WORKSHOP ON REGIONAL AND URBAN ECONOMICS

October 18th-19th 2012

Barcelona

Knowledge spillovers and the geography of duplicated inventions: an analysis from patent citations

Stefano Baruffaldi
École Polytechnique Fédérale de Lausanne
stefano.baruffaldi@epfl.ch

Julio Raffo
World Intellectual Property Organization
julio.raffo@wipo.int

October, 2012

Preliminary version, please do not quote

Keywords: duplication, cumulative innovation, competition, knowledge spillovers, patent citations

JEL:

Abstract

Duplication in R&D is an increasingly interesting phenomenon whose determinants remain largely unexplored. In innovation and economics literature the occurrence of duplicated inventions is mainly treated as a random outcome. We identify and discuss two dynamics that might lead to a duplication of an invention: unawareness and competition. Those are affected by knowledge diffusion and competition incentives, which in turn are both correlated with geographic and time distance among inventors. Therefore we argue and show empirically that the distribution of duplicated inventions in space and time is not random. To test our hypothesis we exploit data recently available in EPO patent bibliographical data which allows identifying trough patent citations whenever a claimed invention duplicates an existing one, according with an examiner. Geographic distance decreases the probability of duplication for recent inventions, for which the incentive to compete is reasonably higher. On the contrary, duplication of less recent inventions occurs more probably at long distances, as a consequence of lack of knowledge regarding the existence of a technology. Partial sectorial differences are encountered and discussed. Our results have implication for the literature on agglomeration, knowledge spillovers and patent system. Our methodology contributes to the debate on the meaning on patent citations.

1. Introduction

The invention of the telescope was claimed by Galileo in 1609, but was also claimed by Della Porta in 1558, Digges in 1571, and Johannides, Metius, Drebbel, Fontana, Jansen, and Lippershey in 1608. Again Galielo claimed the invention of the thermomether around 1592, but it was later claimed also by Van Guericke and Porta (1606), Drebbel in 1608, Sanctorious (1612), Paul and Fludd in 1617. Sir Joseph Swann and Thomas Edison both worked on and solved the problem of electric light. Similar examples involve the invention of the telegraph, the telephone, the electro-magnetic clocks, the typewriter, the discovery of oxygen, the periodical classification of the chemical elements, the Diesel engine, the jet propulsion, and many others. The common sense leads to imagine inventions and discoveries primarily as the unique product of one or a team of innovative inventors and scientists. On the contrary these and other examples leaded Merton to the provocative hypothesis that "far from being odd or curious or remarkable, the pattern of independent multiple discoveries in science is in principle the dominant pattern".

From an economic perspective duplication in research and innovation is potentially a serious matter of concern for firms and policy makers (Dasgupta & David, 1994; Foray, 2009; Jorde & Teece, 1990; Scotchmer, 1991). If knowledge is a non-rival good, the replication of efforts for its creation can be considered as socially suboptimal. Therefore the phenomenon of duplication turns to be of interest for the comprehension of scientific and technological progress dynamics as well as in order to assess the link between investment in R&D and economic growth. The possibility that R&D effort lead to overlapping contributions to technology, leading to diminishing returns on R&D investments, challenges the hypothesis of increasing return based on the notion of knowledge spillovers (Gómez, 2011; C. I. Jones, 1995; C. I. Jones & Williams, 2000; Jones, 2009; Kortum, 1993).

Also, while it was claimed that at least in science the use of technologies for the diffusion of knowledge and information should decrease the rate of duplication (Brannigan & Wanner, 1983), other determinants and evidence suggest that this rate could be increasing in time, especially in technology and R&D. The probability of a duplication increases with the density of inventors in a certain sector. Similarly, the cumulative nature of knowledge makes harder for future generations of inventors to propose novel innovations and discoveries (Jones, 2009). Finally there are reasons to believe that the function of the patent system as a mechanism of knowledge disclosure is severely deteriorating over time. Notably, Bessen & Meurer (2008) find that the number and cost of patent lawsuits has constantly increased over the last 30 years and conclude that "..a significant and growing number of very expensive lawsuits occur each year because firms have invested millions of dollars on the research, development, and commercialization of technology that is legedly owned by others" (Bessen & Meurer, 2008, pg 121).

Few literature contributions have addressed the characteristics and determinants of this phenomenon. We start from the assumption that the occurrence of duplicated inventions is intimately related with the way knowledge flows and economic agents communicate and interact. We refer to the literature on geography of innovation to argue that since knowledge diffusion and knowledge spillovers are geographically and socially bounded, duplications are not distributed randomly in time and space (Feldman & Kogler, 2010; S Breschi & F Lissoni, 2001; Stefano Breschi & Francesco Lissoni, 2009; Jaffe, Trajtenberg, & Henderson, 1993). Therefore we investigate an unexplored link between geography and the occurrence of duplications. Relying on the existing literature on cumulative innovation (Dasgupta & David, 1994; Murray & O'Mahony, 2007) and the equivalent on innovation and patent competition (Aghion, Bloom, Blundell, Griffith, & Howitt, 2002; Denicolò, 1999; Encaoua, Ulph, & others, 2005), we propose to distinguish between two different mechanisms leading to duplication. On the one hand, duplication may arise from an imperfect diffusion of information. As such, laggard agents duplicate inventions without knowing the existence of the original ones. In other words, these uninformed inventors just 'reinvent the wheel'. On the other hand, duplication can also occur if we assume perfect information flows among agents. This is the case of agents consciously competing for the same technological solutions. For this reason, we refer to this case as competitive duplication. In each of these cases, the occurrence of duplication in a certain time and location can be assumed to be not random but rather related to the presence or absence of knowledge spillovers and diffusion Furthermore we build on the assumption that, given the knowledge disclosure in patent documents and patent protection is less than perfect (Atal & Bar, 2010; Walsh, Cohen, & Cho, 2007), a duplicated invention can be captured in patent data whenever a patent application is filed for an invention which is not novel compared to already existing (patented) inventions. Therefore we show how the geographical distribution of duplicated inventions is evident in patent data.

The empirical analysis makes use of recent EPO patent bibliographical data, where detailed information on the citations is available from search reports and examination. In this database, it is possible to distinguish those X, Y or E citations, which reflect according with an EPO examiner that the cited patent document compromises the patentability of the citing patent application. Such information has been used in recent contributions from innovation economics to track the lack of innovativeness of certain technologies (Criscuolo & Verspagen, 2008; D Guellec, Martinez, & Zuniga, 2009). We use these categories of citations as indicators of duplication, being able to link the original and the duplicated invention.

