Krimsky, 2003; Murray and Stern, 2007; Fabrizio and Di Minin, 2008; Azoulay, Ding and Stuart, 2009; Czarnitzki, Glänzel and Hussinger, 2009; Czarnitzki, Hussinger and Schneider, 2011).

Was the introduction of software patents good or bad for universities? The answer crucially hinges on the objectives of universities (Thursby, Jensen and Thursby, 2001; Perkmann et al., 2013). Aside from research and teaching, one third, commonly admitted, mission of universities is the diffusion of knowledge to the economy (Laredo, 2007; Etzkowitz, 2002). By bringing scientists closer to the market, patents help fulfilling this third mission (Eisenberg, 1996; Hall, Link and Scott, 2003). Our results highlight that scientists do not live in isolation from the institutional setting. On the one hand, the introduction of IP rights does increase patenting which is arguably a good thing given the *third mission*. On the other hand, each extra patent produced comes at the "price" of

multiple quality-weighted publications, negatively affecting the university's research mission. Whether the excess patents produced is worth the publication loss is up to decision makers.

Our results also complement evaluations of Bayh-Dole Act type policies, which focus on the effects on the commercialization of university generated inventions (e.g., Henderson, Jaffe and Trajtenberg, 1998; Mowery and Ziedonis, 2002; Sampat, Mowery and Ziedonis, 2003; Hvide and Jones, 2018; Czarnitzki et al., 2016; Ejermo and Toivanen, 2018). With a focus on the U.S., these studies typically conclude a positive effect on the commercialization of university generated inventions. Here, we show that there is also an associated cost in terms of a loss of science. From a policy perspective, this negative effect should be taken into account when assessing the overall benefit/cost of the patent legislation on the economy (on this debate, see, for instance, Boldrin and Levine, 2013; Sampat, 2018; Williams, 2017).

Our findings are also connected to the empirical literature on the interplay between patenting and publishing (Henderson, Jaffe and Trajtenberg, 1998; Azoulay, Ding and Stuart, 2009; Fabrizio and Di Minin, 2008; Czarnitzki, Glänzel and Hussinger, 2009). Overall, this literature suggests a positive relationship, or complementarity, between publication and patent outcomes, for which the exact mechanisms vary across studies. If there is strong complementarity, such that patents would emerge as a side product of science production, one might expect that strengthening IP rights would result in more publications via an increase in the total returns of scientific work. At first glance, this seems to stand in contrast to our empirical results.

Our findings can, however, be reconciled with this literature, for two reasons. First, our study fundamentally departs from this strand of literature, in that we examine a regime shift changing scientists' incentives, as opposed to these studies in which the rules of the game remain static.[29] We therefore avoid the problem of self-selection of scientists' into patenting. That is, we do not neglect the scientists who engage in commercialization activities that are not visible in patents following the IP reform.

Second, our results accommodate complementarity. Within the model outlined in Section 6, the reduction of publications in the presence of complementarity is plausible:

[29] A notable exception can be found in the supplementary appendix of Hvide and Jones (2018).

we empirically find that the complementarity is not strong enough to compensate for the change in incentives implied by the introduction of software patents. Note that in line with previous studies, we still find a strong positive empirical link between productivity in publications and patenting, as evidenced in Figure 5.

Our study also connects to the discussion of the broadening of the patentable subject matter where we see large differences between the U.S. and Europe. The U.S. grants, for instance, patents rights for plants since the Plant Protection Act of 1930 (e.g. van Overwalle, 1999; Fowler, 2000; see provisions of 35 U.S.C. 161). More recently, in June 2013, the U.S. Supreme Court took a clear stance on the patentability of genes by ruling that human genes cannot be protected because DNA is a "product of nature" so that nothing new is created when a gene is discovered (Association for Molecular Pathology versus Myriad Genetics, 569 U.S. 576). With this decision, the U.S. Supreme Court invalidated the more than 4000 gene patents that were granted until then (Ingram, 2014). While certainly, each of the extension of patent rights towards a new topic demands distinct evaluations regarding the multiple possible implications for science and the business sector, by investigating the implications of the broadening of the patentable subject matter towards software we contribute by delivering one piece of evidence on the potential consequences.

Can our results be generalized? Although our empirical setting focuses on a very specific setting, the question we have investigated is of general interest and informs about the costs of patent rights in terms of a loss in basic science. Whether our empirical results are generalizable to other fields depends on three mediators that are field-specific: 1) complementarity, defined as the extent to which patents are a by-product of regular scientific research; 2) the difficulty to obtain a patent; and 3) the expected value of the patent. The question of generalizability is further complicated by the interdependence of these three mediators.[30]

First, the higher the complementarity between patents and publications, the lower the expected negative effect of the introduction of IP rights. In fields in which patents directly follow from major scientific discoveries, such as in biotech, we would expect a lower

[30]Note that these three factors are captured by the model in Section 6: 1) the complementarity corresponds to λ,2) the difficulty to patent to γ^{pat} and 3) the expected patent value can be modeled with a coefficient associated to Pat_{it} in the utility function.

decrease in publications. Evidence from a survey of computer scientists and electrical engineers in U.S. universities conducted by Love (2014) suggests that the complementarity between patents and publications in the field of computer science is not high.

Second, patents for software are known to be among the easiest to obtain (see, e.g., Bessen and Hunt, 2004; Webbink, 2005; Bergstra and Klint, 2007).[31] From our theoretical model, it is expected that increasing the difficulty of obtaining a patent reduces the negative effect of introducing IP rights because incentives are less affected.[32] In this regard, the effect should be stronger in our setting, as opposed to other fields in which patents are more difficult to obtain, such as human genomics (Sampat and Williams, 2019).

Third, patents do not have the same value across fields. Whereas there is agreement in the patent literature that patents in the fields of chemistry and pharmaceuticals have the highest commercial value, software patents appear to be at the other end of the spectrum (see, for instance, Williams, 2017 or Sampat, 2018 for recent reviews). Thus, patents should more strongly distort researchers' incentives following an IP rights introduction in fields in which patents are associated with a high commercial value. However, patents with the highest commercial value also tend to be most "science-based" (Ahmadpoor and Jones, 2017) which in turn increases their complementarity with publications, possibly offsetting the negative effect of IP rights on publications. The balance between the two mediators and the net effect in other fields are open empirical questions.

Another important question is whether the withdrawal of patent rights would increase the production of publications of U.S. computer scientists. Our results suggest so, but this claim must be made with great caution, since the mechanisms involved in adding patent rights may well differ from the mechanisms involved in removing them.[33] This question deserves attention and recent developments in the judicial debate on the patentability of software following *Alice Corp. v. CLS Bank International* (573 U.S. 208 [2014]) may allow for a future empirical assessment.

[31] Regarding the ease of obtaining a software patent Webbink (2005) state that "every trivial combination or extension of prior software technology is being accorded the same protection as a groundbreaking drug" (§6).

[32] In the terminology of the model, increased difficulty to obtain a patent is equivalent to reducing γ^{pat}, the ability to patent.

[33] Due, for instance, to the endowment effect (Kahneman, Knetsch and Thaler, 1990).

An interesting avenue for future research would be to investigate the implications of the introduction of software patents for the mobility of U.S. computer scientists from academia to industry (e.g. Zucker, Darby and Torero, 2002; Crespi, Geuna and Nesta, 2007; Kaiser et al., 2018; Toole and Czarnitzki, 2010). In particular: Was there a brain drain? Who were the scientists moving to the private sector and was the performance of their peers impacted? Did it improve the performance of the private sector? Such questions are out of the scope of this paper since our study focuses on the consequences on *academic* scientists, leading us to restrict the analysis to scientists who stay in academia before and after the reform. We leave these important questions for future research.

A related issue is whether the opportunities to collaborate with industry partners pushes academic scientists to patent (Murray and Stern, 2007). Although survey evidence of top U.S. computer science departments suggests that "*ex ante funding from the government for high-tech research is plentiful on university campuses*" (Love, 2014, p. 319) with 83% of the respondents reporting that they had at least one patent covering research financed by the government, this might not be the case for all U.S. universities and computer scientists (Agrawal and Henderson, 2002; Gans and Murray, 2011). Exploring potential heterogeneity among scientists in this regard would be an interesting avenue for future research. Another factor which might push scientists to patent can be the desire to benefit from royalties (Lach and Schankerman, 2008). Although survey evidence reports that most computer scientists do not even know whether their university had a royalty sharing program and what percentage of the royalty faculty inventors are entitled to (Love, 2014, p. 317), there might again be heterogeneity among scientists which deserves further attention.

Finally, a recent trend possibly affecting the behavior of scientists is the surge of contributions to open source software (OSS) whose importance grows exponentially since the early 2000s (Deshpande and Riehle, 2008). In the private sector some evidence suggest that OSS can be used as a signaling mechanism (Lerner and Tirole, 2002) and lead to faster career progress for computer scientists (Riehle, 2015; Huang and Zhang, 2016). Does OSS affect university scientists in the same way? Akin to patents, contributions to open source software is another output from research activities. In light of the model of Section 6,

the effect of OSS can then be decomposed into a substitution, a complementarity and a research direction effect. The total consequences on the production of scientists and the respective contribution of each component remains an open empirical question worth investigating.

9 Conclusion

We investigate how the introduction of patent rights for software inventions affect the scientific output of U.S. computer scientists. Results from difference-in-difference analysis with European software scientists as a counterfactual for our treatment group of U.S. software scientists show that scientists reallocate their efforts from publishing to patenting in response to the reform. We evidence a 20% reduction in U.S. publication numbers: this suggests publications and patents are two, at least partially, competing tasks. Computer scientists who suffer the biggest drop in publications are those having the lowest level of citation-weighted publications before the reform. We also show that the change had an influence on the content produced by U.S. computer scientists, with their research output becoming more applied following the reform.

In summary, our results cannot rule out concerns about the negative implications for science of a regime shift toward stronger commercialization options due to strengthened IP ownership rights for universities. Furthermore, we show that IP policies which are not focused on universities can have important side effects for science.

References

Adams, James D. 1990. "Fundamental stocks of knowledge and productivity growth." *Journal of Political Economy*, 98(4): 673–702.

Aghion, Philippe, and Peter Howitt. 2005. "Growth with quality-improving innovations: an integrated framework." In *Handbook of economic growth*. Vol. 1A, , ed. P. Aghion and S. Durlauf, 67–110. Elsevier.

Agrawal, A., and R. Henderson. 2002. "Putting patents in context: Exploring knowledge transfer from MIT." *Management Science*, 48(1): 44–60.

