Instituto de Políticas y Bienes Públicos

Early Career Researcher Conference (ECRC)

Science dynamics and research systems:
The role of research for meeting societal challenges”

Madrid, 8-9 April 2013

Inventor's knowledge set as the antecedent of patent importance

Ali Mohammadi

DIG, Department of Management,
Economics and Industrial Engineering,
Politecnico di Milano,
P. Leonardo da Vinci 32,
20133, Milan, Italy.
: ali1.mohammadi@mail.polimi.it

Chiara Franzoni[1]

C/ Albasanz 26-28, 3D
28037 Madrid. SPAIN
http://www.ipp.csic.es

Inventor's knowledge set as the antecedent of patent importance

Ali Mohammadi[1,2] Chiara Franzoni[1]

Abstract

This paper seeks to contribute to the debate on the antecedents of patent importance by looking at the prior knowledge set of the inventors. Using independent methodologies, we distinguish between the scientific knowledge set and technical knowledge set of an academic inventor and separate these from other kinds of prior expertize. We proxy importance by citations received after six years. We find that the patents of the inventors who have a prior scholarly knowledge of the topic are on average more important. Conversely, prior technical relatedness is negatively correlated to patent importance. These results are potentially useful to support the work of practitioners like university technology managers, which often face difficulties in identifying the importance and perspective value of the disclosed inventions, amid high market and legal uncertainly and budget shortages.

Keywords: Patent importance; academic patenting; scientific relatedness; technical relatedness; recombinant innovation; technology transfer;

JEL Classification: O31, O32, O34

---

[1] DIG, Department of Management, Economics and Industrial Engineering, Politecnico di Milano, P. Leonardo da Vinci 32, 20133, Milan, Italy.

[2] Corresponding author: ali1.mohammadi@mail.polimi.it

- We thank participant of 7th Annual Conference of the EPIP Association for helpful comments and discussion.

1. Introduction

Scholarly works on Intellectual Property Rights (IPRs) have for long highlighted that market transactions concerning patents and other technological assets are constrained by problems of asymmetric information and moral hazard (Scotchmer, 2004; Shane, 2002). One critical problem is anticipating the prospective relevance of a new technology, which depends on a combination of technological opportunity, market demand and legal perils (Gittelman & Kogut, 2003). Concerning technological and market opportunity, the effective capacity of a technology to translate into functional products and generate revenues depends on events that may or may not take place in the future, such as the possibility that a substitute or competing technology will prevail on the market. In addition to market uncertainty, transactions are complicated by problems of asymmetric information among the parties that are involved in the negotiation (Gallini & Wright, 1990). For example, the potential acquirer tends to have less information on the quality of a technology than the vendor, particularly if the technology embeds frontier knowledge for which it is even difficult to find an expert of the field that wants to assists the transaction (Franzoni, 2007) Both issues -uncertain outcomes and asymmetric information- are even more severe for technologies that are developed and patented in universities (Shane, 2002; Franzoni, 2007), whose relevance is, therefore, even more difficult to anticipate.