Our results show both competitive and independent duplication to be apparent in patent data. This provides a direct contribution to the literature on knowledge spillovers and their role in agglomeration economies. But it also offers interesting insights for the patent system and the policy related with it. Furthermore, we contribute as well to the literature on patent citations (Alcàcer, Gittelman, & Alcacer, 2006; A B Jaffe, Trajtenberg, & Henderson, 1993; Adam B Jaffe, Trajtenberg, & Fogarty, 2000; Lampe, 2007) by offering an interpretation on a meaningful portion of patent citations of and showing its geographical distribution accordingly.

The remaining is organized as follows. The next section introduces the theoretical base of our contribution. The third section enumerates the hypothesis to be empirically tested. Section 4 addresses in detail the data used and proposes an empirical approach. Section 5 presents the empirical results and final section concludes.

2. Knowledge, competition and duplicated technologies.

The duplication of research efforts is the antithetic phenomenon of the process of knowledge accumulation (Murray & O'Mahony, 2007). It has been said that, when scientists and inventors innovate, they "stay on the shoulders of others" (Merton, 1979). From this perspective, science and technological progress relies on innovators being aware and mastering the existing knowledge before improving them by adding new R&D efforts (Jones, 2009). Nonetheless, "accumulation of knowledge is not inherent to the innovation process" (Murray & O'Mahony, 2007). Inventors, firms and innovative regions struggle to reach and keep the technological frontier and seek to avoid duplication of existing inventions, which would make their investments in R&D valueless (Archibugi, 1992; Jorde & Teece, 1990). At the same time, they compete when trying to be the first to introduce an innovation in the market, as well as when trying to erode the technological advantage of competitors.

Therefore, it is possible to identify two main factors – knowledge and competition – both leading to two different kinds of duplication. The lack of awareness regarding the existence of a certain technology may be the cause of duplication. The access to the knowledge related with the state of the art in a particular sector or field is crucial in this respect. An agent with access to this type of knowledge is expected to be aware of the existing technologies and be able to formulate more accurate evaluations of future opportunities. This affects the probability of posing the "right problem at the right moment", that has not been solved before or is not being solved by someone else. Diffusion failures of this type of knowledge may lead to duplicative efforts for an invention, driven by the wrong belief that it does not exist. It is possible to define an *independent duplication* (ID) as an invention which is duplicated without the awareness of its existence and without the full awareness of the risk to replicate others' research efforts. The occurrence of independent duplication is therefore negatively related with the diffusion of knowledge so that we can state the following:

> *The diffusion of knowledge increases the awareness about the existing (or upcoming) technologies and makes less likely that these are involuntarily replicated, i.e. to incur in an independent duplication (ID).*

The perfect diffusion of knowledge provides the opportunity to give up valueless efforts, improve on the existing inventions or specialize in complementary and differentiated technologies. Nonetheless, an inventor might be willingly to compete against other inventors on the same technological space. The possibility to appropriate the value of an invention – for instance, with a patent, creating market barriers or exploiting a first mover advantage – creates the incentive to compete for the same technology regardless someone else pursing the same invention or, a fortiori, if a direct competitor is likely to reach it. Similarly, the inventor might want to try to "invent around" the existing inventions of a competitor in order to erode its technological advantage.

These dynamics are richly discussed in the patent and patent races literature (e.g. Chang, 1995; Gallini, 1992). These technological races may speed the inventive process if the outcome is such that the 'winner takes all', allowing society to benefit from new technologies earlier than without competing agents. Similarly, these incentives might be beneficial if the laggard inventor brings significant improvements or diversity to the leading technology. Nevertheless, "inventing around" in highly competitive context might lead to inefficiency. For instance, inventors can try to get property rights for very similar technologies, with small improvements or marginal changing, only seeking to limit the scope of the competitor's rights, generating higher costs for society without much benefit.

We call this type of duplication *competitive duplication* (CD) when the inventor is aware of replicating others' research and voluntarily engages in this effort because there are sufficient incentives to do it. When the incentives to compete are high, the presence of knowledge spillovers can increase the likelihood of duplication. Knowledge spillovers regarding an upcoming invention increase the opportunity cost to engage in original research efforts against the possibility to anticipate other inventors on the same technology. In a patent race, the follower can try to catch-up with the leader during the development of the invention, thanks to partial knowledge spillovers (Encaoua et al., 2005). Likewise, the presence of knowledge flows regarding the existing technologies of competitors increase the possibility for an inventor to try to "invent around" these inventions (Guellec et al., 2009). Therefore we can state that:

Higher knowledge spillovers increase the incentives to compete for existing (or soon to exist) technologies, making more likely they are replicated, i.e. incurring in a competitive duplication (CD).

3. Duplication of technologies in the patent system

Disclosure of knowledge, e.g. in patents, scientific or informative documents, is a necessary condition to allow for cumulative innovation (Dasgupta & David, 1994). The patent system is considered one of the main tools to increase the potential for accumulative innovation and avoid duplications. A patent assigns exclusive property rights to the inventor preserving the economic incentives to create new technologies. But it requires in exchange the disclosure of the technical knowledge in it, which allows others to build on it instead of replicating it (Denicolo & Franzoni, 2003; Kitch, 1977). Nonetheless, it has been noticed that this disclosure is not a sufficient condition. Literature on knowledge flows and knowledge spillovers has widely discussed the limit of codified knowledge to convey the tacit content related with it (Feldman & Kogler, 2010). Furthermore the grant of property rights trough the patent system creates the incentive to compete.

Mahoney and O'Mahony (2007) distinguish disclosure, access to knowledge and rewards as conditions to assure cumulative innovation. They underline the difference between the existence of a document disclosing technological knowledge and the possibility to actually have access, master and improve on that knowledge. Furthermore the framework of institutional incentives and system of rewards within a certain organization or context affect both the incentive to compete instead of cooperate and willingness of an inventor to disclose and share his knowledge, also providing additional information, tools and materials to improve on his existing inventions (Furman & Stern, 2011; Murray & O'Mahony, 2007). We expect inventors within the same firm to be subject to incentives, at personal and institutional level, to cooperate to avoid duplications of efforts and improve on the works of other colleagues. The company is also likely to provide information systems of knowledge management to diffuse knowledge within the firm, in order to create the preconditions for cumulative innovation. Therefore we formulate the following hypothesis:

H1: Duplication (ID and CD) is less likely to occur within the same firm.

It can be argued that an organizational and incentive framework to favor diffusion of knowledge is generally missing beyond the firm's boundaries.