Ahmadpoor, Mohammad, and Benjamin F Jones. 2017. "The dual frontier: Patented inventions and prior scientific advance." *Science*, 357(6351): 583–587.

Azoulay, Pierre, Waverly Ding, and Toby Stuart. 2009. "The impact of academic patenting on the rate, quality and direction of (public) research output." *The Journal of Industrial Economics*, 57(4): 637–676.

Bakels, R., A.G. Rishab, S. Torrisi, and G. Thomas. 2008. "Study on the Effects of Allowing Patent Claims for Computer-implemented Inventions. Final Report and Recommendations." *European Commission*.

Bar-Ilan, Judit. 2008. "Which h-index? A comparison of WoS, Scopus and Google Scholar." *Scientometrics*, 74(2): 257–271.

Bar-Ilan, Judit. 2010. "Web of Science with the Conference Proceedings Citation Indexes: the case of computer science." *Scientometrics*, 83(3): 809–824.

Bénabou, Roland, and Jean Tirole. 2016. "Bonus culture: Competitive pay, screening, and multitasking." *Journal of Political Economy*, 124(2): 305–370.

Bergstra, Jan A, and Paul Klint. 2007. "About "trivial" software patents: The IsNot case." *Science of Computer Programming*, 64(3): 264–285.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.

Bessen, James, and Michael J Meurer. 2005. "Lessons for patent policy from empirical research on patent litigation." *Lewis & Clark L. Rev.*, 9: 1.

Bessen, James, and Robert M Hunt. 2004. "The software patent experiment." 247–263. Paris:Éditions OCDE.

Bessen, James, and Robert M Hunt. 2007. "An empirical look at software patents." *Journal of Economics & Management Strategy*, 16(1): 157–189.

Blumenthal, David, Eric G Campbell, Nancyanne Causino, and Karen Seashore Louis. 1996. "Participation of life-science faculty in research relationships with industry." *New England Journal of Medicine*, 335(23): 1734–1739.

Boldrin, Michele, and David K Levine. 2013. "The case against patents." *Journal of Economic Perspectives*, 27(1): 3–22.

Burk, Dan L, and Mark A Lemley. 2003. "Policy levers in patent law." *Virginia Law Review*, 1575–1696.

Callaert, Julie, Mariëtte Du Plessis, J Grouwels, Catherine Lecocq, Tom Magerman, B Peeters, Xiaoyan Song, Bart Van Looy, and Caro Vereyen. 2011. "Patent statistics at eurostat: Methods for regionalisation, sector allocation and name harmonisation." *Eurostat Methodologies and Working Papers*.

Campbell, Eric G, Brian R Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A Holtzman, and David Blumenthal. 2002. "Data withholding in academic genetics: evidence from a national survey." *Journal of the American Medical Association*, 287(4): 473–480.

Cockburn, Iain M, and Megan J MacGarvie. 2011. "Entry and patenting in the software industry." *Management Science*, 57(5): 915–933.

Crespi, Gustavo A, Aldo Geuna, and Lionel Nesta. 2007. "The mobility of university inventors in Europe." *The Journal of Technology Transfer*, 32(3): 195–215.

Czarnitzki, Dirk, Katrin Hussinger, and Cédric Schneider. 2011. "Commercializing academic research: the quality of faculty patenting." *Industrial and Corporate Change*, 20(5): 1403–1437.

Czarnitzki, Dirk, Thorsten Doherr, Katrin Hussinger, Paula Schliessler, and Andrew A Toole. 2016. "Knowledge creates markets: The influence of entrepreneurial support and patent rights on academic entrepreneurship." *European Economic Review*, 86: 131–146.

Czarnitzki, Dirk, Wolfgang Glänzel, and Katrin Hussinger. 2009. "Heterogeneity of patenting activity and its implications for scientific research." *Research Policy*, 38(1): 26–34.

Dasgupta, Partha, and Paul A David. 1994. "Toward a new economics of science." *Research Policy*, 23(5): 487–521.

Deshpande, Amit, and Dirk Riehle. 2008. "The total growth of open source." 197–209, Springer.

Eisenberg, Rebecca. 1996. "Patents: help or hindrance to technology transfer?" In *Biotechnology: Science, Engineering, and Ethical Challenges for the Twenty-First Century.* , ed. Larry V. McIntire and Frederick B. Rudolph. National Academies Press.

Ejermo, Olof, and Hannes Toivanen. 2018. "University invention and the abolishment of the professor's privilege in Finland." *Research Policy*, 47(4): 814–825.

Etzkowitz, Henry. 2002. "Incubation of incubators: innovation as a triple helix of university-industry-government networks." *Science and Public Policy*, 29(2): 115–128.

Fabrizio, Kira R, and Alberto Di Minin. 2008. "Commercializing the laboratory: Faculty patenting and the open science environment." *Research Policy*, 37(5): 914–931.

Fowler, Cary. 2000. "The Plant Patent Act of 1930: A sociological history of its creation." *Journal of the Patent and Trademark Office Society*, 82: 621.

Franceschet, Massimo. 2010. "The role of conference publications in CS." *Communications of the ACM*, 53(12): 129–132.

Galasso, Alberto, Matthew Mitchell, and Gabor Virag. 2018. "A theory of grand innovation prizes." *Research Policy*, 47(2): 343–362.

Gallini, Nancy T. 2002. "The economics of patents: Lessons from recent US patent reform." *Journal of Economic Perspectives*, 16(2): 131–154.

Gans, Joshua S, and Fiona Murray. 2011. "Funding scientific knowledge: Selection, disclosure and the public-private portfolio." In *The Rate and Direction of Inventive Activity Revisited.* 51–103. University of Chicago Press.

González, Andrés Guadamuz. 2006. "The software patent debate." *Journal of Intellectual Property Law & Practice*, 1(3): 196–206.

Graham, Stuart JH, and David C Mowery. 2003. "Intellectual property protection in the US software industry." In *Patents in the Knowledge-Based Economy*. Vol. 219, , ed. W. Cohen and S. Merrill, 219–231. Washington, DC: The National Academies Press.

Griliches, Z. 1979. "Issues in assessing the contribution of R&D to productivity growth." *Bell Journal of Economics*, 10(1): 92–116.

Griliches, Zvi. 1992. "The Search for R&D Spillovers." *Scandinavian Journal of Economics*, 94: S29–47.

Guntersdorfer, Michael. 2003. "Software patent law: United States and Europe compared." *Duke Law & Technology Review*, 2(1): 1–12.

Hall, Bronwyn H, Albert N Link, and John T Scott. 2003. "Universities as research partners." *Review of Economics and Statistics*, 85(2): 485–491.

Hall, Bronwyn H, and Megan MacGarvie. 2010. "The private value of software patents." *Research Policy*, 39(7): 994–1009.

Henderson, Rebecca, Adam B Jaffe, and Manuel Trajtenberg. 1998. "Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988." *Review of Economics and Statistics*, 80(1): 119–127.

Huang, Peng, and Zhongju Zhang. 2016. "Participation in Open Knowledge Communities and Job-Hopping." *MIS Quarterly*, 40(3): 785–806.

Hvide, Hans K, and Benjamin F Jones. 2018. "University Innovation and the Professor's Privilege." *American Economic Review*, 108(7): 1860–98.

Ingram, Tup. 2014. "Association for Molecular Pathology v. Myriad Genetics, Inc.: the product of nature doctrine revisited." *Berkeley Technology Law Journal*, 29: 385.

Jaffe, A.B. 1989. "Real effects of academic research." *American Economic Review*, 79(5): 957–970.

Jones, Charles I. 2005. "Growth and ideas." In *Handbook of economic growth*. Vol. 1, , ed. P. Aghion and S. Durlauf, 1063–1111. Elsevier.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1990. "Experimental tests of the endowment effect and the Coase theorem." *Journal of Political Economy*, 98(6): 1325–1348.

Kaiser, Ulrich, Hans C Kongsted, Keld Laursen, and Ann-Kathrine Ejsing. 2018. "Experience matters: The role of academic scientist mobility for industrial innovation." *Strategic Management Journal*, 39(7): 1935–1958.

Klevorick, Alvin K, Richard C Levin, Richard R Nelson, and Sidney G Winter. 1995. "On the sources and significance of interindustry differences in technological opportunities." *Research Policy*, 24(2): 185–205.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological innovation, resource allocation, and growth." *The Quarterly Journal of Economics*, 132(2): 665–712.

Kortum, Samuel, and Josh Lerner. 1999. "What is behind the recent surge in patenting?" *Research policy*, 28(1): 1–22.

Krimsky, Sheldon. 2003. *Science in the private interest: Has the lure of profits corrupted biomedical research?* Rowman & Littlefield, Lanham, Maryland, U.S.A.

Lach, Saul, and Mark Schankerman. 2008. "Incentives and invention in universities." *The RAND journal of economics*, 39(2): 403–433.

Laredo, Philippe. 2007. "Revisiting the third mission of universities: Toward a renewed categorization of university activities?" *Higher Education Policy*, 20(4): 441–456.

Lerner, Josh, and Jean Tirole. 2002. "Some simple economics of open source." *The Journal of Industrial Economics*, 50(2): 197–234.

Love, Brian J. 2014. "Do University Patents Pay Off? Evidence from a Survey of University Inventors in Computer Science and Electrical Engineering." *Yale Journal of Law & Technology*, 16: 285–343.

Lunney Jr, Glynn S. 2000. "e-Obviousness." *Michigan Telecommication & Technology Law Review*, 7: 363.

Magerman, Tom, Bart Van Looy, and Koenraad Debackere. 2015. "Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology." *Research Policy*, 44(9): 1702–1713.

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

Mowery, David C, and Arvids A Ziedonis. 2002. "Academic patent quality and quantity before and after the Bayh-Dole act in the United States." *Research Policy*, 31(3): 399–418.

Mowery, David C, and Bhaven N Sampat. 2001. "University patents and patent policy debates in the USA, 1925–1980." *Industrial and Corporate Change*, 10(3): 781–814.

Mowery, David C, Richard R Nelson, Bhaven N Sampat, and Arvids A Ziedonis. 2001. "The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole act of 1980." *Research Policy*, 30(1): 99–119.

Mowery, DC, and BN Sampat. 2005. "Bayh-Dole Act of 1980 and university-industry technology transfer: a model for other OECD governments?" *Journal of Technology Transfer*, 30: 115–127.

Murray, Fiona. 2002. "Innovation as co-evolution of scientific and technological networks: exploring tissue engineering." *Research Policy*, 31(8-9): 1389–1403.

Murray, Fiona, and Scott 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(4): 648–687.