In this paper, we focus on the antecedents of patent importance in academic patents, and investigate whether and to what extent highly relevant patents are correlated to certain characteristics of the prior knowledge set of the inventor. In general, the literature on academic patents has documented that the most productive scientists are more likely to become inventors than their less productive colleagues (Stephan, et al., 2007; Fabrizo & Di Minin, 2008; Breschi, et al., 2007). It has also documented that inventions are preceded by a flurry of scholarly publications (Azoulay, et al., 2007; Calderini, et al., 2007). This evidence seems to indicate that inventions are developed as a by-product of a prolific research activity that the scientist has conducted for scientific purposes. In practice, however, to the best of the authors' knowledge, no prior study has established a clear link between the flurry of publications that precedes an invention and the invention itself. In other words, we presume that there is an ideal continuity between *what* a scientist does in a certain year in his/her academic lab and *what* does he/she invents in the same year or shortly after. However, until now we have been capable to draw correlations based only on their being time-contiguous, which is not the same as being contiguous in content. To observe relatedness of content standard techniques based on counts of occurrences are not sufficient. To investigate this issue we need to resort to techniques of content analytics. We do so starting from a sample of US patents that have been invented by faculty members in the area of Physics. We measure patent importance by looking at the number of citations received six years after the year of patent priority. We characterize the inventors in terms of the knowledge sets that they possess. We retrieve information on the inventor's scientific knowledge set by looking at the articles that he/she published in scientific outlets prior to the priority year of the focal patent. We retrieve information on the inventor's technical knowledge set by looking at the inventions that he/she filed at the patent office prior to the priority year of the focal patent. This allows us to characterize the subject matter of an invention in terms of its relatedness to two distinct knowledge sets of the inventor: i) the scientific understanding of the subject from which the invention originates, which can be acquired when the inventor conceives the idea in the course of scholarly research, and ii) the technical understanding of a specific technology that originates from prior inventions in the same domain of technology.

In order to assess technical relatedness, we rely on the standard methodology developed by (Fleming, 2001) and based on looking at the prior occurrence of inventions in the same patent class relative to number of prior patents[3]. Concerning scientific relatedness, no standard methodology could be employed and we therefore resort to semantic analysis. We rely on a technique that allows computing the similarity of two or more documents (patents and publications abstracts), based on the co-occurrence of the same semantics (co-word) and on the same pairs (co word-pairs), if pairs exist. We use this technique to partition the inventor's publications into items whose content is related to that of a focal patent and items whose

[3] In a future version of the paper we will use content analysis for screening the relatedness of prior patents to the focal patent.

content appears unrelated. This methodological choice is novel and it is one of the contributions of this paper.

Based on the current scholarly understanding, we formulate the hypotheses that highly cited patents would be those grounded on the scientific knowledge set and that further exploitation of an existing technical knowledge set would be associated to less-important patents.

We test our hypotheses on a sample of 295 academic patents[4] in all subfields of physics. Our results suggest that patents whose content is related to prior scholarly papers are disproportionately distributed among the more cited. On average, during the first six years of life, a patent invented in a prior area of scholarly investigation receives 49% more citations than a patent invented in areas unrelated to the research scope of a scientist. Prior technical knowledge, as captured by prior inventions in the same patent classes relative to total prior patents, is negatively correlated with patent citations.

Although citations give but a blunt appreciation of patent quality, these results clearly point at a correlation between the inventor's knowledge set and the degree to which inventions are likely to have impact and be valuable. We discuss how these results may bring potentially relevant implications for the practice of technology management at universities, as well as for our understanding of the process that leads from scholarly investigation to the development of practical applications.

The paper is structured as follows. Section 2 reviews the findings of the literature on academic patenting with regard to patent importance and patent citations and defines two theoretical hypotheses for the empirical investigation. Section 3 describes the sample and the data collection procedure and defines the variables and the methodology used in the analysis. Section 4 presents the results of the econometric estimates and Section 5 discusses the implications.

---

[4] Academic patents are defined here as patents invented by a university faculty member.

2. Academic patents

2.1. Assessing disclosed inventions at universities

Amid technology and market uncertainty, evaluating the prospective importance of a university patent is an especially difficult task. The scholars who have investigated technology transfer have extensively documented a number of common and widespread problems that make it difficult to assess the importance of academic inventions. First, a large share of the inventions disclosed by academic professors are still at the proof-of concept stage, or anyway have a long way before they could eventually become marketable (Jensen & Thursby, 2001). Second, academic professors often lack business competences, have little understanding of the market conditions, few links to companies potentially interested in their inventions (Swamidass & Vulasa, 2008) and often overestimate the real market importance of their ideas (Owen-Smith & Powell, 2001).