First, we expect time to have a direct effect both on the diffusion of knowledge and the existence of incentives to compete on a certain technology. In the one hand, the more time passes after an invention, the more information about its existence and contents diffuses through firms, regions or social networks. More importantly, time allows the distribution of goods and services that make use of that technology, which is known to be a critical technology diffusion channel(Keller, 2004). In the other hand, it can be argued that incentives to compete decreases when time passes, as it becomes more difficult to catch up with the leader. Similarly, an old technology is likely to have saturated its commercial value, making it less attractive to invest on it. Therefore, both independent and competitive duplications are aligned with respect to time, where:

H2: Duplication (ID and CD) is less likely to occur along time.

Second, we argue that geographic proximity has a role in explaining duplication. It has been shown that knowledge spillovers among inventors are related to geographic proximity. It increases the likelihood of informal and face to face contacts. Social and professional networks through which information flows more efficiently are to a good extent locally based. Empirical evidence has shown that inventors use local information or knowledge to create novel products and processes (Feldman & Kogler, 2010; Giuri et al., 2007). Finally inventors can monitor each other more closely and actively when they are close geographically searching for information which would not be otherwise available or that the counterpart would not be willingly to disclose. Nonetheless we observed that the effect of knowledge spillovers on duplication is twofold. We use time to disentangle the effect on independent duplication (ID) and competitive duplication (CD). As already discussed, in the short period, for recent and upcoming technologies, the incentive to compete is high and knowledge less diffused. Therefore agents sharing the same pool of knowledge are more likely to compete and replicate the same efforts. So we say:

H3: Duplication (CD) is more likely to occur close in space and close in time.

On the contrary when a technology is not recent the incentive to invest on it fades unless there is unawareness about its existence and its characteristics. As we observed, also the knowledge related with its existence diffuse over time. Nonetheless we can argue that if duplication occurs for a technology which is not recent this has to be due to a lack of information and knowledge, therefore most probably far from the location where the technology was developed. Therefore:

H4: When far in time, duplication (ID) is more likely to occur far in space.

4. Data and methods

Patent citations and duplicated inventions

Traditionally patent citations have been used as a proxy for knowledge flows occurring among inventors. However, already Jaffe et al. (1993) noted how this indicator could be noisy given the presence of examiner citations and citations added for different scopes. More importantly, recent debate on the use of patent citations has acknowledged that not all patent citations are appropriate indicators of knowledge flows (Alcàcer et al., 2006; Alcácer, Gittelman, & Sampat, 2009; Criscuolo & Verspagen, 2008). In particular, Breschi and Lissoni (2005) directly point to the possibility that patent citations refer to duplicative efforts. Examiners have to verify the novelty of an invention compared to existing state of the art in the public domain. Whenever the examiner considers a piece of knowledge as a proof of lack of novelty of the claimed invention, this prior element – typically a document, but not only – has to be cited in the search or examination report. Recent EPO data allows us to identify where a citation comes originally from – i.e. application, search report, examination, opposition, etc. – and also what it stands for. It has to be noted that it is always the EPO examiner who categorizes citations – e.g. as prior art with (X) or without (A) effect on claims – regardless if the citation was already in the original applications. Therefore duplication of inventions can be recorded in patent documents as citations to the original invention when the examiners have categorized the citations accordingly.

-- Insert Figure 1 about here --

It is worth noting that not all duplicated inventions are likely to be captured as citations among overlapping patent documents. First, many inventions are not patented or published. In this case, it is virtually impossible to identify in a systematic way if an invention has been duplicated. Second, a patent application has also to be filed for the replicating invention. In principle, the existence of a patent should discourage the second inventor to file her patent, particularly if it is too similar to the existing one. Therefore, the inventor of the second coming invention can renounce to patent it and the duplication would be unobservable through patent data.

Nonetheless, evidence suggests that this is likely to be the exception. First, if two inventors arrive at the same invention very close in time, it is probable that the patent application for the first invention will not be published yet as EPO takes 18 months from the filing date to publish it. This would be the case of an E citation category. Second, in a context of competition the incentive to file a patent in order to reduce the competitor's scope remains. Third, there is evidence suggesting that inventors remain largely unaware of the existence of patents even if relevant for their projects or related with the knowledge and technologies they are actually using (Walsh et al., 2007). The incentives for the inventor to perform a patent search before and after developing is invention can be low (Atal & Bar, 2010). Examiners have added around 90% of the citations in EPO patent documents. Similarly, other empirical contributions show that patent literature remains in general a limited direct source of knowledge for inventors. Furthermore even in cases when a relevant patent is discovered the research projects of inventors remains often unchanged. Therefore, it is likely that if an inventor has developed a technology with the intention to patent it, she will file a patent application regardless the existence of a similar patent. Whenever an examiner identifies the prior patent we are able to observe a citation linking the two inventions. The process is summarized in figure 1.

-- Insert Table 1 about here --

EPO provides its examiners with precise guidelines on how distinguish citations in several categories[1]. The most relevant for our study are summarized in table 1. The category A corresponds to the typical citation, which describes the state of the art relevant and embedded in the citing patent document without compromising the novelty or inventive step requirements. On the contrary, Y X and E citations refers to citations affecting the patentability of the citing application. Y-cited documents are different from X and E as they refer only to the lack of inventive step and they always need to be combined with at least another citation. On the contrary, each X or E citation is enough to challenge patentability of the citing document. The only difference between E and X citations is that the former links documents very close in time, where the citing application was filed between the filing and the

[1] See "EPO guidelines for Examination in the European Patent Office", http://www.epo.org/law-practice/legal-texts/guidelines.html

publication dates of the cited one. As such, we consider X and E categories as the main indicator of duplication, where the citing application is assumed to replicate the X or E cited patent document. Finally, it is possible to distinguish citations present in the original document from those added by the examiner. As mentioned before, only examiners categorize citations, making all citations relevant for our analysis regardless of the origin.

Data

The sample is built from the patent citations data from EPO's Worldwide Patent Statistics Database (PATSTAT, September 2010) and the information about localization of inventors from the OECD's REGPAT Database (December 2010). Additionally, each NUTS 3 region has been geo-localized in order to construct distance measure between citations. Unfortunately, PATSTAT contains citations categorized mostly for EPO patent documents only. Similarly, REGPAT contains the localization of inventors only those from EPO and PCT patent documents. This means that our sample has to be circumscribed to EPO patent documents citing EPO patent documents (EP-EP). The final sample contains 994,193 EP-EP citations, for a period from 1982 to 2007.