Noel, Michael, and Mark Schankerman. 2013. "Strategic patenting and software innovation." *The Journal of Industrial Economics*, 61(3): 481–520.

Patterson, D., L Snyder, and J Ullman. 1999. "Evaluating Computer Scientists and Engineers For Promotion and Tenure." *Best Practices Memo, Computing Research News*. CRA Computing Research Association.

Perkmann, Markus, Valentina Tartari, Maureen McKelvey, Erkko Autio, Anders Broström, Pablo D'Este, Riccardo Fini, Aldo Geuna, Rosa Grimaldi, Alan Hughes, Stefan Krabel, Michael Kitson, Patrick Llerena, Franceso Lissoni, Ammon Salter, and Maurizio Sobrero. 2013. "Academic engagement and commercialisation: A review of the literature on university–industry relations." *Research policy*, 42(2): 423–442.

Rai, Arti K. 2003. "Engaging facts and policy: A multi-institutional approach to patent system reform." *Columbia Law Review*, 103: 1035–1135.

Rai, Arti K, John R Allison, Bhaven Sampat, and Colin Crossman. 2009. "University software ownership: Technology transfer or business as usual?" *University of North Carolina Law Review*, 87: 1519–1570.

Riehle, Dirk. 2015. "How Open Source Is Changing the Software Developer's Career." *Computer*, 48(5): 51–57.

Sampat, Bhaven, and Heidi L Williams. 2019. "How do patents affect follow-on innovation? Evidence from the human genome." *American Economic Review*, 109(1): 203–36.

Sampat, Bhaven N. 2018. "A survey of empirical evidence on patents and innovation." *National Bureau of Economic Research*. Working Paper 25383.

Sampat, Bhaven N, David C Mowery, and Arvids A Ziedonis. 2003. "Changes in university patent quality after the Bayh–Dole act: a re-examination." *International Journal of Industrial Organization*, 21(9): 1371–1390.

Santos Silva, João M. C., and Silvana Tenreyro. 2006. "The log of gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Scotchmer, Suzanne. 1991. "Standing on the shoulders of giants: cumulative research and the patent law." *Journal of Economic Perspectives*, 5(1): 29–41.

Stephan, Paula E, Shiferaw Gurmu, Albert J Sumell, and Grant Black. 2007. "Who's patenting in the university? Evidence from the survey of doctorate recipients." *Economics of Innovation and New Technology*, 16(2): 71–99.

Sterne, Robert Greene, and Lawrence B Bugaisky. 2004. "The Expansion of Statutory Subject Matter Under the 1952 Patent Act." *Akron Law Review*, 37(2): 217–229.

Thursby, Jerry G, Richard Jensen, and Marie C Thursby. 2001. "Objectives, characteristics and outcomes of university licensing: A survey of major US universities." *The Journal of Technology Transfer*, 26(1): 59–72.

Toole, Andrew A. 2012. "The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry." *Research Policy*, 41(1): 1–12.

Toole, Andrew A, and Dirk Czarnitzki. 2010. "Commercializing science: Is there a university "brain drain" from academic entrepreneurship?" *Management Science*, 56(9): 1599–1614.

van Overwalle, Geertrui. 1999. "Patent protection for plants: a comparison of American and European approaches." *IDEA: The Journal of Law and Technology*, 39: 143.

Visser, Martijn S, and Henk F Moed. 2005. "Developing bibliometric indicators of research performance in computer science." *Proceedings of the 10th international conference of the international society for scientometrics and informetrics*, 275–279.

Webbink, Mark H. 2005. "A new paradigm for intellectual property rights in software." *Duke Law & Technology Review*, 4(1): 1–15.

Williams, Heidi L. 2017. "How do patents affect research investments?" *Annual Review of Economics*, 9: 441–469.

Zucker, Lynne G, Michael R Darby, and Maximo Torero. 2002. "Labor mobility from academe to commerce." *Journal of Labor Economics*, 20(3): 629–660.

Appendix for: How Patent Rights Affect University Science

A Data Appendix

A.1 Data

Our starting point is the Web of Science (WoS) database provided by Thomson Reuters from which we retrieve all publications in the WoS field "computer science" from 1989 to 2004. Computer science is defined by the WoS subfields Artificial Intelligence, Information Systems, Theory & Methods, Software Engineering, Interdisciplinary Applications, Hardware & Architecture, Cybernetics.[34] The WoS database contains both publications in scientific journals and publications in conference proceedings which are frequent in computer science and used for a speedy dissemination of results (Patterson, Snyder and Ullman, 1999). Proceedings form a growing share of computer scientists' output over time. One main feature of a conference proceeding is that its content tends to be more applied than the content of articles in traditional scientific journals. Our initial sample consisted of 655,441 unique publications. The WoS records contain various information such as authors' names and affiliations, the subfields within computer science, the number of citations, etc.

A.2 Unique identifiers

To track scientists' publications over time, we created unique author identifiers mapping the publications to individual authors. We employed a novel disambiguation approach developed by Doherr (2017).[35] In a nutshell, this method uses a "Google like" search algorithm for correcting spelling issues. It uses network analysis to disambiguate namesakes where a network is created in which two articles are connected if their shared features (such as journal name, affiliation, keywords, co-author, etc) are far enough from "chance" in terms of relative probabilities. Network analysis tools are applied to create coherent clusters of articles which are confidently identified as being written by the same author. For further details we refer to Doherr (2017).

[34] There are seven different subfields in computer science. In the full sample of publications, the number of publications per subfield is as follows: Artificial Intelligence: 203,467; Information Systems: 172,991; Theory & Methods: 164,370; Software Engineering: 138,165; Interdisciplinary Applications: 137,959; Hardware & Architecture: 97,506; Cybernetics: 33,751.

[35] The algorithm has been used for disambiguation for the articles Czarnitzki et al. (2015, 2016); Cappelli et al. (2019) among others.

In our data set which consisted of 1,889,740 author-article pairs, the algorithm identified 691,061 unique scientists. From those, we dropped all scientists that had only one publication in computer science. Those can be PhD students that left academia or scientists in related field that ended up on a scientific publication in computer science by collaboration or coincidence. This led to a sample of 199,010 unique scientists.

In the next step, we eliminated all scientists who did not have 100% of their affiliations in our sample years of software patents within the U.S. or Europe.[36] Researchers who change continent in our time window of interest are limited. A total of 5.98% of the European scientists and 8.03% of the U.S. scientists were dropped leading to a sample of 123,509 scientists from both regions. In the next step, we kept only scientists affiliated to universities across all the years, excluding scientists employed at any point by a governmental institution or a private firm. This left us with 81,377 university computer scientists.

A.3 Assigning affiliations

In order to select university scientists, a complex procedure to clean the affiliations was implemented before. We retrieved from WoS the addresses of each scientist in a given year. An address is a formatted character string, such as "Harvard Univ, Aiken Comp Lab, Cambridge, MA 02138, USA", from which we extract the institution name from the first item before the first comma (here "Harvard Univ"). We define an *institution* as a combination of an institution name and a country.[37] WoS unfortunately does not link the addresses/institutions to the individual scientists on the publications. Instead it provides a list of all addresses per publication, as well as an indication of the address of the reprint author. Moreover, 4.2% of the publications report no address and 18% do not provide reprint author information.

In order to assign an institution to each scientist-year, we applied an algorithm consisting of several steps. We started with the simplest case of scientists that have only one publication per year which contains only one institution (which then is valid for all

[36] Our definition of Europe encompasses the EU 28 countries plus Norway and Switzerland.

[37] E.g. AT&T based in Seattle in New Jersey and AT&T based in Denver in Colorado are considered as the same institution since both are in the USA

authors on the publication). In this case, the scientist can be unambiguously assigned to that institution for that year. We moved stepwise forward to the more complex cases involving scientists with several publications per year, each with multiple different institutions. In these cases, we infer the institution based on the previous information we obtained and on the frequency of occurrences of the institutions per author. In the end, we were able to assign an institution to 97% of the scientist-years.

A.3.1 Details of the algorithm

Using the reprint and regular addresses, we create an heuristic to assign each *author-year* to an unique institution. These heuristics can be split in two categories: 1) unequivocal cases, and 2) equivocal cases.

We start with unequivocal cases. We apply these three heuristics successively:

1) The reprint author information is the most reliable source of information. In a given year if: i) an author is at least once a reprint author, ii) all her reprint institutions are the same and iii) this institution is strictly included in the set of institutions of her non-reprint-author articles: Then we assign the author-year to the reprint institution. This leads to an identification of 31.9% of the sample.

With the author-years identified, we can remove them from the sample and update the institutions for each article. This means concretely that if an article contains N authors and K institutions, and if $N-1$ authors are identified and attributed to $K-1$ institutions, then we "*know*" that the N^{th} author is affiliated to the K^{th} institution. We then use this information as follows:

2) In a given year if all of an author's articles display only a unique institution (using the updated article-institution information): Then we assign the author-year to that institution. This methods identifies 76.9% of the sample.

3) In a given year if i) an institution appears in all the articles of an author and ii) that institution appears *strictly* more times than the other institutions: Then we assign the author-year to that majority institution. Now 79.9% of all author-years are identified.

An illustration of these methods are given in Table A1.

Now a scientist may change institutions in a given year, and have two articles in the

Table A1: Example of unequivocal identification of author-year institutions.

Case 1. Assume that in 2000 Jane Doe has only two articles:

Author	Article ID	Institution	Is reprint author
Jane Doe	1	HARVARD UNIV	Yes
Jane Doe	1	UNIV ILLINOIS	No
Jane Doe	2	AT&T BELL LABS	No
Jane Doe	2	HARVARD UNIV	No

Only looking at the addresses we cannot assign her to an institution. However, as she is a reprint author in the first article, in which her institution is Harvard, and Harvard also appears in her second article, then we can assign her to Harvard.

Case 2. In 2000, assume that John Smith has two articles. These articles contain several institutions, but some institutions were identified to other author-years via the previous method and John Smith is the last unidentified author:

Author	Article ID	Institution	Is reprint author
John Smith	1	~~HARVARD UNIV~~	No
John Smith	1	UNIV ILLINOIS	No
John Smith	3	UNIV ILLINOIS	No
John Smith	3	~~UNIV CAROLINA~~	No

Then we assign John Smith to UNIV ILLINOIS.

Case 3. In 2000, Julien Dupond also has two articles but his co-authors were not identified via Case 1:

Author	Article ID	Institution	Is reprint author
Julien Dupond	4	ECOLE POLYTECH	No
Julien Dupond	4	UNIV PARIS	No
Julien Dupond	5	ECOLE POLYTECH	No
Julien Dupond	5	UNIV POLITECN MADRID	No

We assign him to ECOLE POLYTECH as it appears in all his publications and strictly more times than the other institutions.

same year from these two institutions: This is an equivocal case.

