School of Economics Working Paper
2024 -02

Ex-ante Novelty and Invention Quality: A Cross-country Sectoral Empirical Study

Yuan Gao *
Emiliya Lazarova *

* University of East Anglia

University of East Anglia

SCHOOL OF ECONOMICS

School of Economics
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/economics

Ex-ante Novelty and Invention Quality: A Cross-country Sectoral Empirical Study

Yuan Gao*[1] and Emiliya Lazarova†[2]

[1,2]School of Economics, University of East Anglia, Norwich Research Park, Norwich, NTR4 7TJ

September, 2023

Keywords: Innovation quality, Disruptive novelty, Ex-ante novelty, Patent, Network

Abstract

The research on measuring technological innovation quality has evolved with our understanding of the origin of novelty. Patents have been widely used in such studies because they are a form of copyright-protected outcome of inventions deemed to be valuable. The quality of technological innovation can be measured in multiple dimensions. In this paper, we make a methodological con-

*Corresponding author: y.gao4@uea.ac.uk

†e.lazarova@uea.ac.uk

tribution to the literature on ex-ante technological novelty and propose two new indices based on a network approach: the Inverse Recombination Intensity Index (IRII) to capture the extent to which an invention is the outcome of a novel combination of pre-existing technological components; and the New Technology Ratio (NTR) to measure the share of new knowledge elements in the invention. Through an in-depth empirical study of patents filed in the Pharmaceuticals and Computer Technology sectors, we show that our proposed indices are correlated with some of the conventional patent quality indicators and go beyond that to reveal previously unnoticed features of the inventions process, of which some are sector-specific. Moreover, through our regression analysis, we demonstrate that IRII and NTR are important predictors of a patents' potential impact on future inventions, which confirms the ex-ante nature of our indices. In the regression analysis we also include sector-country-specific R&D input variables as controls to test the robustness of our results. Our analysis suggests that the distinct characteristics of each sector affect how the quality of innovation is related to the ex-ante measures of technological novelty. We argue, therefore, that future analysis of the link between ex-ante novelty and ex-post quality of innovation needs to take into consideration the recombinant content of the invention and account for sectoral characteristics.

1 Introduction

Motivations: Unpacking the drivers and effects of radical innovations is of major interest to scholars, as well as firms and policy makers. - Business opportunities - Regulatory management - Greely (2022): "Horizon Scanning" in order to spot and analyze emerging technologies with great potential effects The processes of innovation and realization can be complicated and lengthy. Yet most empirical literature identify groundbreaking innovation through ex-post measures.

Innovation is a widely acknowledged driving force for economic growth and advances of the society. Researchers have been studying innovation in order to gain insights into the status of technology development and its relationships with socio-economic changes. In quantitative assessment of technological novelty and values, patent information has been a widely used data source because patents are directly associated with inventions - the outcome of scientific and technological research and development (R&D) activities. As we entered the digital age, patent data has become particularly popular with the availability of electronic patent database through the Internet, and was embraced by a series of important works by the NBER (National Bureau of Economic Research) researchers [Griliches et al., 1986, Fleming, 2001, Jaffe and Trajtenberg, 2002, Hall et al., 2005]. Most literature uses relatively simple counts statistics, such as the number of patent publications or citations. Some took a step further to develop composite indices based on such basic counts, like the patent quality index by Squicciarini et al. [2013]. However, as Schumpter noted in his original German book in 1911, innovation goes beyond invention - it is more about the scientific or technological novelty embedded in the inventions combined with the application Schumpeter [2017]. Simple counts are insufficient to establish a sound understanding of the degree or nature of novelty. For example, the number of patent applications does not capture the quality in terms of their economic values and impact on the following technological advances. It is also ill-equipped to reveal whether the inventions are driven by similar adaptations of the existing technologies or more ground-breaking methods. Other researchers have also questioned the reliability of conventional patent measures Griliches [1989], J. Acs and Audretsch [1989], Griliches [1998], Shepherd and Shepherd [2003]. Researchers have therefore proceeded to a search of other measures to assess the level of technology development or innovative activities.

One type of innovation that has particularly attracted researchers' interest is *disruptive innovation*; often associated with other descriptions like radical, unconventional, ground-breaking. These types of innovations have the potential to "disrupt" the existing industry and market by destructing the established products or operations while replacing with new ones, or changing the course of current technology development rather significantly. Such changes typically happen quickly in a powerful way, rather than in a gradual progress, and have a fundamental impact within and sometimes out of the original sector. Indeed, among the earliest works, Schumpeter described "Creative Destruction" as a process that "incessantly revolutionizes the economic structure from within, incessantly destroying the old ones, and incessantly creating a new one" Schumpeter [1942]. These ideas were later summarised as the "Innovation Trilogy": invention, innovation and diffusion Kaya and Joseph [2015]. Bower and Christensen [1995] influenced the direction of further research into disruptive technologies by establishing the idea that they are distinguished by a difference in performance attributes that rapidly improve to penetrate established markets rather than by technological complexity or novelty. This view of disruptive innovation served as a theoretical foundation for subsequent researchers to take a posterior perspective and identify and discuss disruptive technologies exclusively based on their commercial applications. While we agree with Bower and

Christensen's view that technological novelty and complexity do not necessarily contribute to disruptiveness, we would argue that it is essential to pursue an ex-ante perspective and study disruptiveness at the invention stage. An ex-ante assessment of the degree of technological disruptiveness has the potential to provide an early indication of the potential of a patent to induce technological change. It also allows to normalize the measure of novelty at the time of invention irrespective of the geographical origin of the invention and does not suffer from biases linked to varying by geography socioeconomic factors that impact on diffusion and adoption. Finally, the R&D phase of innovation is where government expenditure on innovation is concentrated and where national-level policies have most impact. With our work, we therefore focus attention on the technological novelty from an ex-ante perspective.

In this respect, we contribute to the line of research that aims to develop measures of the disruptive technological content of an invention by linking it to the degree to which an invention is the result of recombination of technological components using international patent data [Fleming, 2001, Arts and Veugelers, 2015, Kaplan and Vakili, 2015, Verhoeven et al., 2016, Silvestri et al., 2018]. By its very nature a patented invention presents a novel technology, the challenge authors in this branch of the literature aim to address is to quantify the degree of technological novelty on which it is built. To date, there is no recognised single best measure. Authors put forward statistical tools that vary in their versatility, computational complexity, and the empirical counterparts for foundational concepts. Verhoeven et al. [2016] offer a comprehensive overview of these measures and illustrate this point. For example, the empirical counterpart of a technological component in the development of an invention has being derived from a textual analysis of the abstract on a patent application [Kaplan and Vakili, 2015]; the technological fields of prior inventions cited on the patent's application, e.g. Dahlin and Behrens [2005]; or the author-reported attributes of the patent under the established technological classification system [Fleming, 2001, Strumsky and Lobo, 2015, Arts and Veugelers, 2015]. Arguably, the first two methods are more susceptible to biases: the use of language and choice of algorithm may be culture and norm dependent while the choice of references may be strategic as this information is used by patenting agency to evaluate the originality of the patent application. We, therefore, adopt the approach of using the author-reported technological fields in the description of the patent on its application. The additional advantages of using this information are that it is based on a globally adopted classification system; it is directly linked to the technological content of the invention; and, in accordance with gaining full protection, applicants have the incentive to provide a comprehensive description of their invention.

Even when authors use the same empirical counterpart for a technological input, they may employ it differently in the development of their index of ex-ante novelty with methods varying from simple counts of each component in isolation [Rosenkorf and Nerkar, 2001], distance between technological fields in a classification system of pairs of components [Trajtenberg et al., 1997], or tracking the changes in combinations of components usage [Fleming, 2001]. In our view, focusing on a single or a pair of components at a time, carries the potential to underestimates the complex way in which an invention combines multiple technological components of possible a range of technological fields. In the example of the "Oncomouse" patent application used by Verhoeven et al. [2016], there are eight technological components. If a researcher restricts attention to only pairwise combinations, then they will be looking at 56 possible pairwise combinations of technological components rather than $2^8 - 1 = 255$ possible subsets of any size of the same eight components. On the other hand, judging the ex-ante novelty based on the uniqueness of the entire list, may over-state the recombinant content of a patent if overlapping subsets of the complete list of technological components have been frequently used in previous cohorts of patents.

Moreover, authors also differ in the reference group against which novel use of technological components are identified with many of them referencing the whole historic records [Fleming, 2001, Arts and Veugelers, 2015, Verhoeven et al., 2016] and a handful using a shorter window, e.g. previous give years Silvestri et al. [2018].

What sets our measures apart is that they are grounded in time-indexed comprehensive maps of linkages between technological components that capture the complex multilateral combinations of these components that are used in the development of patents of a specific cohort. Rather than tracing if a combination of technological components has been utilised or not in the past, we develop a statistical measure of the likelihood of the combination combining information on pairwise relations between components and the formation of clusters of components based on the frequency of their use on patent applications. We evaluate the degree of technological recombination based on the whole network of connections rather than the relatedness of two nodes. We would argue that our measures offer a more comprehensive network perspective on the process of re-combination. Furthermore, we utilise a finer level of the classification of technological components and can therefore measure the degrees of recombination and use of novel components with greater accuracy.

In addition, Verhoeven et al. [2016] measure ex-ante novelty of a patent by the use of new components that previously have not been used in a technological field. In the same vein, we put forward two distinct measures of the ex-ante novelty of a patent: the Inverse Recombination Intensity Index (IRII) and the New Technology Ratio (NTR).

We illustrate the value of using our two ex-ante technological novelty measures to look into the process of disruptive innovation by using them to study inventions in two established technological sectors: pharmaceuticals and computer technology over a period of about 37 years. We choose these sectors because, for one, they represent a significant contribution to the volume of patenting activities by investors from a wide range of countries around the globe WIPO [2022]. Our choice of sectors is also because of their distinct processes of innovation. Saha and Bhattacharya [2011], alongside others, point to the extraordinary large share of sales of pharmaceuticals, that the R&D expenditure constitutes. These authors add that the competitive edge in the pharmaceutical sector is predicated on the advancement of scientific knowledge rather than technological know-how. Our ex-ante technological novelty measures, indeed, capture systematic differences between the two sectors that are consistent with these insights from the literature. We observe high rates of recombination of technological components among patents attributed to the pharmaceutical sector while patents attributed to the computer technologies contain a higher ratio of technological components that are new among patents in this sector.

We explore these differences between the two sectors through studying the correlations between the ex-ante and ex-post indicators of technological novelty among the patents. A strong link between the ex-ante and ex-post measures is important to make a successful case in favour of the use of the ex-ante measures as an early indicator of the performance potential of a patent. We benchmark our IRII and NTR against several measures of patent quality and technological value described in Squicciarini et al. [Squicciarini et al., 2013] and widely used in studies on patenting and innovation. When correlated to their ex-ante indicators - patent scope, family size, backward citations, and originality - our proposed measures exhibit low to moderate degrees of linear relation. More significant is the observation that our two measures capture differences in the relation between the ex-ante and ex-post measures of patent novelty between the pharmaceuticals and computer technology sector that remain undetected when using only the established measures of ex-ante indicators.

During this benchmark endeavor, the ex-post indicators are considered measures of the outcome of invention. Meanwhile, there is an equally understandable interest on the input end. [Evenson, 1993, Kim and Marschke, 2004, Singh, 2008, Arts and Veugelers, 2015, Briggs, 2015, Fink et al., 2016], argued that the intensity and performance of innovation dependent on the regional investment and resources on R&D, such as the availability of funds, specialist skill resources and the shares of public and private input, as these are closely related to the economic status and policies of the region where the R&D activities occur. Therefore, we proceed to a country-sector analysis to look into the relationship between IRII and NTR and the output of innovation, considering national and sectoral R&D inputs.

In the following sections, we will first present an overview of the existing research on ex-ante and ex-post measures of disruptive innovation using patent data. We next describe the methodology we use to construct the network, identify technology cohorts clusters, and the development of our proposed indices: IRII and NTR. We will then present the descriptive statistics of these indices through empirical analysis, by using patent data from the pharmaceuticals and the computer technology sectors. Lastly, through correlation study and regression estimation models, we demonstrate results on the relationship between our proposed metrics with the conventional patent indicators at patent level with and without control for country-sector-specific investment and resources in R&D.

2 Literature Review

In the recent literature, researchers use a variety of metrics to capture disruptive innovation through proposed empirical counterparts to novelty, unconventionality, and commercialization potential. There is not yet a consensus on a "best" measure as the proposed indicators capture different stages of the innovation cycle, depend on data availability, or a specific to sector or a selected group of inventions.

Among the indicators that focus on the first stage of innovation, that involves production of new knowledge and inventions, the most widely used is the *originality* index which was developed by Trajtenberg et al. [Trajtenberg et al., 1997] as a measure of knowledge diversification in the development of a patent based on the range of subclasses included in the backward citations of the patent application. The index has been used by researchers in studying the patenting agencies' decisions and economic performance of invention enterprises[Gompers et al., 2005, Harhoff and Wagner, 2009, Stahl, 2010]. It is worth contrasting our IRII and the *originality* index as both measures use the degree of concentration of "same group" technological components inferred from a patent application. The main difference is that while the *originality* index refers to a 'group' as the classification codes belonging to the same subclass among the set of patents cited by the reference one, in the IRII, the 'group' is defined by all subclasses that belong to the same cluster based on the frequency of their co-assignment to a patent in the reference cohort of patent applications. Thus, our index is designed to measure the degree of *novelty in the combination* of technological components compared to the mere breadth of usage, which is why our approach is better tailored to capture the destructive nature of innovation. We measure the novelty vis-à-vis the whole cohort of patent applications, while the breadth is derived from the backward citations on the specific patent application.

Other measures, similar to *originality*, used in the literature include *patent scope*, *backward citations*, and *family size*. *Patent scope* is defined as a simple count of the number of distinct subclasses assigned to the patent on the application. The intuition behind this measure is that a greater number of subclasses is indicative of a wider range of technological components being used and therefore more complex invention or far-reaching impact. The indicator has been used in the literature as a measure for potential of a patent to generate higher technological value as in fundamental invention or economic value through a market return, for example, [Lerner, 1994, Régibeau and Rockett, 2010]. *Backward citations* is also derived exclusively from information listed on the patent application: it is the number of citations of prior art (such as other patents or scientific work) as a source of knowledge on which the authors have relied in the development of their invention. The backward citations are used by the patenting agency to assess the technological novelty of the invention. It should be noted, however, that agencies differ in disclosure rules and therefore comparison across of this variable across patent applications filed in different countries maybe problematic. In the literature[Criscuolo and Verspagen, 2008] backward citations are used to study knowledge transfer and the dynamics of invention within a firm or sector. While some authors find evidence that a large number of backward citations is negatively related to the degree of technological novelty of a patent [Criscuolo and Verspagen, 2008, Lanjouw and Schankerman, 2001, 2004], others find it being positively correlated to the invention's market value [Harhoff et al., 2003].

Lastly, *family size* is a count of the number of jurisdictions at which a given invention has been protected. This indicator is linked to the rights of an inventor to seek wider geographical protection for their invention via related applications to other than their home country patenting offices within 12 months of the first priority filing. In the literature, a larger patent family size is found to be positively correlated with the invention's potential to generate economic value via wider geographical market capture. [Lanjouw et al., 1998, Harhoff et al., 2003].

Two sets of authors, as we discuss above, Verhoeven et al. [2016] and Silvestri et al. [2018] develop ex-ante technological novelty indicators that are specifically designed to capture the disruptive nature of innovation. We re-call the earlier discussion from the Introduction about the distinct features of our methodology that aims at higher accuracy and more comprehensive approach to measuring re-combination and novel use of technological components. Verhoeven et al. [2016] demonstrate the validity of their measures by finding a strong positive correlation with the likelihood that a patent is in a group of award-winning patents, on the one hand, and a strong negative correlation with the likelihood that a patent is refused by the European Patent Office, on the other. Silvestri et al. [2018], instead, offer a time-series analysis of the correlation between business-cycle fluctuations and fluctuations in the degree of unconventionality and do not conduct any analysis of the link to between their ex-ante measure of technological novelty and any ex-post indicator.

There are several established ex-post indicators introduced in Squicciarini et al. [2013]: *forward citations*, *generality*, and *breakthrough*. *Forward citations*, similar to *backward citations*, is the number of citations on subsequent patent applications that a patent receives within five to seven years after its publication date. Intuitively, it is thought to reflect the foundational value of a patent in developing new technologies. Several authors have indeed found a positive correlation between the number forward citations and the economic value of a patent [Trajtenberg, 1990, Hall et al., 2005, Harhoff et al., 2003]. [Lanjouw and Schankerman, 2004]. *Generality* is the ex-post counter part of the ex-ante indicator, *originality*, by using forward citations to capture the scope and degree of general-purpose technology that a patent enables. In the literature, this index has been utilised to understand the commercialization potential of inventions and how innovation meets the market [Henderson et al., 1998, Layne-Farrar and Lerner, 2011, Galasso, 2011]. *Breakthrough* is also derived from the number of forward citations a published patent received: it is an indicator variable which equals 1 for patents in the the top 1% by the number of forward citations among those filed in the same year; and 0 otherwise. It was first put forward by Ahuja and Morris Lampert [2001] to identify inventions that have a significant impact on future technological development. In their seminal work Ahuja and Lampert found that familiarity, maturity and propinquity are three "traps" that could hinder the creation of a breakthrough invention in firm organizations [Ahuja and Morris Lampert, 2001]. More recently, Srivastava and Gnyawali [2011] found that the quality and diversity of a firm's portfolio of technological resources have a positive impact on the probability of a breakthrough innovation. Kerr [2010], Popp et al. [2013] provided evidence that the occurrence of breakthrough innovations could stimulate subsequently regional and sectoral innovation activities.

Other authors, Arts and Veugelers [2015] and Briggs [2015] modify the standard definition of a breakthrough innovation and introduces an endogenous threshold of citations that depends on the observed distribution among the patents in each cohort to allow for a time-varying sharing of patents in each cohort to be breakthrough [Arts and Veugelers, 2015, Briggs, 2015]. Briggs [2015] further motivate this methodology by referring to sectoral differences in the volume of citations and show that co-ownership of a patent is an important factor in determining the breakthrough potential of the patent as defined in their work. Given its well-recognized indicative significance as a ex-post patent quality indicator, *breakthrough* is also used in our study to validate the power of our *ex-ante* novelty measures, IRII and NTR, in predicting future technological novelty impact.

3 Measures of Ex-ante Technological Novelty

To measure the extent of technological novelty embedded in a single patent, we must first establish what is the existing state of technological knowledge in the sector to which the patent belongs. In this respect, we build on our previous work, [Gao and Lazarova, 2022], where as part of a framework for quantifying the technological evolution at sectoral level, we offer a methodology for mapping the frontier of current technological knowledge as captured by a cohort of patent applications. In our work, we represent the frontier as a network of technological components, their interconnectedness and the strength of pairwise connections. Within this complex diagrammatic representation of the state of technological knowledge, we proceed to identify patterns of combinations of technological components usage which occur with a high frequency. Equipped with this information, we are able to gauge the degree of novelty in the combination of technological components listed on a new patent application as compared to those present in a cohort of patent application from a most recent reference period. In addition, we can identify among the patent characteristics any technological components which have not been listed in an application in the reference cohort.

In summary, our methodology consists of two stages: mapping of technological knowledge use and identifying high frequency patterns of usage in a sector;

and, measuring the ex-ante technological novelty of a patent in the sector. The first stage uses information from the whole cohort of patent applications. The second stage quantifies two distinct aspects of technological novelty: the intensity of novel combinations in the use of established technological components in a patent (IRII) and the proportion of technological components in the patent application that are new to the sector (NTR). As we have discussed in detail the network-level analysis that constitutes stage one in our previous work[Gao and Lazarova, 2022], we only provide an overview of these parts of the methodology below. The novel part of this methodology is in the patent level measures of ex-ante technological novelty that follows from that.

3.1 Network Construction and Clusters Identification

We follow the network construction method developed in [Gao and Lazarova, 2022] to build a map of the frontier of technological knowledge in a sector in a given time period. This exercise uses information from the set of all patent applications filed in the cohort linked to a specific sector. In the description of an invention, a patent application contains a list of technological components on which the invention is built. These technological components are well-defined categories in technological classification systems published by patenting authorities. Here we adopt the International Patent Classification (IPC) scheme, a hierarchic system assigning technical fields as a patent attribute developed and released by the World Intellectual Property Organization (WIPO)[1]. We employ two tiers of the IPC scheme: the first tier is the 4-digit level IPC codes, known as

[1]The IPC scheme can be accessed at https://www.wipo.int/classifications/ipc/en/

subclasses; and the 2nd-tier is the 8 to 11 digits IPC codes labeled as *subgroups*. In the network representation of the technology encoded in the patents, we use the subclasses listed on all patent applications in a cohort as the network nodes. We define a link between two nodes to exist in the network if the two subclasses corresponding to these nodes are co-listed on at least one patent application in the cohort. The weight of the link between the two nodes is calculated by using the 2nd-tier IPC codes at the subgroup level and aggregating this information across the whole cohort. In particular, We take the strength of the technological complementarity between any two subclasses in the development of a patent to be proportional to the number of pairwise combinations of subgroups listed under each subclass. For example, consider two patents, A and B, that both list subclassess 1 and 2 as their technological attributes. In patent A, subclass 1 lists one subgroup and in patent B subclass 1 lists three subgroups. Let subclass 2 list two subgroups as attributed to both patents A and B. Then, the strength of the complementarity between subclasses 1 and 2 is calculated to be two in the development of patent A (there are only two distinct pairs of subgroups between the two subclasses) and six in patent B (there are 6 distinct pairs). If two subclasses are not co-listed on a given patent, then their technological complementarity in the development of that patent is zero. So to derive a measure of the strength of the technological link between any two subclasses present in a cohort of patents, we sum up the number of pairwise combinations of subgroups listed under these two subclassess for all patent applications in the cohort.[2]

[2]We acknowledge that the IPC scheme is imbalanced in the sense the number of subgroups listed under each subclass varies. This implies that subclasses with a smaller number of subgroups, theoretically, can form fewer pairwise links. However, empirically, we do not observe that subclasses with a larger number of subgroups list more subgroups on a patent.**Can we add some data to support this statement?** Since our method tracks only the number of subgroups and not the variety of subgroups, we think there is no empirical bias that underestimates the degree of complementarity between subclasses with a smaller number of subgroups

The resulting weighted network provides a comprehensive snapshot of the interconnectedness between the subclasses used in the development of the cohort of inventions. We will use that as a benchmark against which we aim to measure the technological novelty of an invention that occurs in the future period. Our task is to measure how close the technology use in a new patent is to those which were used in the development of all the patents filed in the reference window. The next step towards answering this question, given the complexity of the information on technology use captured by the network, is to identify groupings of technological components based on the high frequency of their co-usage in the cohort.

As in [Gao and Lazarova, 2022], we use Carlo Piccardi's lumped Markov chains network community identification method [Piccardi, 2011] to identify naturally formed clusters in the network. When the sample data size is sufficiently large, the clustering method results in a distinguishable network partition with the definition of each cluster being directly related to the strength of links between any two nodes within the cluster. As technologies evolve in every new cohort of patent applications, the composition of clusters, their size, and connectedness strength vary. From a pure probability point of view, where the reference network is partitioned into more clusters, a new patent application is more likely to utilize a combination of subclasses that spans different clusters compared to a reference network partition with fewer clusters. To configure the networks of consecutive cohorts in a temporally comparable way, we choose to fix the number of clusters in the partition of each network.[3]

in the IPC scheme.

[3] In Gao and Lazarova [2022] we examine different values for a fixed number clusters as part of a robustness check in the construction of the technological frontier.

The resulting output from the stage of the methodology is a partition of the set of subclasses listed on all patent applications in a given cohort based on the frequency and strength of their co-listings based on the cluster identification method employed. We capture this output in the following notation, which we will subsequently use in the formal definition of our patent-level technological novelty indices. We will denote a generic patent as k and the set of all patents filed in a period t as N_t. We will denote the set of subclasses listed on a patent application k as S_k and the collection of all subclasses listed on all patent applications filed in a reference window of size s time periods, i.e., from year $t-s+1$ up to year t as $\mathcal{C}_{s,t} = \cup_{j \in N_{t-s+1} \cup \ldots \cup_t} S_j$. We denote the resulting partition of the set $\mathcal{C}_{s,t}$ into n clusters as $P_{s,t} = \{C_{s,t}(1), \ldots, C_{s,t}(n)\}$.

3.2 Inverse Recombination Intensity Index

Our Inverse Recombination Intensity Index (IRII) characterises a patent application by the degree to which the grouping of technological components on which it is based presents a novel way of combining these IPC subclasses compared to their mode of usage in the preceding cohort of patent applications in the same sector. The index is designed to measure the degree of a radical ex-ante technological innovation carried by an individual invention which is benchmarked against the sector-wide practice. We first introduce some notation that we will use in the formal definition of the index.

For a patent k, we recall that the collection of subclasses listed on the k's application is denoted as S_k. Then, the IRII of a patent k filed in a period t

vis-à-vis the reference period $t - s, \ldots, t - 1$ is formally defined as:

$$\mathrm{IRII}_k = \sum_{j=1}^{n} \left(\frac{\mid C_{s,t-1}(j) \cap S_k \mid}{\mid S_k \mid} \right)^2 \tag{1}$$

where $C_{s,t-1}(j)$ is the jth cluster of the partition $P_{s,t-1}$ of subclasses listed on patent applications that were filed in the period from year $t - s$ to $t - 1$ in the same sector. First, we note that there is at least one subclass which is listed in common on k application and the application of patents filed in the reference window since these patents below to the same technology sector. It follows that the lowest value that IRII can attain is bounded by $\frac{1}{|S_k|}$. This is obtained when all subclasses listed on k's application but the sector-definition one are not elements of the set $\mathcal{C}_{s,t-1}$. The maximum value that the index can obtain, instead, is 1. This is when all the subclasses listed on k's applications belong to the same cluster in the reference window, i.e. there has been no radical recombination in the use of technological components used in the development of k compared to those combination used in the development of the cohort of patents in the reference window.