Third, faculty members naturally are moved by sets of priorities and goals typical of academia, which are not always or not necessarily aligned to those of the potential investors (Jensen, et al., 2003). For example, some scientists are interested in the personal reputation that they think they can gain from patenting (Göktepe-Hulten and Mahagaonkar 2010). Other scientists may be more concerned about the delays in publication (Blumenthal, et al. 1996, Campbell, et al. 2002) and other conflict of interests that a commercialization strategy may bring about (Ambos, et al., 2008; Murray & Stern, 2007; Haeussler, et al., 2009; Walsh, et al., 2005).

These circumstances partly explain the difficulties encountered by universities in getting sizable revenues from the commercialization of technologies. The available figures point at a situation in which few of the many universities active in technology commercialization generate a substantial stream of income. Those that have income generate their revenues from a handful of highly successful patents (AUTM 2011). For example, the University of California –known to be top or tied-to-top for profit generation- makes about half of its revenue with just five patents, all in the medical or biotech sector (Farrell, 2008). In more common situations, technology managers strive to find interested potential buyers. A large survey conducted in eight European countries highlighted that the share of unlicensed patents at academic institutions is almost double than that of unlicensed patents at large and medium firms (EC, 2006). Amid budget shortages, many offices report having just enough resources to accomplish the legal and procedural aspects of patent filing, leaving too few for marketing and commercialization (Swamidass & Vulasa, 2008).

2.2. Importance of academic patents

Some early works on patent citations have been conducted between 1998 and 2003. The focus of these analyses was to determine if the institutionalization of the patenting activities from universities in the aftermath of the Bayh-Dole act had lead to a decline of patent quality (Sampat, et al., 2003). The purpose of this body of works is different from the scope of our analysis. Nonetheless, a summary of the basic findings provides a helpful background for our research. First, academic patents (defined as those assigned to universities or having at least one scientist among the inventors) on average tend to be more highly cited and broader in scope than non-university patents (Henderson, et al., 1998; Mowery, et al., 2001; Czarnitzki, et al., 2011). Second, in terms of trends, there is a clear increase in the number of patents issued to universities, but it appears less clear if the above-average citation rate has remained unchanged. The most recent findings seem to suggest that there has been no sensible decline in citations received, although citations come at a slower pace, due to slowdowns in the patent application procedure (Sampat, et al., 2003).

Concerning non-US universities, the issue is more controversial, partly because many patents of university faculty members are assigned to companies and more difficult to keep track of (Lissoni, et al., 2010). Recent works show similar citation patterns experienced by university and non-university patents in Europe (Crespi, et al., 2011) and Japan (Bacchiocchi & Montobbio, forthcoming). In terms of trends, Czarnitzki, et al., (2011) find a decline in citations since the mid-1980 in a sample of German patents.

2.3. Scientific knowledge set as an input to the invention process

A body of empirical investigations has convincingly established a link between scientific and inventive productivity, both at the level of firms and the level of single individuals.

At the level of individuals, cross-sectional investigations have shown that the most productive scientists are more likely to become inventors than their less productive colleagues (Stephan, et al., 2007; Fabrizo & Di Minin, 2008; Breschi, et al., 2007). The correlation has also been documented in longitudinal studies, showing that inventions are preceded by a flurry of scholarly publications (Azoulay, et al., 2007; Calderini, et al., 2007).

When looking at licenses of academic patents, Elfenbein (2007) showed the signalling nature of publications on increasing the probability of licencing a new technology.

When looking at the level of firms, rather than at the level of the individuals, the positive relationship is confirmed. Gittleman and Kogut (2003) studied the effect of firms' scientific publishing on patent importance and provided evidence that patents based on scientific research are more likely to be cited, although highly-cited publications are negatively correlated with good patents. Fleming and Sorensen (2004) find that patents citing publications are more likely to be broader in scope. Zucker et al. (2002) studied the role of collaboration with star scientists and its effect on patent importance. Their empirical results support the hypothesis that the existence of publications of firms' employees with star scientists is positively correlated with the citation rate of the firm's patents.