For many records, the simple cases described previously do not occur anymore: authors have different institutions in a given year and we cannot directly discriminate which one is really her affiliation – maybe simply because she is indeed affiliated to several institutions. Thus we proceed as follows:

1) If in a given year if an author's institution appears *strictly* more times than other institutions: Then we assign the author-year to that institution. 83.5% of the author-years becomes identified. Note that this case differs from the 3^{rd} case of the "unequivocal cases" because the requirement of that institution appearing in all articles of the year is not here anymore. Thus we simply consider that if an author has more publications in one institution, it is likely that it is the place where she was the most active.

2) In a given year if i) an author's institution has already been identified in other years via the previous methods and ii) that institution has been assigned to other author-years a *strict* majority of times: Then we assign this author-year to this institution, leading to a 89.6% of the sample identified.

3) In a given year if i) an author's institution also appears in other years and ii) that institution is the most frequent institution across *all* years: Then we assign the author-year to that institution. We end with 90.8% of the sample being identified.

Finally, we apply ex post modifications: After these two classifications are done, we correct the sequences of institutions for possible mistakes. We consider two causes of problems:

1. Timing of publications: a scientist changes institutions, but the publications timing is not appropriate.

2. Missing information in our raw data due to formatting issues or mere misreporting.

To cure these two problems, we apply a simple rule: when an institution is surrounded by two identical institutions, we replace it by the surrounding institution, as illustrated by Table A2. We apply 21,303 such modifications (in the initial full sample of 1,208,600 scientist-years, it is a 1.7% frequency).

Finally when a scientist doesn't have a publication in a given year, we recursively assign her to the identified institution of the previous year. If no previous year information is

Table A2: Corrections of institutional spells.

	1990	1992	1993	1994	1996	1998
Case 1	A	A	**B**	**A**	B	B
Case 1 Corrected	A	A	**A**	**B**	B	B
Case 2	A	A	**B**	A	A	A
Case 2 Corrected	A	A	**A**	A	A	A

Notes: A and B represent two different institutions. In case 1, it is very likely that the scientist was in institution A in 1993 (and left to institution B during 1993) and is in institution B from 1994 on.
In case 2, it is very likely that the institution reported in 1993 is a mistake (usually it is the institution of a coauthor). For example, in the article Simultaneous Fitting Of Several Planes To Point Sets Using Neural Networks published in Computer Vision Graphics Image Processing in 1990, the author Behrooz Kamgar-Parsi is affiliated to Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, Maryland 20742, USA. However, in our data, the only two addresses showing up in that publication record are: "George Mason Univ, Dept Comp Sci, Fairfax, VA 22030", and "USN, Res Lab, Washington, DC 20375" which are the two institutions of his two co-authors.

provided, we assign them recursively to the identified institution of the next year.

A.3.2 Identification of universities

To identify which institution is a university, we apply a pattern matching on the institution name. Thanks to manual identification, the following classification guarantees that no institution with more than 150 publications is left unidentified.

An institution is considered as a university if:

- it contains one of the following words: caltech, coll, cuny, ecole, ens, epfl, eth, fac, faculty, harvard, kth, mit, nyu, politecn, polytech, sch, school, scuola, stanford, suny, supelec, tu, univ, university, upmc,
- or it contains the following patterns: inst technol, inst sci technol, virginia tech,
- or it is equal to: ensieg, enst, enst bretagne, georgia tech, iit, imag, imag lab grenoble, inst eurecom, itesm, lirmm, rhein westfal th aachen, telecom paris, th darmstadt, tima lab, ucl, ufrgs, ufrj, umist, unicamp, verimag.

We further exclude from this classifications some companies filling the previous criteria.[38] In total, there are 67,305 different institution names. We identify 14,705 of them as

[38]It concerns the following: abb ens, mit gmbh, samsung adv inst technol, lg corp inst technol, samsung adv inst sci technol and tu elect co.

universities. Looking at the 859,862 publication-institution pairs, 74% are universities.

A.4 Patent information

Although the main focus of the paper is on scientists' publication output, we complement each scientist's information with her patent records. We extract all granted patents applied for at the USPTO between 1989 to 2004 from the Patstat database. We identify inventors using the same disambiguation algorithm as described in Section A.2. Then we match the patent data information to the publication data using the disambiguated inventor/scientist names as well as all other complementary information, such as the institutions' names contained in the patent/publication documents. We find a total of 680 university software scientists having at least one patent over the period (460 U.S. scientists).

The application of the disambiguation algorithm provided us with three different career realms: the U.S. inventor, the European inventor and the author publishing in the field of computer science. The distinction between the inventor careers stems from the separation of the patent authorities. The term career refers to a sequence of documents produced by an individual with a high probability. A career is labeled with a name, it has an entry and an exit point defined by the first and last published document (patent or publication) and is associated with affiliations of applicants. Of course, there are more properties to a career, but given out specific setup, we refrain from imposing additional restrictions to avoid a bias towards state dependency, i.e. ignoring job changes or the diffusion of the research field engendered by participation in larger projects.

As a first step, for the three realms, we create a table containing all name and country code combinations encountered in the respective data. The country codes stem from the associated affiliations or applicants. By applying the search tool "SearchEngine"[39] configured for n-grams (see Doherr, 2017), we link the publication names with the two patent name tables. Every author-name-country combination produces a list of inventor name candidates with the same country and a high similarity to the author name. Linking merely by the name and country seems to be a recipe for disaster given the high degree

[39]The tool can be downloaded at ftp://ftp.zew.de/tools/searchengine.zip.

of homonymy in our data exacerbated by the fact, that the publication data does only provide initials instead of proper first names, defining the lowest common denominator for our matching effort.

Fortunately, the author names come attached with additional meta information transferred from the disambiguation routine. We can directly exploit the estimated number of namesakes to assess the matching capabilities of a name. Further, by observing the number of careers associated with a name for the three realms, we can exclude one-to-many or many-to-many career intersections, implicitly solving the issue of having multiple matched inventor name variants for an author name. As long as the variants are accrued within one inventor career, the one-to-one tenet is not violated. We can relax this tenet, by including affiliation to applicant linkage. In a separate step, we matched the affiliations to patent applicants using the "SearchEngine" configured for frequency-based heuristics to filter filler words like legal forms. This linkage introduces additional criteria potentially separating multiple career assignments into one-on-one matches. Of course, not every one-on-one assignment represents a distinctive career switch from author to inventor or vice-versa, but the inclusion of the estimated namesake count and the juxtaposition of the respective career periods represented by the entry and exit points allow for fine-tuning of recall vs. precision. A career switch of an author with a high namesake count, overlapping career periods or an implausible stretch of inactivity is deemed to be dropped during an explorative phase of sample adjustments. Under the assumption of conditional independence of names to career paths, we are confident to introduce not any bias due to our arbitrary decisions.

A.5 Matching patents' non-patent literature to publications

The main challenge in using references to the non-patent literature (henceforth NPL) contained in patents is that these references' format is highly unstructured, making them difficult to match to publication data. For instance, here is a sample of the different existing formats:

1. Gunji et al. Correlation between the serum level of hepatitis C virus RNA and disease activities in acute and chronic hepatitis C. Int. J. Cancer 52(5):726-

730 (1992).

2. J. Bacteriol., vol. 172, No. 12, Dec. 1990; Norihiko Misawa et al.: Elucidation of the Erwinia uredovora Caretenoid Biosynthetic Pathway by Functional Analysis of Gene Products Expressed In Escherichia coli, pp. 6704-6712.

3. Daniell et al., Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology, Trends in Plant Science, 2002, 84-91, 7.

4. Doranz et al., “A small-molecule inhibitor directed against the chemokine receptor CXCR4 prevents its use as an HIV-1 coreceptor,” Journal of Experimental Medicine (1997a), vol. 186, pp. 1395-1400.

To perform the matching with the publications of our sample, we use two fields: the title and the year. There is a match if the years are identical and, after some preprocessing,[40] if the titles are also identical. The difficulty resides in extracting the title from the NPL records. When the record contains quotes we extract the title as the quoted sentence.[41] Otherwise, we proceed by step-by-step deleting information we know are not related to publication titles (journal names, pages, volume, authors, etc...), then apply a pattern-based algorithm to find the title from the leftovers. Out of 23,029,136 observations containing a date, we were able to find a title for 18,197,624 of them (79%).[42]

The number of patents-publications pairs matched with the 1989-2004 worldwide WoS computer science publications and in a 5-years window is 522,484. This number is 42,309 when considering only the U.S. software scientists of our working sample.

B Additional Estimations

B.1 Different definitions of active computer scientists

The main results of this paper are based on computer scientists active before and after the introduction of patent rights for software inventions. We defined an active computer

[40]The preprocessing includes: cleaning any special character and html markup, lowering the case, deleting all punctuation.

[41]This is the case in Example 4, since the special characters “ and ” represent quotes.

[42]In the four examples above, all titles are recovered appropriately. Further, note that NPL citations can refer to items that are not journal/conference articles.

scientist as a scientist with a least two publications in two different years before 1996 and at least one publication after 1996. Here, we show robustness of our results for different definitions of active scientists. First, we consider scientists either active in two, or more restrictively, in three different years before 1996. Second, we consider scientists: a) without restriction, or having at least b) one, or c) two active years after 1996.

Table B1 reports the results of the 6 estimations. All coefficients are in range of our main result in the paper (which is reported in column 2). It is worth noting that the estimates are even larger in magnitude when we do not make restrictions for the production after the law change.

Table B1: Main estimates for varying definitions of active scientists.

At least...	2 Active Years Before 1996			3 Active Years Before 1996		
# of Active Years After 1996	≥ 0	≥ 1	≥ 2	≥ 0	≥ 1	≥ 2
Dep. Variable:	Yearly # Publications					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Treat×Post	-0.1955***	-0.1645***	-0.1594***	-0.1810***	-0.1761***	-0.1716***
	(0.0306)	(0.0272)	(0.0266)	(0.0337)	(0.0313)	(0.0304)
Fixed-Effects						
Scientist	✓	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓	✓
Fit statistics						
Observations	196,683	109,029	79,672	90,954	66,324	53,379
# Scientist	14,986	8,133	5,894	6,537	4,744	3,804
Adj-pseudo R^2	0.26415	0.21675	0.20121	0.23821	0.20952	0.19946
Log-Likelihood	-175,561.0	-128,272.4	-104,801.2	-105,750.7	-88,075.2	-75,873.2

*Clustered (Scientist) standard-errors in parentheses. Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Fixed-effects Poisson estimations. An active year is a year with a publication.
Sources: Authors' own calculations based on publication data from Web of Science.