3.3 New Technology Ratio

Our New Technology Ratio (NTR) characterises a patent application by the degree to which it employs technological components that have not been listed on any patent applications in the previous cohort in the sector. Patent data analysts have used IPC classifications at different hierarchic levels for their purposes. Since our aim is to detect any new technological elements compared to the previous cohort, we define the ratio at the subgroup level of the IPC hierarchy as this is a more refined measure with a higher degree of variability compared to a similar measure based on subclasses. Like Verhoeven et al. [2016] who used the 7-digit subgroups to identify novelty in technological knowledge origins, we take a further step from there to measure the intensity of such novelty. Given a patent k, we denote the set of subgroups listed on k's application as G_k. We denote the set of all subgroups listed on a patent application in a given sector in the period $s-t+1$ to t as $\Gamma_{s,t} = \cup_{j \in N_{t-s+1} \cup \ldots \cup_t} G_j$. By exclusion, the subset of subgroups listed on k's application filed in period t which had not been referenced on a patent application in the k's sector during the reference window is given by $G_k \setminus \Gamma_{s,t-1}$.

$$\mathrm{NTR}_k = \frac{\mid \{G_k \setminus \Gamma_{s,t-1}\} \mid}{\mid G_k \mid}. \tag{2}$$

Notice that NTR is higher as a patent uses new subgroups within the hierarchy of the technological classes on which the sector is defined. It may also increase if new subgroups under the hierarchical classification of other sectors are employed in the development of the patent. The maximum value of NTR is 1; this is when none of the subgroups listed on a patent application are used by any patent in this sector in the reference window. The minimum value, conversely, is 0; this is when all the subgroups listed on a patent application are in the set of subgroups from the reference window.

4 Data and Empirical Statistics

4.1 Patent-level Data

To illustrate the use of our novel measures, we obtain data on patent applications from the REGPAT database [Maraut et al., 2008] in the COMP and PHARM sectors. Measures of patent quality are taken from the OECD's Patent Quality Indicators database [Squicciarini et al., 2013]. We use the February 2022 release by OECD for both datasets. Based on the information in the REGPAT dataset, we can identify and select all patent applications that can be attributed to each of the two sectors. As per established practice in technological field studies, [Fink et al., 2016], patents are classified into these sectors in accordance with the definition of the WIPO.[4] We note that the IPC scheme has undergone regular updates to keep up with the latest scientific and technological developments. With each reform, WIPO re-classify patent files to reflect the changes made to the IPC scheme through the revision. By downloading the data in one batch, we ensure that the IPC classification information is consistent and coherent across cohorts of patent applications.

While longer time-series are available in these databases, we select the sample period from 1980 to 2018 for patents in PHARM, and from 1981 to 2018 for the COMP. The samples are selected on the basis that the volume of applications is consistently above 500 in each consecutive year. We need such large cohorts of patents in order to construct a network with a sufficiently large

[4]Sector definition for Pharmaceuticals and Computer Technology can be found at `https://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_41/ipc_ce_41_5-annex1.doc`. (last accessed December 2022).

number of nodes and high enough density of connections to identify persistent clusters in each consecutive year at the first stage of our methodology; therefore, derive reliable values for IRII and NTR at the second stage.

To ensure the robustness of our analysis, we adopt a similar approach to Gao and Lazarova [2022] and calculate IRII and NTR with three different reference windows: 1 year, 3 years and 5 years, labelled as IRII11, IRII31 and IRII51; and NTR11, NTR31, and NTR51, respectively. For IRII and NTR with three-year and five-year reference periods it is feasible to calculate IRII and NTR from 1981 and 1983, respectively, for both COMP and PHARM. In addition, to render our results less sensitive to the choice of the number of clusters at the first stage of our methodology, we construct four different network partitions using 8, 12, 16 and 20 clusters, respectively. We then obtain the average IRII value of a patent over the 4 different partition configurations.[5] We perform the same average calculation for the IRII in each of the there reference window: IRII11, IRII31, and IRII51.

As an illustration of our two-stage methodology, we provide a series of graphs that allow us to visualise the network clustering, distribution of new subgroups, and the recombination process. In all network graphs, we choose to present the network partitions of the 2005 and 2006 cohorts because a new version of the IPC scheme (the eighth edition) was released on January 1, 2006 which presented a major revision [Makarov, 2006]. This allows us to detect a possible impact of the process of revision on our results.

We start with Figure 1, where panels (a) and (b) show the 8-cluster network

[5]The number of clusters in a partition does not impact on the definition of NTR.

partitions constructed based on PHARM patents filed in years 2005 and 2006, respectively. Figure 2 shows the partitions of the COMP patent networks in the same two consecutive years. Nodes highlighted in blue color are the subclasses containing subgroups that are not present among patents of the sectoral cohort of the previous year, i.e., the new subclasses used in the calculations of the NTR. We note that in neither Figures 1 or 2, the network partitions show a structural change between the two years. The distribution and portion of these nodes are also similar in the temporally consecutive network partitions. While Figures 1 and 2 provide snapshots, in Figure 3 we present the temporal trend by sector of the share of new subgroups in the total number of unique subgroups, and the share of patents containing such new subgroups, whereby the new subgroups are identified using a 1-year reference window. Both sectors show a decreasing trend of the two metrics. Similar to the cluster structure in Figures 1 and 2, the major IPC scheme update in 2006 does not appear to introduce a structural break in these series.

Next, we present Figures 4 and 5 which illustrate the process of recombination with reference windows of one year and five years. Using PHARM patent data, Figure 4-(a) shows how the network clusters of subclasses used by patents filed in 2006 are recombined vis-à-vis the clusters identified through patent use among those filed in 2005. Compared to Figure 4-(c) where the reference is the network partition built on patents filed in the 5-year window from 2001 to 2005, some differences are visually detectable. For example, the largest cluster in 2006 is broken down to more evenly distributed in size sub-clusters in (c) compared to (a), which indicates that the degree of recombination of that cluster is higher against the 5-year window network partition. Since our algorithm for identifying network clusters implies a likely positive correlation between the persistence probability and the cluster size, in this example, IRII51 is likely to be smaller than IRII11 thanks to the higher extent of recombination in the largest cluster in the network partition of 2006. Similarly, Figure 5 provides the resulting networks for COMP using data for the same two years. The recombinant degree in Figure 5-(a) is not so different from the one exhibited in 5-(c). This can also be seen in the intermediate stage in (b). In Figure 4-(b) about half of the nodes in the core blue sub-cluster within the largest cluster are in red color - i.e. they below to a different cluster in the 5-year window, while in Figure 5-(b) a much smaller number of nodes in the blue sub-cluster are red, showing that the network partition of COMP patents in 2005 is not that different from that in 2001-2005. In Figures 4 and 5, we start noting important differences in the composition of ex-ante technological novelty between PHARM and COMP. We will explore these further throughout our empirical case study.

The left part of Table 1-(a) and 1-(b), under the header "whole sample", provides a summary of the descriptive statistics of IRII and NTR calculated for the two sectors, as detailed above, using different reference windows. During the sample period PHARM and COMP are comparable in the number of patent filings. There are, however, important sectoral differences. PHARM, on the one hand, has lower average NTR, which is indicative of a lower proportion of new technological components being introduced in this sector compared to COMP. COMP, on the other hand, has higher average IRII, suggesting that the pattern of usage of established technological components is more persistent and inventions in the field are likely to rely, to a less extent than in PHARM, on a novel combination of the established technologies. We have already seen an indication of this observation in the two-year snapshots presented in Figures 4 and 5, which is now demonstrated again in the summary statistics of the entire sample. Meanwhile, the minimum values of IRII in PHARM are larger than those in COMP for all indices irrespective of the reference window. This indicates that the patents with the lowest IRII in COMP have a higher level of technological recombination than those in PHARM. There are six COMP patents with IRII11 equal to zero, the minimum value. They all have NTR11 equal to one, the maximum feasible value. Each of these patents has only one or two subclasses which are all new technological components compared to the previous year.

Figure 6 shows how the annual average IRIIs and NTRs, when calculated using different reference windows, change over time. For both sectors, IRIIs fluctuate around a constant level and NTRs exhibit a decreasing trend. We deduce, therefore, that while the recombination of technologies is a permanent feature of inventions, the introduction of new technologies diminishes as a sector matures; an observation which is consistent with Figure 3. Using different reference windows for calculating the IRII and NTR result in similar values and trends. This supports the robustness of our method. Comparing the two sectors, we note that the IRIIs of COMP not only have the highest all-time average values, but also exhibit the lowest level of fluctuations, which indicate a more stable rate of technology recombination in this sector.

We further assess IRII and NTR in comparison with several measures of patent quality and technological value described in Squicciarini et al. [Squicciarini et al., 2013]. For the sample of patents for which we have calculated IRII and NTR, we retrieve individual patent data from the OECD Quality Indicators dataset using the unique patent identifiers.[6] We discussed these indicators in our literature review. In Table 2-(a) we provide their definitions. Some of these patent quality indicators are designed to measure ex-ante technological novelty (patent scope, family size, backward citations, originality) similar to IRII and NTR. Others - generality, breakthrough rate, and forward citations[7] - are measures of ex-post quality.

We will further investigate the statistical power of correlation between the ex-ante and ex-post measures in the next sections. Here, we take the opportunity to illustrate that this relation is not trivial by using the first stage of our methodology and the data on breakthrough rate. In Figures 7 and 8, for PHARM and COMP, respectively, we highlight the nodes with subclasses that belong to patents designated as *breakthrough* patents in the network clusters in years 2005 and 2006. As shown in these figures, while the larger clusters contain more subclasses that belong to *breakthrough* patents, such subclasses can be found in a cluster of any size and their distribution in different clusters varies from year to year.

The left part of Table 1 also includes the summary statistics of the variables listed above for PHARM and COMP. By focusing on a sectoral comparison of these variables, we note that PHARM has a larger *patent scope* and *family size*

[6] The OECD patent quality indicator dataset provides two data tables: one at patent level; and the other at cohort level by year of filing and technology field. We use the patent-level data set.

[7] We note that the OECD dataset provides two metrics on *forward citations*: one counts citations within five years after patent's publication, and the other, within seven years. The publication date of a patent is usually within 18 months of the patent application filing date. Thus, patents with a more recent application date are expected to have fewer forward citations. In our study, we use the five-year post-publication forward citation numbers to utilise a longer time-series sample with a more accurate count of *forward citations*.

on average than COMP. A smaller *patent scope* (smaller S_k) could be a potential contributor to a larger IRII, but we cannot conclude that this is the cause of the higher average IRII values in COMP as shown in Table 1.[8] PHARM patents also tend to cite more prior arts and receive more citations in five years after patent publication.[9] Despite a lower average value, COMP patents have a wider range of *patent scope* than PHARM. Indeed, the kurtosis of COMP *patent scope* is 24.0196, much higher than that of PHARM: 6.4531. Meanwhile, PHARM is wider in range than COMP in *family size*, *backward citations*, and *forward citations*. However, only with *forward citations* PHARM has higher skewness (PHARM: 37.1322, COMP: 22.8177) and kurtosis (PHARM: 2935.5600, COMP: 1067.0120).[10] In summary, the COMP sector, with lower mean values on these variables, has a more positively skewed and more leptokurtic distribution than PHARM in *family size* and *backward citations*, and the opposite is true for *patent scope* and *forward citations*. PHARM patents also demonstrate higher mean values in the *originality* and *generality* indicators. Both sectors have negative skewness and positive kurtosis values in these two variables, with the absolute values larger in PHARM.

[8]The minimum value of *patent scope* is reported as zero in both sectors. This may be taken as typo. Instead, in the original OECD data source there is one PHARM patent and four COMP patents codes with *patent scope* equal to zero. We have manually looked up these patents using the European Patent Office's patent search service, Espacenet, and found that each of them actually has one IPC subgroup/subclass. Therefore, the correct value of *patent scope* by definition should be one. After removing these patents from each sector's sample, the difference in the mean value of all the variables is at the 5th or 6th place after decimal point (See Supplement A in the Supplementary Datasheet attachment for the summary statistics excluding the zero-patent-scope observations.). We further manually computed the number of unique subclasses of each patent in the sample data and compared with the OECD dataset. Out of the 278,990 observations in PHARM, 4,454 show different values from *patent scope*, and the average difference is -0.0172. For COMP, 5,794 patents out of 282,506 have different *patent scope* values, with an average difference of -0.0232. So, we consider the impact due to this potential data inaccuracy to be minimal and continue to use the OECD dataset in the subsequent analysis.

[9]Both forward and backward citations include citations to and from patents within and outside of the sector.

[10]See Supplement B in the Supplementary Datasheet attachment for the statistics for all the variables.

The summary statistics show that overall inventions in the two sectors carry different ex-ante and ex-post characteristics. PHARM patents tend to have a wider technological breadth and a larger set of patents filed in international patent jurisdictions that are related to the same priority filings. Both indicators have been used as measures of the potential of the invention to generate higher commercial value for the patent owner [Lerner, 1994, Lanjouw et al., 1998, Harhoff et al., 2003]. In addition, largers numbers of backward and forward citations indicate that knowledge spillover among patents plays a bigger role in PHARM inventions than in COMP; while higher average values of *originality* and *generality* of PHARM patents point to the likelihood of inventions being more original [Trajtenberg et al., 1997] and more general-purpose [Hall and Trajtenberg, 2004], but less fundamentally novel [Lanjouw and Schankerman, 2001]. Among all the variables based on simple counts, except for *forward citations*, COMP summary statistics have distributions with higher peak and thinner tails.

Finally, we note that significantly fewer patents have data on *originality* and *generality*. In particular, data availability of *generality* varies significantly over time and across sector. We present the number of observations over the sample period split by sector in Figure 9. The figure clearly illustrates that the *generality* time series drop to rather low levels by 2018 for both PHARM and COMP. The lower number of observations is most likely linked to the increasingly shorter window over which forward citations can be observed.

As for *originality*, about 4.4% PHARM patents and 6.2% COMP patents don't have this data in the OECD dataset. We present the summary statistics of patents with *originality* data in the right part of Table 1-(a) and Table 1-(b) for each sector, respectively. Comparison of the mean values in the left and right parts of the tables shows some different traits between the two sectors. In PHARM, IRII of patents with *originality* data is larger on average than those without, while the opposite is true in COMP. The gap in NTR is negligible in PHARM, but in COMP NTR is on average larger on the right side. Among the patent quality indicators, the mean value of *patent scope* is larger in the whole sample, while in COMP it is larger in the sample with *originality* data. Overall, by excluding patents without *originality* data, the PHARM sample shows a lower level of technological recombination and smaller patent scope, and the opposite is true for COMP. All the other variable values are larger for both sectors in the sample where *originality* data are available.

To check that the limited availability of these two variables does not introduce a selection bias in our analysis, we will present computations both including and excluding these variables in the next subsection.

4.2 Patent-level Correlations

We begin our analysis with a discussion of the correlation matrices of the continuous variables that we introduced in the previous section for each sector under investigation: COMP and PHARM. In Tables 3 and 4 we present the pairwise correlation coefficients in two parts: sub-tables (a) in each table presents results based on the sample where data for all the variables except *originality* and *generality* are available, and sub-tables (b) include the pairwise correlations for the full list of variables. The sample size for the computations of sub-tables (a) is much larger than the ones that is used for sub-tables (b) due to the limited data availability for *generality* as shown in Figure 9 and to a lesser extent for *originality* as revealed in Table 3.

Tables 3 and 4 both include the pairwise correlation coefficients for IRII and NTR computed using three different reference windows. Overall, the pairwise correlations between our technological novelty indices and the patent quality indicators decrease or stay at the same level as the reference time window increases with the exception of the pairwise correlations between *family size* and IRII11 and IRII13 in PHARM where the strength of the pairwise correlation increases with the length of the reference window. Based on this observation, we will not make a distinction between the same index computed with different reference windows in the discussion below.

In addition, comparing parts (a) and (b) of each table, we do not detect substantial differences between the two panels despite the difference in the number of observations used to compute these pairwise correlations. We can point out that *originality* is only weakly correlated with IRII in PHARM, but the correlation is much stronger in COMP. Instead, the correlation between *originality* and NTR doesn't differ as much. With respect to *generality*, IRII exhibits the stronger correlation in both sectors compared to that with NTR. Finally, for both IRII and NTR, the correlations are stronger in COMP compared to PHARM.

Focusing on the pairwise correlations of IRII and NTR with the other patent quality indicators for which data is more widely available, we can identify some clear patterns. IRII is consistently negatively correlated with the other variables except for a weak positive correlation with *backward citations* in the PHARM sector. Among these negative correlations, the strongest in absolute value is with *patent scope* and the second-strongest is with NTR. The weakest correlation for PHARM is with *family size*, and for COMP with *backward citations* and *forward citations*. A comparison across the sectors reveals a general tendency for the strength of the pairwise correlations between IRII and the other patent quality indicators to be weaker in PHARM and stronger in COMP.

In relation to the other NTR-based correlations, the two sectors are notably distinct. In PHARM, NTR is only positively correlated with *patent scope* and *generality*, while it has weak negative correlations with the other patent quality variables. In COMP, NTR, as we would expect a priory, it is positively correlated with all the conventional patent quality measures. In COMP the pairwise correlations with NTR are also larger in absolute values compared to those in PHARM.

Yuan: We need to mention the comparison/complementary contribution to the literature in the Introduction and Conclusion. To understand these sectoral differences in the sign of correlations between NTR and the other patent quality indicators, we could draw a comparison with the work of Verhoeven et al. [2016]. In this authors' work, the presence of Novelty in Technological Origins in patent families, on average, increases the number of forward citations of a patent; a result which is seemingly consistent with our pairwise correlations in COMP but opposite to the negative correlation between NTR and *forward citations* in PHARM. In contrast to our continuous variable (NTR), however, Verhoeven et. al. use a binary indicator to measure novelty in technological origins which is equal to one when there is at least one new technological origin in a patent family and zero otherwise. To better compare our findings to those of Verhoeven et. al., we compute the average value of *forward citations* when NTR11 is equal to zero and when it is strictly positive. The results show that in PHARM, the average number of *forward citations* is 1.568 when $NTR11 > 0$ and 1.268 when $NTR11 = 0$; and for COMP, the average value is 1.522 when $NTR11 > 0$ and 0.837 when $NTR11 = 0$. Thus, on the basis of this replication practice using our dataset, we can draw conclusions which are consistent with the findings in the literature.

We would argue that our approach of using a continuous measure for the intensity of new knowledge origins that are employed in the development of invention allows us to identify more complex relations between variables compared to when an indicator variable is used. Our further investigation, indeed, points towards the potential presence of a non-linear relationship between NTR and *forward citations* and a possible explanation for this in the interaction with *patent scope*. We demonstrate this point on Figure 10 where we present a scatter plot with NTR11 masured on the horizontal axis and number of forward citations measure on the vertical axis in each sector. In both COMP and PHARM, the majority of patents utilise the same knowledge origins as those used in patents filed in the previous year, thus we observe a concentration of observations at NTR11 equal zero in Figure 10. When NTR11 is positive, the scatter plot on the left side, where the data for PHARM is displayed, shows a relatively larger number of patents within the range of $NTR11 < 0.2$, with the largest number of forward citations around $NTR11 = 0.05$. Considering Equation 2, we can infer that this range would require a total number of unique subgroups of or above 5. *Forward citations* tends to increase as the number of unique subgroups approaches to around 20, followed by a decrease, on average, when the number increases beyond this point. In COMP, instead, the pivoting point is when NTR11 is around 0.5, equivalent to a minimum number of unique subgroups of 2. Meanwhile, we find a strong correlation between the number of unique subgroups and subclasses (i.e. *patent scope*): 0.5994 in PHARM and 0.6753 in COMP). This investigation further reveals the sectoral distinctions reflected in patent properties. The influence of the interaction between *patent scope* and our indices will be further analyzed in the regression analysis that follows.

Overall, the discussion on the pairwise correlations presented in Tables 3 and 4 suggest that IRII - our proposed new ex-ante technological recombination novelty index - has the expected signs of correlation coefficients with the established indicators of patent quality. As IRII is an *inverse* index, the negative correlation coefficients are in line with the expectation that a higher intensity of re-combination of technological components, i.e. lower IRII, is associated with higher patent quality as captured by one of the established indicators. For NTR the picture is more obscure and it is hard to draw a unifying summary of the different coefficient signs and strength of correlations, especially in relation to *family size*, *backward* and *forward citations*, and *originality*. We could say that the novelty brought by new technologies in COMP tend to be more aligned with the establish indicators of ex-ante patent quality. Overall, for both sectors, the correlation coefficients between IRII and NTR, respectively, and the other patent quality indicators are low with the notable exception of the pairwise correlation between IRII and *patent quality* where at its highest - in Table 4-(b) -

it can categorised moderately high, suggesting that IRII and NTR capture different information sets. We also note that the degree of correlation varies across sectors which, along with the discussion of earlier figures and tables, points to a need for a sector-specific empirical analysis. To render this analysis more accurate, we endeavour to include variations across socio-economic environments by including sector-country-level controls.

4.3 Sector-Country-level Data and Summary Statistics

We source country-sector-level data from the OECD MSTI database which covers a wide range of sector-country-level variables for the OECD member states and seven non-member ones starting from 1981 onward. The MSTI database contains a wide range of variables from which we have chosen a selection of controls that fall into one of the three categories that are relevant to this study: (1) three sector-level variables: Business and Enterprise R&D expenditure (BERD), trade balance, and export market shares (defined in Table 2-(b) as B_COMP, B_PHARM, TD_COMP, TD_PHARM, TD_XCOMP and TD_XPHARM); (2) country-level capital R&D expenditure variables such as the R&D expenditure in three major segments: Business and Enterprise, Government Intramural, and Higher Education; all measured both in current Purchasing Power Parity (PPP) $ (defined in Table 2-(b) as B_PPP, GV_PPP, and H_PPP); and (3) a country-level human resources in research variable measured in full-time equivalent unit (FTE) (defined in Table 2-(b) as TP_RS).

We use the patent filing date and applicants' residence information, both included in the OECD REGPAT data, to control for cohort effects and country-fixed effects. The information allows us to control for factors that are common for all patents but vary from cohort-to-cohort such as the state of the world economy and world-wide technological frontier; as well as account for differences in regulatory environment and policy at the national level, which are invariant over time in the period under investigation. Since our key independent variables are characteristics of individual patents, the sector-country-year variables obtained from the OECD MSTI database need to be transformed to patent level. We do so by defining patent-level MSTI variables as the weighted average of the sector-country level variable where the weights are the share of applicants residing in each unique country listed on the patent application. For example, consider a PHARM patent that was filed in 2000 and listed two applicants located in Germany and one applicant in Japan. For this patent, each of the MSTI variables mentioned in the first paragraph of this subsection will be computed using the MSTI variable of PHARM-Germany-2000, weighted by the applicants' country share of 2/3, added to the MSTI variable of PHARM-Japan-2000, weighted by Japan's share among applicants' residency of 1/3.

The datasets for both sectors cover the period from 1981 to 2018 for 25 countries [11]. Table 5 provides the descriptive statistics of the patent-level MSTI variables for both sectors during the period from 1981 to 2018. We note that PHARM and COMP have very similar values for both the mean and standard deviation statistics of all MSTI variables, thus, any sectoral differences in patent quality is unlikely to be driven by any of these factors.

[11] The dataset includes information from the following countries: Australia, Austria, Belgium, Canada, Switzerland, China, Germany, Denmark, Spain, Finland, France, the United Kingdom, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, the Netherlands, Norway, Russia, Singapore, Sweden, Taiwan and the United States.

In Table 6 we present country-level data on the total number of patent applications, number of breakthrough patents, and the percentage of breakthrough patents per sector for the OECD countries in the sample period.[12]. The data in Table 6 suggests an imperfect, at best, correlation, between patent volume and breakthrough rate. In both sectors the countries with the top breakthrough rates, those above 1%, are all placed in the bottom half of the table in terms of applications volume; in PHARM the countries with the most and least patent applications, USA (ISO code US) and Portugal (ISO code PT), have comparable breakthrough rates: 0.427% and 0.420%, respectively; and in COMP, Germany (ISO code DE), the country with the third highest volume, has the forth lowest breakthrough rate of 0.09%. Based on the data in Table 6 in the following sections we will further explore country-specific influences on the ex-ante and ex-post novelty of patenting activities.

5 Relationship between IRII and NTR and Ex-post Invention Quality

We choose to use *forward citations* and *breakthrough* as measures for ex-post technological quality of an invention due to their widely recognized significance in predicting a patent's influence on future technological development and wide availability in large databases of patent applications.**EL: we need to include some references here**. Our main objective is to study how our two ex-ante novelty measures, IRII and NTR, correlate with these ex-post patent quality

[12]Country codes are taken from the ISO 3166 alpha-2 standard definition issued by the International Organization for Standardization, which can be accessed at: `https://www.iso.org/iso-3166-country-codes.html`

indicators while controlling for the explanatory power of other established factors. Among these controls we include patent-level ex-ante measures of quality and relevant sector-country level variables; all of which we discuss extensively in Section 4.