This evidence is consistent to the scenario in which scientific research is a valuable input for the production of innovation. Fleming & Sorenson (2004) interpret these findings in light of the theory of the invention as a process of knowledge recombination (Nelson & Winter, 2002; Kogut & Zander, 1992; Hargadon & Sutton, 1997; Weitzman, 1998; Fleming, 2001), which argues that all inventions at a closer look are the result of a combination of pre-existing discrete elements and components. They maintain that the search for new components is conducted locally, in the sense that it usually occurs in areas that are proximal in cognitive and semantic terms to the prior experience of the inventor. Their view is that experience of scientific research helps inventors in orienting their local search towards more fruitful and less exhausted combinations.

It is important to note that all these studies investigate numerical relationships between counts (or dummies) of publications, and counts of citations, but are uncapable to account for the degree to which two documents (e.g. a patent and a paper) have similar or dissimilar content. As such, quantitative relationships have limitations. For example, if we observe that inventors who publish more papers are more likely to patent later on, we cannot tell if this is so because scientific research was an input to the inventive process, or simply because the inventor was going through an exceptionally fruitful period of his/her life, and/or had more resources, hence more output of any kind. If the scientific research conducted by an inventor and disclosed in articles is a valuable input, for example because it orients the search of new technical solutions towards unexplored pathways In this paper, we improve this literature by coding separately for documents of similar content. Our measure of relatedness is based on observing a high degree of semantic overlap between the patent and the published research by the same inventor/author. We distinguish this variable from the numerical count of prior publications that we use as a control, thus clearing-off the impact of scientific research used as an input, from the effect of mere variance in the productivity of the inventor in a given period of time.

Our first research hypothesis is therefore formulated as follows:

H1: inventions that rely on a prior scientific knowledge set (are built on the results of the prior scientific research of the academic inventor) are more likely to be important than inventions that do not rely on a prior scientific knowledge set

2.4. Technical knowledge set as an input to the invention process

Many academic scholars produce inventions quite continuously during their career (Azoulay, et al., 2007). In line with the theory of invetions as the product of knowledge recombination mentioned, several scholars have investigated the impact that the background of experience accumulated by serial inventors has on the characteristics of the new patents invented. This literature has generally used patent classes to characterize the degree to which patents are similar/dissimilar, a technique that has encountered increasing criticism (Strumsky, Lobo and van der Leeuw 2012). Two patents are considered similar if they are assigned to the same patent class, or to the same combination of patent classes. Fleming (2001) shows that patents that are based on non-familiar components or on a novel recombination of familiar components are more likely to be a breakthrough. Conversely, patents exploiting the exact same combination used in the past are less likely to be a breakthrough. Audia and Goncalo (2007) broaden these findings by showing that inventors who were more successful in the past are more likely to move along incremental lines, rather than experimenting on new trajectories. They speculate that people who experienced success are more productive of new ideas, but these ideas are less likely to be divergent, since people anticipate the payoffs from exploiting known areas instead of engaging in new exploration (Audia & Goncalo, 2007). Conti et al. (2010) supported this argument by studying a rich database of European patents. They conclude that continuing to use known technological components increases the productivity in numerical terms, but reduces the probability of producing high quality inventions. In conclusion, there is growing evidence that serial inventors are likely to produce more patents, but the new patents would be of lower impact (Gambardella, et al., 2011).