B.2 Different definitions of software scientists

This section reports estimates for varying definitions of software computer scientists. To identify software computer scientists, we relied on the information contained in the subfield of each journal. We define a hardware publication as a publication containing the subfield "Hardware & Architecture" and *not containing* the subfield "Software Engineering". We then defined a software scientists as someone having no hardware publication.

We now replicate the main estimation, varying the threshold to qualify as a software

scientist. Table B2 reports the estimates for thresholds ranging from 0 hardware publication (like in the paper) to 3. The estimates for different thresholds are all in line with the main estimates.

Table B2: Varying the threshold defining software scientists.

Total Hardware Publications (1989-2004)	≤ 0	≤ 1	≤ 2	≤ 3
Dependent Variable:		Yearly # Publications		
Model:	(1)	(2)	(3)	(4)
Variables				
Treat×Post	-0.1645***	-0.1436***	-0.1381***	-0.1445***
	(0.0272)	(0.0236)	(0.0225)	(0.0217)
Fixed-Effects				
Scientist	✓	✓	✓	✓
Year	✓	✓	✓	✓
Fit statistics				
Observations	109,029	129,253	139,506	146,132
# Scientist	8,133	9,636	10,399	10,888
Adj-pseudo R^2	0.21675	0.22160	0.22796	0.23211
Log-Likelihood	-128,272.4	-156,220.7	-171,620.6	-181,704.8

Clustered (Scientist) standard-errors in parentheses.
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Fixed-effect Poisson estimations.
Sources: Authors' own calculations based on publication data from Web of Science.

B.3 Replication with an alternative data base: DBLP

The dblp computer science bibliography (DBLP) is a data base dedicated to collecting all publications in the field of computer science. Contrary to our main source, Web of Science, this one only focuses on computer science and may therefore be more accurate. In particular, DBLP is known to record a comprehensive number of publications in conference proceedings, as opposed to WoS. To ensure our analysis is not dependent on the source of the data, in this section we replicate the main estimations with the DBLP data set, when possible.

Although DBLP is curated for computer science, it suffers from a major drawback that impedes us from using it in our main analysis: the affiliation of the authors is not reported. Since we, however, have this information from WoS, we match the two data sets to replicate our estimations.

Since the matching can be performed on the author names only, we ensure that the names are unique enough to be confidently attributed to an unique identity. We proceed as follows:

1. Out of the 8133 scientists from our main analysis, we select the ones that have no homonym (same first letter of the first name and same last name) in the master data set so that a publication from that name can be confidently attributed to one and only one person. We are left with 3804 authors without homonyms.

2. We then attach each of these 3804 authors without homonym in WoS to the DBLP records, with the match based on the first letter of the first name and the last name. At he end of this process, 3292 authors remain.

Once the authors are matched, the number of publication from the two sources are aggregated at the scientist-year level. Specifically, we create three variables: the total number of publications, the number of journal articles and the number of conference proceedings.

Table B3 reports the correlations and the descriptive statistics of the matched WoS-DBLP sample. The correlation between WoS publication numbers and their DBLP counterpart is high (above 59% in all three cases) although not equal to 1. The descriptive statistics show that the main difference is driven by the number of conference proceedings per year since WoS records about half the numbers of in DBLP (0.41 versus 0.79). The number of journal articles is also smaller in WoS, 0.6 versus 0.72, but also the discrepancy is smaller. As illustrated, there are differences between these two data sources, the main question now becomes whether our results are sensitive to these differences.

Table B4 and Figure B1 replicate the paper's estimations for this matched WoS-DBLP sample. The magnitude of the negative effect of the reform is even higher for the DBLP data, reaching a reduction of 18% (coefficient of -0.20) when looking at the total number of publications. Regarding the yearly effects, the coefficient estimates from DBLP almost map the ones from WoS.

B.4 Replication with an alternative citation measure

Citation patterns can be specific to subfields. Changes in citations could thus be the reflection of changing dynamics in subfields of computer science instead of changes in

Table B3: Comparison between publication data from WoS and DBLP: descriptive statistics.

(a) Correlations between WoS and DBLP publication variables.

			1	2	3	4	5	6
WoS	Yearly # of Publications	1	1	0.77	0.77	0.7	0.54	0.6
	Yearly # of Articles	2	0.77	1	0.18	0.55	0.62	0.33
	Yearly # of Proceedings	3	0.77	0.18	1	0.53	0.22	0.59
DBLP	Yearly # of Publications	4	0.7	0.55	0.53	1	0.75	0.88
	Yearly # of Articles	5	0.54	0.62	0.22	0.75	1	0.34
	Yearly # of Proceedings	6	0.6	0.33	0.59	0.88	0.34	1

(b) Descriptive statistics of WoS and DBLP publication data, based on a matched sample of scientists found in the two data sets.

	WoS							DBLP						
	Min	Median	Q3	90%	Max	Mean	SD	Min	Median	Q3	90%	Max	Mean	SD
Yearly # of Publications	0	1	1	3	42	1	1.6	0	1	2	4	65	1.5	2.7
Yearly # of Articles	0	0	1	2	18	0.6	1	0	0	1	2	26	0.72	1.4
Yearly # of Proceedings	0	0	0	1	33	0.41	1	0	0	1	2	58	0.79	1.9
# Scientist-year	44,323													
# Scientists	3292													

Table B4: Comparison of DiD estimates between data based on WoS or DBLP.

Dep. Variables:	Yearly # of Publications		Yearly # of Article		Yearly # of Proceedings	
Data source:	WoS	DBLP	WoS	DBLP	WoS	DBLP
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Treat× Post	-0.1432***	-0.2002***	-0.2258***	-0.3382***	-0.0094	-0.0449
	(0.0407)	(0.0423)	(0.0431)	(0.0425)	(0.0591)	(0.0595)
Fixed-effects						
Scientist	✓	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓	✓
Fit statistics						
Observations	44,323	43,701	43,026	41,349	35,220	33,127
Pseudo R^2	0.22054	0.39130	0.19090	0.26180	0.25974	0.37526
Log-Likelihood	-53,809.3	-60,061.1	-39,697.6	-42,326.5	-28,772.3	-37,864.3

*Clustered (*Scientist*) standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: Poisson fixed-effects estimations. Matched WoS-DBLP sample. The number of observations varies across columns dues to the removal of scientists with only 0 outcomes across the whole period linked to the Poisson estimation.

Source: Authors' own calculations based on publication data from Web of Science and DBLP.

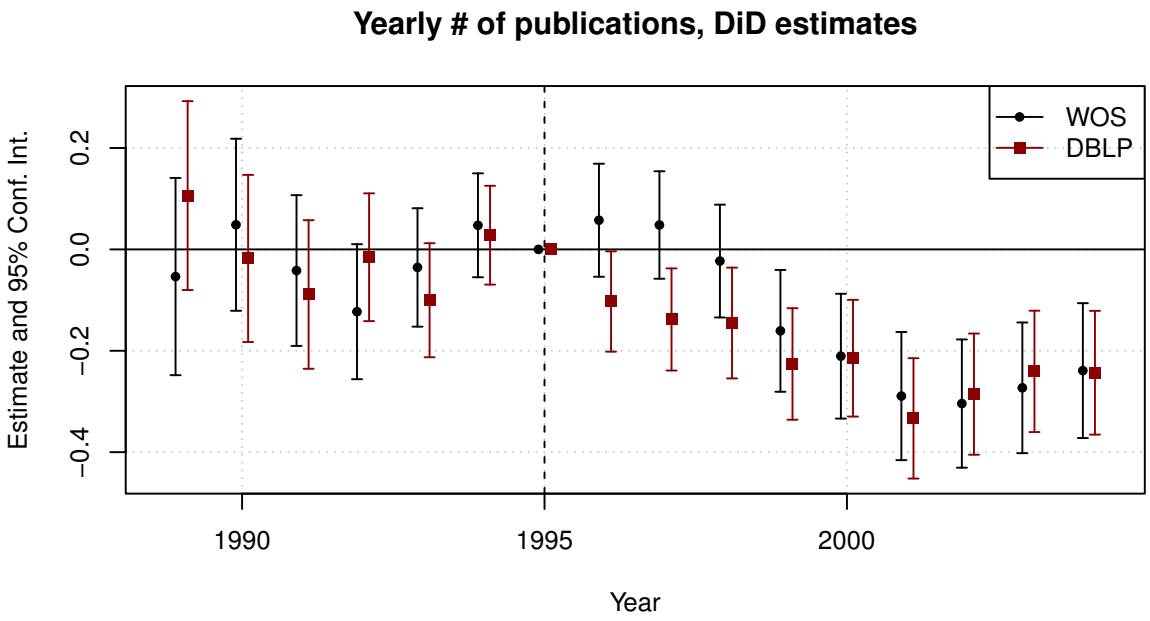

Figure B1: Yearly treatment effect for WoS versus DBLP publication data.

quality. To control for this potential problem, we recompute the citation variables by normalizing per subfield. Ninety percent of the publications in our dataset report a subfield, these include Artificial Intelligence, Information Systems, etc. The 26 subfields report over 1000 publications. To substract the subfield component from the citation measure, we apply the following correction:

$$cites_{i,f,t}^{corrected} = cites_{i,f,t} \times \frac{\frac{1}{n_{f,t}} \sum_{i'} cites_{i',f,t}}{\frac{1}{n_t} \sum_{f'} \sum_{i'} cites_{i',f',t}},$$

with $cites_{i,f,t}$ being the number of citations received by publication i in subfield f and year t; $n_{f,t}$ the number of publications in subfield f in year t and n_t the total number of publications in year t. When a publication is asscociated with several subfields, the numerator of the previous equation is the average across subfields. Finally, for the 11% publications without subfield, we consider missingness as a specific subfield. This modification effectlively corrects for subfield differences in means.

Table B5 replicates the main estimations for the subfield-corrected citation measures. The results are almost identical for the citation-weighted number of publications while the coefficient is lower in magnitude for the number of top 5% publications but still sizeably negative and significant. Figure B2 displays the yearly treatment effects. The estimates

are in line with the ones of the main text. Importantly, the results for the citation-weighted number of publications are stronger: the absence of pre-trend is more salient, and the effect after the reform has a larger magnitude.

Table B5: Replicaiton of the main estimation with subfield-corrected citation measures.