Formally, the patent-level regression models for patent k of cohort t in sector $j = \{COMP, PHARM\}$ are presented below. We start with the Poisson regression model where *forward citations*$_k$ is the dependent variable:

$$\begin{aligned}\log(\textit{forward citations})_k &= \alpha_0 + \alpha_1 \mathrm{IRIIX}_k + \alpha_2 \mathrm{NTRX}_k + \boldsymbol{\zeta_1}' \mathbf{QI}_k + \\ &\quad + \boldsymbol{\zeta_2}' \mathbf{MSTI}_{jkt} + \mu_{jk} + \lambda_{kt} \end{aligned} \tag{3}$$

Next, we present the Probit regression model where *breakthrough*$_k$ is the dependent variable and Φ is the cumulative standard normal distribution function:

$$\begin{aligned}Pr[\textit{breakthrough}_k = 1] &= \Phi\big(\beta_0 + \beta_1 \mathrm{IRIIX}_k + \beta_2 \mathrm{NTRX}_k + \boldsymbol{\phi_1}' \mathbf{QI}_k + \\ &\quad + \boldsymbol{\phi_2}' \mathbf{MSTI}_{jkt} + \nu_{jk} + \theta_{kt}\big) \end{aligned} \tag{4}$$

In regression models (3) and (4), the main regressors of interest are IRIIX_k, the Inverse Recombination Intensity Index of patent k and NTRX_k, the New Technology Ratio of patent k, both computed for one of three reference windows $X \in \{11, 31, 51\}$. The additional controls are the following: $\mathbf{QI}_k$ is a vector of ex-ante patent quality indicators defined in Section 4.1, namely, *patent scope*$_k$, *family size*$_k$, *backward citations*$_k$, *originality*$_k$; $\mathbf{MSTI}_k$ is a vector of sector-country specific variables from the OECD MSTI database discussed in Section 4.1 and listed in Table 2-(b); μ_k and ν_k are vectors of five country dummy variables, one for each of the top five countries by volume of patent applications by sector (as reported in Table 6) that equals 1 if at least one of the applicants listed on patent k's application resides in this top five country; and λ_t and θ_t are patent k's year of application, t, cohort fixed effect.

We perform estimations of the above regression models for each sector separately and test the robustness of the results by imposing a variety of restrictions and introducing non-linear effects. In particular, we remind the reader that *originality* is a variable that is not available for nearly 40,000 patent applications across the two sector. Informed by the discussion on the summary statistics of the left-hand and right-hand sections of Table 1, we therefore present estimations with and without *originality* to control for differences in effects due to a different set of controls and due to a different sample set. Similarly, we conduct estimations with and without the sector-country variables derived from the MSTI dataset due to the gaps in data availability. Moreover, following the findings in Section 4.2 on the non-linear relationship between NTR and *forward citations*, we augment the regression models presented in (3) by adding the interaction terms $\text{IRIIX}_k \times \text{patent scope}_k$ and $\text{NTRX}_k \times \text{patent scope}_k$.

5.1 Empirical Results

We start our discussion of the estimates of our regression models with the largest data sample available to us. Initially, we therefore impose the restrictions of $\boldsymbol{\zeta_2} = 0$ and $\boldsymbol{\phi_2} = 0$on Equations (3) and (4), respectively, i.e., we exclude from the estimations the MSTI-controls.

We first present the estimates for the pharmaceutical sector. In Tables 7 to 9 we present the estimations of Equation (3); Table 8 we augment this model with the interaction term between *patent scope* and IRII and in Table 9 we add to be base model the interaction between *patent scope* and NTR. Table 10 presents the estimates of Equation (4). In the same order, Tables 10 to 13 provide the analogous results of COMP. We note that in this manuscript we only report results with the interaction terms between *patent scope* and our ex-ante technological novelty measures when the independent variable is *forward citations* because the impact of these interaction terms on *breakthrough* is negligible.[13] All estimates include the cohort year fixed effects and the patent quality control variables described above. The country fixed effects μ_k and ν_k are also included to control for the potential relationship between the regional level of sectoral technological innovation and individual invention's influence.

In every table there are three variants of the base model: columns (1) to (3) use the largest sample of available data and exclude *originality* as a regressor, with IRII and NTR computed using the different reference windows–one year, three years, and five years–respectively; columns (4) to (6) use the same set of controls as in the first three columns but with the sample for which *originality* data are available; and (7) to (9) present the result with *originality* among the controls. We present our results in this order to make more transparent the impacts of data availability versus set of controls.

Focusing on the regression models using *forward citations* as a dependent

[13]The estimation results of Equation (4) augmented with the interaction terms are available upon request.

variable and its relationship with IRII, we note that the magnitude and significance of the IRII11's coefficient in the basic model (column 1, Tables 7 and 11) are remarkably similar: -0.22 and -0.24 for PHARM and COMP, respectively. These estimates suggest that an increase in the recombinant novelty of a patent would significantly increase the number of citations it generates in the next 5 years, Yet, there are some notable differences between the two sectorsa. The estimated marginal effect of IRII is the largest in absolute value and significance for IRII11 and then decreases in magnitude and significance for IRII31 and IRII51 in PHARM; whereas the opposite is true in COMP whereby the largest marginal effect in magnitude and significance is estimated for the model using IRII51. From the discussion of the summary statistics presented in Table 1, we learned that the average values and standard deviations of the indices using the three window estimations are remarkably similar for both COMP and PHARM, however, COMP exhibits on average a lower level of re-combination across all three windows, and, as displayed in the time series plot in Figure 6, a lower degree of fluctuation. The tables with the pairwise correlation coefficients (Tables 3 and 4) confirm the same observation that the information content of IRII11, IRII31, IRII51 is much more overlapping in COMP and more distinct across the three indices in PHARM. This would imply that the network of IPC subclasses and its partition, constructed at the first stage of the IRII methodology, using the three different sets of patent filings (1-year window, 3-year window, and 5-year window) identify very similar groupings of IPC subclasses by utilization in COMP and that the groupings are more dependent on the cohort of patent filings in PHARM, a scenario similar to the examples in Figure 4 and 5. This could explain why in COMP we observe similar statistical significance across

IRII of the three reference periods; while in PHARM the statistical significance varies.

Next, we compare the results across the different columns within PHARM and COMP. In PHARM (Table 7) the coefficient of IRII11 remains significant and negative in all three variants of the model with a decrease in its magnitude from columns (1) to (7). The latter observation is likely linked to the increase in IRII11's average value as shown in Table 1. A pattern of gradual changes is observed on IRII31 and IRII51 when the sample is restricted to patents with available data on *originality* (columns (5) and (6)), and further when *originality* is included in the estimation - both coefficients are significant and positive in columns (8) and (9). In contrast, in COMP (Table 11), IRII consistently shows a significant and negative effect on the dependent variable. We can draw the conclusion that while the results in COMP are robust to data availability and the inclusion of *originality* as a regressor, for PHARM, conditioning or not on *originality* affects the relationship between *forward citations* and technological recombination against longer reference periods which is partly due to the changes in the statistical characteristics of the sample and partly due to the interrelation between IRII, *originality* and *forward citations*.

From Table 1 we know that among the patents with *originality* data, the average level of technological recombination is lower in PHARM and higher in COMP, while the average number of *forward citations* is higher in both sectors compared to the whole sample. We remind our readers that Tables 3 and 4 show a stronger negative correlation between IRII and *originality* in COMP than in PHARM. These results point to differences in the link between ex-ante and ex-post novelty or perhaps different sample selection bias in the availability of data on originality between the two sectors.

Now turning our attention to the relation between *forward citations* and NTR, we note that across Tables 7 and 11 we see a significant negative coefficient of NTR that is robust in all the variant of the models and when using different reference time windows the calculation of NTR in both sector. This appears to indicate that novelty through introducing technological elements out of the recent sectoral common practice is highly likely to be associated with a reduced number of forward citations received in the five years following patent publication. This result is in line with the observations made based on the pair-wise correlations in Section 4.2 where we also point to a possible role of *patent scope* in the relationship between NTR and *forward citations*. We explore this further next.

In PHARM (Table 9), when we include the interaction term between *patent scope* and NTR, the coefficient of NTR becomes positive in the PHARM sector, with a high statistical significance at the 0.001 level when the reference window is three or five years. In contrast, the inclusion of the interaction term does not affect the estimated effect of NTR in COMP (Table 13). The interaction effect is robust regardless of the sample composition or the inclusion of *originality*. Overall, when controlling for the interaction between *patent scope* and NTR, NTR shows a positive significant impact on the number of forward citations in five years for the PHARM sector, explicitly opposite to the results without the interaction term. Given the non-linear relationship between NTR11 and *forward citations* shown in Figure 10, we further strengthen the conjecture that the interaction has such an influence on the coefficient of NTR in PHARM because the non-linear relationship corresponds to a unique subgroup number around the range of 5 to 20. While for COMP, the pivoting number is around 2. This would give the PHARM patent cohort a much larger range for the interaction term to affect the coefficient of NTR on *forward citations*.

For consistency, we perform a similar analysis with respect to IRII and test for a possible non-linear relationship through the interaction of IRII and *patent scope*. In PHARM (Table 8) when we include the interaction variable, we note a distinct change in the coefficients of IRII31 and IRII51. Now IRII presents a significant and negative relationship with *forward citations* across all variants of the model with the exception of the one presented in column (8). The coefficient of the interaction term is significantly positive in all the three-and-five-year reference window estimations. In COMP Table 12, the interaction terms also shows a significant and positive effect and the magnitude of the absolute value of IRII's coefficients increase compared to the corresponding columns in Table 11.

We now look at the estimation of the probit models using *breakthrough* as the dependent variable. In PHARM (Table 10) IRII has a negative effect at the statistical significance level of above 0.01 while NTR does not carry statistical significance on the probability that a patent is in the top 1% of cited patents in its cohort. In the COMP sector, IRII still maintains its significant negative coefficients similar to the estimations using *forward citations* as dependent variable, however, NTR (as in PHARM) is estimated not to have a significant effect on the probability of a patent being rated as *Breakthrough*. As a measure of ex-post patent quality, *breakthrough* is designed to capture similar information as *forward citations* - the influence on future inventions - but it is more polarising as it only qualifies the top 1% most cited patents. Thus *breakthrough* differentiates patents across the full range of ex-post qualities to a lower extent compared to *forward citations* which is a continuous variable. Therefore, it is unsurprising that an estimation with the same regressors yields results in the similar direction but varying significance.

It is fair to summarize based on the results described above that we are looking at a complex picture. Factoring in the sectoral differences which take form in the influence from including *originality* and the interaction terms, the recombination of technological elements from the previous cohort of technologies within the field is significantly related to an increased number of citations in future inventions. NTR demonstrates a non-linear relationship with *forward citations* that depends on the distribution of the number of unique technological elements as a patent attribute. Our analysis using NTR shows that introducing more technologies not used in inventions filed in the last few years in the sector does not necessarily increase the likelihood of receiving more citations. It depends on a more in-depth understanding of the status and trends of technological development in the sector.

It is worth noting that across all estimations in both sectors, we observe that the established measures of ex-ante patent quality which we use as control variables –*patent scope*, *family size*, *backward citations*, and *originality*–all show a significant and positive relationship with the number of forward citations as well as the likelihood of a patent being Breakthrough. These results confirm findings in the literature.

We also need to comment on the effects of country-specific dummy variables. Controlling for an author's residence in a country with one of the five highest patent application volumes does not effect the estimated coefficients of the other variables. Notably, however, the country-specific coefficients reveal statistically significant differences depending on the location of the applicants. In particular, a patent application with at least one applicant residing in the U.S. or Japan has a higher likelihood of an increase in its number of forward citations compared to an average application where no applicant is from one of the top 5 countries in both sectors with the magnitude of the effects being larger in COMP compared to PHARM. In PHARM, having an applicant residing in France or the UK could lower the number of received forward citations. In COMP, the South Korea indicator variable has the largest statistically significant and positive coefficient, while having an applicant from France or Germany is estimated to have a negative effect in a statistically significant sense. These effects are also found in the estimation of *breakthrough* with some differences: In the PHARM sector (Table 10), having an applicant from Japan now has the largest and most statistically significant and positive coefficient, overtaking the U.S., while the negative significance of the France and the UK indicators decrease. In the COMP sector (Table 14), the U.S. country indicator surpasses that of South Korea in the magnitude of effect. The results suggest that even among the most prolific countries there are important country-level factors that could affect the potential of their resident innovators to produce a more influential or even breakthrough patent. These findings motivate us to take a closer look at the country-specific ingredients in the next sub-section.

5.2 Analysis with R&D input variables

We now discuss the estimations of Equations (3) and (4) that include the vector of country R&D input variables that we group under the heading of MSTI (defined in Section 4.3) labelled after the source of data. Tables 15-19 contain the results for PHARM and Tables 20-24 the corresponding results for COMP. As shown in Table 5, the sample sizes of *B_PHARM* and *B_COMP* are much smaller than the other MSTI variables. Similar to the approach we took with respect to *originality*, we perform a set of robustness checks to de-couple the effects of sample size from the inclusion of these variables among the controls. To this end, in Tables 15 (PHARM) and 20 (COMP), the first three columns provide results based on Equation (3) including all the vector of $\mathbf{MSTI}_k$, except for *B_PHARM* and *B_COMP*, respectively. Columns (4) - (6) provide results from the same model but based on patents for which the sectoral BERD data are available, thus showing a drop in the sample size - nearly 20% in PHARM and over 25% in COMP. In the last three columns, all the variables in the vector $\mathbf{MSTI}_k$ are included.

In Table 15 (PHARM), the significant and negative coefficients of IRI11 and all variants of NTR remain robust as in the estimations without the R&D input variables (Table 7, columns 1-3). The coefficient of IRII31 is unstable switching from positive and significant to negative and significant (columns (2), (5), (8)) while the coefficient of IRII51 is positive as in Table 7 but more strongly significant. With respect to COMP, instead, we observe robustness of the IRII coefficients but changes in some of the NTR coefficients. In particular, when comparing estimations without and with the vector of MSTI variables (Table

11, columns (1)-(3) and Table 20, columns (1)-(9)), we see that the coefficients of NTR31 and NTR51 become positive and in some cases strongly significant, once we control for MSTI variables and (the availability of) *B_COMP*.

In the following tables with respect to each sector, the estimations are presented in the same order as in the tables discussed in Section 5.1 with the only difference being the inclusion of the MSTI variables. The results show that, controlling for R&D resources does not substantially affect the estimates of IRII and NTR for PHARM, across all models: including or not *originality* or the interaction terms with *patent scope*. The most significant effect is in Table 19 where the dependent variable is *breakthrough* where, compared to the estimations presented in Table 10, none of the coefficients of IRII11 are statistically significant, indicating that IRII is not a strong predictor of whether a PHARM patent is ranked among the top 1% most cited in the sector in a given cohort conditional on controlling for the level of R&D inputs associated with the applicants' place of residence.

For COMP, on the other hand, the notable change concerns NTR and it is evident across all the estimations using *forward citations* as a dependent variable. The interaction between NTR and *patent scope*, as shown in Table 22, has a significant and positive relationship with *forward citations* and sees significant and negative coefficients of NTR of all reference windows. The results in Table 24 are largely consistent with those in Table 14, which means the effects on *breakthrough*, instead, are not affected.

Overall, controlling for the vector of MSTI variables, we observe mixed robustness levels of our benchmarks estimations discussed in Section 5.1. The results are most sensitive to the inclusion of different sets of controls and sample size in the effects of the longer-reference windows of IRII in PHARM and of NTR in COMP.

With respect to the R&D input variables, we note that in PHARM with *forward citations* as the dependent variable, government intramural expenditure on R&D shows a significant and positive effect while the expenditures in the business enterprise and higher education segments are negatively related to *forward citations*, which point to a difference in the efficacy of the inputs. In addition, we note that more researchers' time, business enterprise research and development, and a higher sectoral export market share are all strongly likely to be related to an increase in the number of forward citations at the statistical significance level of 0.001; while the sectoral trade balance shows no statistical significance.

The parallel estimation of the effects on *forward citations* in the COMP sector differ. Among the non-sectoral variables, expenditures in the higher education segment plays a significantly positive role and the total researcher FTE appears to be negative; among the sectoral factors, adding to the same significant regressors in PHARM, the negative significance of trade balance indicates that having a trade deficit in a country's computer, electronic and optical market means a patent filed by at least one applicant in the country is more likely to receive more citations.

Turning to *breakthrough* as the dependent variable, expenditure in the business enterprise segment is negative with a weak significance in PHARM and a strong one in COMP. *B_PHARM* for the PHARM sector, and *GV_PPP* for

COMP, show a weak to moderate significant and positive relationship with the likelihood of being among the top 1% most cited patent.

Over all, the individualized MSTI variables demonstrate robust effects within each sector on patent influence in terms of the likelihood to be cited more. but some variables play different roles in the two sectors. As we are modeling the regression using weighted country-level data on individual patents, and the MSTI variables are included as controls, we are not going to further investigate how the national R&D investment conditions could affect innovation output. It is still worth-mentioning that these estimates again show that the two sectors under analysis have their unique characteristics, and that IRII remains significance with *breakthrough* in COMP where most of the MSTI variables appear to be insignificant, as shown in Table 24.

6 Discussion and Conclusion

Throughout all the analysis using the novelty indices we propose in this paper, from the descriptive statistics of IRII and NTR to the regression results, using the network visualization, the correlation between IRII and NTR and the conventional patent quality indicators, and estimations of two different dependent variables, we could confirm the framework of novelty origins in former works by Verhoeven et al. [2016]. The network-based approach allows us to identify naturally formed technology clusters and capture how they change over time, providing us with the tools to measure technological recombination more accurately. And through a thorough and in-depth quantitative analysis, we reveal rather meaningful complementary findings on the more complex dynamics of innovation.

Within the two sectors under analysis in this paper, recombining technological elements that have been used in inventions during the preceding period and bringing in elements not contained in patents filed in the reference time window both exhibit statistically significant relationship with a patent's future influence, measured in the number of forward citations or the eligibility to get into the elite top 1%. After clearing out the "noises" - factors interfering the analysis either due to changes in the data sample limited by the availability of certain variables, notably *originality* and the sectoral BERD, or due to the statistical distribution of *patent scope* - both proposed novelty indices demonstrate rather non-trivial roles in indicating patents' potential impact on innovation in the next five years, not only in the same sector but generally across all the technological fields, as *forward citations* is recorded regardless of the sector classification. Such roles are consistent with the additional consideration of the national policy and business environment over all as well as the expenditure and trade conditions specific to the sector.

The so-called noises are actually clues leading to rather relevant characteristics. Locating their relationship with IRII and NTR is not only important to isolate their effects on the analysis results, but also to understand the innovation pattern unique to each sector. It prompts us to remind the readers that comparison between different sectors and interpretation of results must be done in the context of sector-specific processes of new technology accumulation and creation. Such sectoral differences could be due to a mixture of factors, including but not limited to the nature of scientific disciplines, stage of development, the required inputs and lengths of the innovation life cycles, regulatory environment, market dynamics, the sector classification based on which our data samples have been constructed, and the practices in patenting process. For example, research and development in pharmaceuticals typically involves a lengthy and rigorous process that follows a sequential path, while facing high regulatory criteria due to safety and efficacy concerns [DiMasi et al., 1991, 2003, Grabowski et al., 2002, Scherer, 2010, Pammolli et al., 2011]. As a result of the high costs and high risks, PHARM patents tend to involve significant health implications through new compounds, novel delivery mechanisms, or groundbreaking treatments. In contrast, as discussed in literature [Banu Goktan and Miles, 2011] ***(this needs to be refined, we should echo in the Discussion about the relationship between incremental v.s. radical innovation)***, innovation cycles in computer technology are usually shorter. They range from incremental changes like software and hardware upgrades at a more regular basis to highly disruptive technologies, such as the development of solid state drive. The investment and risks associated also vary accordingly. While we are by no means qualified to make insightful comments as experts working in any of the sectors, the method and indices we propose allow us to detect features and changes in technological innovation in a more generic and relatively low-cost manner.

The ex-ante significance demonstrated through the regressional analysis could serve as guiding signals of potentially promising technological trends. From there, targeted studies can be carried out and stakeholders will be able to seize opportunities and/or identify risks and gaps early on to facilitate technological advances. Indeed, that's the purpose of one of the two important aspects of technology governance discussed in [Greely, 2022]: the so-called "Horizon Scanning", the process of identifying and analyzing emerging technologies with great potential.

(Need to associate this with the literature)The effects of country-specific variables we report could provide support for national policy making on research, education and innovation. These issues are not fully explored in this paper but present promising dimensions for future research.

Figures

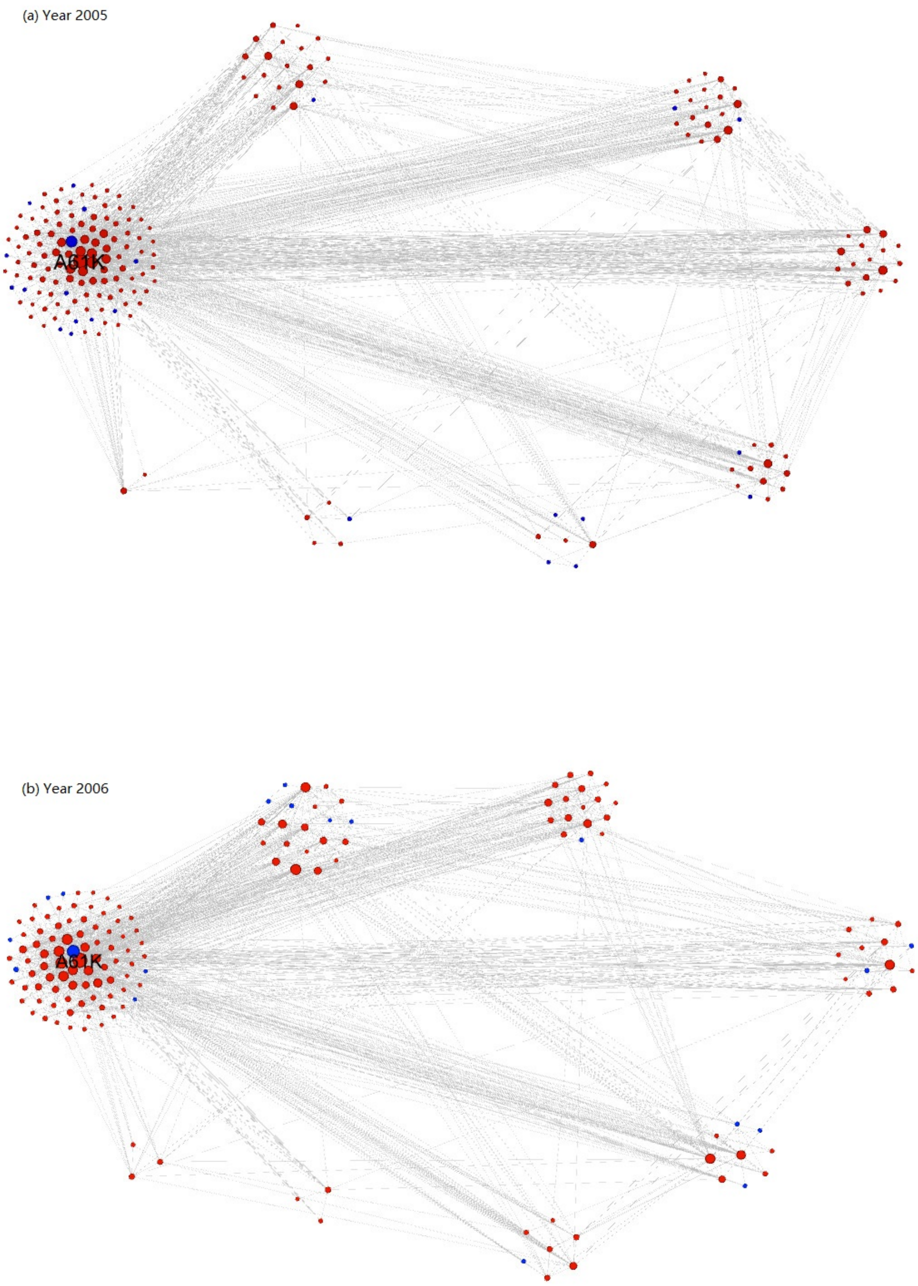


Figure 1: 8-cluster PHARM network partition highlighting subclasses containing new subgroups with 1-year reference window

Nodes in red color represent the subclasses containing new subgroups that are not found in patents of the PHARM sector in the previous year, and the the blue nodes do not contain new subgroups. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Labels are shown for nodes with degree above the median value. The edge lengths are not indicative of the connection strength.

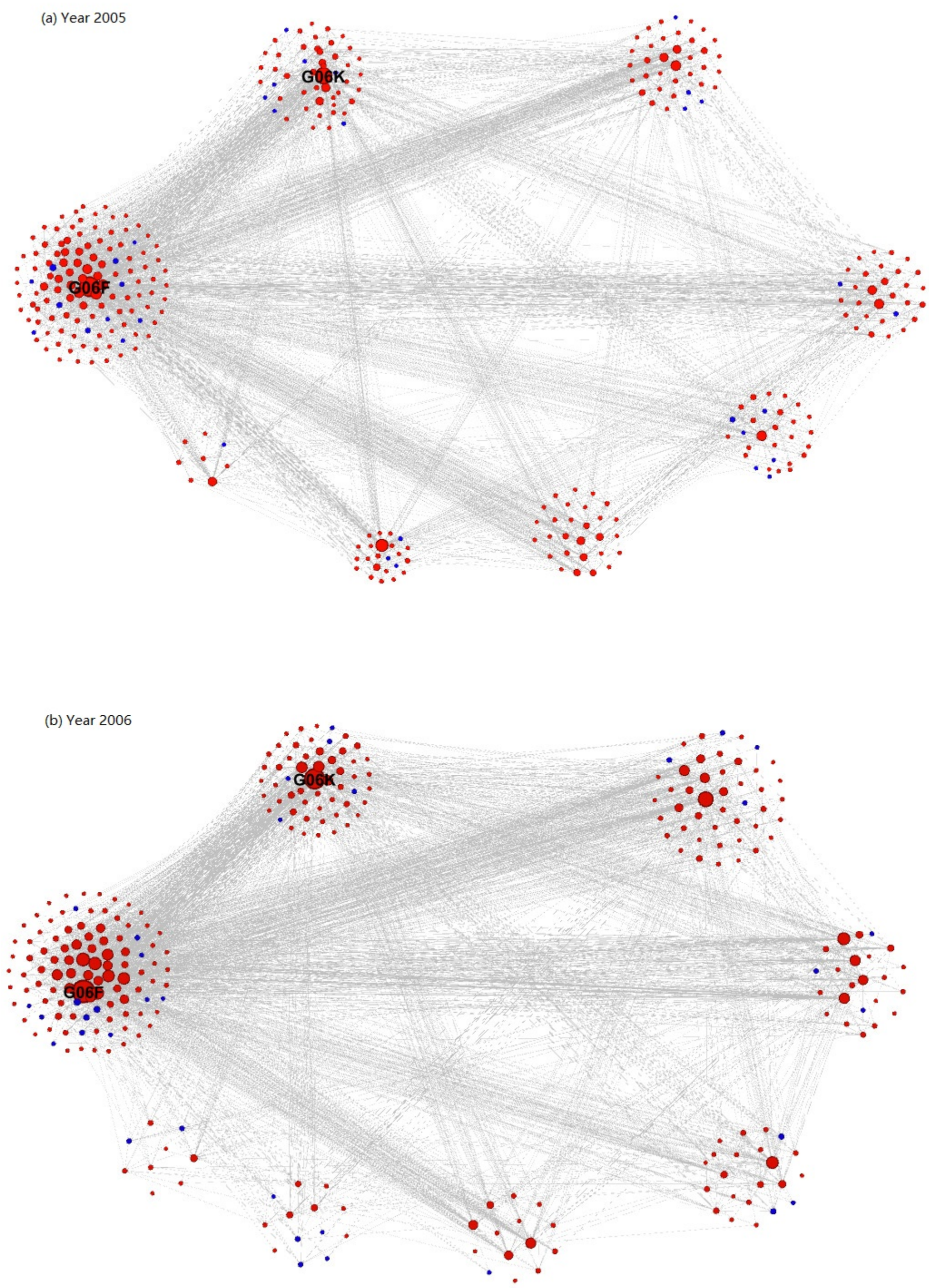


Figure 2: 8-cluster COMP network partition highlighting subclasses containing new subgroups with 1-year reference window

Nodes in red color represent the subclasses containing new subgroups that are not found in patents of the COMP sector in the previous year, and the the blue nodes do not contain new subgroups. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Labels are shown for nodes with degree above the median value. The edge lengths are not indicative of the connection strength.