Descriptive statistics on citation categories

We report the percentage of each citation category in our sample in table 2. There is a high share of X (26%) and Y (13%) citations, although a smaller one of E (1.8%) citations. More than half of the patent documents in our sample have received at least one X (56%) and one third received an Y (33%) citation. Also the share of E citations (4%) is relevant if we take into account that these are citations occurring only in the first 18 months after the filing date of the cited patent. It is important to highlight that the presence of these citations does not directly imply that it won't be granted (Tan & Roberts, 2010). Nonetheless, it certainly increases the probability of a rejection or can alternatively result in a reduction of claims in the granted patent, with a direct negative impact on the private value of the patent for the applicant (Dominique Guellec & van Pottelsberghe de la Potterie, 2000). As a further indication other studies suggests that the number of non-granted patents is also a considerable high and it is increasing over time.

The large majority of X citations (92%) and the totality of E citations are added by the examiner. Nonetheless, a considerable number of citations are already present in the original application (15%) and later categorized by the examiner as X. There are many reasons why X citations may appear in the original document. First, these citations could have been added after the development of the project, probably during the process preparing the patent application. If this is the case then they would not be conceptually different from X citations added by the examiner. Second, if the X citation was known before the development of the invention, we must argue that the inventor tried to improve the related technology without success (according with the examiner). Nevertheless, under EPO rules, the citations introduced by the examiner do not necessarily mean that the inventor was not aware of their existence. On a similar note, what it is often called an inventor citation might not be the case. Not only inventors, but also applicant, patent attorneys and others might be behind a patent citation.

-- Insert Table 2 about here --

On the contrary, applicants add more often Y citations compared to the examiner. The explanation of this evidence relies on the fact of Y-cited documents being relevant for novelty only if combined with other documents. Therefore, one possible case is that the invention was inspired by the knowledge of the Y cited document but it is not novel compared to a third related invention of which the inventor was not aware or that she decided not to cite. Similarly the examiner can be triggered from these citations added by the applicant to search for certain documents that later imply the lack of novelty of the invention (Criscuolo & Verspagen, 2008).

-- Insert Figure 2 about here --

Finally, figure 2 shows the amount of patents which have received at least a citation of a certain category over years. The same graph for the share of citations would present a similar trend. It is evident how the number of X citations increases dramatically over the years while the share of A citations decreases. If an increase in the number of examiner citation is generally encountered (Criscuolo & Verspagen, 2008), this is probably strongly related with this increase in citations relevant for the novelty of the patents. From one hand one might be tempted to conclude that this is due to an increase of duplications of inventions. Nonetheless many factors can contribute to this result: e.g. a more rigid and precise examination by the patent examiners along time, a change in the patenting strategies of firms. Addressing this specific topic and this issues go beyond the focus of our present analysis. However we conclude this descriptive section pointing out that the presence of overlapping patent is a consistent and increasing phenomenon in patent data. Citations identifying and linking these patents offer the opportunity to study the determinants of duplication in R&D activities.

Model

In our analysis we want to study the probability to observe a duplicated invention. We frame the empirical approach as the probability that a certain invention is replicating another one with respect to their geographical and temporal distance.

It is important to note that we cannot refer to all possible pairs of inventions as most of them are totally unrelated technologies. A random comparison of two inventions in our sample will completely skew the probability of observing duplication. Therefore, we need to limit our sample to those pairs of related technologies. In other words, we study the probability of duplication as conditional on the actual distribution of the relevant knowledge for a certain invention. In practical terms, we analyze the probability to observe a citation X or E with respect to observing a different one (mainly A citations) for each citing patent application. Hence, our dependent variable is a binary dummy taking the value of one if the citation linking two patents is an X or E and zero otherwise[2].

A second challenge is to handle the heterogeneity across citing patent applications. This can result from many factors. For instance, we can expect spillovers to differ across industries or technological fields, with respect to concentration, practices, etc. Similarly, we can expect to observe heterogeneity on the specific technology, such as patentability or quality. In order to avoid these sources of bias, we control for any fixed effect from the citing patent application. This will not only control for sector and patent specific heterogeneity, but also for any trend on duplication with respect to filing date.

Subsequently the model is specified as follows:

$$P\left(Y_{ij} = 1 \mid X_{ij} \right) = \beta_0 + \beta_1 A_{ij} + \beta_2 I_{ij} + \beta_3 T_{ij} + \beta_4 G_{ij} + \beta_5 T_{ij} \times G_{ij} + a_i + \varepsilon_{ij}$$

Where the left hand side represents the probability of the citation from patent document *i* to patent document *j* is categorized as X or E conditional to a set of independent variables (X_{ij}) relating to the pair of patent documents. On the right side we have the parameterization of it as linear function of X_{ij}, where β_k are the parameter of interest, a_i are the fixed effects and ε_{ij} the error term. Within X_{ij}, A_{ij} accounts for those pairs of patent documents from the same applicant and I_{ij} for those from the same inventor; T_{ij} stands for the filling time distance between documents; and, G_{ij} refers to the geographical distance. Additionally, a term is added to capture the interaction between time and geographical distance.

For robustness purposes, we specify the geographic distance in different manners. First, we consider it as a continuous variable measuring the minimum great-circle distance – in kilometers – between all possible pairs of inventors across patent documents. Nevertheless, results were consistent for the average or maximum distance of such pairs. As an alternative, we consider geographical distance as a set of three dichotomous variables describing if at least one of the possible pair of inventors come from the same country, the same NUTS2 region or the NUTS3 region, respectively.

The model has been specified as a linear probability model and estimated accordingly, although the alternative of conditional logit has been also tested. We keep the former to allow a direct interpretation of the coefficients.

All variables used are summarized in table 3.

-- Insert Table 3 about here --

[2] We do not consider Y citations in the dependent variable. The fact that Y-cited documents shows lack of inventive step only if combined with others creates ambiguity. Nonetheless our results are robust to include Y citations in the dependent variable or even when fully excluded from the sample.

5. Results

Table 4 reports five variations of the model described above. Model 1 studies the effect of distance without including the interaction effect with the time lag. Model 2 introduces the interaction effect. Model 3, 4 and 5 replicate the estimation in Model 2 for the different measures of distance, respectively: same country, same NUTS2 region and same NUTS3 region. In table 5 we report mainly the same model for a reduced sample without self-citations.

In accordance with our first hypothesis, citations from the same applicant have a consistently significant and negative effect on the probability of duplication across the different specifications. Puzzlingly, the effect of citations from the same inventor is significant and positive. While it is odd that an inventor replicates her own work, the result can be explained by the inventor's willingness to multiply patents for the same technology, as well as the inventor's difficulties to create radically new technologies.