Dep. Variables:	Yearly # of	Yearly # of Top 5%
	Citations-Weighted Pub.	Cited Articles (Worldwide)
	Citations corrected for subfields	
Model:	(1)	(2)
	Neg. Bin.	Poisson
Variables		
Treat × Post	-0.4450***	-0.1699***
	(0.0505)	(0.0576)
Fixed-effects		
Scientist	✓	✓
Year (16)	✓	✓
Fit statistics		
Observations	106,907	49,172
# Scientist	7,959	3,595
Pseudo R^2	0.06091	0.14879
Log-Likelihood	-235,435.6	-23,587.8
Over-dispersion	0.16500	

Clustered (Scientist) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Sources: Authors' own calculations based on publication data from Web of Science.

B.5 Quantification estimates

The estimates used to quantify the gain in patents in Section 8.1 of the paper are reported in Table B6.

B.6 Dotcom bubble: estimation with only stayers

Table B7 reports the main estimates for the sample of researchers still active in academia in 2003-2004. This estimate is discussed in Section 8.2 of the paper.

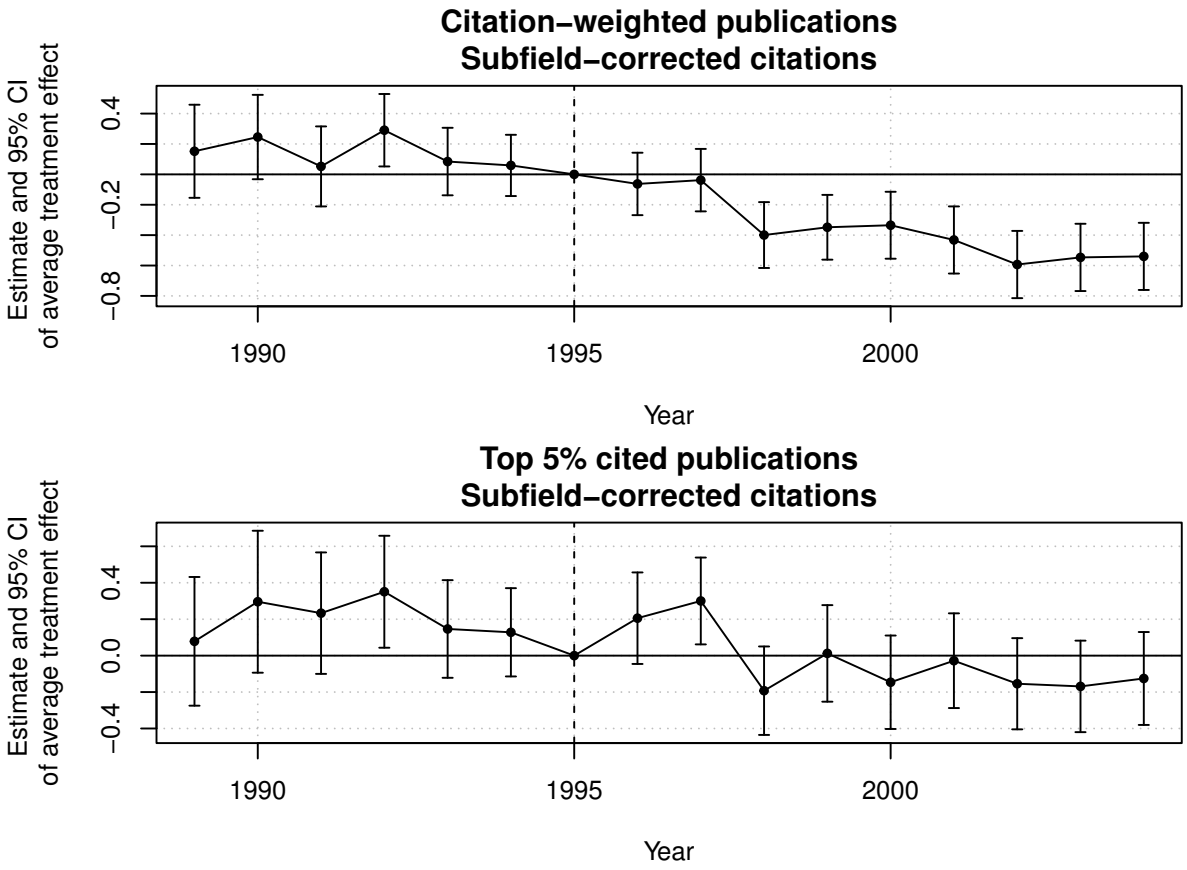

Figure B2: Estimation of yearly treatment effect for subfield-corrected citation measures.

Table B6: Estimation of the increase in patent production for U.S. university software scientists.

Dependent Variable:	Yearly # Patents
Variables	
Post	0.2839**
	(0.1350)
Time Trend (log)	0.1694
	(0.1239)
Fixed-Effects	
Scientist	✓
Fit statistics	
Observations	6,236
# Scientist	460
Adj-pseudo R^2	0.21492
Log-Likelihood	-3,198.3
Clustered (Scientist) standard-errors in parentheses.	
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*	

Notes: Fixed-effect Poisson estimation.
Sources: Authors' own calculations based on patent data from the USPTO.

Table B7: Estimation for computer scientists active in 2003-2004.

Dependent Variables:	Yearly # of Publications	Yearly # of Citations-Weighted Publications	Yearly # of Top 10% Ranked Articles (JIF)	Yearly # of Top 5% Cited Articles (Worldwide)
	Poisson	Neg. Bin.	Poisson	Poisson
Model:	(1)	(2)	(3)	(4)
Variables				
Treat×Post	-0.1472***	-0.4959***	-0.2266***	-0.2252***
	(0.0284)	(0.0575)	(0.0489)	(0.0635)
Fixed-Effects				
Scientist	✓	✓	✓	✓
Year	✓	✓	✓	✓
Fit statistics				
Observations	59,176	58,870	41,949	31,897
# Scientist	4,360	4,335	3,060	2,312
Adj-pseudo R^2	0.22564	0.05901	0.19741	0.16542
Log-Likelihood	-80,773.8	-161,443.7	-28,234.7	-16,465.9
Over-dispersion		0.21389		

*Clustered (Scientist) standard-errors in parentheses. Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The coefficients correspond to maximum likelihood Poisson and Negative Binomial (column 2) estimates. The sample consists of active computer scientists working in software. An active computer scientist is defined by having at least two publication before 1996 (with one before 1994), and at least one publication in the period 2003-2004. Observations correspond to *scientist* × *year*.
Although the same sample is used across all regressions, the number of observations vary because, due to the Poisson/Negative Binomial fixed-effects setup, all scientists whose dependent variable is equal to 0 across all periods are dropped.
Sources: Authors' own calculations based on publication data from Web of Science.

C Model: Extension and proof

C.1 Including quality-weighted publication output

Consider that each scientist has the possibility to distribute her effort across two types of publication projects: *i*) low risk, low quality (Pub^L), or *ii*) high risk, high quality (Pub^H). To keep the exposition simple, and contrary to the main model, we assume no complementarity ($\lambda = 0$). The production process of these two types of publication is as follows:

$$Pub_{it}^L \sim Poisson\left(\gamma_{it}^{pub} e_{it}^{pub^L}\right), \quad Pub_{it}^H \sim Poisson\left(r_H \gamma_{it}^{pub} e_{it}^{pub^H}\right).$$

They both depend on the effort invested and on the publication ability of the scientist, γ_{it}^{pub}. Further, the term $r_H < 1$ reflects the fact that the production of higher quality projects requires more efforts. Now the utility becomes:

$$\begin{aligned} U_{it} &= \left(1 - u_i^H\right) \times Pub_{it}^L + u_i^H \times Pub_{it}^H + Pat_{it} \\ &\quad - \left[e_{it}^{pub^L}\right]^2 - \left[e_{it}^{pub^H}\right]^2 - \left[e_{it}^{pat}\right]^2 \\ &\quad - e_{it}^{pub^L} \times e_{it}^{pat} - e_{it}^{pub^H} \times e_{it}^{pat} - e_{it}^{pub^L} \times e_{it}^{pub^H} \end{aligned}$$

with $u_i^H \in [0, 1]$ an idiosyncratic preference for more difficult but higher quality projects.[43]

Using the same logic as in the main model, with scientist k the counterfactual, following the reform the change in publication volume becomes:

$$\begin{aligned} \log\left(\widehat{Pub_{it}^L + Pub_{it}^H}\right) &\approx \log\left(\widehat{Pub_{kt}^L + Pub_{kt}^H}\right) \\ &\quad - \frac{(1 + r_H)\,\Delta}{\left[3\left(1 + (1 - r_H^2)\,u_i^H\right) - r_H\right]\gamma_{it}^{pub} - (1 + r_H)\,\gamma_{it}^{pat}}, \end{aligned}$$

while the change in high quality publications is:

$$\log\left(\widehat{Pub_{it}^H}\right) \approx \log\left(\widehat{Pub_{kt}^H}\right) - \frac{\Delta}{\left[3u_i^H r_H - (1 - u_i^H)\right]\gamma_{it}^{pub} - \gamma_{it}^{pat}}.$$

Hence according to this model the reform would reduce the production of both publication

[43] Note that the term u_i^H could also embody the scientists' institution's valuation for high quality publications.

volume and quality.

C.2 Proof of Proposition 1

Omitting indices for clarity, the expected utility without reform writes:

$$\begin{aligned} E(U) &= r(\tau)\gamma^{pub}e^{pub} + \gamma^{pat}\left(e^{pat} + \lambda(\tau)e^{pub}\right) \\ &\quad - \left[e^{pub}\right]^2 - e^{pub} \times e^{pat} - \left[e^{pat}\right]^2 \end{aligned}$$

Assume τ_0 is the optimal topic choice in the absence of reform. The first order condition is:

$$r'(\tau_0)\gamma^{pub} + \lambda'(\tau_0)\gamma^{pat} = 0. \tag{5}$$

As we can see the topic choice depends only on the abilities and the form of the functions r and λ. Following the reform, the new first order condition is:

$$r'(\tau)\gamma^{pub} + \lambda'(\tau)\left(\gamma^{pat} + \Delta\right) = 0 \tag{6}$$

Let us find ϵ such that $\tau = \tau_0 + \epsilon$ solves the previous equation. A first order approximation of the previous equation writes:

$$\begin{aligned} (r'(\tau_0) + \epsilon r''(\tau_0))\gamma^{pub} + (\lambda'(\tau_0) + \epsilon\lambda''(\tau_0))\left(\gamma^{pat} + \Delta\right) &= 0 \\ \Leftrightarrow \left[r'(\tau_0)\gamma^{pub} + \lambda'(\tau_0)\gamma^{pat}\right] + \epsilon r''(\tau_0)\gamma^{pub} + \epsilon\lambda''(\tau_0)\left(\gamma^{pat} + \Delta\right) &= -\lambda'(\tau_0)\Delta \end{aligned}$$

$$\Leftrightarrow \epsilon = \frac{\lambda'(\tau_0)\Delta}{-r''(\tau_0)\gamma^{pub} - \lambda''(\tau_0)(\gamma^{pat} + \Delta)},$$

where we used the result from Equation (5) for simplification. Since the second derivatives of r and λ are negative, we obtain that $\epsilon > 0$. Further, we can clearly see that $\partial\epsilon/\partial\gamma^{pub}$ is negative. □

D Appliedness measure

The objective of this section is to define a measure that captures the content of the scientists' production. In particular, we would like to assess whether the content produced is more or less applied following the reform.