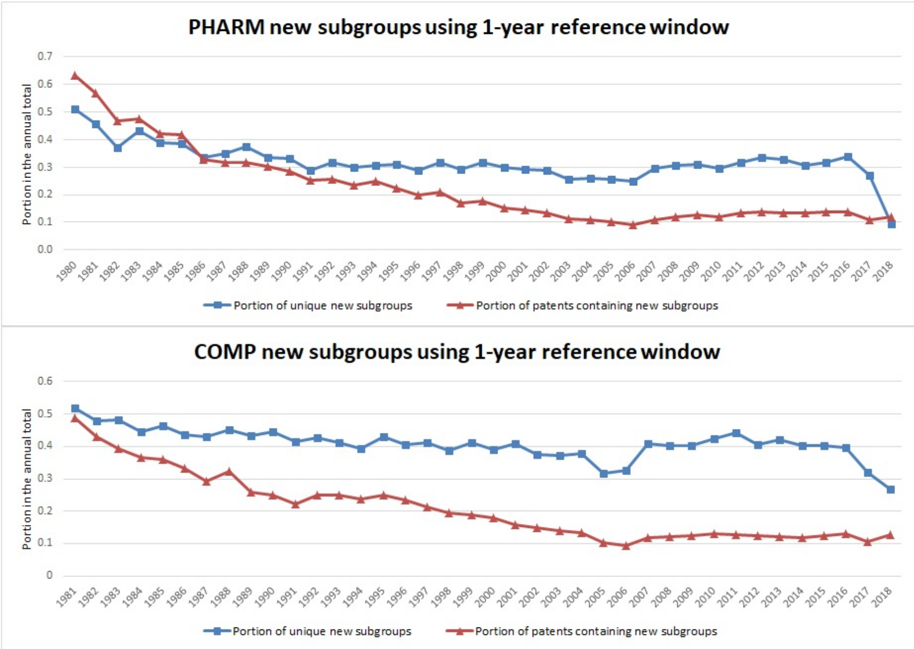


Figure 3: Subgroups portion in annual total quantities with one-year reference window

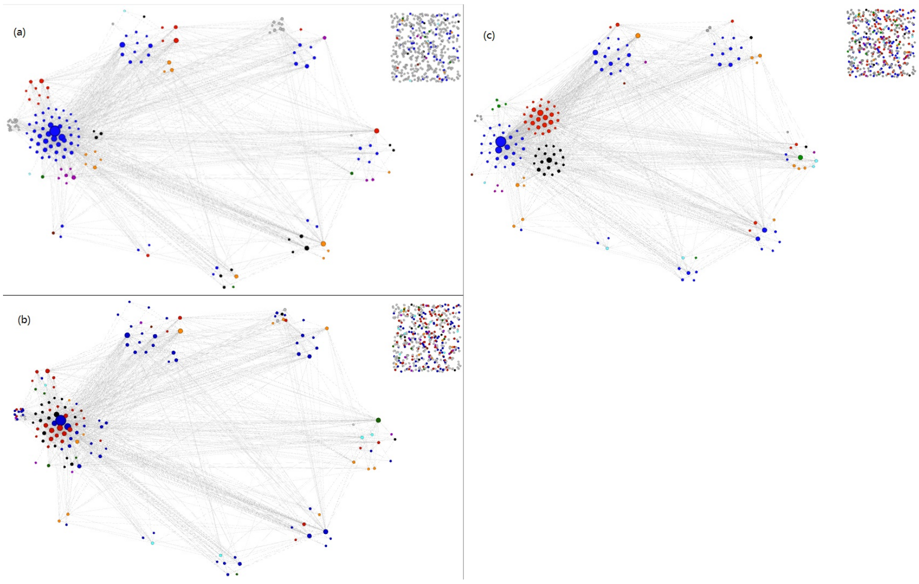


Figure 4: 8-cluster PHARM network partition of 2006 showing recombination in reference to the previous 1-year and 5-year windows

Each panel of the figures shows the partition with eight clusters generated from the network constructed using he cohort of PHARM patents filed in 2006, plus a square cluster of nodes at the upper-right corner that represents unconnected nodes and nodes (subclasses) not used in the patent cohort. In Panel (a), each cluster is further divided into eight sub-clusters, each with a distinct color representing the network partition generated using the cohot of PHARM patents filed in 2005. Panel (b) is in the same layout as Panel (a), but the color palette represents the network partition of the patent cohort in the 5-year reference window, 2001-2005. Nodes in Panel (c) have the same color representation as Panel (b), but with the sub-clusters visually grouped.

The color palette follows the cluster sizes in each partitioning: light grey for unconnected or unused nodes, blue for nodes in the largest cluster, red for nodes in the second largest cluster, black the third, and yellow, purple, green, light blue, and brown. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Node labels are omitted for visual clearance. The edge lengths are not indicative of the connection strength.

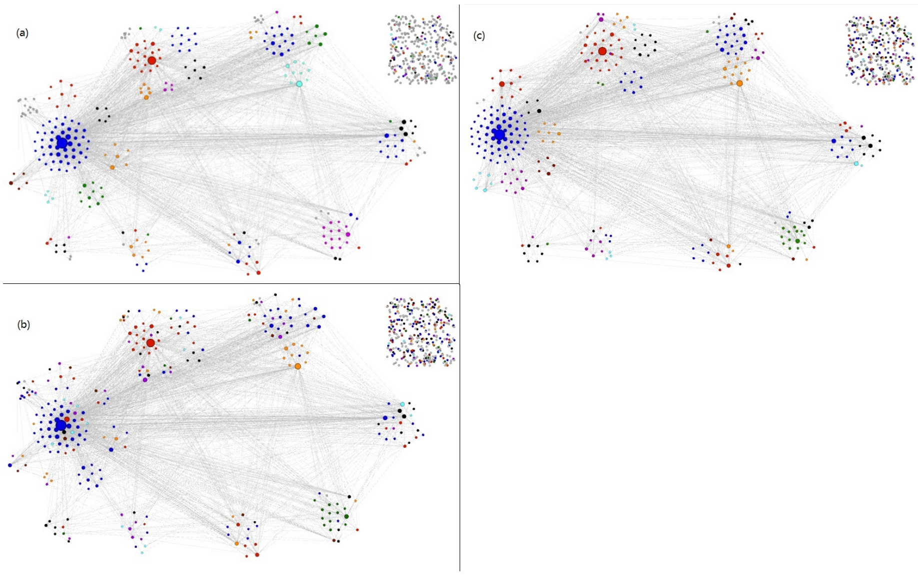


Figure 5: 8-cluster COMP network partition of 2006 showing recombination in reference to the previous 1-year and 5-year windows

Each panel of the figures shows the partition with eight clusters generated from the network constructed using he cohort of COMP patents filed in 2006, plus a square cluster of nodes at the upper-right corner that represents unconnected nodes and nodes (subclasses) not used in the patent cohort. In Panel (a), each cluster is further divided into eight sub-clusters, each with a distinct color representing the network partition generated using the cohort of COMP patents filed in 2005. Panel (b) is in the same layout as Panel (a), but the color palette represents the network partition of the patent cohort in the 5-year reference window, 2001-2005. Nodes in Panel (c) have the same color representation as Panel (b), but with the sub-clusters visually grouped.

The color palette follows the cluster sizes in each partitioning: light grey for unconnected or unused nodes, blue for nodes in the largest cluster, red for nodes in the second largest cluster, black the third, and yellow, purple, green, light blue, and brown. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Node labels are omitted for visual clearance. The edge lengths are not indicative of the connection strength.

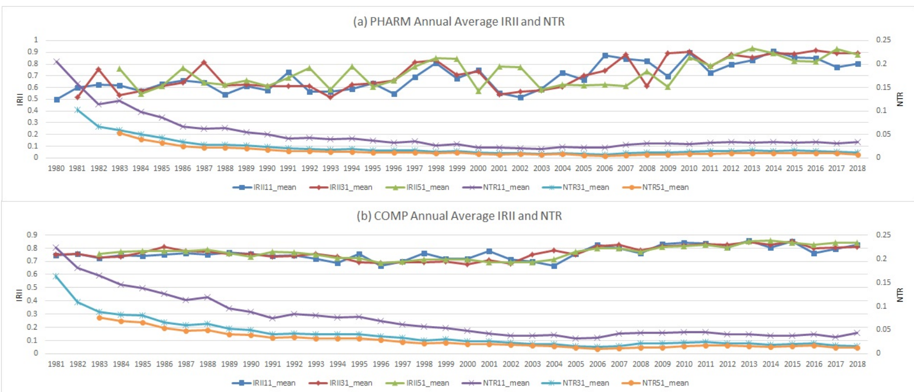
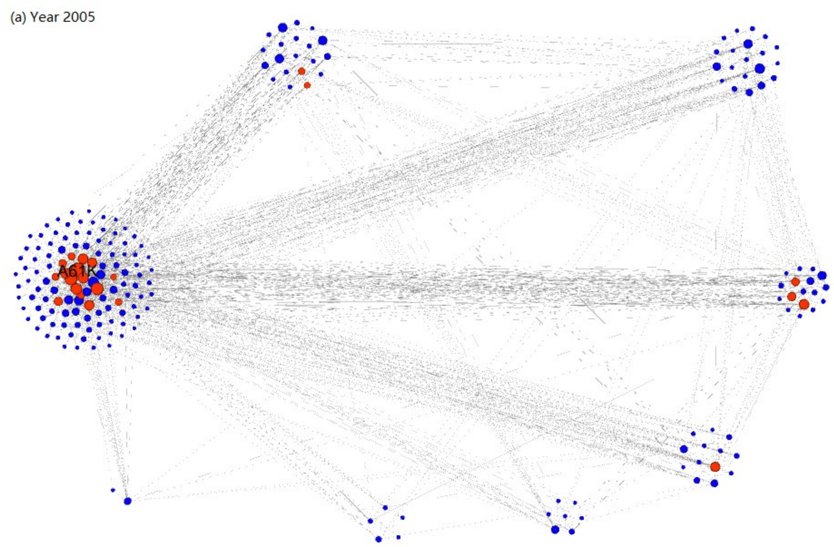


Figure 6: Annual sectoral average IRII and NTR with different reference windows



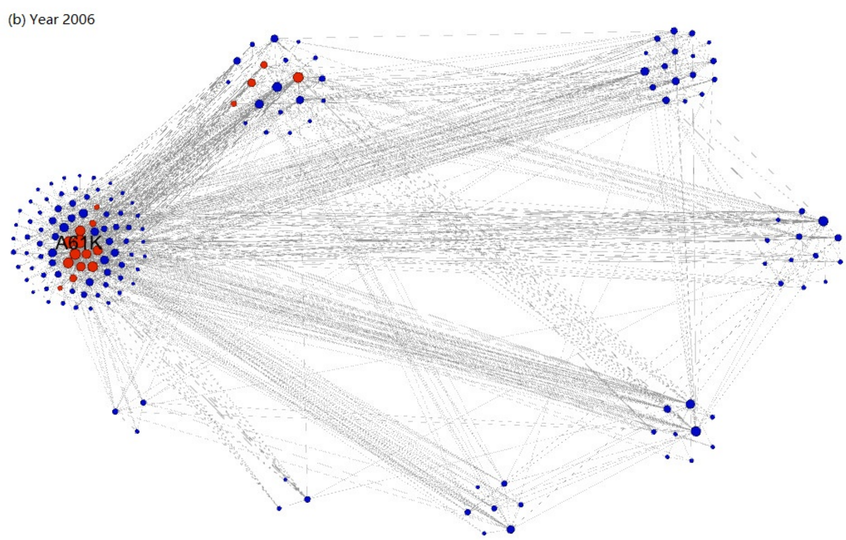
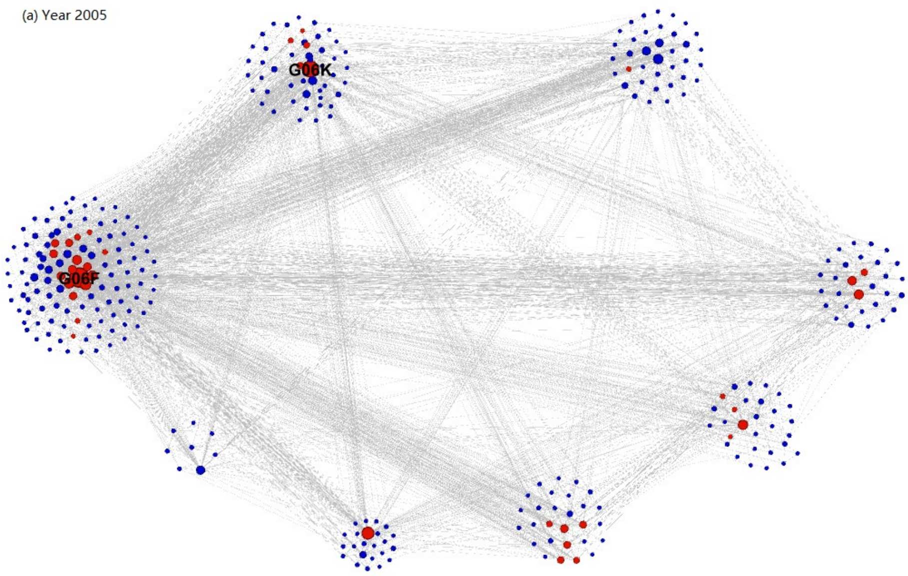


Figure 7: 8-cluster PHARM network partition highlighting subclasses of Breakthrough patents

Nodes in red color represent the subclasses assigned to Breakthrough PHARM patents of the years, and the blue nodes are not assigned to Breakthrough patents. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Labels are shown for nodes with degree above the median value. The edge lengths are not indicative of the connection strength.



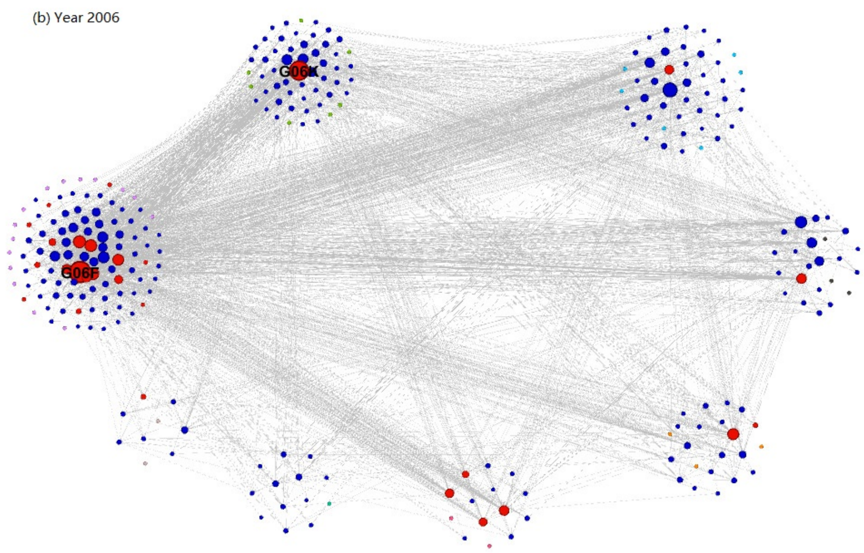


Figure 8: 8-cluster COMP network partition highlighting subclasses of Breakthrough patents

Nodes in red color represent the subclasses assigned to Breakthrough COMP patents of the years, and the blue nodes are not assigned to Breakthrough patents. Node size is proportional to the node degree in the network, i.e. the number of connections to other nodes. Labels are shown for nodes with degree above the median value. The edge lengths are not indicative of the connection strength.

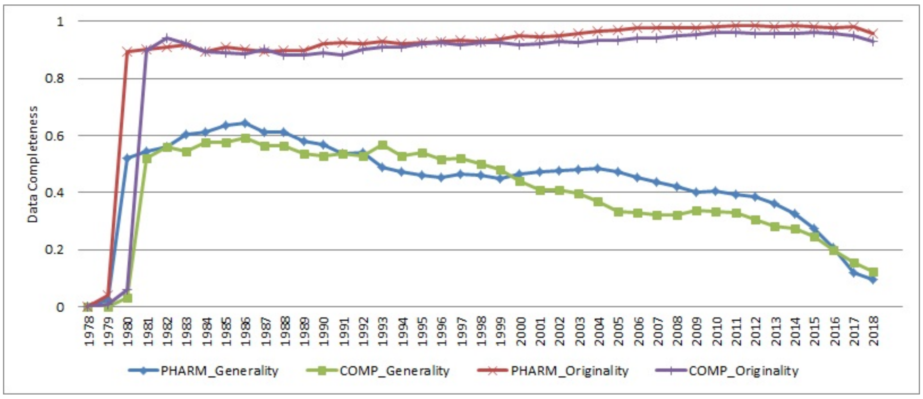


Figure 9: Originality and Generality data completeness of each sector

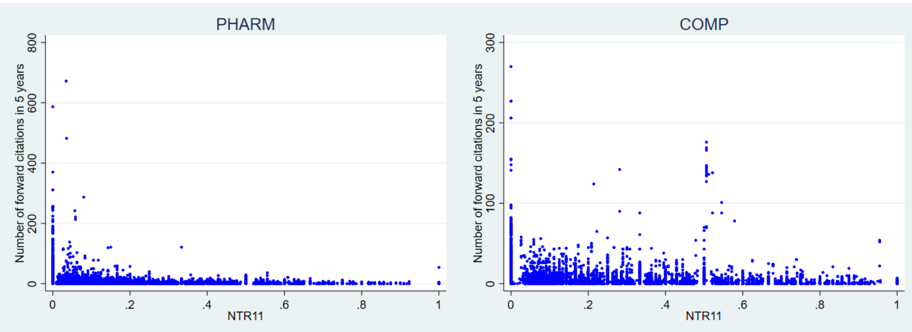


Figure 10: Relationship between forward citation number and NTR of each sector

Tables

References

Gautam Ahuja and Curba Morris Lampert. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, 22(6-7):521–543, 2001.

Sam Arts and Reinhilde Veugelers. Technology familiarity, recombinant novelty,

Table 1: Patent-level summary statistics by sector

(a) PHARM

	whole sample					patents with originality data				
Variable	Obs	Mean	Std. dev.	Min	Max	Obs	Mean	Std. dev.	Min	Max
IRII11	278,990	0.7171	0.2314	0.0278	1	266,616	0.7202	0.2318	0.0278	1
IRII31	277,922	0.7295	0.2298	0.0556	1	265,646	0.7322	0.2299	0.0556	1
IRII51	275,003	0.7272	0.2341	0.0625	1	262,979	0.7285	0.2346	0.0625	1
NTR11	278,990	0.0347	0.1000	0	1	266,616	0.0347	0.1001	0	1
NTR31	277,922	0.0156	0.0653	0	1	265,646	0.0156	0.0654	0	1
NTR51	275,003	0.0105	0.0532	0	1	262,979	0.0106	0.0533	0	1
patent scope	278,990	3.0289	1.4706	0	21	266,616	3.0046	1.4491	0	21
family size	278,990	9.8403	7.6757	1	57	266,616	9.9168	7.7261	1	57
backward citations	278,990	7.8111	20.9826	0	1013	266,616	8.1729	21.3951	1	1013
forward citations	278,990	1.3189	4.9999	0	672	266,616	1.3318	5.0693	0	672
generality	122,563	0.5056	0.2247	0	0.9388	117,834	0.5036	0.2245	0	0.9388
breakthrough	278,990	0.0034	0.0586	0	1	266,616	0.0035	0.0586	0	1
originality	266,616	0.7955	0.1626	0	0.9863	266,616	0.7955	0.1626	0	0.9863

(b) COMP

	whole sample					patents with originality data				
Variable	Obs	Mean	Std. dev.	Min	Max	Obs	Mean	Std. dev.	Min	Max
IRII11	282,506	0.7726	0.2540	0	1	264,878	0.7694	0.2546	0	1
IRII31	282,506	0.7769	0.2490	0.0278	1	264,878	0.7741	0.2494	0.0278	1
IRII51	280,401	0.7763	0.2475	0.0400	1	262,896	0.7732	0.2479	0.0400	1
NTR11	282,506	0.0502	0.1393	0	1	264,878	0.0514	0.1405	0	1
NTR31	282,506	0.0263	0.0983	0	1	264,878	0.0269	0.0992	0	1
NTR51	280,401	0.0195	0.0834	0	1	262,896	0.0200	0.0841	0	1
patent scope	282,506	2.0965	1.2761	0	30	264,878	2.1191	1.2872	0	30
family size	282,506	4.7957	2.8728	1	45	264,878	4.8235	2.8842	1	45
backward citations	282,506	4.5430	6.5689	0	498	264,878	4.8451	6.6753	1	498
forward citations	282,506	0.9442	3.1078	0	270	264,878	0.9591	3.1735	0	270
generality	103,344	0.3529	0.2803	0	0.9378	97,544	0.3552	0.2802	0	0.9378
breakthrough	282,506	0.0041	0.0639	0	1	264,878	0.0042	0.0646	0	1
originality	264,878	0.6798	0.2272	0	0.9823	264,878	0.6798	0.2272	0	0.9823

Table 2: List of variables used in analysis

(a) Patent-Level Quality Indicators		
Variable	Definition	Type
Patent scope	The number of distinct 4-digit IPC subclasses assigned to the patent	Integer
Family size	The number of patent offices operating in different jurisdictions at which a given invention has been protected	Integer
Backward citation	The number of citations of prior art listed on a patent applications as a source of knowledge in the development of the invention	Integer
Forward citation	The number of citations a patent receives within five years after the publication date	Integer
Breakthrough	A binary variable which equals 1 for patents in the the top 1% by the number of forward citations among those filed in the same year within the next 5 years; and 0 otherwise.	Binary (0/1)
Originality	A measure of knowledge diversification in the development of a patent based on the range of subclasses included in the backward citations of the patent application	decimal (0-1)
Generality	The ex-post counterpart of Originality, by using forward citations to capture the scope and degree of general-purpose technology that a patent enables	decimal (0-1)

(b) Country-Level MSTI Variables		
Variable	Definition	Unit
B_COMP	BERD (Business enterprise Expenditure on R&D) performed in the computer, electronic and optical industry (current PPP $)	USD $MM
B_PHARM	BERD performed in the pharmaceutical industry (current PPP $)	USD $MM
TD_BCOMP	Trade Balance: Computer, electronic and optical industry (current prices)	USD $MM
TD_BPHARM	Trade Balance: Pharmaceutical industry (current prices)	USD $MM
TD_XCOMP	Export market share: Computer, electronic and optical industry	%
TD_XPHARM	Export market share: Pharmaceutical industry	%
B_PPP	Business Enterprise Expenditure on R&D (BERD) at current PPP $	USD $MM
GV_PPP	Government Intramural Expenditure on R&D (GOVERD) at current PPP $	USD $MM
H_PPP	Higher Education Expenditure on R&D (HERD) at current PPP $	USD $MM
TP_RS	Total researchers (FTE)	FTE

Table 3: PHARM patent-level IRII and NTR correlations with quality indicators

(a) (1980-2018)

N=275,003	IRII11	IRII31	IRII51	NTR11	NTR31	NTR51	patent scope	family size	backward citations	forward citations
IRII11	1.0000									
IRII31	0.7032	1.0000								
IRII51	0.5136	0.5449	1.0000							
NTR11	-0.1992	-0.1731	-0.1637	1.0000						
NTR31	-0.1841	-0.1590	-0.1475	0.7702	1.0000					
NTR51	-0.1767	-0.1524	-0.1387	0.6866	0.8958	1.0000				
patent scope	*-0.4982*	*-0.4475*	*-0.3606*	*0.1861*	*0.1630*	*0.1514*	1.0000			
family size	*-0.0197*	*-0.0505*	*-0.0247*	*-0.0392*	*-0.0336*	*-0.0297*	0.0731	1.0000		
backward citations	*0.0424*	*0.0540*	*0.0143*	*-0.0178*	*-0.0129*	*-0.0132*	-0.0256	0.0588	1.0000	
forward citations	*-0.0619*	*-0.0630*	*-0.0443*	*-0.0064*	*-0.0051*	*-0.0013*	0.0977	0.1530	0.1044	1.0000

(b) (1980-2014)