-- Insert Table 4 about here --

As expected, we also find a significantly negative effect of time distance between documents. This is a robust result across all our models and confirmed in the composite marginal effect analysis detailed below. Roughly speaking, each year that passes between the two inventions reduces little more than 1% the likelihood of duplication.

Similarly, geographical distance shows an overall positive effect on duplication (see Table 4, model 1). This suggests a dominance of the independent duplication over the competitive duplication with respect to geographical distance, as depicted in hypothesis H2a. Nevertheless, the effect is not completely robust across specifications and, more importantly, the magnitude of it is negligible. For instance, the maximum distance registered in our sample – little less than 20 thousand km – increases as much as 1.1% the likelihood of duplication.

Finally, when introducing the interaction effect we obtain evidence supporting both H4a and H4b. In particular, the coefficient is negative for distance alone, meaning that the effect of distance is negative for citations close in time. In other words, when the time lag between the filing dates of patents is short, distance decreases the probability to observe duplication. On the contrary the coefficient on the interaction effect is positive, meaning that for patents which are not close in time, distance does increase the probability of duplication. The result is robust both without considering self-citations and for the different measures of distance. In model 5 of table 5, the coefficient on the dummy representing citations within the same NUTS3 region is positive but not significant, suggesting that the "competition effect" is more appreciable for medium distances than it is for very short distances.

-- Insert Table 5 about here --

In order to confirm our results we computed two sets of marginal effects for each model in table 5 (model 2, 3, 4 and 5). First, we computed the marginal effects of distance on the probability to observe duplication for the range of time distance in years in our sample. Second, we computed the marginal effect of time as a function of geographical distance.

$$\frac{\partial P(Y_{ij} = 1 \mid X_{ij})}{\partial G} = \beta_4 + \beta_5 T$$

$$\frac{\partial P(Y_{ij} = 1 \mid X_{ij})}{\partial T} = \beta_3 + \beta_5 G$$

This latter estimation for time lag marginal effect confirms further H2 showing that the effect of time, despite being less negative over distance, is always significantly negative for the geographic distances in our sample[3]. To some extent this trend is coherent with the idea that the probability of duplication decreases over time but decreases faster at short distances than at long distances due to the effect of knowledge diffusion. Finally, table 6 shows the marginal effect of geographic distance for different time lags. These estimations confirms the results discussed and allows to be more precise: the effect of distance appears to be significantly negative for a distance in time from 0 to 2 years, from 2 to 4 years is negative but not significant, from more than 4 years on it is positive but significant only from a time distance of 6 years or more (from Table 6, estimations for Model 2).

-- Insert Table 6 about here --

5.1. Technological heterogeneity

In order to understand the technological heterogeneity we tested our main model in different subsamples. First, we split the main into two differentiating complex and discrete technologies[4]. These are depicted in Table 7 as models 1 and 2, respectively.

Second, we estimated our regressions for all the 35 different sectors of the citing patents in our sample. 21 sectors presented the same (significant) or coherent results (equal in signs but weakly or only almost significant) with those discussed. These sectors account for roughly the 70% percent of citations in our sample. 4 sectors, accounting for around 14% of the sample, presented no effect or only a positive and weak effect of distance. Finally the remaining 10 sectors, for a 16% of the sample,

[3] Result are not shown, but available upon request
[4] Note: This classification follows G. von Graevenitz, S. Wagner and D. Harhoff (2008), "Incidence and Growth of Patent Thickets - The Impact of Technological Opportunities and Complexity", CEPR Discussion Paper No. DP6900

showed coefficients opposite in sign, in most of the cases not significant but in some weakly significant. Models from 3 to 5 show one example for each of these three cases, where semiconductors represent the case where our results are confirmed, mechanical elements those where distance show a positive and weak effect and basic materials chemistry those where coefficients are found with the opposite in sign.

Those sectors where results were found less consistent with our hypotheses, appear to be quite concentrated in chemistry and mechanical engineers sectors. We bring three main possible explanations for this outcome. First, to the extent to which patent documents can be an actual and direct source of knowledge – which is expected in the case of such sectors – inventors might be able to monitor constantly the patent literature in order to evaluate technological opportunities and avoid costly applications for patents. In this case not only it would be harder to observe duplications in the patent data but both time and geographical proximity would be less relevant in the diffusion of knowledge. Second, sectors characterized by high concentration – like the case of big multinationals in chemistry and pharmaceutical sectors – may be competing and sourcing for knowledge on a global scale, as opposed to local spillovers. Again the geography dimension would be less relevant in this case, fading our results. Third, our results appear to be robust for sectors were patenting has been also an active arena for competition (e.g. semiconductors). In sectors were also other strategies are used, results might be biased against our hypothesis. Finally, our assumptions are built on a concept of innovation as a cumulative and sequential process. Despite there are reasons to believe that is increasingly the case in several sectors, this assumption can be more or less valid depending on the nature of the technology involved. To conclude, our results are robust for the majority of the industries considered in our sample. Nonetheless these differences constitute both a limitation in our estimations and an interesting insight for future investigations.

-- Insert Table 7 about here --

6. Robustness

In Table 8, we test if our main model on two different subsamples: one only with citations added by the examiner (Model 1) and the other with only citations added by the applicant (Model 2).. These results are qualitatively equivalent if considering self-citations or not. Therefore we report only the estimations for the sample with both self and non self-citations.

The results are unchanged for the sample with only examiner citations. In the case of inventor citations on the contrary, we do not find a negative effect of distance for citations short in time, in our framework we lose the effect of competition (Model 2). This result can be explained in different ways. : 1. As we mentioned inventors racing for a technology might be not aware of the existence of a patent or of a patent application being filed by another inventor, because this is not published yet or only recently published; 2. Even when aware of an existing document, inventor and applicant can be less incline to add citations to patent of competitors, especially if they are directly competing for a similar technology. Therefore our results suggests that if an inventor add an X citation (which we remind being the case in the 2% of all citations and 15% of inventor citations) this is probably the result of an independent duplication and of prior art search posterior to the development of the invention. Alternatively it could also be the case of an attempt to improve an existing and known technology, which, according to our results, is more likely to fail when the original source of knowledge is far in time and space, since this would prevent further flows of knowledge a part from the patent document.