We introduce a simple measure of appliedness, or patentability, of research. It can loosely be defined by whether the keywords of the publications appear in the abstracts of the software patents.

To construct this indicator, we first extract software patents (as defined à la Bessen and Hunt, 2007) and then look at the frequency at which keywords from publications appear in patents. The index is at the publication level for which we categorize each keyword into one of three categories: i) applied, ii) neutral and iii) basic. The index of a publication is then the number of applied keywords, minus the number of basic keywords, divided by the total number of keywords. We end up with a measure in between -1 and 1.

Formally this index can be defined as:

$$applied_p = \frac{\sum_{w \in W_p} \left(1\{w \in W^{applied}\} - 1\{w \in W^{basic}\} \right)}{|W_p|}, \tag{7}$$

where $applied_p$ is the appliedness index for publication p, W_p is the set of all keywords in publication p, and $W^{applied}$ and W^{basic} are the sets of applied and basic keywords.

The key element of this index is the categorization of the keywords, which we hereby describe.

D.1 Categorizing keywords into basic or applied

To identify whether a keyword is basic or applied, we will use a source of information *external to the publications*: the patents. First we extract all patents relating to software in the sample period (1989–2004). We use Bessen and Hunt (2007) methodology to identify which patent is software related.[44] We end up with 39,017 such patents.

[44] It corresponds to all USPTO patents containing "software" or "computer program" in the title and not containing a) "chip", "bus", "circuit" or "circuitry" in the title, nor b) "antigen", "antigenic" or "chromatography" in the description.

The second step is to map the keywords from the publications to the patents to identify which keyword is more patent related. There are important challenges to this task:

1. several publication keywords can refer to the same concept but be written differently,
2. the patents do not contain keywords.

We tackle these challenges in turn.

D.1.1 Turning publication keywords into concepts

Of the 655,441 publications of our full data set extracted from WoS, less than half contain at least one keyword (269,566 or 41%). The mode is four keywords per publications, leading to 1,238,023 publication-keyword pairs. The first year for which publication have keywords in our sample is 1990 (i.e. 1989 is fully missing).

The publication keywords from our data set are the ones provided by the authors and are not formatted. Hence keywords referring to the same concept can vary substantially, for example the keywords "3-D", "three dimension", "3 dimensional", etc, all refer to the same idea but cannot be used "as such" since they would erroneously be considered different.

Another issue relates to the scope of the keywords. Indeed, keywords can refer to very precise ideas, or instead, general ones, the latter nesting possibly many precise ideas. For instance, it would be wrong to consider that "relative convex hull" is completely different from "convex hull", but we could consider that the latter nests the former.

We now describe how we perform the cleaning and grouping of keywords with the objective to keep the most signal from them and remove as much noise as possible.

In the first step, we clean the keywords according to the following steps:

1. removing all terms in parentheses (e.g. "self-organizing map (SOM)" $\Rightarrow$ "self-organizing map")
2. cleaning all punctuation (except the point when numbers are attached, like in "IEEE 802.11") and putting everything in lowercase, and manually taking care of the "3-D" vs "three D" case (same for "2D").

3. we stem all words, i.e. we remove the suffixes to keep only the radical. For instance "dimensional" becomes "dimension", "spaces" become "space", etc.

4. grouping 2-grams when appropriate. Sometimes some keywords have spaces and others not, like in "non linear" and "nonlinear". We create the set of all 2-grams (combination of two consecutive words) and look at whether their frequency in the original keywords is higher when split or merged. We then split or merge the 2-grams accordingly. For instance, "nonlinear" was more frequent than "non linear", so we transformed any "non linear" into "nonlinear".

The second step aims to group keywords that may be too precise into more general keywords. We define a **concept** as a keyword that appears in at least 20 publications, there are 7,723 such concepts. We define a fuzzy keyword as a keyword appearing in 50 or less publications (note that a fuzzy keyword can be a concept). We then look at whether fuzzy keywords are textually included in concepts which are more frequent. For example we transform "discontinuous Galerkin finite element methods" into "finite element method" since the latter is more frequent and includes the former. We transform 74,683 keywords in that way.

At the end of this process, we end up with 277,454 unique keywords, 7,707 of which are concepts.

D.1.2 Turning patent abstracts into concepts

Now we turn to the 39,017 patents identified as dealing with software. The aim is to turn the text of the patents into a set of keywords similar to the ones of the publications.

We first clean the patent abstracts in the same way as for the publication keywords, the only difference being that we also delete common stop-words (e.g. "a", "is", "are", etc[45]).

We then create the set of all 1-, 2-, 3- and 4-grams for each abstract. For example: "A computer program dealing with user identification..." would become "computer program deal user identify" after cleaning, then we would have, e.g., "computer program" and "user identify" as 2-grams, and e.g., "computer program deal user" as 4-gram.

[45]The full list of stopwords used can be found here: http://snowball.tartarus.org/algorithms/english/stop.txt.

This process leads to a large set of 1- to 4-grams for each patent which we consider as equivalent to the publication keywords obtained in the previous section.

D.1.3 Definition of basic and applied concepts

We take a very simple approach to categorize the publication keywords. A keyword is:

- **applied** if it is a "concept" that appears in at least 40 patents (0.1% of the sample),
- **neutral** if it is a "concept" that appears in 5 to 39 patents or if it is not a "concept",
- **basic** if it is a "concept" that appears in less than 4 patents.

Of course the terms "applied" and "basic" are only a shorthand; the terms "software-patent-related" and "non-software-patent-related" would have been closer to the reality they catch but are much less convenient. We continue with this simplification.

We end up with 580 applied keywords, 6,511 basic keywords and 255,046 neutral keywords (note that 91% of the keywords appear less than 5 times across the entire corpus, even after cleaning, and hence don't qualify as concepts and are thus considered as neutral).

From these sets of keywords, we can then build the appliedness measure defined in Equation (7) for each publication reporting at least one keyword. The applied score ranges from -1 (basic) to 1 (applied). Across all publications of U.S. and E.U. authors, the average of this measure is -0.0533 with a standard-deviation of 0.373.

D.2 Practical relevance, limitations and advantages

To assess whether the measure we constructed makes sense, we make a simple relevance test: do the publications of private sector scientists score "high" in appliedness? Indeed, we expect that, on average, scientists working in the private sector publish work that are more applied in nature. From the full data set, based on the affiliations, we group the scientists into either university or private sector scientists.

Figure D3 shows the evolution of the average applied score per publication from 1990 to 2004, split by sector. We can see a clear difference between publications authored by private sector scientists and the ones authored by university scientists. In the private

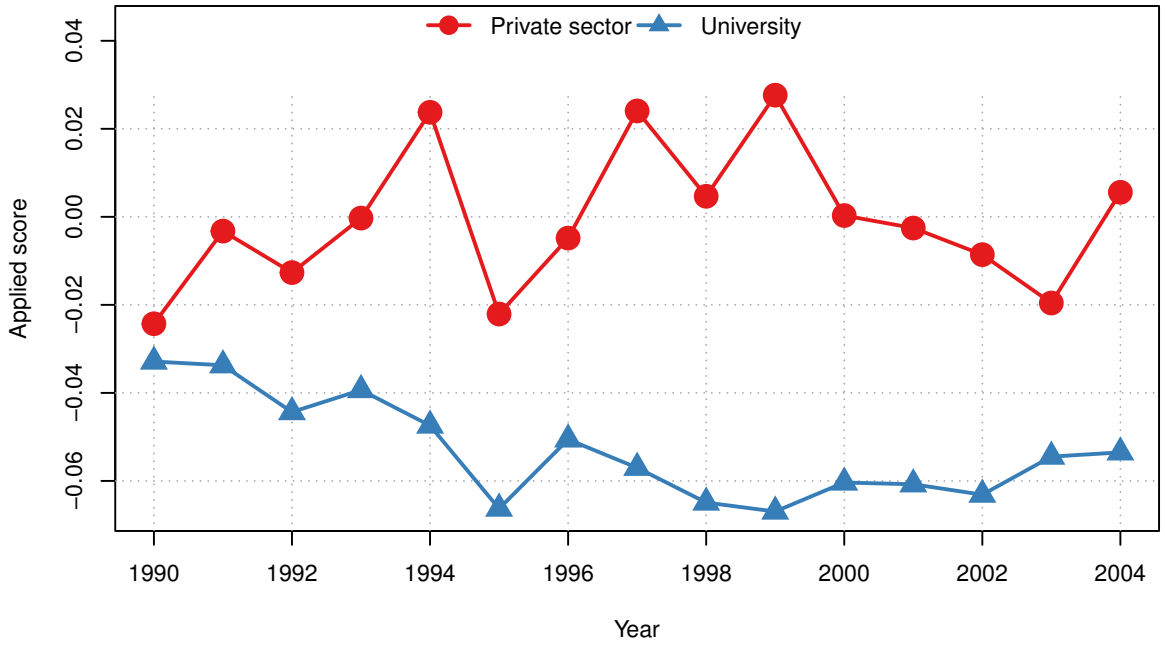

Figure D3: Evolution of the appliedness score: private sector vs university publications.

sector, the applied score is higher for any year. Its value hovers around 0 with several highs at 0.02. On the other hand, for university scientists, the measure is slightly decreasing up to 1995 and stays at -0.06 from that point.

This suggests that this measure is able to capture some information on the content of the publications. Although the range of variation of the measure is not important, of about 0.08, the difference between the sectors is clearly marked.

Limitations. This measure has many limitations that one should consider:

- about 60% of all publications do not contain any keyword.
- general concepts, like "algorithm" for instance, tend to be based on only one or two common words. Hence they have a higher chance of appearing in a patent abstract although possibly used as regular words and not as "concepts". Stated differently, general concepts have higher chances to be false positives.
- many ideas contained in publication keywords may be written differently in an abstract. Hence, many keywords may not be detected as being "applied".