	IRII11	IRII31	IRII51	NTR11	NTR31	NTR51	patent scope	family size	backward citations	forward citations	originality	generality
IRII11	1.0000											
IRII31	0.7072	1.0000										
IRII51	0.4936	0.5401	1.0000									
NTR11	-0.2077	-0.1750	-0.1653	1.0000								
NTR31	-0.1870	-0.1603	-0.1478	0.7780	1.0000							
NTR51	-0.1765	-0.1493	-0.1386	0.6902	0.8981	1.0000						
patent scope	*-0.4986*	*-0.4485*	*-0.3492*	*0.1809*	*0.1586*	*0.1496*	1.0000					
family size	*-0.0135*	*-0.0390*	*-0.0107*	*-0.0357*	*-0.0331*	*-0.0290*	0.0671	1.0000				
backward citations	*0.0478*	*0.0494*	*0.0154*	*-0.0196*	*-0.0143*	*-0.0129*	-0.0244	0.0621	1.0000			
forward citations	*-0.0550*	*-0.0519*	*-0.0342*	*-0.0061*	*-0.0050*	*-0.0015*	0.0943	0.1508	0.1128	1.0000		
originality	*-0.0529*	*-0.0711*	*-0.0722*	*-0.0213*	*-0.0226*	*-0.0191*	0.1407	0.0722	0.1461	0.0468	1.0000	
generality	*-0.1939*	*-0.1795*	*-0.1514*	*0.0272*	*0.0107*	*0.0060*	0.3294	0.0732	0.0257	0.1414	0.1405	1.0000

Table 4: COMP patent-level IRII and NTR correlations with quality indicators

(a) (1981-2018)

N=280,401	IRII11	IRII31	IRII51	NTR11	NTR31	NTR51	patent scope	family size	backward citations	forward citations
IRII11	1.0000									
IRII31	0.8003	1.0000								
IRII51	0.7903	0.8516	1.0000							
NTR11	-0.2857	-0.2543	-0.2400	1.0000						
NTR31	-0.2463	-0.2235	-0.2089	0.7852	1.0000					
NTR51	-0.2319	-0.2103	-0.1978	0.7107	0.9071	1.0000				
patent scope	*-0.6721*	*-0.6511*	*-0.6564*	*0.2977*	*0.2606*	*0.2467*	1.0000			
family size	*-0.1219*	*-0.1189*	*-0.1260*	*0.0749*	*0.0701*	*0.0704*	0.1835	1.0000		
backward citations	*-0.0624*	*-0.0545*	*-0.0573*	*0.0399*	*0.0335*	*0.0307*	0.0821	0.0330	1.0000	
forward citations	*-0.0931*	*-0.0957*	*-0.0982*	*0.0483*	*0.0443*	*0.0467*	0.1590	0.1805	0.0589	1.0000

(b) (1981-2014)

	IRII11	IRII31	IRII51	NTR11	NTR31	NTR51	patent scope	family size	backward citations	forward citations	originality	generality
IRII11	1.0000											
IRII31	0.7984	1.0000										
IRII51	0.7968	0.8510	1.0000									
NTR11	-0.2877	-0.2534	-0.2361	1.0000								
NTR31	-0.2485	-0.2237	-0.2047	0.7902	1.0000							
NTR51	-0.2319	-0.2091	-0.1932	0.7124	0.9077	1.0000						
patent scope	*-0.6735*	*-0.6527*	*-0.6621*	*0.2963*	*0.2605*	*0.2473*	1.0000					
family size	*-0.1265*	*-0.1214*	*-0.1261*	*0.0753*	*0.0710*	*0.0727*	0.1988	1.0000				
backward citations	*-0.1005*	*-0.0907*	*-0.0952*	*0.0624*	*0.0541*	*0.0500*	0.1264	0.0665	1.0000			
forward citations	*-0.0951*	*-0.0959*	*-0.0964*	*0.0489*	*0.0446*	*0.0482*	0.1689	0.1824	0.0548	1.0000		
originality	*-0.2182*	*-0.2123*	*-0.2241*	*0.0692*	*0.0612*	*0.0572*	0.2577	0.0509	0.2824	0.0374	1.0000	
generality	*-0.2744*	*-0.2791*	*-0.2736*	*0.1142*	*0.0943*	*0.0864*	0.3241	0.0836	0.0575	0.1835	0.2202	1.0000

Table 5: Summary statistics of weighted MSTI variables at patent level for each sector

PHARM (1981-2018)					
Variable	Sample size	Mean	Std. dev.	Min	Max
GV_PPP	259,786	19362.5200	18181.0000	16.5214	84124.8000
B_PPP	256,906	114276.6000	110427.3000	10.6821	429134.4000
H_PPP	259,428	21850.8700	20196.3200	0.9419	74722.0000
TP_RS	246,689	616141.3000	452805.4000	708.2000	1866109.0000
B_PHARM	211,376	16235.6500	19105.7300	0.2389	66202.0000
TD_BPHARM	270,283	-3210.1860	14102.5800	-67899.7300	47548.0700
TD_XPHARM	270,239	8.2620	4.5362	0.0010	19.6955

COMP (1981-2018)					
Variable	Sample size	Mean	Std. dev.	Min	Max
GV_PPP	275,046	21432.7100	19334.5700	16.5214	84124.8000
B_PPP	274,203	129885.9000	115365.9000	24.0575	429134.4000
H_PPP	274,913	24064.1600	20798.5800	0.9419	74722.0000
TP_RS	266,437	681733.3000	449777.6000	815.1000	1866109.0000
B_COMP	207,172	33336.0400	24746.7300	0.5354	78575.0000
TD_BCOMP	278,922	-24781.9600	77526.7600	-212256.7000	186830.1000
TD_XCOMP	278,847	9.4136	7.3226	0.0001	31.2690

Table 6: Total number of patent filings and Breakthrough patents of OECD countries, ranked by number of patent filings

PHARM				COMP			
	Patent No.	Breakthrough No.	Breakthrough%		Patent No.	Breakthrough No.	Breakthrough%
US	111,930	478	0.427%	US	109,172	702	0.643%
DE	32,552	91	0.280%	JP	56,127	181	0.322%
JP	26,420	108	0.409%	DE	24,368	22	0.090%
FR	20,067	38	0.189%	FR	16,414	32	0.195%
GB	17,373	44	0.253%	KR	13,773	53	0.385%
CH	14,518	46	0.317%	NL	11,630	24	0.206%
NL	7,189	34	0.473%	GB	8,657	20	0.231%
IT	6,774	24	0.354%	CN	7,771	4	0.051%
CA	5,531	27	0.488%	CA	5,563	14	0.252%
SE	5,143	8	0.156%	SE	4,764	10	0.210%
BE	4,040	16	0.396%	FI	4,544	48	1.056%
DK	3,673	28	0.762%	CH	4,338	11	0.254%
IL	3,669	3	0.082%	TW	2,842	3	0.106%
KR	3,465	2	0.058%	IT	2,681	1	0.037%
ES	3,182	5	0.157%	IL	2,466	20	0.811%
AU	3,112	11	0.353%	AU	1,510	3	0.199%
CN	3,038	1	0.033%	BE	1,363	6	0.440%
IN	2,717	3	0.110%	IE	1,043	0	0.000%
AT	2,671	11	0.412%	AT	915	1	0.109%
IE	1,550	2	0.129%	SG	812	0	0.000%
NO	1,048	3	0.286%	IN	751	0	0.000%
TW	922	0	0.000%	DK	750	3	0.400%
FI	896	0	0.000%	ES	693	7	1.010%
LU	770	11	1.429%	NO	520	1	0.192%
HU	744	0	0.000%	RU	311	0	0.000%
RU	582	6	1.031%	LU	294	0	0.000%
SG	547	0	0.000%	TR	259	0	0.000%
TR	538	0	0.000%	HU	93	0	0.000%
SI	451	0	0.000%	PT	78	0	0.000%
CZ	311	0	0.000%	CZ	53	0	0.000%
PT	238	1	0.420%	SI	40	0	0.000%

Notes: ISO 3166 alpha-2 country codes are used in the table. The full definition can be accessed at: `https://www.iso.org/iso-3166-country-codes.html`.

Table 7: PHARM Poisson regression with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	278,990	277,922	275,003	266,616	265,646	262,979	266,616	265,646	262,979
chi2	194220.4737	193484.2034	191697.0658	189440.4553	188756.9870	187062.3299	192751.1737	192106.1557	190340.5336
IRII11	-0.2158*** (0.0100)			-0.2027*** (0.0102)			-0.1810*** (0.0102)		
NTR11	-0.7437*** (0.0196)			-0.7800*** (0.0200)			-0.7472*** (0.0200)		
IRII31		-0.0109 (0.0102)			0.0093 (0.0103)			0.0405*** (0.0104)	
NTR31		-0.9284*** (0.0300)			-0.9957*** (0.0308)			-0.9369*** (0.0308)	
IRII51			0.0175 (0.0093)			0.0386*** (0.0094)			0.0568*** (0.0095)
NTR51			-0.8315*** (0.0358)			-0.9084*** (0.0367)			-0.8395*** (0.0367)
patent scope	0.1468*** (0.0011)	0.1552*** (0.0011)	0.1552*** (0.0011)	0.1505*** (0.0011)	0.1592*** (0.0011)	0.1591*** (0.0011)	0.1403*** (0.0012)	0.1491*** (0.0011)	0.1485*** (0.0011)
family size	0.0491*** (0.0002)	0.0491*** (0.0002)	0.0492*** (0.0002)	0.0484*** (0.0002)	0.0483*** (0.0002)	0.0484*** (0.0002)	0.0475*** (0.0002)	0.0474*** (0.0002)	0.0476*** (0.0002)
backward citations	0.0043*** (0.0000)	0.0044*** (0.0000)	0.0044*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)
originality							0.6999*** (0.0127)	0.7080*** (0.0128)	0.7107*** (0.0130)
APPC_US	0.2489*** (0.0041)	0.2500*** (0.0042)	0.2476*** (0.0042)	0.2442*** (0.0042)	0.2451*** (0.0042)	0.2429*** (0.0042)	0.2369*** (0.0042)	0.2375*** (0.0042)	0.2350*** (0.0042)
APPC_DE	0.1089*** (0.0060)	0.1061*** (0.0060)	0.1038*** (0.0060)	0.1002*** (0.0061)	0.0970*** (0.0061)	0.0950*** (0.0061)	0.1014*** (0.0061)	0.0979*** (0.0061)	0.0965*** (0.0061)
APPC_JP	0.2711*** (0.0063)	0.2704*** (0.0064)	0.2690*** (0.0064)	0.2694*** (0.0065)	0.2686*** (0.0065)	0.2666*** (0.0065)	0.2568*** (0.0065)	0.2555*** (0.0065)	0.2533*** (0.0065)
APPC_FR	-0.0943*** (0.0077)	-0.0967*** (0.0078)	-0.0949*** (0.0078)	-0.0983*** (0.0079)	-0.1016*** (0.0079)	-0.0994*** (0.0080)	-0.0835*** (0.0079)	-0.0873*** (0.0079)	-0.0842*** (0.0080)
APPC_GB	-0.1254*** (0.0078)	-0.1297*** (0.0078)	-0.1391*** (0.0079)	-0.1415*** (0.0080)	-0.1461*** (0.0080)	-0.1554*** (0.0081)	-0.1448*** (0.0080)	-0.1494*** (0.0080)	-0.1585*** (0.0081)
cons	-0.7251*** (0.0281)	-0.7148*** (0.0226)	-0.5418*** (0.0191)	-0.7165*** (0.0293)	-0.7195*** (0.0235)	-0.5376*** (0.0197)	-1.1767*** (0.0305)	-1.1977*** (0.0252)	-1.0200*** (0.0217)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 8: PHARM Poisson regression including interaction between IRII and Patent Scope, with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	278,990	277,922	275,003	266,616	265,646	262,979	266,616	265,646	262,979
chi2	194221.1924	193527.1499	192222.5137	189444.6481	188820.1924	187717.4287	192754.0366	192124.6792	190856.9657
IRII11	-0.2267*** (0.0163)			-0.2293*** (0.0165)			-0.1588*** (0.0166)		
NTR11	-0.7414*** (0.0198)			-0.7744*** (0.0202)			-0.7518*** (0.0202)		
IRII31		-0.0976*** (0.0167)			-0.0970*** (0.0170)			-0.0176 (0.0171)	
NTR31		-0.9008*** (0.0303)			-0.9612*** (0.0311)			-0.9184*** (0.0311)	
IRII51			-0.2881*** (0.0164)			-0.3073*** (0.0167)			-0.2527*** (0.0168)
NTR51			-0.7014*** (0.0360)			-0.7589*** (0.0369)			-0.7082*** (0.0369)
patent scope	0.1448*** (0.0027)	0.1385*** (0.0028)	0.0935*** (0.0030)	0.1455*** (0.0027)	0.1388*** (0.0028)	0.0893*** (0.0030)	0.1444*** (0.0027)	0.1380*** (0.0028)	0.0864*** (0.0030)
IRII11×patent scope	0.0037 (0.0044)			0.0091* (0.0045)			-0.0076 (0.0045)		
IRII31×patent scope		0.0283*** (0.0043)			0.0349*** (0.0044)			0.0190*** (0.0044)	
IRII51×patent scope			0.0959*** (0.0042)			0.1090*** (0.0043)			0.0973*** (0.0043)
family size	0.0491*** (0.0002)	0.0490*** (0.0002)	0.0490*** (0.0002)	0.0484*** (0.0002)	0.0483*** (0.0002)	0.0482*** (0.0002)	0.0476*** (0.0002)	0.0474*** (0.0002)	0.0474*** (0.0002)
backward citations	0.0043*** (0.0000)	0.0044*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)
originality							0.7013*** (0.0128)	0.7045*** (0.0128)	0.6953*** (0.0130)
APPC_US	0.2488*** (0.0041)	0.2497*** (0.0042)	0.2446*** (0.0042)	0.2442*** (0.0042)	0.2449*** (0.0042)	0.2395*** (0.0042)	0.2370*** (0.0042)	0.2374*** (0.0042)	0.2322*** (0.0042)
APPC_DE	0.1089*** (0.0060)	0.1059*** (0.0060)	0.1029*** (0.0060)	0.1002*** (0.0061)	0.0968*** (0.0061)	0.0939*** (0.0061)	0.1015*** (0.0061)	0.0978*** (0.0061)	0.0955*** (0.0061)
APPC_JP	0.2709*** (0.0063)	0.2691*** (0.0064)	0.2676*** (0.0064)	0.2689*** (0.0065)	0.2668*** (0.0065)	0.2647*** (0.0065)	0.2572*** (0.0065)	0.2546*** (0.0065)	0.2519*** (0.0065)
APPC_FR	-0.0944*** (0.0077)	-0.0972*** (0.0078)	-0.0960*** (0.0078)	-0.0985*** (0.0079)	-0.1022*** (0.0079)	-0.1005*** (0.0080)	-0.0833*** (0.0079)	-0.0876*** (0.0079)	-0.0855*** (0.0080)
APPC_GB	-0.1255*** (0.0078)	-0.1310*** (0.0078)	-0.1432*** (0.0079)	-0.1418*** (0.0080)	-0.1476*** (0.0080)	-0.1596*** (0.0081)	-0.1445*** (0.0080)	-0.1502*** (0.0080)	-0.1623*** (0.0081)
cons	-0.7185*** (0.0291)	-0.6588*** (0.0242)	-0.3293*** (0.0213)	-0.7003*** (0.0303)	-0.6513*** (0.0251)	-0.2982*** (0.0219)	-1.1912*** (0.0317)	-1.1580*** (0.0268)	-0.7952*** (0.0239)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 9: PHARM Poisson regression including interaction between NTR and Patent Scope, with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	278,990	277,922	275,003	266,616	265,646	262,979	266,616	265,646	262,979
chi2	194786.3	193971.7	192093.2	190133.5	189332.4	187523.1	193372.6	192628.5	190758.2446
IRII11	-0.2091*** (0.0100)			-0.1938*** (0.0102)			-0.1732*** (0.0102)		
NTR11	0.0705 (0.0398)			0.1443*** (0.0409)			0.1310** (0.0410)		
IRII31		-0.0060 (0.0102)			0.0158 (0.0103)			0.0461*** (0.0104)	
NTR31		0.2522*** (0.0618)			0.3278*** (0.0637)			0.3274*** (0.0639)	
IRII51			0.0221* (0.0093)			0.0445*** (0.0094)			0.0622*** (0.0095)
NTR51			0.4365*** (0.0730)			0.5026*** (0.0753)			0.5081*** (0.0755)
patent scope	0.1577*** (0.0012)	0.1631*** (0.0011)	0.1616*** (0.0011)	0.1631*** (0.0012)	0.1681*** (0.0012)	0.1663*** (0.0011)	0.1523*** (0.0012)	0.1576*** (0.0012)	0.1555*** (0.0012)
NTR11×patent scope	-0.1907*** (0.0085)			-0.2165*** (0.0088)			-0.2057*** (0.0088)		
NTR31×patent scope		-0.2613*** (0.0128)			-0.2926*** (0.0133)			-0.2796*** (0.0133)	
NTR51×patent scope			-0.2712*** (0.0146)			-0.3013*** (0.0152)			-0.2879*** (0.0152)
family size	0.0492*** (0.0002)	0.0491*** (0.0002)	0.0492*** (0.0002)	0.0484*** (0.0002)	0.0483*** (0.0002)	0.0484*** (0.0002)	0.0475*** (0.0002)	0.0475*** (0.0002)	0.0476*** (0.0002)
backward citations	0.0043*** (0.0000)	0.0044*** (0.0000)	0.0044*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)
originality							0.6919*** (0.0127)	0.7020*** (0.0128)	0.7057*** (0.0130)
APPC_US	0.2473*** (0.0041)	0.2488*** (0.0042)	0.2466*** (0.0042)	0.2424*** (0.0042)	0.2439*** (0.0042)	0.2417*** (0.0042)	0.2354*** (0.0042)	0.2364*** (0.0042)	0.2340*** (0.0042)
APPC_DE	0.1096*** (0.0060)	0.1064*** (0.0060)	0.1038*** (0.0060)	0.1010*** (0.0061)	0.0973*** (0.0061)	0.0949*** (0.0061)	0.1020*** (0.0061)	0.0982*** (0.0061)	0.0964*** (0.0061)
APPC_JP	0.2699*** (0.0063)	0.2689*** (0.0064)	0.2677*** (0.0064)	0.2678*** (0.0065)	0.2668*** (0.0065)	0.2650*** (0.0065)	0.2556*** (0.0065)	0.2540*** (0.0065)	0.2520*** (0.0065)
APPC_FR	-0.0940*** (0.0077)	-0.0966*** (0.0078)	-0.0946*** (0.0078)	-0.0979*** (0.0079)	-0.1015*** (0.0079)	-0.0992*** (0.0080)	-0.0833*** (0.0079)	-0.0873*** (0.0079)	-0.0841*** (0.0080)
APPC_GB	-0.1256*** (0.0078)	-0.1297*** (0.0078)	-0.1391*** (0.0079)	-0.1416*** (0.0080)	-0.1461*** (0.0080)	-0.1553*** (0.0081)	-0.1448*** (0.0080)	-0.1493*** (0.0080)	-0.1584*** (0.0081)
cons	-0.7737*** (0.0282)	-0.7554*** (0.0227)	-0.5733*** (0.0192)	-0.7737*** (0.0294)	-0.7654*** (0.0236)	-0.5727*** (0.0198)	-1.2252*** (0.0306)	-1.2365*** (0.0252)	-1.0496*** (0.0218)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 10: PHARM Probit regression with Breakthrough probability as dependent variable

	Dependent Variable: Breakthrough								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	278,671	277,603	274,684	266,310	265,340	262,673	266,310	265,340	262,673
chi2	1053.9714	1044.6836	1038.4280	1014.2670	1005.2716	1000.6533	1029.8242	1021.3654	1016.2295
IRII11	-0.2163** (0.0695)			-0.1830** (0.0708)			-0.1742* (0.0712)		
NTR11	-0.1509 (0.1233)			-0.1159 (0.1251)			-0.1008 (0.1252)		
IRII31		-0.0341 (0.0697)			-0.0061 (0.0711)			0.0084 (0.0716)	
NTR31		0.0036 (0.1751)			0.0107 (0.1788)			0.0389 (0.1788)	
IRII51			-0.0015 (0.0639)			0.0189 (0.0650)			0.0271 (0.0654)
NTR51			0.1932 (0.1962)			0.1989 (0.2000)			0.2301 (0.2000)
patent scope	0.0807*** (0.0075)	0.0875*** (0.0073)	0.0881*** (0.0072)	0.0813*** (0.0077)	0.0880*** (0.0075)	0.0884*** (0.0075)	0.0765*** (0.0078)	0.0832*** (0.0076)	0.0834*** (0.0076)
family size	0.0267*** (0.0011)	0.0266*** (0.0011)	0.0267*** (0.0011)	0.0264*** (0.0011)	0.0263*** (0.0011)	0.0263*** (0.0011)	0.0260*** (0.0011)	0.0259*** (0.0011)	0.0259*** (0.0011)
backward citations	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0026*** (0.0002)	0.0026*** (0.0002)	0.0027*** (0.0002)
originality							0.3473*** (0.0925)	0.3547*** (0.0930)	0.3522*** (0.0938)
APPC_US	0.0801** (0.0280)	0.0784** (0.0281)	0.0746** (0.0281)	0.0771** (0.0285)	0.0761** (0.0285)	0.0721* (0.0286)	0.0740** (0.0286)	0.0729* (0.0286)	0.0688* (0.0286)
APPC_DE	-0.0056 (0.0423)	-0.0059 (0.0423)	-0.0142 (0.0427)	-0.0253 (0.0436)	-0.0263 (0.0437)	-0.0315 (0.0440)	-0.0232 (0.0437)	-0.0245 (0.0437)	-0.0293 (0.0440)
APPC_JP	0.1790*** (0.0409)	0.1762*** (0.0410)	0.1745*** (0.0412)	0.1771*** (0.0417)	0.1736*** (0.0417)	0.1717*** (0.0419)	0.1722*** (0.0417)	0.1683*** (0.0418)	0.1664*** (0.0420)
APPC_FR	-0.1434* (0.0582)	-0.1448* (0.0583)	-0.1428* (0.0583)	-0.1572** (0.0600)	-0.1592** (0.0601)	-0.1567** (0.0601)	-0.1495* (0.0600)	-0.1517* (0.0601)	-0.1488* (0.0602)
APPC_GB	-0.1050 (0.0556)	-0.1051 (0.0556)	-0.1024 (0.0557)	-0.1367* (0.0584)	-0.1376* (0.0584)	-0.1350* (0.0585)	-0.1362* (0.0584)	-0.1371* (0.0584)	-0.1344* (0.0585)
cons	-3.3149*** (0.2024)	-3.3999*** (0.1685)	-3.2036*** (0.1273)	-3.4445*** (0.2416)	-3.4499*** (0.1843)	-3.1777*** (0.1290)	-3.6773*** (0.2505)	-3.6976*** (0.1973)	-3.4165*** (0.1446)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with an address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 11: COMP Poisson regression with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	282,506	282,506	280,401	264,878	264,878	262,896	264,878	264,878	262,896
chi2	139922.1502	140139.5781	139914.0625	134065.4100	134264.4570	134041.9757	134662.4517	134826.1557	134575.5770
IRII11	-0.2426*** (0.0091)			-0.2462*** (0.0093)			-0.2110*** (0.0094)		
NTR11	-0.1720*** (0.0133)			-0.1899*** (0.0136)			-0.1960*** (0.0136)		
IRII31		-0.2750*** (0.0090)			-0.2775*** (0.0092)			-0.2434*** (0.0093)	
NTR31		-0.1999*** (0.0177)			-0.2201*** (0.0181)			-0.2281*** (0.0181)	
IRII51			-0.2956*** (0.0091)			-0.2952*** (0.0093)			-0.2610*** (0.0094)
NTR51			-0.1299*** (0.0199)			-0.1510*** (0.0204)			-0.1588*** (0.0204)
patent scope	0.1419*** (0.0010)	0.1410*** (0.0010)	0.1386*** (0.0010)	0.1397*** (0.0010)	0.1388*** (0.0010)	0.1366*** (0.0010)	0.1367*** (0.0011)	0.1357*** (0.0010)	0.1336*** (0.0011)
family size	0.0846*** (0.0004)	0.0847*** (0.0004)	0.0847*** (0.0004)	0.0843*** (0.0004)	0.0844*** (0.0004)	0.0844*** (0.0004)	0.0839*** (0.0004)	0.0840*** (0.0004)	0.0840*** (0.0004)
backward citations	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)
originality							0.2407*** (0.0100)	0.2335*** (0.0100)	0.2290*** (0.0100)
APPC_US	0.3254*** (0.0054)	0.3265*** (0.0054)	0.3273*** (0.0054)	0.3236*** (0.0055)	0.3247*** (0.0055)	0.3257*** (0.0055)	0.3232*** (0.0055)	0.3241*** (0.0055)	0.3251*** (0.0055)
APPC_JP	0.2375*** (0.0063)	0.2368*** (0.0063)	0.2371*** (0.0063)	0.2330*** (0.0064)	0.2324*** (0.0064)	0.2328*** (0.0065)	0.2299*** (0.0064)	0.2295*** (0.0064)	0.2297*** (0.0065)
APPC_DE	-0.0877*** (0.0092)	-0.0854*** (0.0092)	-0.0830*** (0.0092)	-0.0804*** (0.0094)	-0.0780*** (0.0094)	-0.0748*** (0.0095)	-0.0802*** (0.0094)	-0.0782*** (0.0094)	-0.0753*** (0.0095)
APPC_FR	-0.1406*** (0.0106)	-0.1381*** (0.0106)	-0.1345*** (0.0106)	-0.1373*** (0.0109)	-0.1345*** (0.0109)	-0.1309*** (0.0109)	-0.1339*** (0.0109)	-0.1314*** (0.0109)	-0.1280*** (0.0109)
APPC_KR	0.4469*** (0.0100)	0.4508*** (0.0100)	0.4525*** (0.0100)	0.4572*** (0.0102)	0.4616*** (0.0102)	0.4632*** (0.0102)	0.4594*** (0.0102)	0.4634*** (0.0102)	0.4651*** (0.0102)
cons	-0.6244*** (0.0312)	-0.6043*** (0.0311)	-0.4716*** (0.0255)	-0.6101*** (0.0320)	-0.5917*** (0.0319)	-0.4697*** (0.0264)	-0.7788*** (0.0328)	-0.7545*** (0.0327)	-0.6248*** (0.0273)