-- Insert Table 8 about here --

7. Conclusions

To our knowledge this work is the first attempt to study empirically the determinants and occurrence of duplicated inventions. Our results show a clear pattern in the geographic and temporal distribution of duplicated inventions. These evidences provide support for both the existence of a large number of independent and competitive duplications in R&D, in Europe. Duplication occurs more likely at short distances and within national boarders for recent technologies. On the contrary, the effect of distance is inverted for not recent technologies: longer distance among inventors increase the probability that these replicate existing inventions. Importantly the effect of knowledge spillovers (proximity) is twofold: they increase duplication in presence of competitive incentives and on the contrary decrease it when there are no incentives to compete. Noteworthy, if the disclosure in patent were supposed to be perfect, there would be no space for our hypothesis. Therefore the evidence we report rise further concern regarding the potential of patents as a mean of knowledge disclosure.

Finally our methodology constitutes a contribution to the debate on the meaning of patent citations. Our interpretation is mainly based on the definition of the patent citation categories. In our framework X or E citations are by definition the indication (based on the opinion of an examiner) of an overlap between two inventions which might be or not the result of knowledge spillovers. Our results are coherent with the notion of knowledge spillovers being geographically localized but cast further doubts on the use of patent citations as direct indicators of knowledge flows. In this sense we confirm that examiner citations can present significant geographical pattern with respect to the inventor location (Alcàcer et al., 2006) and our theoretical framework explain the reason of this distribution. If one can still argue that inventor citations can be correlated with knowledge flows, this is probably truer for inventions that do not compromise the novelty of the patent. Also, knowledge spillovers coming from direct competitors are more likely to be captured in examiner citations. This is not only because applicants might be more reluctant in adding these citations (Lampe, 2007), but also because development of technologies among competitors is likely to be simultaneous in time As a consequence it remains quite ambiguous which category is more appropriate as an indicator of knowledge flows.

Limitations of our analysis concern first fall the lack of a direct measure of the awareness of a certain inventor regarding the existence of a certain invention or of a competitor working on it (Lampe, 2007). Second of all, we intend to include in future analysis a measure of the social proximity of inventors which is likely to affect both access to knowledge spillovers and competitive incentives

(Breschi & Lissoni, 2009). Finally, our results show differences across sectors that deserve further attention.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2002). Competition and innovation: An inverted U relationship.

Alcàcer, J., Gittelman, M., & Alcacer, J. (2006). How do I know what you know? Patent examiners and the generation of patent citations. *Review of Economics and Statistics*, *88*(4), 774–779.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, *38*(2), 415–427. doi:DOI: 10.1016/j.respol.2008.12.001

Archibugi, D. (1992). Specialization and size of technological activities in industrial countries: The analysis of patent data. *Research Policy*, *21*(1), 79–93. doi:10.1016/0048-7333(92)90028-3

Atal, V., & Bar, T. (2010). Prior art: To search or not to search. *International Journal of Industrial Organization*, *28*(5), 507–521. doi:10.1016/j.ijindorg.2009.12.002

Bessen, J., & Meurer, M. J. (2008). *Patent failure: How judges, bureaucrats, and lawyers put innovators at risk*. Princeton Univ Pr.

Brannigan, A., & Wanner, R. A. (1983). Multiple Discoveries in Science: A Test of the Communication Theory. *The Canadian Journal of Sociology / Cahiers canadiens de sociologie*, *8*(2), pp. 135–151. Retrieved from http://www.jstor.org/stable/3340123

Breschi, S., & Lissoni, F. (2005). Knowledge Networks from Patent Data. In H. Moed, W. GlÃ¤nzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 613–643). Springer Netherlands. Retrieved from http://dx.doi.org/10.1007/1-4020-2755-9_29

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*(4), 439–468. doi:10.1093/jeg/lbp008

Chang, H. F. (1995). Patent Scope, Antitrust Policy, and Cumulative Innovation. *The RAND Journal of Economics*, *26*(1), pp. 34–57. Retrieved from http://www.jstor.org/stable/2556034

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, *37*(10), 1892–1908. doi:DOI: 10.1016/j.respol.2008.07.011

Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521.

Denicolò, V. (1999). The optimal life of a patent when the timing of innovation is stochastic. *International Journal of Industrial Organization*, *17*(6), 827–846. doi:DOI: 10.1016/S0167-7187(97)00061-1

Encaoua, D., Ulph, D., & others. (2005). Catching-up or Leapfrogging: The effects of competition on innovation and growth.

Foray, D. (2009). Understanding "Smart Specialisation." *The Question of R&D Specialisation: Perspectives and policy implications'. Pontikakis, D., Kyriakou, D., and van Bavel, R.,(Editors). Joint Research Centre Scientific and Technical Report, by Joint Research Centre and Directorate General for Research. Lux.*

Gallini, N. T. (1992). Patent Policy and Costly Imitation. *The RAND Journal of Economics*, *23*(1), pp. 52–63. Retrieved from http://www.jstor.org/stable/2555432

Guellec, D, Martinez, C., & Zuniga, M. P. (2009). Pre-emptive patenting: securing market exclusion and freedom of operation. *OECD Science, Technology and Industry Working Papers*.

Guellec, Dominique, & van Pottelsberghe de la Potterie, B. (2000). Applications, grants and the value of patent. *Economics Letters*, *69*(1), 109–114. doi:DOI: 10.1016/S0165-1765(00)00265-2

Jaffe, A B, Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, *108*(3), 577–598.

Jaffe, Adam B, Trajtenberg, M., & Fogarty, M. S. (2000). The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees, (7631). Retrieved from http://www.nber.org/papers/w7631

Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *Review of Economic Studies*, *76*(1), 283–317. doi:10.1111/j.1467-937X.2008.00531.x

Jorde, T. M., & Teece, D. J. (1990). Innovation and Cooperation: Implications for Competition and Antitrust. *The Journal of Economic Perspectives*, *4*(3), pp. 75–96. Retrieved from http://www.jstor.org/stable/1942930

Lampe, R. (2007). Strategic citation. *The Review of Economics and Statistics*, (0).

Merton, R. K. (1979). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press. Retrieved from http://books.google.ch/books?hl=fr&lr=&id=zPvcHuUMEMwC&oi=fnd&pg=PR9&dq=merton+the+sociology+of+science&ots=x4NRTnhbtQ&sig=a5ZO-l4rf9ELsDK7E_JCWeT9WZ8#v=onepage&q=merton the sociology of science&f=false

Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006–1021.