The first limitation reduces the relevance of a panel data analysis with this measure as the dependent variable. Since this measure is conditional on production, missing values

will in effect remove scientists-years from the estimation and introduce artifacts. This is why, instead of using scientists-years in the empirical analysis, we will use aggregate the panel into two periods: pre and post. This ensures that the data will be complete and greatly limits the influence of the missingness on the estimates.[46]

The second and third limitations are acknowledgments of the, possibly large, noise present in the variable. This will very likely lead to a bias towards 0 if we use this variable in our analyses. However, the specific pattern for the private sector previously reported shows that this measure contains relevant information.

The main advantage of this measure is its reliance on data *external* to publications. To identify appliedness, we look at the proximity in *content* to software patents. Hence, the measure is not influenced by the identity of who patents and there is no connection between the measure and the patenting status of the scientists. This is in contrast to other measures which capture the proximity in content by using the publications of patenting scientists, such as for instance in Azoulay, Ding and Stuart (2009). The main assumption of these measures is that the scientists who patent *do* more applied science *per se*. This is a strong assumption that we can avoid to make thanks to our measure.

D.3 Appliedness of journal articles and conference proceedings

Computer scientists publish in academic journals and in conference proceedings, the content of the latter tends to be more applied. Indeed, in the words of the Computer Research Association: "experimental research is at variance with conventional academic publication tradition" and "experimentalists [prefer] conference publications" (Patterson, Snyder and Ullman, 1999, page A), suggesting that conferences tend to host more applied research than traditional journals. We now use the appliedness measure to document whether the proceedings are indeed more applied in nature than the articles. Figure D4 reports the evolution of the appliedness measure for articles and proceedings. We can clearly see a difference between the articles and the proceedings, the latter scoring consistently higher than the former apart for the two years 1991 and 1994. From 1996 onwards, the

[46]The hypothesis that we make is that publications without any keyword reported in WoS (which does not mean that the publication did not contain keywords) have an applied score similar to the other publications without missing values for the current period (pre or post).

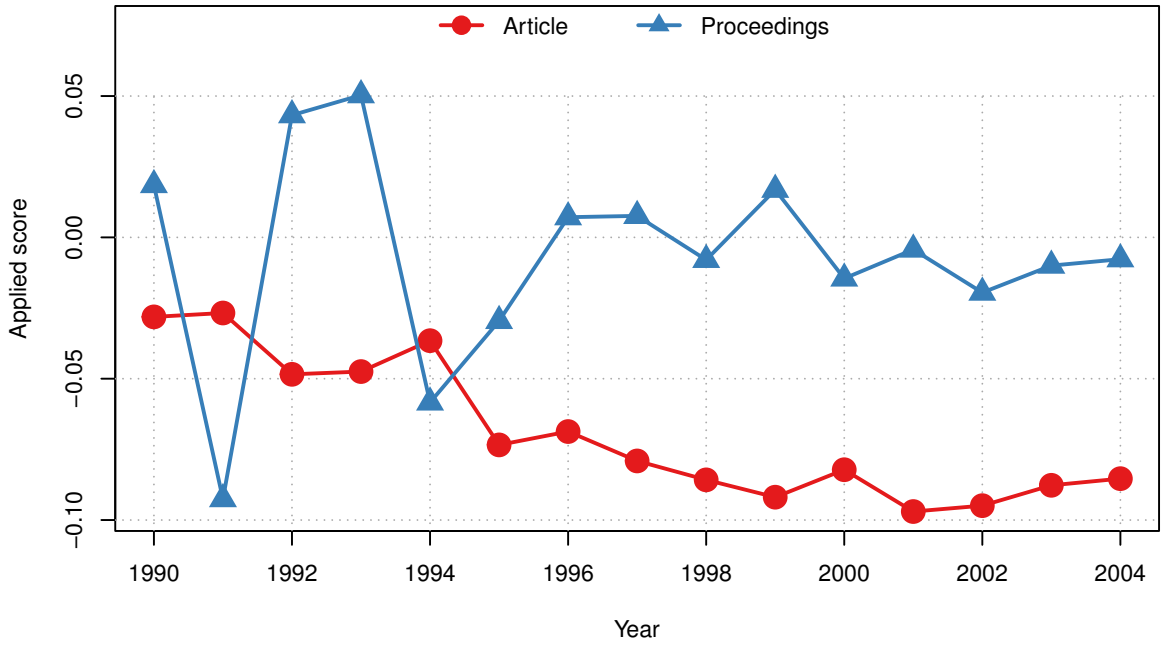

Figure D4: Evolution of the applied score for articles and proceedings.

appliedness of the conference proceedings appear to be stable over time, slightly below 0. In contrast, the appliedness of the journal articles decreases over time, reaching almost -0.1 at the end of the period.

The gap between the two types of publications appear to be larger than the gap between private sector and university publications, as illustrated in the previous section. This is in line with our interpretation of the proceedings being more applied *on average*; or equivalently, that computer science articles in journals are more basic on average.

We now look at the change in the production of conference proceedings and journal articles to approximate the *applied content* of researcher's publications, in complement to the direct appliedness measure used in Section 7.2 of the main text. As in Section 7.2, we estimate DiD models for five groups of scientists categorized by their position in the citations distribution, using the number of journal articles and conference proceedings as two separate dependent variables. Figure D5 reports the results for the two variables. Across all groups, the decrease in the number of conference proceedings produced is lower in magnitude than the one for journal articles. As expected, the largest gap between the coefficients of conference proceedings and journal articles is for the lowest quartile of the publication ability distribution. Scientists in this group produce 35% less journal articles after the law change while this number is only 20% for conference proceedings. Although

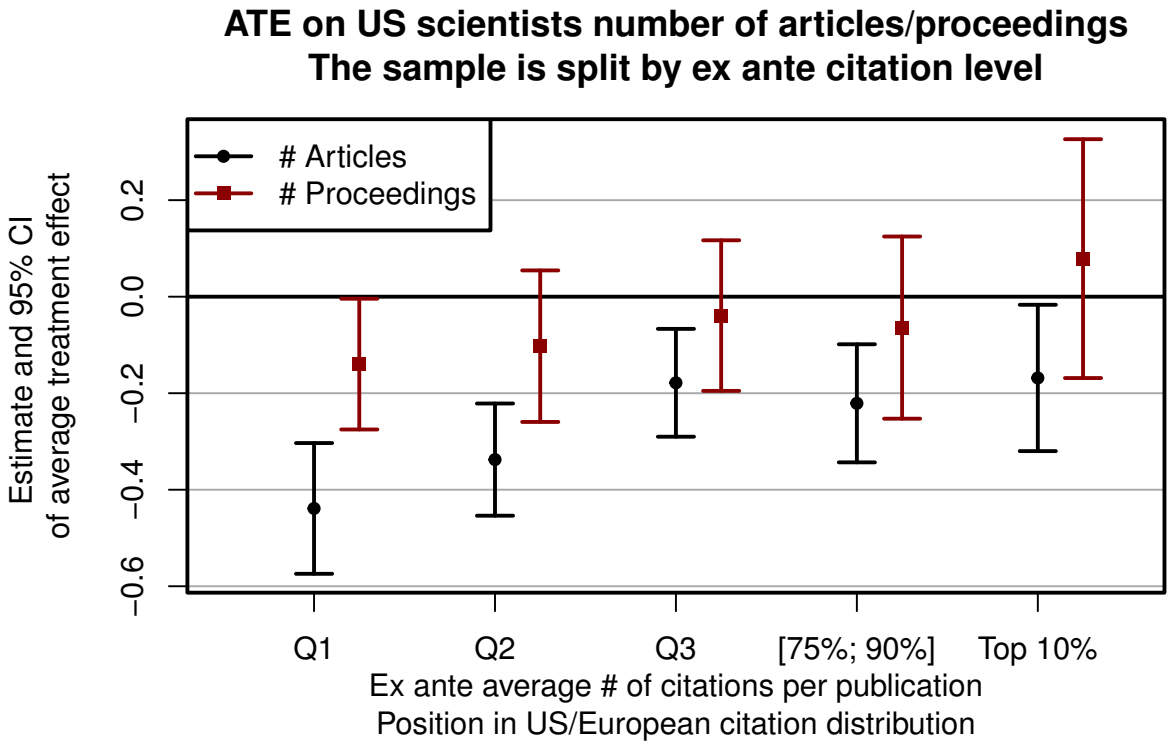

Figure D5: Average treatment effect for publications in journals and publications in conference proceedings, mediated by an *ex ante* measure of ability.
Notes: For each of the two dependent variables, the graph reports the estimates and 95% confidence intervals of the average treatment effect for 5 separate regressions, where the full sample is split according to the ex ante (1989-1996) average number of citations received per publication. Each regression is a Poisson fixed-effects estimation with scientist and year fixed-effects.
Sources: Authors' own calculations based on Web of Science data.

an imperfect measure of appliedness, these results suggest that software scientists tend to prioritize applied over basic research after the introduction of software patent rights, especially scientists at the left tail of the publication ability distribution.

Appendix references

Azoulay, Pierre, Waverly Ding, and Toby Stuart. 2009. "The impact of academic patenting on the rate, quality and direction of (public) research output." *The Journal of Industrial Economics*, 57(4): 637–676.

Bessen, James, and Robert M Hunt. 2007. "An empirical look at software patents." *Journal of Economics & Management Strategy*, 16(1): 157–189.

Cappelli, Riccardo, Dirk Czarnitzki, Thorsten Doherr, and Fabio Montobbio. 2019. "Inventor mobility and productivity in Italian regions." *Regional Studies*, 53(1): 43–54.

Czarnitzki, Dirk, Thorsten Doherr, Katrin Hussinger, Paula Schliessler, and Andrew A Toole. 2015. "Individual versus institutional ownership of university-

discovered inventions." ZEW-Centre for European Economic Research Discussion Paper.

Czarnitzki, Dirk, Thorsten Doherr, Katrin Hussinger, Paula Schliessler, and Andrew A Toole. 2016. "Knowledge creates markets: The influence of entrepreneurial support and patent rights on academic entrepreneurship." *European Economic Review*, 86: 131–146.

Doherr, Thorsten. 2017. "Inventor Mobility Index: A Method to Disambiguate Inventor Careers." *ZEW Discussion Paper*, , (17-018). Mannheim.

Patterson, D., L Snyder, and J Ullman. 1999. "Evaluating Computer Scientists and Engineers For Promotion and Tenure." *Best Practices Memo, Computing Research News*. CRA Computing Research Association.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim
ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de