In line with these findings, we formulate our second research hypothesis as follows:

H2: inventions that rely on the prior technical knowledge set (are based on other inventions within the same technological domain) are less likely to be important than inventions that do not rely on a prior technical knowledge set

3. Data and Methodology

3.1. Sample

We draw our analysis based on a dataset of 373 unique patent-academic inventor combinations. Academic inventors were identified as the scientists in all areas of physics who were appointed "Fellows of the American Physical Society" (APS) prior to 2005, had a US academic affiliation at that date and had been inventors of at least one patent deposited at the US Patent and Trademark Office (USPTO) between 1992 and 2005.[5] Full patent records were retrieved from the archives of Thompson Innovation at the beginning of 2012, along with

[5] Full information on the sampling procedure is given in (C. Franzoni, Do scientists get fundamtal ideas by solving practical problems? 2009).

patent citation records. Information on the scholarly activity of the academic inventors was retrieved from the archives of ISI-Web of Knowledge and complemented with individual information on the inventor (gender, PhD year, affiliation, specialty, etc.). These were compiled from faculty WebPages and CVs available on the web.

In order to avoid using patent documents that may constitute substantial duplicates, we worked carefully to clean-away records from the same patent family, as defined by those having at least one common priority code. For each patent family, we choose to keep only the one record that had the largest number of citations and, with equal citations, the one issued at the earliest date. This cleanout resulted in 98 drops, leaving us with 295 unique patent-inventor combinations from 95 different inventors.

3.2. Dependent variable

During spring 2012, we retrieved the citations received by the 295 patents and restrict the citation window to six years since the year of priority. Recall that our aim is to investigate correlations between the inventor knowledge set and the prospective importance of a patent, as captured by patent citations. Prior contributions have indicated that citations are positively correlated to patent importance (Jaffe, Trajtenberg and Henderson 2003) and patents that are cited more frequently are related to innovations with higher social (Trajtenberg, 1990) and private (Hall, et al., 2005; Griliches, 1981) value. It has also been shown that the citations received in the first years are a strong predictor of later citations (Fleming, 2001; Gittelman & Kogut, 2003). It seems therefore safe to adopt 6-years citation counts as a proxy for the importance of an invention.

Since our patents are deposited in different years, the window of observation of patent citations is dynamic. For example, for the patents deposited in 1992, we computed the citations received until the end of 1998. For those deposited in 1993, we computer the citations received until the end of 1999 and so on. For the most recent patents in our dataset, i.e. those issued in 2005, citations were computed until the end of 2011.

A description of the variables used in the analysis is provided in Table1.

3.3. **Independent Variables**

We have two independent variables of prior scientific and technical knowledge set.

1) *Prior scientific knowledge set:* Prior empirical investigations have devised a commonly accepted methodology to account for the prior technical knowledge set of an inventor. Unfortunately, a standard methodology does not exist when it comes to accounting for the prior scientific knowledge set. One contribution of this paper is to fill this methodological gap. We explain the information collection and the methodology for treatment in this sub-section.

We begin by retrieving information on the scholarly activity of inventors. One advantage of using patents of faculty members is that the prior scientific knowledge set of the inventor is well documented. We searched the articles published on scientific outlets by the individuals between 1990 and 2005 in ISI Web of Knowledge. For each article, title and the full abstract were coded in chronological order. Our aim is to study if the inventions that have emerged in domains of expertise that a scientist masters because of his/her scholarly knowledge are more cited on average than those that have emerged in other areas. We therefore want to know the degree to which the inventor is an expert (in scientific terms) of the domain employed in a certain invention. To gain this information we compare the content of the patent to the content of the scholarly articles that the person has published until the time of patent priority. Comparison is based on semantic similarity. Rather than doing this screening manually, we use an algorithm of semantic analysis contained in a commercial software for scientific research, called Crawdad Text Analysis System 1.2. The software compares a single document against a group of documents (the set of all publications authored by the inventor) and outputs a score of resonance for each pair of documents based on the co-occurrence of the same semantics and on the co-occurrence of word pairs in the title and abstract of the patent and the title and abstract of each article. Based on the work of Franzoni and Scellato (2010) that used the same algorithm for a similar purpose and validated the scores by means of a panel of experts, we choose a quite restrictive threshold of similarity (0.1). We code patents as related to the scholarly work of the inventor if the patent was similar ("resonant" in the wording of the software) to one or more scientific articles, or unrelated in case we find no scholarly publication with a similar content. Our threshold and consequent measure of relatedness is quite conservative. We find a total of 71 (24.1%) patents that are related to at least one article and 224 (75.9%) patents that resulted to be unrelated to prior scientific research.