Notes:
Sample of patents filed from 1981 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 12: COMP Poisson regression including interaction between IRII and Patent Scope, with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	282,506	282,506	280,401	264,878	264,878	262,896	264,878	264,878	262,896
chi2	140462.0767	140875.9223	140844.1938	134563.4587	134946.2554	134927.2748	135037.2731	135373.8563	135290.7146
IRII11	-0.4202*** (0.0120)			-0.4216*** (0.0123)			-0.3684*** (0.0125)		
NTR11	-0.1522*** (0.0133)			-0.1704*** (0.0136)			-0.1783*** (0.0136)		
IRII31		-0.5136*** (0.0128)			-0.5134*** (0.0131)			-0.4612*** (0.0133)	
NTR31		-0.1740*** (0.0177)			-0.1951*** (0.0181)			-0.2049*** (0.0181)	
IRII51			-0.5373*** (0.0123)			-0.5376*** (0.0126)			-0.4878*** (0.0129)
NTR51			-0.1193*** (0.0200)			-0.1412*** (0.0204)			-0.1484*** (0.0204)
patent scope	0.0977*** (0.0022)	0.0787*** (0.0026)	0.0769*** (0.0024)	0.0966*** (0.0022)	0.0779*** (0.0026)	0.0754*** (0.0024)	0.0993*** (0.0022)	0.0810*** (0.0026)	0.0783*** (0.0024)
IRII11×patent scope	0.1006*** (0.0043)			0.0984*** (0.0044)			0.0860*** (0.0045)		
IRII31×patent scope		0.1252*** (0.0046)			0.1227*** (0.0047)			0.1109*** (0.0048)	
IRII51×patent scope			0.1358*** (0.0045)			0.1350*** (0.0046)			0.1227*** (0.0046)
family size	0.0850*** (0.0004)	0.0848*** (0.0004)	0.0844*** (0.0004)	0.0847*** (0.0004)	0.0846*** (0.0004)	0.0841*** (0.0004)	0.0843*** (0.0004)	0.0842*** (0.0004)	0.0838*** (0.0004)
backward citations	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0080*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)
originality							0.2162*** (0.0101)	0.2051*** (0.0100)	0.1909*** (0.0101)
APPC_US	0.3242*** (0.0054)	0.3243*** (0.0054)	0.3239*** (0.0054)	0.3223*** (0.0055)	0.3224*** (0.0055)	0.3221*** (0.0055)	0.3221*** (0.0055)	0.3221*** (0.0055)	0.3219*** (0.0055)
APPC_JP	0.2311*** (0.0063)	0.2292*** (0.0063)	0.2275*** (0.0063)	0.2267*** (0.0064)	0.2249*** (0.0064)	0.2231*** (0.0065)	0.2247*** (0.0064)	0.2230*** (0.0064)	0.2215*** (0.0065)
APPC_DE	-0.0876*** (0.0092)	-0.0867*** (0.0092)	-0.0865*** (0.0092)	-0.0805*** (0.0094)	-0.0797*** (0.0094)	-0.0788*** (0.0094)	-0.0803*** (0.0094)	-0.0797*** (0.0094)	-0.0789*** (0.0094)
APPC_FR	-0.1316*** (0.0106)	-0.1281*** (0.0106)	-0.1256*** (0.0106)	-0.1286*** (0.0109)	-0.1249*** (0.0109)	-0.1224*** (0.0109)	-0.1267*** (0.0109)	-0.1232*** (0.0109)	-0.1208*** (0.0109)
APPC_KR	0.4374*** (0.0100)	0.4376*** (0.0100)	0.4369*** (0.0100)	0.4480*** (0.0102)	0.4487*** (0.0102)	0.4476*** (0.0102)	0.4511*** (0.0102)	0.4517*** (0.0102)	0.4507*** (0.0102)
cons	-0.5446*** (0.0314)	-0.4714*** (0.0315)	-0.3438*** (0.0259)	-0.5312*** (0.0322)	-0.4602*** (0.0323)	-0.3419*** (0.0268)	-0.6923*** (0.0331)	-0.6156*** (0.0332)	-0.4820*** (0.0278)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 13: COMP Poisson regression including interaction between NTR and Patent Scope, with forward citation number as dependent variable

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	282,506	282,506	280,401	264,878	264,878	262,896	264,878	264,878	262,896
chi2	140317.8353	140533.1343	140378.7013	134459.6406	134634.6913	134477.3963	135084.7706	135222.1788	135038.6022
IRII11	-0.2654*** (0.0092)			-0.2695*** (0.0094)			-0.2350*** (0.0096)		
NTR11	-0.5070*** (0.0215)			-0.5327*** (0.0220)			-0.5528*** (0.0221)		
IRII31		-0.2961*** (0.0091)			-0.2983*** (0.0093)			-0.2647*** (0.0094)	
NTR31		-0.6785*** (0.0300)			-0.6965*** (0.0307)			-0.7229*** (0.0308)	
IRII51			-0.3209*** (0.0092)			-0.3201*** (0.0094)			-0.2863*** (0.0096)
NTR51			-0.7140*** (0.0338)			-0.7317*** (0.0346)			-0.7595*** (0.0347)
patent scope	0.1295*** (0.0012)	0.1317*** (0.0011)	0.1290*** (0.0012)	0.1271*** (0.0013)	0.1296*** (0.0012)	0.1272*** (0.0012)	0.1233*** (0.0013)	0.1259*** (0.0012)	0.1236*** (0.0012)
NTR11×patent scope	0.0867*** (0.0042)			0.0882*** (0.0043)			0.0918*** (0.0043)		
NTR31×patent scope		0.1127*** (0.0054)			0.1118*** (0.0056)			0.1161*** (0.0056)	
NTR51×patent scope			0.1293*** (0.0057)			0.1281*** (0.0058)			0.1327*** (0.0059)
family size	0.0838*** (0.0004)	0.0839*** (0.0004)	0.0838*** (0.0004)	0.0835*** (0.0004)	0.0836*** (0.0004)	0.0835*** (0.0004)	0.0830*** (0.0004)	0.0831*** (0.0004)	0.0831*** (0.0004)
backward citations	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0081*** (0.0001)	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)
originality							0.2466*** (0.0100)	0.2391*** (0.0100)	0.2351*** (0.0101)
APPC_US	0.3244*** (0.0054)	0.3271*** (0.0054)	0.3284*** (0.0054)	0.3226*** (0.0055)	0.3253*** (0.0055)	0.3268*** (0.0055)	0.3220*** (0.0055)	0.3247*** (0.0055)	0.3262*** (0.0055)
APPC_JP	0.2404*** (0.0063)	0.2398*** (0.0063)	0.2404*** (0.0063)	0.2360*** (0.0064)	0.2353*** (0.0064)	0.2360*** (0.0065)	0.2328*** (0.0064)	0.2324*** (0.0064)	0.2329*** (0.0065)
APPC_DE	-0.0865*** (0.0092)	-0.0828*** (0.0092)	-0.0794*** (0.0092)	-0.0790*** (0.0094)	-0.0753*** (0.0094)	-0.0710*** (0.0095)	-0.0788*** (0.0094)	-0.0755*** (0.0094)	-0.0715*** (0.0095)
APPC_FR	-0.1365*** (0.0106)	-0.1330*** (0.0106)	-0.1290*** (0.0106)	-0.1329*** (0.0109)	-0.1293*** (0.0109)	-0.1253*** (0.0109)	-0.1293*** (0.0109)	-0.1260*** (0.0109)	-0.1222*** (0.0109)
APPC_KR	0.4488*** (0.0100)	0.4531*** (0.0100)	0.4550*** (0.0100)	0.4590*** (0.0102)	0.4638*** (0.0102)	0.4657*** (0.0102)	0.4612*** (0.0102)	0.4657*** (0.0102)	0.4676*** (0.0102)
cons	-0.5599*** (0.0314)	-0.5455*** (0.0313)	-0.4213*** (0.0256)	-0.5443*** (0.0322)	-0.5333*** (0.0320)	-0.4199*** (0.0265)	-0.7137*** (0.0329)	-0.6972*** (0.0328)	-0.5768*** (0.0274)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 14: COMP Probit regression with Breakthrough probability as dependent variable

	Dependent Variable: Breakthrough								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	282,506	282,506	280,401	264,878	264,878	262,896	264,878	264,878	262,896
chi2	1811.4562	1819.0067	1803.0888	1723.8036	1731.7056	1714.0385	1749.7856	1756.1848	1737.8805
IRII11	-0.1969*** (0.0515)			-0.1794*** (0.0527)			-0.1416** (0.0533)		
NTR11	-0.0226 (0.0702)			-0.0285 (0.0718)			-0.0349 (0.0720)		
IRII31		-0.2372*** (0.0509)			-0.2292*** (0.0520)			-0.1926*** (0.0526)	
NTR31		0.0052 (0.0905)			-0.0265 (0.0932)			-0.0335 (0.0934)	
IRII51			-0.2932*** (0.0522)			-0.2851*** (0.0534)			-0.2464*** (0.0541)
NTR51			-0.0092 (0.1049)			-0.0446 (0.1083)			-0.0518 (0.1086)
patent scope	0.1054*** (0.0061)	0.1039*** (0.0060)	0.1008*** (0.0061)	0.1048*** (0.0062)	0.1030*** (0.0061)	0.0999*** (0.0062)	0.1005*** (0.0063)	0.0985*** (0.0062)	0.0956*** (0.0063)
family size	0.0558*** (0.0023)	0.0559*** (0.0023)	0.0562*** (0.0023)	0.0560*** (0.0024)	0.0561*** (0.0024)	0.0564*** (0.0024)	0.0558*** (0.0024)	0.0559*** (0.0024)	0.0562*** (0.0024)
backward citations	0.0031*** (0.0007)	0.0032*** (0.0007)	0.0031*** (0.0007)	0.0031*** (0.0007)	0.0031*** (0.0007)	0.0031*** (0.0007)	0.0027*** (0.0008)	0.0027*** (0.0008)	0.0027*** (0.0008)
originality							0.3061*** (0.0624)	0.2974*** (0.0624)	0.2989*** (0.0636)
APPC_US	0.2525*** (0.0298)	0.2529*** (0.0299)	0.2385*** (0.0314)	0.2503*** (0.0305)	0.2507*** (0.0305)	0.2369*** (0.0320)	0.2502*** (0.0305)	0.2505*** (0.0305)	0.2369*** (0.0321)
APPC_JP	0.0669 (0.0369)	0.0658 (0.0369)	0.0443 (0.0384)	0.0711 (0.0376)	0.0693 (0.0376)	0.0488 (0.0391)	0.0676 (0.0376)	0.0662 (0.0376)	0.0454 (0.0391)
APPC_DE	-0.3299*** (0.0714)	-0.3278*** (0.0714)	-0.3557*** (0.0747)	-0.3297*** (0.0733)	-0.3273*** (0.0733)	-0.3552*** (0.0767)	-0.3311*** (0.0736)	-0.3292*** (0.0735)	-0.3576*** (0.0770)
APPC_FR	-0.1521* (0.0672)	-0.1502* (0.0672)	-0.1673* (0.0688)	-0.1447* (0.0687)	-0.1432* (0.0686)	-0.1594* (0.0703)	-0.1383* (0.0686)	-0.1371* (0.0685)	-0.1531* (0.0702)
APPC_KR	0.2117*** (0.0547)	0.2156*** (0.0547)	0.2014*** (0.0554)	0.2240*** (0.0555)	0.2277*** (0.0555)	0.2139*** (0.0562)	0.2262*** (0.0556)	0.2294*** (0.0556)	0.2159*** (0.0563)
cons	-2.7131*** (0.1223)	-2.6870*** (0.1213)	-2.9883*** (0.1418)	-2.7035*** (0.1235)	-2.6661*** (0.1225)	-2.9672*** (0.1432)	-2.9175*** (0.1315)	-2.8734*** (0.1305)	-3.1749*** (0.1507)

Notes:
Sample of patents filed from 1981 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 15: PHARM Poisson regression with forward citation number as dependent variable, MSTI variables included in estimation, without Originality

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with B_COMP data			patents with B_COMP data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	246,003	246,003	243,843	198,111	198,111	198,111	198,111	198,111	198,111
chi2	180260.7861	179679.9170	178251.1438	156692.3360	156210.8057	155903.3220	157881.0363	157389.2689	157089.0649
IRII11	-0.1757*** (0.0106)			-0.2244*** (0.0119)			-0.2258*** (0.0119)		
NTR11	-0.7852*** (0.0213)			-0.8515*** (0.0251)			-0.8553*** (0.0251)		
IRII31		0.0268* (0.0107)			-0.0544*** (0.0121)			-0.0555*** (0.0121)	
NTR31		-0.9565*** (0.0323)			-1.1477*** (0.0395)			-1.1497*** (0.0395)	
IRII51			0.0453*** (0.0098)			0.0811*** (0.0111)			0.0897*** (0.0111)
NTR51			-0.8396*** (0.0381)			-1.0155*** (0.0458)			-1.0111*** (0.0458)
patent scope	0.1550*** (0.0012)	0.1622*** (0.0011)	0.1616*** (0.0011)	0.1534*** (0.0013)	0.1597*** (0.0013)	0.1630*** (0.0012)	0.1551*** (0.0013)	0.1614*** (0.0013)	0.1651*** (0.0012)
family size	0.0490*** (0.0002)	0.0489*** (0.0002)	0.0490*** (0.0002)	0.0510*** (0.0002)	0.0508*** (0.0002)	0.0508*** (0.0002)	0.0510*** (0.0002)	0.0508*** (0.0002)	0.0508*** (0.0002)
backward citations	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0043*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0043*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)
GV_PPP	0.2769*** (0.0127)	0.2752*** (0.0127)	0.2769*** (0.0127)	0.3659*** (0.0154)	0.3646*** (0.0154)	0.3654*** (0.0154)	0.3111*** (0.0157)	0.3101*** (0.0156)	0.3105*** (0.0156)
B_PPP	-0.0262*** (0.0019)	-0.0260*** (0.0019)	-0.0258*** (0.0019)	-0.0467*** (0.0023)	-0.0462*** (0.0023)	-0.0460*** (0.0023)	-0.0801*** (0.0025)	-0.0794*** (0.0025)	-0.0793*** (0.0025)
H_PPP	0.0254** (0.0084)	0.0278*** (0.0084)	0.0290*** (0.0084)	0.0347*** (0.0094)	0.0363*** (0.0094)	0.0362*** (0.0094)	-0.1881*** (0.0116)	-0.1856*** (0.0115)	-0.1865*** (0.0115)
TP_RS	-0.0078*** (0.0004)	-0.0079*** (0.0004)	-0.0080*** (0.0004)	-0.0080*** (0.0005)	-0.0081*** (0.0005)	-0.0082*** (0.0005)	0.0019*** (0.0005)	0.0017** (0.0005)	0.0017** (0.0005)
B_PHARM							0.2717*** (0.0078)	0.2705*** (0.0078)	0.2715*** (0.0078)
TD_BPHARM	-0.0445*** (0.0043)	-0.0440*** (0.0043)	-0.0428*** (0.0044)	-0.0401*** (0.0051)	-0.0392*** (0.0051)	-0.0394*** (0.0051)	-0.0028 (0.0052)	-0.0021 (0.0052)	-0.0023 (0.0052)
TD_XPHARM	0.0101*** (0.0010)	0.0101*** (0.0010)	0.0102*** (0.0010)	0.0030* (0.0014)	0.0031* (0.0014)	0.0030* (0.0014)	0.0046*** (0.0014)	0.0047*** (0.0014)	0.0046*** (0.0014)
APPC_US	0.3163*** (0.0136)	0.3209*** (0.0136)	0.3186*** (0.0137)	0.4585*** (0.0175)	0.4631*** (0.0174)	0.4658*** (0.0174)	0.5510*** (0.0173)	0.5551*** (0.0173)	0.5577*** (0.0173)
APPC_DE	0.0542*** (0.0113)	0.0500*** (0.0113)	0.0476*** (0.0114)	0.1275*** (0.0135)	0.1238*** (0.0135)	0.1224*** (0.0134)	0.1344*** (0.0132)	0.1305*** (0.0132)	0.1291*** (0.0132)
APPC_JP	0.5834*** (0.0143)	0.5857*** (0.0143)	0.5906*** (0.0145)	0.6756*** (0.0188)	0.6789*** (0.0187)	0.6812*** (0.0187)	0.5666*** (0.0184)	0.5703*** (0.0184)	0.5718*** (0.0184)
APPC_FR	-0.1730*** (0.0093)	-0.1766*** (0.0093)	-0.1731*** (0.0094)	-0.2744*** (0.0175)	-0.2757*** (0.0175)	-0.2781*** (0.0175)	-0.2007*** (0.0176)	-0.2026*** (0.0176)	-0.2053*** (0.0176)
APPC_GB	-0.1672*** (0.0097)	-0.1694*** (0.0097)	-0.1749*** (0.0098)	-0.2001*** (0.0270)	-0.2008*** (0.0270)	-0.2018*** (0.0270)	-0.2127*** (0.0269)	-0.2132*** (0.0269)	-0.2147*** (0.0269)
cons	-0.6190*** (0.0260)	-0.7842*** (0.0257)	-0.6669*** (0.0230)	-0.5723*** (0.0257)	-0.7099*** (0.0264)	-0.8217*** (0.0255)	-0.7478*** (0.0264)	-0.8847*** (0.0271)	-1.0047*** (0.0262)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XPHARM have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 16: PHARM Poisson regression with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	198,111	198,111	198,111	190,285	190,285	190,285	190,285	190,285	190,285
chi2	157881.0363	157389.2689	157089.0649	154949.7970	154486.7932	154228.3732	157726.1555	157315.6931	157126.4013
IRII11	-0.2258*** (0.0119)			-0.2060*** (0.0120)			-0.1825*** (0.0121)		
NTR11	-0.8553*** (0.0251)			-0.9016*** (0.0256)			-0.8628*** (0.0256)		
IRII31		-0.0555*** (0.0121)			-0.0251* (0.0122)			0.0101 (0.0123)	
NTR31		-1.1497*** (0.0395)			-1.2103*** (0.0402)			-1.1383*** (0.0401)	
IRII51			0.0897*** (0.0111)			0.1197*** (0.0112)			0.1438*** (0.0113)
NTR51			-1.0111*** (0.0458)			-1.0760*** (0.0466)			-0.9932*** (0.0466)
patent scope	0.1551*** (0.0013)	0.1614*** (0.0013)	0.1651*** (0.0012)	0.1608*** (0.0013)	0.1674*** (0.0013)	0.1710*** (0.0013)	0.1496*** (0.0014)	0.1564*** (0.0013)	0.1594*** (0.0013)
family size	0.0510*** (0.0002)	0.0508*** (0.0002)	0.0508*** (0.0002)	0.0502*** (0.0002)	0.0500*** (0.0002)	0.0500*** (0.0002)	0.0491*** (0.0002)	0.0490*** (0.0002)	0.0489*** (0.0002)
backward citations	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0040*** (0.0000)	0.0040*** (0.0000)	0.0041*** (0.0000)
originality							0.8265*** (0.0165)	0.8354*** (0.0165)	0.8453*** (0.0165)
GV_PPP	0.3111*** (0.0157)	0.3101*** (0.0156)	0.3105*** (0.0156)	0.3368*** (0.0157)	0.3355*** (0.0157)	0.3365*** (0.0157)	0.3371*** (0.0157)	0.3354*** (0.0157)	0.3372*** (0.0157)
B_PPP	-0.0801*** (0.0025)	-0.0794*** (0.0025)	-0.0793*** (0.0025)	-0.0819*** (0.0025)	-0.0810*** (0.0025)	-0.0811*** (0.0025)	-0.0826*** (0.0025)	-0.0817*** (0.0025)	-0.0820*** (0.0025)
H_PPP	-0.1881*** (0.0116)	-0.1856*** (0.0115)	-0.1865*** (0.0115)	-0.1879*** (0.0116)	-0.1851*** (0.0116)	-0.1864*** (0.0116)	-0.1824*** (0.0116)	-0.1796*** (0.0116)	-0.1815*** (0.0116)
TP_RS	0.0019*** (0.0005)	0.0017** (0.0005)	0.0017** (0.0005)	0.0015** (0.0005)	0.0012* (0.0005)	0.0012* (0.0005)	0.0015** (0.0005)	0.0012* (0.0005)	0.0012* (0.0005)
B_PHARM	0.2717*** (0.0078)	0.2705*** (0.0078)	0.2715*** (0.0078)	0.2693*** (0.0079)	0.2677*** (0.0079)	0.2692*** (0.0079)	0.2670*** (0.0079)	0.2656*** (0.0079)	0.2675*** (0.0079)
TD_BPHARM	-0.0028 (0.0052)	-0.0021 (0.0052)	-0.0023 (0.0052)	-0.0014 (0.0052)	-0.0007 (0.0052)	-0.0010 (0.0052)	-0.0002 (0.0052)	0.0005 (0.0052)	-0.0000 (0.0052)
TD_XPHARM	0.0046*** (0.0014)	0.0047*** (0.0014)	0.0046*** (0.0014)	0.0063*** (0.0014)	0.0065*** (0.0014)	0.0064*** (0.0014)	0.0059*** (0.0014)	0.0060*** (0.0014)	0.0059*** (0.0014)
APPC_US	0.5510*** (0.0173)	0.5551*** (0.0173)	0.5577*** (0.0173)	0.5339*** (0.0176)	0.5379*** (0.0176)	0.5407*** (0.0176)	0.5261*** (0.0176)	0.5299*** (0.0176)	0.5330*** (0.0176)
APPC_DE	0.1344*** (0.0132)	0.1305*** (0.0132)	0.1291*** (0.0132)	0.1001*** (0.0133)	0.0958*** (0.0133)	0.0949*** (0.0133)	0.1064*** (0.0133)	0.1019*** (0.0133)	0.1020*** (0.0133)
APPC_JP	0.5666*** (0.0184)	0.5703*** (0.0184)	0.5718*** (0.0184)	0.5865*** (0.0186)	0.5901*** (0.0186)	0.5920*** (0.0186)	0.5701*** (0.0185)	0.5728*** (0.0185)	0.5752*** (0.0185)
APPC_FR	-0.2007*** (0.0176)	-0.2026*** (0.0176)	-0.2053*** (0.0176)	-0.2168*** (0.0178)	-0.2190*** (0.0178)	-0.2226*** (0.0178)	-0.2027*** (0.0178)	-0.2046*** (0.0178)	-0.2083*** (0.0178)
APPC_GB	-0.2127*** (0.0269)	-0.2132*** (0.0269)	-0.2147*** (0.0269)	-0.2102*** (0.0271)	-0.2109*** (0.0271)	-0.2125*** (0.0271)	-0.2028*** (0.0271)	-0.2038*** (0.0271)	-0.2051*** (0.0271)
cons	-0.7478*** (0.0264)	-0.8847*** (0.0271)	-1.0047*** (0.0262)	-0.7438*** (0.0272)	-0.8940*** (0.0279)	-1.0105*** (0.0270)	-1.3302*** -0.0297	-1.5000*** -0.0305	-1.6102*** -0.0296