Scotchmer, S. (1991). Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. *The Journal of Economic Perspectives*, *5*(1), pp. 29–41. Retrieved from http://www.jstor.org/stable/1942700

Tan, D., & Roberts, P. W. (2010). Categorical coherence, classification volatility and examiner-added citations. *Research Policy*, *39*(1), 89–102. doi:10.1016/j.respol.2009.11.001

Walsh, J. P., Cohen, W. M., & Cho, C. (2007). Where excludability matters: Material versus intellectual property in academic biomedical research. *Research Policy*, *36*(8), 1184–1203. doi:10.1016/j.respol.2007.04.006

Figures & Tables

Figure 1: Revealed duplications

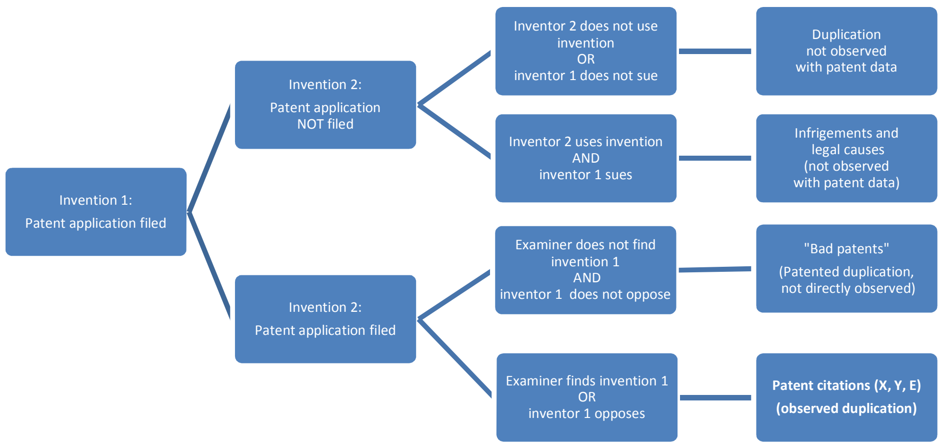

Table 1: Citations categories

Categories	*Description*
A	Documents defining the state of the art and not prejudicing novelty or inventive step
Y	Particularly relevant documents if combined with another document, such a combination proving the lack of an inventive step.
X	Citations classified under this category are such that when taken alone a claimed invention cannot be considered novel or cannot be considered to involve an inventive step.
E	Any patent document relevant for novelty (same as X citation) bearing a filing or priority date earlier than the filing date of the application searched but published later than the that date.
D	Documents cited in the original application (usually referred as to “applicant or inventor citations”).

Table 2: Citations categories shares

Column % (Total %)

	Examiner citations	*Applicant citations - D*	*All*	*% Patents with at least one cit. of the category*
A	58% (50.5%)	67% (8.7%)	59.2%	77%
Y	12% (10.4%)	17% (2.2%)	13%	33%
X	27% (23.5%)	15% (2%)	26%	56%
E	2% (1.8%)	0% (0%)	1.8%	4%
All	100% (87%)	100% (13%)	100%	-

Figure 2: Citations categories over years
(Share of patents with at least one cit. of the category)

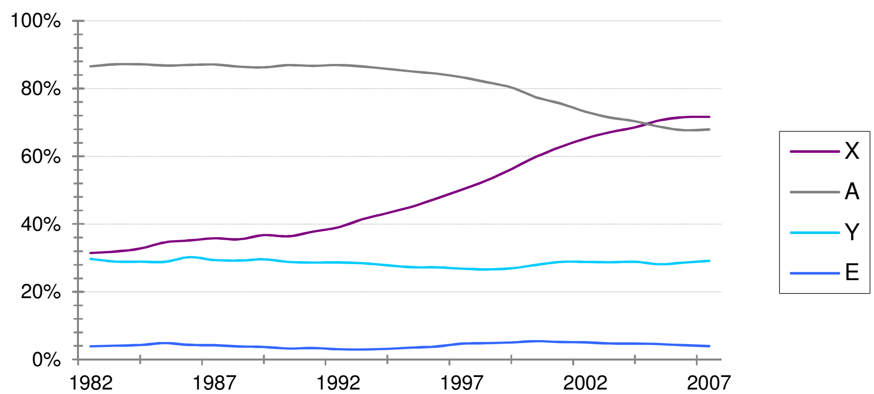

Table 3: Variable definition and statistics

Variable	*Description*	*Mean*	*Std. Dev.*	*Min*	*Max*
same_inventor	=1 if at least one inventor is the same in the citing and cited document	0.1435214	0.350604	0	1
same_applicant	=1 if at least one applicant is the same in the citing and cited document	0.2623535	0.439914	0	1
prior_diffy	Number of years between the priority dates of the two patents	4.935919	3.999751	0	31.2
distmin	Minimum distance in km for all pairs of inventors between the two patents	3370.798	4197.679	0	19915
same_ctry_any	=1 if at least two inventors, each from one of the two patents, are in the same country	0.5192171	0.4996308	0	1
same_reg2_any	=1 if at least two inventors, each from one of the two patents, are in the same NUTS2 region	0.3362516	0.4724264	0	1
same_reg3_any	=1 if at least two inventors, each from one of the two patents, are in the same NUTS3 region	0.2795091	0.4487583	0	1

Table 4: Liner probability model with fixed effects (full sample)

	Model 1	***Model 2***	***Model 3***	***Model 4***	***Model 5***
same_inventor	0.0575***	0.0535***	0.0523***	0.0475***	0.0441***
	(0.00225)	(0.00227)	(0.00228)	(0.00235)	(0.00241)
same_applicant	-0.0117***	-0.0139***	-0.0107***	-0.0121***	-0.0157***
	(0.00196)	(0.00197)	(0.00202)	(0.00222)	(0.00223)
prior_diffy	-0.0113***	-0.0136***	-0.00829***	-0.00895***	-0.00917***
	(0.000181)	(0.000232)	(0.000231)	(0.000203)	(0.000197)
Distmin	5.50e-07***	-2.71e-06***			
	(1.80e-07)	(2.73e-07)			
c.prior_diffy#c.distmin		6.00e-07***			
		(3.77e-08)			
same_ctry_any			0.0241***		
			(0.00239)		
same_ctry_any#c.prior_diffy			-0.00686***		
			(0.000324)		
same_reg2_any				0.0382***	
				(0.00272)	
same_reg2_any#c.prior_diffy				-0.00941***	
				(0.000368)	
same_reg3_any					0.0498***
					(0.00290)
same_reg3_any#c.prior_diffy					-0.0110***
					(0.000401)
Constant	0.332***	0.345***	0.323***	0.324***	0.324***
	(0.00137)	(0.00157)	(0.00153)	(0.00127)	(0.00122)
F	1359.92***	1139.36***	1189.14***	1223.15***	1239.52***
Observations	626,726	626,726	626,726	626,726	626,726
Number of groups	237,714	237,714	237,714	237,714	237,714