2) *Prior technical knowledge:* Unlike for the scientific knowledge set, in this case we can stand on a standard and well-codified methodology. Following the methodology of Fleming (2001), we compare the main International Patent Classification (IPC) of the patent to those of prior patents by the same person. The technical knowledge set has been compiled by retrieving information on the prior patenting activity of the inventors. For each inventor, we collected all patents issued prior to the years of observation. We characterize inventors in terms of the number of patents that they had previously in the same IPC class of the focal patent and we normalized it by total number of prior patents. In case of patents with multiple IPCs, we used for comparisons the main (primary) IPC code[6].

3.4. Control variables

We controlled for a number of potentially confounding factors that may concur to determine the importance of a patent and/or affect our measure of patent importance. Potentially confounding factors relate to characteristics of the inventors, characteristics of the assignee, patent characteristics and annual trends. In terms of individual characteristics, we control for gender and age, which were found to correlate –among other things- with patent performances (Azoulay, et al., 2007; Bacchiocchi & Montobbio, forthcoming; Stephan, et al., 2007). We also control for the individual productivity in terms of both patents (Gambardella, et al., 2011) and scholarly works (Fleming and Sorenson 2004), whether or not the inventor is at his/her first patent, and how productive he/she has been in the last years in terms of both patents and scholarly articles. Concerning the characteristics of the assignees and in line with the findings of extant literature (Lissoni, et al., 2010), we control for the assignee type (whether it is a university; a firm; or other types), and for the assignee's prior patenting experience. Concerning patent features, we control for a number of patent characteristics that

[6] We have however performed the same analysis by comparing all IPC codes and the results hold invariant. The analysis was also repeated for the first 4 digit IPC classes, and the results are the same.

may be correlated to unobserved heterogeneity. These are the number of claims, the number of backward citation (from patents and scientific articles separately),[7] the number of inventors and the broadness or scope of the patents (number of IPC classes assigned to the patent) (Trajtenberg, et al., 1997). In order to account for trends in the number of patents issued year after year, which likely alters the amount of potentially citing documents, hence of total citations, we control for the number of patents granted by the USPTO in the 5 years period after the priority year. Moreover we account for heterogeneity caused by the geographic distribution of patents by controlling for number of patents granted in the same year and in the same US state where the assignee is located. We additionally control for the year in which the patent was filed (priority year dummies), and the technological domain to which the patent belongs (2-digits IPC classes dummies).

[Table 1 about here]

3.5. Descriptive Statistics

Summary statistics of all variables used in the analysis are reported in Table 2. Several variables are transformed in natural logarithm to correct for skewed distributions. A preliminary inspection of the data shows the expected high variance and skewness of forward citations. 32 patents (10.58%) have not received any citation in the 6 years after priority. 26 had received 1 (8.8%). 59 received between 2 and 5 citations (20.0%). The median number of citation is 7 and the mean is 18.6. There are 9 patents that have more than 100 citations and some outliers with a maximum of 421 citations received by one patent.

[Table 2 about here]

The inventor experience is also highly variable, ranging from inventors with no prior patents and a maximum of 76 prior patents. Table 3 reports distribution of patents among inventors.

---

[7] It has been shown patents which cite published material –peer review or commercial- receive more citations (Sorenson and Fleming 2004).