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XPHARM have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 17: PHARM Poisson regression including interaction between IRII and Patent Scope, with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	198,111	198,111	198,111	190,285	190,285	190,285	190,285	190,285	190,285
chi2	157881.8490	157442.6219	157825.2829	154950.4756	154582.1455	155147.7258	157733.0632	157356.0075	157899.4273
IRII11	-0.2121*** (0.0192)			-0.2185*** (0.0194)			-0.1421*** (0.0195)		
NTR11	-0.8584*** (0.0254)			-0.8987*** (0.0258)			-0.8719*** (0.0258)		
IRII31		-0.1692*** (0.0198)			-0.1784*** (0.0200)			-0.0906*** (0.0202)	
NTR31		-1.1073*** (0.0399)			-1.1525*** (0.0405)			-1.1013*** (0.0405)	
IRII51			-0.3346*** (0.0195)			-0.3598*** (0.0197)			-0.2995*** (0.0198)
NTR51			-0.8068*** (0.0460)			-0.8437*** (0.0467)			-0.7837*** (0.0467)
patent scope	0.1576*** (0.0031)	0.1397*** (0.0033)	0.0786*** (0.0035)	0.1585*** (0.0031)	0.1382*** (0.0033)	0.0734*** (0.0036)	0.1571*** (0.0031)	0.1374*** (0.0033)	0.0696*** (0.0036)
IRII11×patent scope	-0.0046 (0.0051)			0.0043 (0.0052)			-0.0137** (0.0052)		
IRII31×patent scope		0.0369*** (0.0051)			0.0499*** (0.0051)			0.0326*** (0.0052)	
IRII51×patent scope			0.1329*** (0.0050)			0.1505*** (0.0051)			0.1388*** (0.0051)
family size	0.0510*** (0.0002)	0.0508*** (0.0002)	0.0505*** (0.0002)	0.0502*** (0.0002)	0.0499*** (0.0002)	0.0497*** (0.0002)	0.0492*** (0.0002)	0.0489*** (0.0002)	0.0487*** (0.0002)
backward citations	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0040*** (0.0000)	0.0040*** (0.0000)	0.0040*** (0.0000)
originality							0.8293*** (0.0166)	0.8285*** (0.0166)	0.8228*** (0.0165)
GV_PPP	0.3111*** (0.0157)	0.3099*** (0.0156)	0.3141*** (0.0156)	0.3368*** (0.0157)	0.3350*** (0.0157)	0.3402*** (0.0157)	0.3372*** (0.0157)	0.3352*** (0.0157)	0.3407*** (0.0157)
B_PPP	-0.0801*** (0.0025)	-0.0796*** (0.0025)	-0.0814*** (0.0025)	-0.0819*** (0.0025)	-0.0812*** (0.0025)	-0.0834*** (0.0025)	-0.0825*** (0.0025)	-0.0819*** (0.0025)	-0.0841*** (0.0025)
H_PPP	-0.1883*** (0.0116)	-0.1837*** (0.0115)	-0.1858*** (0.0115)	-0.1877*** (0.0116)	-0.1824*** (0.0116)	-0.1855*** (0.0116)	-0.1830*** (0.0116)	-0.1779*** (0.0116)	-0.1809*** (0.0116)
TP_RS	0.0019*** (0.0005)	0.0017** (0.0005)	0.0018*** (0.0005)	0.0014** (0.0005)	0.0012* (0.0005)	0.0013* (0.0005)	0.0015** (0.0005)	0.0012* (0.0005)	0.0013* (0.0005)
B_PHARM	0.2717*** (0.0078)	0.2705*** (0.0078)	0.2769*** (0.0078)	0.2693*** (0.0079)	0.2677*** (0.0079)	0.2755*** (0.0079)	0.2670*** (0.0079)	0.2656*** (0.0079)	0.2733*** (0.0079)
TD_BPHARM	-0.0029 (0.0052)	-0.0015 (0.0052)	-0.0013 (0.0052)	-0.0014 (0.0052)	0.0003 (0.0052)	0.0003 (0.0052)	-0.0004 (0.0052)	0.0011 (0.0052)	0.0011 (0.0052)
TD_XPHARM	0.0046*** (0.0014)	0.0048*** (0.0014)	0.0051*** (0.0014)	0.0064*** (0.0014)	0.0066*** (0.0014)	0.0069*** (0.0014)	0.0059*** (0.0014)	0.0061*** (0.0014)	0.0065*** (0.0014)
APPC_US	0.5512*** (0.0173)	0.5536*** (0.0173)	0.5560*** (0.0173)	0.5338*** (0.0176)	0.5357*** (0.0176)	0.5384*** (0.0176)	0.5265*** (0.0176)	0.5286*** (0.0176)	0.5313*** (0.0176)
APPC_DE	0.1347*** (0.0132)	0.1290*** (0.0132)	0.1264*** (0.0132)	0.0999*** (0.0133)	0.0936*** (0.0133)	0.0917*** (0.0133)	0.1071*** (0.0133)	0.1004*** (0.0133)	0.0986*** (0.0133)
APPC_JP	0.5667*** (0.0184)	0.5692*** (0.0184)	0.5760*** (0.0183)	0.5864*** (0.0186)	0.5884*** (0.0186)	0.5962*** (0.0185)	0.5704*** (0.0185)	0.5719*** (0.0185)	0.5797*** (0.0185)
APPC_FR	-0.2006*** (0.0176)	-0.2026*** (0.0176)	-0.2034*** (0.0176)	-0.2169*** (0.0178)	-0.2188*** (0.0178)	-0.2202*** (0.0178)	-0.2024*** (0.0178)	-0.2045*** (0.0178)	-0.2066*** (0.0178)
APPC_GB	-0.2126*** (0.0269)	-0.2133*** (0.0269)	-0.2139*** (0.0269)	-0.2102*** (0.0271)	-0.2110*** (0.0271)	-0.2114*** (0.0271)	-0.2027*** (0.0271)	-0.2039*** (0.0271)	-0.2044*** (0.0271)
cons	-0.7558*** (0.0279)	-0.8141*** (0.0288)	-0.7105*** (0.0285)	-0.7364*** (0.0286)	-0.7992*** (0.0296)	-0.6775*** (0.0292)	-1.3560*** (0.0313)	-1.4329*** (0.0323)	-1.2865*** (0.0318)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XPHARM have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 18: PHARM Poisson regression including interaction between NTR and Patent Scope, with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	198,111	198,111	198,111	190,285	190,285	190,285	190,285	190,285	190,285
chi2	158446.9490	157967.0985	157604.7571	155619.1830	155145.4095	154811.3174	158321.9543	157913.3206	157658.3896
IRII11	-0.2155***			-0.1932***			-0.1710***		
	(0.0119)			(0.0120)			(0.0121)		
NTR11	0.1930***			0.2579***			0.2326***		
	(0.0514)			(0.0524)			(0.0524)		
IRII31		-0.0489***			-0.0167			0.0173	
		(0.0121)			(0.0122)			(0.0123)	
NTR31		0.5699***			0.6608***			0.6452***	
		(0.0825)			(0.0842)			(0.0843)	
IRII51			0.0967***			0.1282***			0.1517***
			(0.0111)			(0.0112)			(0.0113)
NTR51			0.8538***			0.9480***			0.9426***
			(0.0939)			(0.0960)			(0.0962)
patent scope	0.1673***	0.1710***	0.1733***	0.1748***	0.1781***	0.1801***	0.1629***	0.1667***	0.1683***
	(0.0014)	(0.0013)	(0.0013)	(0.0014)	(0.0014)	(0.0013)	(0.0015)	(0.0014)	(0.0014)
NTR11×patent scope	-0.2391***			-0.2647***			-0.2503***		
	(0.0109)			(0.0112)			(0.0111)		
NTR31×patent scope		-0.3722***			-0.4053***			-0.3867***	
		(0.0172)			(0.0176)			(0.0176)	
NTR51×patent scope			-0.3923***			-0.4257***			-0.4076***
			(0.0191)			(0.0196)			(0.0196)
family size	0.0510***	0.0508***	0.0508***	0.0501***	0.0500***	0.0500***	0.0491***	0.0490***	0.0489***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
backward citations	0.0042***	0.0042***	0.0042***	0.0042***	0.0042***	0.0042***	0.0040***	0.0040***	0.0041***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
originality							0.8149***	0.8258***	0.8373***
							(0.0165)	(0.0165)	(0.0165)
GV_PPP	0.3106***	0.3090***	0.3087***	0.3363***	0.3343***	0.3346***	0.3366***	0.3343***	0.3352***
	(0.0157)	(0.0156)	(0.0156)	(0.0157)	(0.0157)	(0.0157)	(0.0157)	(0.0157)	(0.0157)
B_PPP	-0.0806***	-0.0798***	-0.0796***	-0.0825***	-0.0815***	-0.0814***	-0.0831***	-0.0822***	-0.0822***
	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
H_PPP	-0.1854***	-0.1834***	-0.1849***	-0.1847***	-0.1825***	-0.1846***	-0.1795***	-0.1773***	-0.1799***
	(0.0116)	(0.0115)	(0.0115)	(0.0116)	(0.0116)	(0.0116)	(0.0116)	(0.0116)	(0.0116)
TP_RS	0.0020***	0.0018***	0.0017**	0.0015**	0.0013*	0.0012*	0.0015**	0.0013*	0.0013*
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
B_PHARM	0.2729***	0.2717***	0.2728***	0.2706***	0.2690***	0.2706***	0.2682***	0.2669***	0.2688***
	(0.0078)	(0.0078)	(0.0078)	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)
TD_BPHARM	-0.0024	-0.0014	-0.0017	-0.0009	0.0001	-0.0003	0.0003	0.0012	0.0007
	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)
TD_XPHARM	0.0048***	0.0048***	0.0047***	0.0066***	0.0066***	0.0065***	0.0062***	0.0062***	0.0060***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
APPC_US	0.5469***	0.5518***	0.5547***	0.5293***	0.5340***	0.5372***	0.5220***	0.5264***	0.5298***
	(0.0173)	(0.0173)	(0.0173)	(0.0176)	(0.0176)	(0.0176)	(0.0176)	(0.0176)	(0.0176)
APPC_DE	0.1326***	0.1287***	0.1275***	0.0980***	0.0937***	0.0931***	0.1043***	0.0998***	0.1002***
	(0.0132)	(0.0132)	(0.0132)	(0.0133)	(0.0133)	(0.0133)	(0.0133)	(0.0133)	(0.0133)
APPC_JP	0.5661***	0.5683***	0.5691***	0.5859***	0.5879***	0.5891***	0.5698***	0.5710***	0.5727***
	(0.0184)	(0.0184)	(0.0184)	(0.0185)	(0.0186)	(0.0186)	(0.0185)	(0.0185)	(0.0185)
APPC_FR	-0.2006***	-0.2018***	-0.2043***	-0.2168***	-0.2181***	-0.2215***	-0.2029***	-0.2039***	-0.2074***
	(0.0176)	(0.0176)	(0.0176)	(0.0178)	(0.0178)	(0.0178)	(0.0178)	(0.0178)	(0.0178)
APPC_GB	-0.2122***	-0.2136***	-0.2151***	-0.2094***	-0.2111***	-0.2128***	-0.2023***	-0.2041***	-0.2055***
	(0.0269)	(0.0269)	(0.0269)	(0.0271)	(0.0271)	(0.0271)	(0.0271)	(0.0271)	(0.0271)
cons	-0.8006***	-0.9256***	-1.0402***	-0.8028***	-0.9392***	-1.0490***	-1.3776***	-1.5360***	-1.6414***
	(0.0265)	(0.0272)	(0.0263)	(0.0273)	(0.0279)	(0.0270)	(0.0298)	(0.0305)	(0.0296)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XPHARM have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 19: PHARM Probit regression with Breakthrough probability as dependent variable, MSTI variables included in estimation

	Dependent Variable: Breakthrough								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	198,062	198,062	198,062	190,239	190,239	190,239	190,239	190,239	190,239
chi2	892.2427	888.5972	890.7000	860.4586	857.6043	860.1882	871.8403	869.4935	872.6406
IRII11	-0.1583 (0.0834)			-0.1411 (0.0849)			-0.1311 (0.0854)		
NTR11	-0.1017 (0.1533)			-0.0914 (0.1553)			-0.0748 (0.1554)		
IRII31		-0.0224 (0.0835)			-0.0018 (0.0851)			0.0165 (0.0858)	
NTR31		-0.0014 (0.2210)			0.0031 (0.2243)			0.0370 (0.2241)	
IRII51			0.1076 (0.0772)			0.1195 (0.0784)			0.1323 (0.0791)
NTR51			0.1625 (0.2482)			0.1750 (0.2510)			0.2120 (0.2507)
patent scope	0.0829*** (0.0090)	0.0882*** (0.0088)	0.0925*** (0.0086)	0.0836*** (0.0093)	0.0891*** (0.0091)	0.0931*** (0.0089)	0.0786*** (0.0095)	0.0842*** (0.0092)	0.0879*** (0.0091)
family size	0.0272*** (0.0014)	0.0270*** (0.0014)	0.0270*** (0.0013)	0.0271*** (0.0014)	0.0270*** (0.0014)	0.0269*** (0.0014)	0.0267*** (0.0014)	0.0265*** (0.0014)	0.0264*** (0.0014)
backward citations	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0027*** (0.0002)	0.0026*** (0.0002)	0.0026*** (0.0002)	0.0026*** (0.0002)
originality							0.3799** (0.1186)	0.3891** (0.1190)	0.3984*** (0.1192)
GV_PPP	0.0743 (0.1256)	0.0733 (0.1254)	0.0754 (0.1253)	0.1299 (0.1199)	0.1289 (0.1197)	0.1309 (0.1195)	0.1304 (0.1198)	0.1293 (0.1196)	0.1316 (0.1195)
B_PPP	-0.0439* (0.0177)	-0.0435* (0.0177)	-0.0440* (0.0177)	-0.0456* (0.0181)	-0.0453* (0.0180)	-0.0458* (0.0181)	-0.0458* (0.0181)	-0.0455* (0.0181)	-0.0460* (0.0181)
H_PPP	0.0164 (0.0935)	0.0187 (0.0934)	0.0173 (0.0933)	0.0192 (0.0895)	0.0216 (0.0893)	0.0202 (0.0892)	0.0207 (0.0895)	0.0231 (0.0893)	0.0214 (0.0892)
TP_RS	0.0002 (0.0042)	0.0001 (0.0042)	0.0001 (0.0042)	-0.0006 (0.0043)	-0.0007 (0.0043)	-0.0007 (0.0043)	-0.0006 (0.0043)	-0.0007 (0.0043)	-0.0007 (0.0043)
B_PHARM	0.1602** (0.0579)	0.1590** (0.0579)	0.1613** (0.0580)	0.1475* (0.0587)	0.1464* (0.0587)	0.1488* (0.0588)	0.1463* (0.0588)	0.1454* (0.0588)	0.1479* (0.0588)
TD_BPHARM	-0.0369 (0.0352)	-0.0366 (0.0352)	-0.0371 (0.0352)	-0.0340 (0.0352)	-0.0337 (0.0352)	-0.0342 (0.0352)	-0.0337 (0.0352)	-0.0335 (0.0352)	-0.0340 (0.0352)
TD_XPHARM	0.0025 (0.0098)	0.0025 (0.0098)	0.0023 (0.0098)	0.0060 (0.0099)	0.0060 (0.0099)	0.0059 (0.0099)	0.0058 (0.0099)	0.0057 (0.0099)	0.0056 (0.0099)
APPC_US	0.2114 (0.1261)	0.2129 (0.1259)	0.2140 (0.1259)	0.1269 (0.1288)	0.1278 (0.1287)	0.1285 (0.1286)	0.1266 (0.1285)	0.1275 (0.1284)	0.1283 (0.1283)
APPC_DE	0.1253 (0.0944)	0.1237 (0.0943)	0.1236 (0.0943)	0.0516 (0.0971)	0.0496 (0.0970)	0.0496 (0.0970)	0.0574 (0.0971)	0.0554 (0.0970)	0.0560 (0.0970)
APPC_JP	0.3626** (0.1357)	0.3633** (0.1356)	0.3635** (0.1357)	0.3833** (0.1383)	0.3835** (0.1383)	0.3838** (0.1384)	0.3788** (0.1382)	0.3783** (0.1382)	0.3789** (0.1382)
APPC_FR	-0.5317* (0.2238)	-0.5318* (0.2237)	-0.5328* (0.2235)	-0.5513* (0.2245)	-0.5516* (0.2245)	-0.5525* (0.2241)	-0.5409* (0.2240)	-0.5412* (0.2240)	-0.5414* (0.2235)
APPC_GB	-0.3991 (0.2859)	-0.4022 (0.2862)	-0.4056 (0.2866)	-0.3926 (0.2867)	-0.3961 (0.2870)	-0.3995 (0.2874)	-0.3883 (0.2869)	-0.3921 (0.2872)	-0.3950 (0.2876)
cons	-3.2960*** (0.1851)	-3.3966*** (0.1895)	-3.4985*** (0.1839)	-3.2794*** (0.1883)	-3.3862*** (0.1929)	-3.4798*** (0.1870)	-3.5500*** (0.2074)	-3.6721*** (0.2130)	-3.7654*** (0.2067)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XPHARM have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. DE: Germany, JP: Japan, FR: France, GB: United Kingdom.

Table 20: COMP Poisson regression with forward citation number as dependent variable, MSTI variables included in estimation, without Originality

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with B_COMP data			patents with B_COMP data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	266,230	266,230	264,396	199,354	199,354	199,354	199,354	199,354	199,354
chi2	135217.5580	135412.1248	135150.8959	107263.1881	107263.6990	107341.3357	107489.6526	107497.1431	107574.9565
IRII11	-0.2311***			-0.2119***			-0.2139***		
	(0.0094)			(0.0112)			(0.0112)		
NTR11	-0.1471***			-0.0500**			-0.0525**		
	(0.0139)			(0.0166)			(0.0166)		
IRII31		-0.2652***			-0.2069***			-0.2112***	
		(0.0093)			(0.0111)			(0.0111)	
NTR31		-0.1448***			0.0028			-0.0018	
		(0.0184)			(0.0219)			(0.0219)	
IRII51			-0.2796***			-0.2216***			-0.2261***
			(0.0094)			(0.0112)			(0.0112)
NTR51			-0.0540**			0.0977***			0.0924***
			(0.0206)			(0.0242)			(0.0242)
patent scope	0.1449***	0.1436***	0.1415***	0.1543***	0.1548***	0.1527***	0.1543***	0.1548***	0.1526***
	(0.0011)	(0.0010)	(0.0011)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
family size	0.0857***	0.0858***	0.0856***	0.0854***	0.0854***	0.0852***	0.0856***	0.0856***	0.0854***
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
backward citations	0.0082***	0.0082***	0.0082***	0.0079***	0.0079***	0.0079***	0.0079***	0.0080***	0.0079***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
GV_PPP	-0.0294*	-0.0292*	-0.0314*	0.0430*	0.0406*	0.0412*	0.1480***	0.1471***	0.1477***
	(0.0135)	(0.0135)	(0.0137)	(0.0182)	(0.0182)	(0.0182)	(0.0195)	(0.0195)	(0.0195)
B_PPP	-0.0419***	-0.0418***	-0.0417***	-0.0279***	-0.0277***	-0.0275***	-0.0692***	-0.0697***	-0.0695***
	(0.0020)	(0.0020)	(0.0020)	(0.0029)	(0.0029)	(0.0029)	(0.0040)	(0.0040)	(0.0040)
H_PPP	0.3994***	0.3966***	0.3992***	0.1776***	0.1777***	0.1749***	0.0886***	0.0876***	0.0846***
	(0.0122)	(0.0122)	(0.0123)	(0.0245)	(0.0244)	(0.0244)	(0.0251)	(0.0251)	(0.0251)
TP_RS	-0.0058***	-0.0058***	-0.0058***	-0.0069***	-0.0068***	-0.0068***	-0.0052***	-0.0050***	-0.0050***
	(0.0004)	(0.0004)	(0.0005)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
B_COMP							0.1602***	0.1627***	0.1628***
							(0.0106)	(0.0106)	(0.0106)
TD_BCOMP	0.0277***	0.0274***	0.0282***	-0.0067*	-0.0064*	-0.0068*	-0.0206***	-0.0205***	-0.0209***
	(0.0019)	(0.0019)	(0.0019)	(0.0030)	(0.0030)	(0.0030)	(0.0031)	(0.0031)	(0.0031)
TD_XCOMP	0.0157***	0.0158***	0.0156***	0.0287***	0.0283***	0.0285***	0.0272***	0.0267***	0.0269***
	(0.0008)	(0.0008)	(0.0008)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
APPC_US	0.4846***	0.4842***	0.4946***	0.4091***	0.4108***	0.4116***	0.2107***	0.2091***	0.2098***
	(0.0171)	(0.0171)	(0.0173)	(0.0313)	(0.0313)	(0.0313)	(0.0343)	(0.0343)	(0.0343)
APPC_JP	0.1633***	0.1628***	0.1625***	0.0928***	0.0908***	0.0926***	0.0515*	0.0486*	0.0503*
	(0.0173)	(0.0173)	(0.0174)	(0.0241)	(0.0241)	(0.0241)	(0.0243)	(0.0243)	(0.0243)
APPC_DE	-0.1487***	-0.1467***	-0.1456***	-0.1476***	-0.1457***	-0.1443***	-0.1095***	-0.1070***	-0.1056***
	(0.0104)	(0.0104)	(0.0104)	(0.0152)	(0.0152)	(0.0152)	(0.0155)	(0.0155)	(0.0155)
APPC_FR	-0.1063***	-0.1038***	-0.1012***	-0.1595***	-0.1593***	-0.1596***	-0.1443***	-0.1439***	-0.1442***
	(0.0114)	(0.0114)	(0.0114)	(0.0189)	(0.0189)	(0.0189)	(0.0189)	(0.0189)	(0.0189)
APPC_KR	0.4267***	0.4302***	0.4286***	0.4600***	0.4614***	0.4631***	0.3368***	0.3365***	0.3382***
	(0.0124)	(0.0124)	(0.0125)	(0.0149)	(0.0149)	(0.0148)	(0.0170)	(0.0170)	(0.0170)
cons	-0.8797***	-0.8633***	-0.7259***	-1.0506***	-1.0527***	-1.0463***	-1.1382***	-1.1392***	-1.1329***
	(0.0344)	(0.0343)	(0.0300)	(0.0418)	(0.0418)	(0.0418)	(0.0421)	(0.0421)	(0.0420)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XCOMP have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 21: COMP Poisson regression with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	199,354	199,354	199,354	187,809	187,809	187,809	187,809	187,809	187,809
chi2	107489.6526	107497.1431	107574.9565	103639.3794	103643.2670	103706.4921	103824.8922	103823.0035	103876.5030
IRII11	-0.2139*** (0.0112)			-0.2144*** (0.0115)			-0.1899*** (0.0116)		
NTR11	-0.0525** (0.0166)			-0.0616*** (0.0168)			-0.0670*** (0.0169)		
IRII31		-0.2112*** (0.0111)			-0.2113*** (0.0113)			-0.1862*** (0.0115)	
NTR31		-0.0018 (0.0219)			-0.0151 (0.0223)			-0.0218 (0.0223)	
IRII51			-0.2261*** (0.0112)			-0.2234*** (0.0115)			-0.1981*** (0.0116)
NTR51			0.0924*** (0.0242)			0.0822*** (0.0246)			0.0755** (0.0246)
patent scope	0.1543*** (0.0012)	0.1548*** (0.0012)	0.1526*** (0.0012)	0.1522*** (0.0013)	0.1527*** (0.0012)	0.1506*** (0.0013)	0.1501*** (0.0013)	0.1505*** (0.0013)	0.1486*** (0.0013)
family size	0.0856*** (0.0005)	0.0856*** (0.0005)	0.0854*** (0.0005)	0.0856*** (0.0005)	0.0856*** (0.0005)	0.0854*** (0.0005)	0.0854*** (0.0005)	0.0854*** (0.0005)	0.0851*** (0.0005)
backward citations	0.0079*** (0.0001)	0.0080*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)
originality							0.1646*** (0.0122)	0.1622*** (0.0122)	0.1578*** (0.0122)
GV_PPP	0.1480*** (0.0195)	0.1471*** (0.0195)	0.1477*** (0.0195)	0.1673*** (0.0198)	0.1663*** (0.0198)	0.1670*** (0.0198)	0.1638*** (0.0198)	0.1631*** (0.0198)	0.1638*** (0.0198)
B_PPP	-0.0692*** (0.0040)	-0.0697*** (0.0040)	-0.0695*** (0.0040)	-0.0704*** (0.0040)	-0.0708*** (0.0040)	-0.0707*** (0.0040)	-0.0701*** (0.0040)	-0.0705*** (0.0040)	-0.0704*** (0.0040)
H_PPP	0.0886*** (0.0251)	0.0876*** (0.0251)	0.0846*** (0.0251)	0.0622* (0.0256)	0.0614* (0.0255)	0.0587* (0.0255)	0.0645* (0.0256)	0.0637* (0.0255)	0.0612* (0.0255)
TP_RS	-0.0052*** (0.0008)	-0.0050*** (0.0008)	-0.0050*** (0.0008)	-0.0049*** (0.0008)	-0.0048*** (0.0008)	-0.0048*** (0.0008)	-0.0049*** (0.0008)	-0.0048*** (0.0008)	-0.0048*** (0.0008)
B_COMP	0.1602*** (0.0106)	0.1627*** (0.0106)	0.1628*** (0.0106)	0.1645*** (0.0109)	0.1669*** (0.0109)	0.1670*** (0.0109)	0.1640*** (0.0109)	0.1660*** (0.0109)	0.1661*** (0.0109)
TD_BCOMP	-0.0206*** (0.0031)	-0.0205*** (0.0031)	-0.0209*** (0.0031)	-0.0234*** (0.0031)	-0.0233*** (0.0031)	-0.0237*** (0.0031)	-0.0230*** (0.0031)	-0.0228*** (0.0031)	-0.0232*** (0.0031)
TD_XCOMP	0.0272*** (0.0028)	0.0267*** (0.0028)	0.0269*** (0.0028)	0.0260*** (0.0028)	0.0256*** (0.0028)	0.0258*** (0.0028)	0.0260*** (0.0028)	0.0256*** (0.0028)	0.0258*** (0.0028)
APPC_US	0.2107*** (0.0343)	0.2091*** (0.0343)	0.2098*** (0.0343)	0.2096*** (0.0351)	0.2078*** (0.0351)	0.2083*** (0.0351)	0.2124*** (0.0351)	0.2109*** (0.0351)	0.2113*** (0.0351)
APPC_JP	0.0515* (0.0243)	0.0486* (0.0243)	0.0503* (0.0243)	0.0582* (0.0247)	0.0554* (0.0247)	0.0574* (0.0247)	0.0505* (0.0247)	0.0483 (0.0247)	0.0505* (0.0247)
APPC_DE	-0.1095*** (0.0155)	-0.1070*** (0.0155)	-0.1056*** (0.0155)	-0.0929*** (0.0158)	-0.0905*** (0.0158)	-0.0894*** (0.0158)	-0.0925*** (0.0158)	-0.0906*** (0.0158)	-0.0897*** (0.0158)
APPC_FR	-0.1443*** (0.0189)	-0.1439*** (0.0189)	-0.1442*** (0.0189)	-0.1246*** (0.0195)	-0.1244*** (0.0195)	-0.1246*** (0.0195)	-0.1203*** (0.0195)	-0.1202*** (0.0195)	-0.1205*** (0.0195)
APPC_KR	0.3368*** (0.0170)	0.3365*** (0.0170)	0.3382*** (0.0170)	0.3543*** (0.0175)	0.3542*** (0.0175)	0.3557*** (0.0175)	0.3531*** (0.0175)	0.3533*** (0.0175)	0.3550*** (0.0175)
cons	-1.1382*** (0.0421)	-1.1392*** (0.0421)	-1.1329*** (0.0420)	-1.1319*** (0.0437)	-1.1324*** (0.0437)	-1.1291*** (0.0436)	-1.2439*** (0.0445)	-1.2444*** (0.0445)	-1.2381*** (0.0444)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XCOMP have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 22: COMP Poisson regression including interaction between IRII and Patent Scope, with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	199,354	199,354	199,354	187,809	187,809	187,809	187,809	187,809	187,809
chi2	107630.1503	107828.2568	108119.3285	103753.1410	103926.3992	104196.4658	103904.5612	104057.3278	104294.5330
IRII11	-0.3237*** (0.0146)			-0.3158*** (0.0150)			-0.2777*** (0.0153)		
NTR11	-0.0377* (0.0166)			-0.0479** (0.0168)			-0.0550** (0.0169)		
IRII31		-0.4088*** (0.0157)			-0.3986*** (0.0160)			-0.3616*** (0.0164)	
NTR31		0.0227 (0.0219)			0.0076 (0.0223)			-0.0002 (0.0223)	
IRII51			-0.4544*** (0.0151)			-0.4457*** (0.0154)			-0.4120*** (0.0158)
NTR51			0.1039*** (0.0242)			0.0928*** (0.0246)			0.0871*** (0.0246)
patent scope	0.1266*** (0.0027)	0.1025*** (0.0032)	0.0931*** (0.0029)	0.1269*** (0.0027)	0.1037*** (0.0032)	0.0933*** (0.0030)	0.1289*** (0.0027)	0.1059*** (0.0032)	0.0954*** (0.0030)
IRII11×patent scope	0.0619*** (0.0052)			0.0566*** (0.0053)			0.0478*** (0.0054)		
IRII31×patent scope		0.1023*** -0.0056			0.0961*** -0.0057			0.0881*** -0.0058	
IRII51×patent scope			0.1266*** -0.0054			0.1221*** -0.0055			0.1141*** -0.0056
family size	0.0857*** (0.0005)	0.0855*** (0.0005)	0.0849*** (0.0005)	0.0857*** (0.0005)	0.0856*** (0.0005)	0.0850*** (0.0005)	0.0855*** (0.0005)	0.0853*** (0.0005)	0.0848*** (0.0005)
backward citations	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0080*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)	0.0078*** (0.0001)
originality							0.1501*** (0.0123)	0.1394*** (0.0123)	0.1211*** (0.0123)
GV_PPP	0.1540*** (0.0195)	0.1572*** (0.0195)	0.1570*** (0.0195)	0.1730*** (0.0198)	0.1762*** (0.0198)	0.1765*** (0.0198)	0.1690*** (0.0198)	0.1725*** (0.0198)	0.1734*** (0.0198)
B_PPP	-0.0695*** (0.0040)	-0.0701*** (0.0040)	-0.0689*** (0.0040)	-0.0707*** (0.0040)	-0.0712*** (0.0040)	-0.0701*** (0.0040)	-0.0704*** (0.0040)	-0.0709*** (0.0040)	-0.0699*** (0.0040)
H_PPP	0.0865*** (0.0250)	0.0851*** (0.0250)	0.0843*** (0.0250)	0.0605* (0.0255)	0.0594* (0.0255)	0.0585* (0.0255)	0.0628* (0.0255)	0.0616* (0.0255)	0.0606* (0.0255)
TP_RS	-0.0051*** (0.0008)	-0.0050*** (0.0008)	-0.0051*** (0.0008)	-0.0049*** (0.0008)	-0.0048*** (0.0008)	-0.0048*** (0.0008)	-0.0049*** (0.0008)	-0.0048*** (0.0008)	-0.0048*** (0.0008)
B_COMP	0.1596*** (0.0106)	0.1601*** (0.0106)	0.1569*** (0.0106)	0.1639*** (0.0109)	0.1643*** (0.0109)	0.1611*** (0.0109)	0.1635*** (0.0109)	0.1638*** (0.0109)	0.1608*** (0.0109)
TD_BCOMP	-0.0200*** (0.0031)	-0.0195*** (0.0031)	-0.0189*** (0.0031)	-0.0228*** (0.0031)	-0.0224*** (0.0031)	-0.0218*** (0.0031)	-0.0225*** (0.0031)	-0.0221*** (0.0031)	-0.0216*** (0.0031)
TD_XCOMP	0.0257*** (0.0028)	0.0247*** (0.0028)	0.0241*** (0.0028)	0.0246*** (0.0028)	0.0236*** (0.0028)	0.0230*** (0.0028)	0.0249*** (0.0028)	0.0239*** (0.0028)	0.0232*** (0.0028)
APPC_US	0.2181*** (0.0343)	0.2232*** (0.0343)	0.2318*** (0.0343)	0.2159*** (0.0351)	0.2206*** (0.0351)	0.2288*** (0.0350)	0.2174*** (0.0350)	0.2219*** (0.0350)	0.2295*** (0.0350)
APPC_JP	0.0498* (0.0243)	0.0483* (0.0242)	0.0474 (0.0242)	0.0569* (0.0247)	0.0555* (0.0247)	0.0551* (0.0246)	0.0502* (0.0247)	0.0495* (0.0247)	0.0500* (0.0247)
APPC_DE	-0.1098*** (0.0155)	-0.1084*** (0.0155)	-0.1081*** (0.0154)	-0.0936*** (0.0158)	-0.0925*** (0.0158)	-0.0926*** (0.0158)	-0.0932*** (0.0158)	-0.0925*** (0.0158)	-0.0928*** (0.0158)
APPC_FR	-0.1459*** (0.0189)	-0.1465*** (0.0189)	-0.1464*** (0.0189)	-0.1264*** (0.0195)	-0.1275*** (0.0195)	-0.1276*** (0.0195)	-0.1222*** (0.0195)	-0.1234*** (0.0195)	-0.1241*** (0.0195)
APPC_KR	0.3297*** (0.0170)	0.3268*** (0.0170)	0.3259*** (0.0170)	0.3480*** (0.0175)	0.3454*** (0.0175)	0.3444*** (0.0175)	0.3480*** (0.0175)	0.3455*** (0.0175)	0.3446*** (0.0175)
cons	-1.0718*** (0.0425)	-1.0076*** (0.0427)	-0.9823*** (0.0425)	-1.0710*** (0.0441)	-1.0085*** (0.0443)	-0.9837*** (0.0441)	-1.1826*** (0.0450)	-1.1150*** (0.0453)	-1.0766*** (0.0451)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 23: COMP Poisson regression including interaction between NTR and Patent Scope, with forward citation number as dependent variable, MSTI variables included in estimation