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Liner probability model with fixed effects (without self-citations)

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
prior_diffy	-0.00881***	-0.00985***	-0.00794***	-0.00855***	-0.00869***
	(0.000224)	(0.000308)	(0.000261)	(0.000234)	(0.000229)
distmin	2.29e-07	-9.85e-07***			
	(2.08e-07)	(3.23e-07)			
c.prior_diffy#c.distmin		2.22e-07***			
		(4.53e-08)			
same_ctry_any			0.00970***		
			(0.00297)		
same_ctry_any#c.prior_diffy			-0.00274***		
			(0.000417)		
same_reg2_any				0.00813**	
				(0.00410)	
same_reg2_any#c.prior_diffy				-0.00232***	
				(0.000603)	
same_reg3_any					0.00549
					(0.00517)
same_reg3_any#c.prior_diffy					-0.00186**
					(0.000765)
Constant	0.331***	0.337***	0.329***	0.332***	0.332***
	(0.00158)	(0.00195)	(0.00166)	(0.00139)	(0.00134)
F	775.33***	524.97***	533.18***	522.14***	518.96***
Observations	441,394	441,394	441,394	441,394	441,394
Number of groups	206,748	206,748	206,748	206,748	206,748

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Marginal effects of distance (without self-citations)

Time lag (years)	*Model 2 (Km distance)*	*Model 3 (same country)*	*Model 4 (same region)*	*Model 5 (same province)*
0	-9.85e-07***	0.009704***	0.00813**	0.005487
	(3.23E-07)	(0.002967)	(0.004099)	(0.005174)
1	-7.63e-07***	0.006964***	0.00581	0.003632
	(2.90E-07)	(0.002666)	(0.003661)	(0.004616)
2	-5.41e-07**	0.004223*	0.003491	0.001777
	(2.60E-07)	(0.0024)	(0.003276)	(0.004124)
3	-3.19e-07	0.001483	0.001172	-7.8E-05
	(2.36E-07)	(0.002182)	(0.002965)	(0.003725)
4	-9.66e-08	-0.00126	-0.00115	-0.00193
	(2.18E-07)	(0.002027)	(0.002753)	(0.003453)
5	1.26e-07	-0.004	-0.00347	-0.00379
	(2.09E-07)	(0.001951)	(0.002663)	(0.003337)
6	3.48e-07*	-0.00674**	-0.00579**	-0.00564*
	(2.09E-07)	(0.001963)	(0.002709)	(0.003393)
7	5.70e-07***	-0.00948***	-0.0081***	-0.0075**
	(2.19E-07)	(0.002061)	0.002882	(0.003615)
8	7.92e-07***	-0.01222***	-0.01042***	-0.00935**
	(2.37E-07)	(0.002233)	(0.003163)	(0.003974)
10	1.24e-06***	-0.0177***	-0.01506***	-0.01306**
	(2.92E-07)	(0.002741)	(0.003948)	(0.004975)
12	1.68e-06***	-0.02318***	-0.0197***	-0.01677***
	(3.61E-07)	(0.003381)	(0.004907)	(0.006196)
15	2.35e-06***	-0.0314***	-0.02666***	-0.02234***
	(4.79E-07)	(0.004463)	(0.006502)	(0.008223)
18	3.01e-06***	-0.03962***	-0.03361***	-0.02791***
	(6.04E-07)	(0.005616)	0.008187	(0.010363)
21	3.68e-06***	-0.04784***	-0.04057***	-0.03347***
	(7.33E-07)	(0.006804)	(0.009916)	(0.012558)
25	4.57e-06***	-0.0588***	-0.04985***	-0.04089***
	(9.08E-07)	(0.008417)	(0.012258)	(0.015529)
31	5.90e-06***	-0.07524***	-0.06376***	-0.05202***
	(1.17E-06)	(0.010867)	(0.015811)	(0.020036)

Delta method standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 7: Liner probability model with fixed effects
(Sample)

	Model 1 ***(Complex technologies)***	***Model 2*** ***(Discrete technologies)***	***Model 3*** ***(Semiconductors)***	***Model 4*** ***(Mechanical elements)***	***Model 5*** ***(Basic materials chemistry)***
prior_diffy	-0.0146***	-0.0122***	-0.0213***	-0.0110***	-0.0116***
	(0.00043)	(0.00048)	(0.00186)	(0.00166)	(0.000853)
distmin	-3.35e-06***	-1.07e-06*	-6.89e-06***	1.04e-06	6.26e-07
	(4.95e-07)	(6.30e-07)	(1.74e-06)	(2.27e-06)	(1.15e-06)
c.prior_diffy#c.distmin	8.03e-07***	3.16e-07***	1.65e-06***	5.56e-07*	-1.01e-07
	(7.11e-08)	(7.80e-08)	(2.68e-07)	(2.91e-07)	(1.40e-07)
same_inventor	0.0619***	0.0493***	0.0605***	0.0143	0.0309***
	(0.00440)	(0.00486)	(0.0179)	(0.0181)	(0.00850)
same_applicant	-0.0205***	-0.0027	-0.0398***	-0.00997	-0.0166**
	(0.00360)	(0.00438)	(0.0145)	(0.0158)	(0.00757)
Constant	0.3498***	0.3460***	0.402***	0.285***	0.346***
	(0.00284)	(0.00376)	(0.0113)	(0.0122)	(0.00639)
F	352.43***	236.11***	34.5***	12.41***	64.84***
Observations	206,039	109,483	16,033	9,292	31,435
Number of groups	81,622	39,144	6,379	3,849	10,417

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Liner probability model with fixed effects
(Sample)

	Model 1 ***(Examiner citations)***	***Model 2*** ***(Inventor citations)***
prior_diffy	-0.0141***	-0.00380***
	(0.000281)	(0.000410)
distmin	-3.67e-06***	3.35e-07
	(3.07e-07)	(7.47e-07)
c.prior_diffy#c.distmin	7.19e-07***	1.20e-07*
	(4.51e-08)	(7.00e-08)
same_inventor	0.0628***	0.0406***
	(0.00271)	(0.00472)
same_applicant	-0.00936***	-0.00538
	(0.00231)	(0.00435)
Constant	0.364***	0.172***
	(0.00179)	(0.00386)
F	905.05***	44.12***
Observations	539,711	87,015
Number of groups	227,705	54,149

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1