The scientists received their PhD between 1950 and 1991 and, accordingly, their career age at the priority year ranges from a minimum of 4 years to a maximum of 55. The scientists also vary considerably for their scientific production. On average, each scientist published 22.6 papers during the 3 years preceding the patent, with considerable variability. Regarding the assignee type, 61.1% of patents are owned by universities (or intermediary institutions owned by universities). This result is aligned with prior studies of American universities. For example, Thursby et al, (2009) showed that around 70% of academic patents in the US are owned by universities. 29.0% of patents in the sample are assigned to firms, and the rest is assigned to individuals (6.44%), US governmental bodies (1.36) or both universities and a firm (2.03%).

[Table 3 about here]

Concerning the number of patents that the assignees had prior to the current patent, we go from a low of zero, to a high of more than 10 thousands. Patents are coming from 21 different technological domains measured by the IPC class at the first two digits. Patents on average cited 15.99 prior patents and 10.39 scientific articles. 95% of inventors in the sample are men while only about 5% are women.

Concerning prior knowledge of the inventor, our methodology let us distinguish patents that rely on prior scientific knowledge and patents that rely on prior technical knowledge. In terms of proportions, the majority of patents (61.7%) are not directly relatable to a specific knowledge of the inventor. About one in four patents (24.0%) are related to prior scientific knowledge visible in publications and about one in six (17.3%) is related to the prior technical knowledge visible in inventions.[8] The mean of normalized technical relatedness is 0.06. Virtually no inventions rely at the same time on some specific technical and scientific knowledge of the inventor. Table 4 shows distribution of technical and scientific knowledge set of inventors.

[Table 4 about here]

[8] We find a maximum of 10 prior patents related to prior documents.

In terms of trends, we see a general surge in the number of patent (utility) granted, with approximately three step increases of 50 thousand each: the first step increase happened in 1993 and lasts until 1998, the second is from 1998 to 2009 and the third is in 2010-2011 (Figure 1).

[Figure 1 about here]

3.1. Methodology and Estimation

We model the quality of patents (proxied by forward citation) as a function of the technical knowledge set (measured as the count of patents in technologically related classes relative to total number of prior patents), and of the scientific knowledge set (measured by the existence of publications whose content is in a very similar area of the patent –i.e. "resonant" to the patent content-), plus a number of controls and an error component.

Since the independent variable is a positive integer, we model the data using count models. Given the high dispersion of our variable, the general assumptions underlying Poisson models that the conditional mean be equal to the variance is violated, suggesting to adopt either Negative Binomial or Quasi-Maximum-Likelihood (QML) Poisson. Both techniques lead to unbiased estimates in case of overdispersion and QML Poisson is deemed to imply fewer restrictions than Negative Binomial (Wooldridge, 2002). However, in our case the Hausmann test indicates that QML Poisson is inconsistent (p=0.000). We therefore chose to comment the result of Negative Binomial estimation and show QML for comparison. The results are almost invariably consistent.

To cope with hetereokedasticity caused by having repetitive inventors, we use robust standard errors adjusted for clustering at inventor level in order to allow for non-independence of the observations for the same inventor. This implies that the conditional mean is:

$$E(FC|X) = \exp\left[\beta_i . X_i + \sum \alpha_i . \mathrm{Year}_{\mathrm{Dummy}_i} + \sum \gamma_i . \mathrm{IPC} - \mathrm{Dummy}_i + \sum \delta_i . Z_i\right]$$

Where X_i is a vector of independent variables (*technical knowledge set*; *scientific knowledge set*). Z_i is a vector of controls variables that is meant to clear-off the effect of potentially confounding factors. To control for the personal characteristics of the inventor, we include variables such as the career age (plain and squared, to account for known curvilinear effects), gender, prior experience of patents (log), count of scientific articles published during the last 3 years (log) and a dummy variable capturing inventors at their first patent. We control for patent characteristics that may capture unobserved heterogeneity such as the number of claims (log), the number of inventors (log), the number of backward citations (patents and scientific articles separately (log)), and the broadness of patent scope (number of IPC classes to which the patent has been assigned). We also control for the prior experience of the assignee with patents at the time of the deposit. Collectively, these characteristics may affect the degree to which patents get cited in ways which do not reflect the importance of the invention, but rather the expertise of the assignee with patents, the accuracy with which the patent document has been prepared, etc.