	Dependent Variable: Number of forward citations in 5 years								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	199,354	199,354	199,354	187,809	187,809	187,809	187,809	187,809	187,809
chi2	108422.9355	108329.5101	108495.3824	104601.7664	104499.2752	104651.5240	104813.5813	104701.6260	104845.2874
IRII11	-0.2569*** (0.0115)			-0.2586*** (0.0118)			-0.2343*** (0.0119)		
NTR11	-0.7654*** (0.0292)			-0.7989*** (0.0297)			-0.8179*** (0.0298)		
IRII31		-0.2482*** (0.0113)			-0.2495*** (0.0116)			-0.2242*** (0.0117)	
NTR31		-1.0183*** (0.0424)			-1.0668*** (0.0432)			-1.0900*** (0.0433)	
IRII51			-0.2684*** (0.0114)			-0.2673*** (0.0117)			-0.2417*** (0.0118)
NTR51			-1.1103*** (0.0476)			-1.1624*** (0.0486)			-1.1868*** (0.0486)
patent scope	0.1316*** (0.0016)	0.1388*** (0.0014)	0.1368*** (0.0014)	0.1287*** (0.0016)	0.1361*** (0.0014)	0.1342*** (0.0014)	0.1256*** (0.0016)	0.1333*** (0.0015)	0.1315*** (0.0015)
NTR11 x P.S.	0.1735*** (0.0056)			0.1787*** (0.0056)			0.1820*** (0.0057)		
NTR31 x P.S.		0.2254*** (0.0075)			0.2325*** (0.0076)			0.2362*** (0.0076)	
NTR51 x P.S.			0.2527*** (0.0079)			0.2605*** (0.0080)			0.2643*** (0.0081)
family size	0.0840*** (0.0005)	0.0843*** (0.0005)	0.0840*** (0.0005)	0.0839*** (0.0005)	0.0842*** (0.0005)	0.0840*** (0.0005)	0.0837*** (0.0005)	0.0840*** (0.0005)	0.0837*** (0.0005)
backward citations	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0079*** (0.0001)	0.0078*** (0.0001)	0.0079*** (0.0001)	0.0078*** (0.0001)	0.0077*** (0.0001)	0.0077*** (0.0001)	0.0077*** (0.0001)
originality							0.1764*** (0.0122)	0.1724*** (0.0122)	0.1688*** (0.0122)
GV_PPP	0.1440*** (0.0195)	0.1415*** (0.0195)	0.1416*** (0.0195)	0.1631*** (0.0198)	0.1605*** (0.0199)	0.1606*** (0.0198)	0.1597*** (0.0199)	0.1572*** (0.0199)	0.1574*** (0.0199)
B_PPP	-0.0675*** (0.0040)	-0.0687*** (0.0040)	-0.0687*** (0.0040)	-0.0685*** (0.0040)	-0.0698*** (0.0040)	-0.0699*** (0.0040)	-0.0683*** (0.0040)	-0.0695*** (0.0040)	-0.0696*** (0.0040)
H_PPP	0.0935*** (0.0251)	0.0920*** (0.0251)	0.0887*** (0.0251)	0.0671** (0.0256)	0.0658* (0.0256)	0.0626* (0.0256)	0.0692** (0.0256)	0.0680** (0.0256)	0.0651* (0.0256)
TP_RS	-0.0052*** (0.0008)	-0.0050*** (0.0008)	-0.0049*** (0.0008)	-0.0049*** (0.0008)	-0.0047*** (0.0008)	-0.0047*** (0.0008)	-0.0049*** (0.0008)	-0.0047*** (0.0008)	-0.0047*** (0.0008)
B_COMP	0.1549*** (0.0106)	0.1608*** (0.0106)	0.1618*** (0.0107)	0.1589*** (0.0109)	0.1649*** (0.0109)	0.1660*** (0.0109)	0.1583*** (0.0109)	0.1639*** (0.0109)	0.1650*** (0.0109)
TD_BCOMP	-0.0198*** (0.0031)	-0.0200*** (0.0031)	-0.0205*** (0.0031)	-0.0227*** (0.0031)	-0.0228*** (0.0031)	-0.0234*** (0.0031)	-0.0221*** (0.0031)	-0.0223*** (0.0031)	-0.0229*** (0.0031)
TD_XCOMP	0.0263*** (0.0028)	0.0262*** (0.0028)	0.0264*** (0.0028)	0.0251*** (0.0028)	0.0251*** (0.0028)	0.0253*** (0.0028)	0.0250*** (0.0028)	0.0251*** (0.0028)	0.0253*** (0.0028)
APPC_US	0.2056*** (0.0343)	0.2004*** (0.0344)	0.1983*** (0.0344)	0.2045*** (0.0351)	0.1986*** (0.0351)	0.1962*** (0.0351)	0.2084*** (0.0351)	0.2028*** (0.0351)	0.2002*** (0.0351)
APPC_JP	0.0494* (0.0243)	0.0430 (0.0243)	0.0432 (0.0243)	0.0562* (0.0247)	0.0497* (0.0247)	0.0500* (0.0247)	0.0480 (0.0247)	0.0422 (0.0247)	0.0426 (0.0247)
APPC_DE	-0.1138*** (0.0155)	-0.1070*** (0.0155)	-0.1047*** (0.0155)	-0.0973*** (0.0158)	-0.0903*** (0.0158)	-0.0882*** (0.0158)	-0.0968*** (0.0158)	-0.0903*** (0.0158)	-0.0885*** (0.0158)
APPC_FR	-0.1501*** (0.0189)	-0.1472*** (0.0189)	-0.1474*** (0.0189)	-0.1304*** (0.0195)	-0.1275*** (0.0195)	-0.1278*** (0.0195)	-0.1260*** (0.0195)	-0.1232*** (0.0195)	-0.1235*** (0.0195)
APPC_KR	0.3408*** (0.0171)	0.3376*** (0.0171)	0.3386*** (0.0171)	0.3584*** (0.0175)	0.3553*** (0.0175)	0.3560*** (0.0175)	0.3571*** (0.0175)	0.3543*** (0.0175)	0.3551*** (0.0175)
cons	-0.9999*** (0.0425)	-1.0295*** (0.0423)	-1.0195*** (0.0423)	-0.9887*** (0.0441)	-1.0192*** (0.0439)	-1.0124*** (0.0439)	-1.1048*** (0.0448)	-1.1352*** (0.0447)	-1.1261*** (0.0446)

Notes:
Sample of patents filed from 1980 to 2018. Year fixed effects are included in all the estimations.
Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

Table 24: COMP Probit regression with Breakthrough probability as dependent variable, MSTI variables included in estimation

	Dependent Variable: Breakthrough								
	whole sample			patents with originality data			patents with originality data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N	199,354	199,354	199,354	187,809	187,809	187,809	187,809	187,809	187,809
chi2	1586.6443	1589.2683	1587.6486	1510.6349	1513.4399	1512.1835	1522.6240	1524.9499	1523.7859
IRII11	-0.2624*** (0.0637)			-0.2462*** (0.0653)			-0.2154** (0.0659)		
NTR11	0.0536 (0.0863)			0.0473 (0.0883)			0.0420 (0.0885)		
IRII31		-0.2744*** (0.0627)			-0.2673*** (0.0642)			-0.2354*** (0.0649)	
NTR31		0.1099 (0.1114)			0.0727 (0.1151)			0.0677 (0.1152)	
IRII51			-0.2689*** (0.0637)			-0.2619*** (0.0652)			-0.2285*** (0.0660)
NTR51			0.1373 (0.1240)			0.1035 (0.1280)			0.0971 (0.1282)
patent scope	0.1134*** (0.0077)	0.1141*** (0.0075)	0.1143*** (0.0075)	0.1127*** (0.0078)	0.1133*** (0.0076)	0.1133*** (0.0076)	0.1089*** (0.0079)	0.1094*** (0.0077)	0.1095*** (0.0077)
family size	0.0583*** (0.0028)	0.0586*** (0.0028)	0.0584*** (0.0028)	0.0585*** (0.0029)	0.0588*** (0.0028)	0.0586*** (0.0029)	0.0584*** (0.0029)	0.0586*** (0.0029)	0.0585*** (0.0029)
backward citations	0.0027** (0.0008)	0.0028*** (0.0008)	0.0027*** (0.0008)	0.0027** (0.0008)	0.0028*** (0.0008)	0.0027*** (0.0008)	0.0024** (0.0009)	0.0025** (0.0009)	0.0024** (0.0009)
originality							0.2548*** (0.0759)	0.2501** (0.0760)	0.2513*** (0.0761)
GV_PPP	0.2416* (0.1063)	0.2404* (0.1063)	0.2399* (0.1063)	0.2611* (0.1078)	0.2598* (0.1078)	0.2596* (0.1078)	0.2566* (0.1080)	0.2555* (0.1080)	0.2554* (0.1080)
B_PPP	-0.0771*** (0.0212)	-0.0781*** (0.0212)	-0.0777*** (0.0212)	-0.0778*** (0.0216)	-0.0789*** (0.0216)	-0.0784*** (0.0216)	-0.0771*** (0.0216)	-0.0780*** (0.0216)	-0.0776*** (0.0216)
H_PPP	-0.0416 (0.1468)	-0.0423 (0.1465)	-0.0433 (0.1466)	-0.0677 (0.1492)	-0.0675 (0.1489)	-0.0688 (0.1490)	-0.0642 (0.1492)	-0.0640 (0.1490)	-0.0652 (0.1490)
TP_RS	0.0062 (0.0042)	0.0064 (0.0041)	0.0064 (0.0041)	0.0058 (0.0043)	0.0060 (0.0043)	0.0059 (0.0043)	0.0058 (0.0043)	0.0060 (0.0043)	0.0059 (0.0043)
B_COMP	0.0926 (0.0637)	0.0949 (0.0637)	0.0948 (0.0637)	0.1068 (0.0649)	0.1091 (0.0649)	0.1090 (0.0649)	0.1046 (0.0649)	0.1066 (0.0649)	0.1065 (0.0649)
TD_BCOMP	-0.0124 (0.0181)	-0.0123 (0.0181)	-0.0123 (0.0181)	-0.0143 (0.0185)	-0.0141 (0.0184)	-0.0141 (0.0184)	-0.0133 (0.0184)	-0.0131 (0.0184)	-0.0132 (0.0184)
TD_XCOMP	-0.0279 (0.0164)	-0.0283 (0.0164)	-0.0282 (0.0164)	-0.0281 (0.0167)	-0.0286 (0.0167)	-0.0285 (0.0167)	-0.0284 (0.0167)	-0.0288 (0.0167)	-0.0287 (0.0167)
APPC_US	0.3581 (0.1909)	0.3575 (0.1912)	0.3603 (0.1910)	0.3620 (0.1953)	0.3612 (0.1956)	0.3640 (0.1955)	0.3691 (0.1949)	0.3683 (0.1952)	0.3708 (0.1950)
APPC_JP	0.0062 (0.1329)	0.0013 (0.1328)	0.0041 (0.1326)	0.0292 (0.1356)	0.0229 (0.1355)	0.0263 (0.1353)	0.0191 (0.1356)	0.0138 (0.1355)	0.0170 (0.1353)
APPC_DE	-0.4076*** (0.1212)	-0.4022*** (0.1212)	-0.4016*** (0.1214)	-0.4128** (0.1261)	-0.4068** (0.1261)	-0.4064** (0.1262)	-0.4132** (0.1262)	-0.4079** (0.1262)	-0.4076** (0.1263)
APPC_FR	-0.2136 (0.1217)	-0.2137 (0.1215)	-0.2161 (0.1219)	-0.1781 (0.1232)	-0.1790 (0.1230)	-0.1811 (0.1233)	-0.1737 (0.1235)	-0.1745 (0.1233)	-0.1765 (0.1236)
APPC_KR	0.3106** (0.0989)	0.3101** (0.0989)	0.3107** (0.0987)	0.3230** (0.1006)	0.3218** (0.1005)	0.3223** (0.1004)	0.3216** (0.1007)	0.3207** (0.1006)	0.3212** (0.1004)
cons	-3.2521*** (0.2373)	-3.2398*** (0.2368)	-3.2482*** (0.2366)	-3.3039*** (0.2498)	-3.2804*** (0.2493)	-3.2902*** (0.2492)	-3.4674*** (0.2539)	-3.4428*** (0.2535)	-3.4539*** (0.2535)

Notes:
Due to the inclusion of MSTI variables, only data from 1987 to 2018 are available for the estimations. Year fixed effects are included in all the estimations. All the MSTI variables except TD_XCOMP have been divided by 10,000 from their original values.
Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
APPC_US is a dummy variable defined to be 1 when at least one of the applicants of the patent is registered with address in the United States. JP: Japan, DE: Germany, FR: France, KR: South Korea.

and breakthrough invention. *Industrial and Corporate Change*, 24(6):1215–1246, 2015.

A Banu Goktan and Grant Miles. Innovation speed and radicalness: are they inversely related? *Management Decision*, 49(4):533–547, 2011.

Joseph L Bower and Clayton M Christensen. Disruptive technologies: catching the wave. 1995.

Kristie Briggs. Co-owner relationships conducive to high quality joint patents. *Research Policy*, 44:1566–1573, 2015.

Paola Criscuolo and Bart Verspagen. Does it matter where patent citations come from? inventor vs. examiner citations in european patents. *Research policy*, 37(10):1892–1908, 2008.

Kristina B Dahlin and Dean M Behrens. When is an invention really radical? defining and measuring technological radicalnes. *Research Policy*, 34:717–737, 2005.

Joseph A DiMasi, Ronald W Hansen, Henry G Grabowski, and Louis Lasagna. Cost of innovation in the pharmaceutical industry. *J Health Econ*, 1991.

Joseph A DiMasi, Ronald W Hansen, and Henry G Grabowski. The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185, 2003.

Robert E Evenson. Patents, r&d, and invention potential: international evidence. *The American Economic Review*, 83(2):463–468, 1993.

Carsten Fink, Mosahid Khan, and Hao Zhou. Exploring the worldwide patent surge. *Economics of Innovation and New Technology*, 25(2):114–142, 2016.

Lee Fleming. Recombinant uncertainty in technological search. *Management science*, 47(1):117–132, 2001.

Alberto Galasso. Cep discussion paper no 1072 august 2011 trading and enforcing patent rights alberto galasso, mark schankerman and carlos j. serrano. 2011.

Yuan Gao and Emiliya Lazarova. A new empirical index to track the technological novelty of inventions: A sector level analysis. Technical report, School of Economics, University of East Anglia, Norwich, UK., 2022.

Paul Gompers, Josh Lerner, and David Scharfstein. Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999. *The journal of Finance*, 60(2):577–614, 2005.

Henry Grabowski, John Vernon, and Joseph A DiMasi. Returns on research and development for 1990s new drug introductions. *Pharmacoeconomics*, 20 (3):11–29, 2002.

Henry T Greely. Governing emerging technologies—looking forward with horizon scanning and looking back with technology audits. *Global Public Policy and Governance*, 2(3):266–282, 2022.

Zvi Griliches. Patents: Recent trends and puzzles, 1989.

Zvi Griliches. Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence*, pages 287–343. University of Chicago Press, 1998.

Zvi Griliches, Ariel Pakes, and Bronwyn H Hall. The value of patents as indicators of inventive activity, 1986.

Bronwyn H Hall and Manuel Trajtenberg. Uncovering gpts with patent data. Technical report, National Bureau of Economic Research, 2004.

Bronwyn H Hall, Adam Jaffe, and Manuel Trajtenberg. Market value and patent citations. *RAND Journal of economics*, pages 16–38, 2005.

Dietmar Harhoff and Stefan Wagner. The duration of patent examination at the european patent office. *Management Science*, 55(12):1969–1984, 2009.

Dietmar Harhoff, Frederic M Scherer, and Katrin Vopel. Citations, family size, opposition and the value of patent rights. *Research policy*, 32(8):1343–1363, 2003.

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

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

Adam B Jaffe and Manuel Trajtenberg. *Patents, citations, and innovations: A window on the knowledge economy.* MIT press, 2002.

Sarah Kaplan and Keyvan Vakili. The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36:1435–1457, 2015. URL `DOI:10.1002/smj.2294`.

PH Kaya and A Joseph. Schumpeter's perspective on innovation international'. *Journal of Economics, Commerce and Management*, 3(8):25–37, 2015.

William R Kerr. Breakthrough inventions and migrating clusters of innovation. *Journal of urban Economics*, 67(1):46–60, 2010.

Jinyoung Kim and Gerald Marschke. Accounting for the recent surge in us patenting: changes in r&d expenditures, patent yields, and the high tech sector. *Economics of Innovation and New Technology*, 13(6):543–558, 2004.

Jean O Lanjouw and Mark Schankerman. Characteristics of patent litigation: a window on competition. *RAND journal of economics*, pages 129–151, 2001.

Jean O Lanjouw and Mark Schankerman. Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495):441–465, 2004.

Jean O Lanjouw, Ariel Pakes, and Jonathan Putnam. How to count patents and value intellectual property: The uses of patent renewal and application data. *The journal of industrial economics*, 46(4):405–432, 1998.

Anne Layne-Farrar and Josh Lerner. To join or not to join: Examining patent pool participation and rent sharing rules. *International Journal of Industrial Organization*, 29(2):294–303, 2011.

Joshua Lerner. The importance of patent scope: an empirical analysis. *The RAND Journal of Economics*, pages 319–333, 1994.

Mikhail Makarov. The eighth edition of the ipc. *World Patent Information*, 28 (2):122–126, 2006.

Stéphane Maraut, Hélène Dernis, Colin Webb, Vincenzo Spiezia, and Dominique Guellec. The oecd regpat database: a presentation. *OECD Science, Technology and Industry Working Papers*, 2008(2):0_1, 2008.

Fabio Pammolli, Laura Magazzini, and Massimo Riccaboni. The productivity crisis in pharmaceutical r&d. *Nature reviews Drug discovery*, 10(6):428, 2011.

Carlo Piccardi. Finding and testing network communities by lumped markov chains. *PloS one*, 6(11):e27028, 2011.

David Popp, Nidhi Santen, Karen Fisher-Vanden, and Mort Webster. Technology variation vs. r&d uncertainty: What matters most for energy patent success? *Resource and Energy Economics*, 35(4):505–533, 2013.

Pierre Régibeau and Katharine Rockett. Innovation cycles and learning at the patent office: does the early patent get the delay? *The Journal of Industrial Economics*, 58(2):222–246, 2010.

Lori Rosenkorf and Atul Nerkar. Beyond local search: Boundary-spanning, exploration, and impact in the optical disc industry. *Strategic Management Journal*, 22:287–306, 2001.

Chandra Nath Saha and Sanjib Bhattacharya. Intellectual property rights: An overview and implications in pharmaceutical industry. *Journal of Advanced Pharmaceutical Technology & Research*, 2:88–93, 2011. doi: 10.4103/2231-4040.82952.

Frederic M Scherer. Pharmaceutical innovation. *Handbook of the Economics of Innovation*, 1:539–574, 2010.

Joseph Schumpeter. Creative destruction. *Capitalism, socialism and democracy*, 825:82–85, 1942.

Joseph A Schumpeter. *Theory of economic development.* Routledge, 2017.

William G Shepherd and Joanna M Shepherd. *The economics of industrial organization.* Waveland Press, 2003.

Daniela Silvestri, Massimo Riccaboni, and Antonio Della Malva. Sailing in all winds: Technological search over the business cycle. *Research Policy*, 47(10): 1933–1944, 2018.

Jasjit Singh. Distributed r&d, cross-regional knowledge integration and quality of innovative output. *Research Policy*, 37(1):77–96, 2008.

Mariagrazia Squicciarini, Hélène Dernis, and Chiara Criscuolo. Measuring patent quality. 2013.

Manish K Srivastava and Devi R Gnyawali. When do relational resources matter? leveraging portfolio technological resources for breakthrough innovation. *Academy of Management Journal*, 54(4):797–810, 2011.

Jessica C Stahl. Mergers and sequential innovation: evidence from patent citations. 2010.

Deborah Strumsky and Jose Lobo. Identifying the sources of technological novelty in the process of invention. *Research Policy*, 44:1445–1461, 2015.

Manuel Trajtenberg. A penny for your quotes: patent citations and the value of innovations. *The Rand journal of economics*, pages 172–187, 1990.

Manuel Trajtenberg, Rebecca Henderson, and Adam Jaffe. University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and new technology*, 5(1):19–50, 1997.

Dennis Verhoeven, Jurriën Bakker, and Reinhilde Veugelers. Measuring technological novelty with patent-based indicators. *Research Policy*, 45(3):707–723, 2016.

WIPO. World intellectual property indicators. Technical report, 2022. DOI:10.34667/tind.47082.