Moreover we control for changes in patenting trends, as captured by the number of patents granted by the USPTO (log) in the 5 years after the priority and for the geographic distribution of patenting activities, as captured by the number of patents granted in the same U.S state of the assignee (log) and in the same year. To account for potential effects of time and market trends, we include a set of single-year dummy variables, based on the year of priority (Year_Dummy). We finally include a set of 21 dummy variables capturing the propensity of patents in different technological classes to receive citations.

Table 5 reports the correlation matrix.

[Table 5 about here]

4. Results

Table shows the result of the estimates using Negative Binomial and reporting QML Poisson for comparison.

Model 1 reports the results for the full sample of 295 patents and model 2 reports the results for a restricted sample of 267 patents in which we have omitted the patents of the inventors that had no prior patent experience. The econometric models, regardless of the estimation method, strongly supports our hypothesis 1 that patents which are related to the prior scientific knowledge set have on average more citations. By looking at the coefficient of the Negative Binomial model, we can see that patents which are related to prior publications receive on average 49% more citations than the ones which are not related to prior publications.[9]

[Table 6 about here]

Concerning our hypothesis 2 that the prior experience of patents within the same technological classes is associated to a reduced importance of the focal patent; the results of the unrestricted sample confirm the prediction. The coefficient of the variable is negative, as expected and significantly different from zero. This holds for both the QML Poisson and the Negative Binomial estimate. However, when we restrict the sample to include only the patents that had a prior track of technical knowledge, the coefficient is still negative, but the coefficient looses part of its significance. QML Possion remains estimate remain negative and significant.

In Model 3 of Table 6, we include two additional variables of patenting trend and spatial distribution. Since they are missing values in assignee address the sample is limited to 263 observations. The results are qualitatively similar to model 1 regardless of estimation model and support both the hypothesis 1 and 2.

Results of the control variables are aligned to those commonly established in the literature, although not always significant when different estimation methods are used. We can observe that (undifferentiated) experience with patents is associated to more citations. More claims are associated to higher importance. Men produce on average better cited patents than women. In addition, we can see that the number of patents and scientific articles cited in the focal patent are positively related to forward citations. Interestingly, patents assigned to universities on average receive more citations than the baseline. The baseline here is

[9] = exp(0.399)-1

constituted of "other assignees types", and includes individual assignees, as well as governmental bodies. Similarly, patents assigned to firms exhibit a citation premium, compared to the baseline.

4.1. Robustness check

In order to check the robustness of the estimates to the choice of model, we have tried several alternative functional forms. Appendix 1 report the results of these robustness check.

First, we include the square of the technical knowledge set to be used in combination to the plain value (Model 1) of alone (Model 2). Second, we use the share of patent citations and scientific articles citations over the total, rather than the sheer counts (Model 3). Fourth, we exclude claims as control variable since a recent article (Strumsky et al, 2012) has criticized use of number of patent claims as antecedent of patent importance. We therefore show the estimate of the model after excluding patent claims (Model 4). Lastly we account for the potential effect of the changes in disclosure policy of the USPTO occurred since November 2000 which allows publication of pending patent applications after 18 months. The change policy can impact citation trends in ways that are not captured by year dummies. In Model 5 included a dummy which takes the value of one if the patent was filed after November 2000. Results are robust to these alternative specifications.

In Appendix 2 we test the robustness of our estimates to the choice of treatment for patent families. Recall that in the paper we choose to use a single focal patent in a family, as the one that had more citations or had an earlier priority date, in case the number of citations was the same. In order to check robustness of the result to this strategy, we re-estimate the models 1 and 2 of Table 6 with all patents in the family (373 in model 1 and 348 in model 2). Both models confirm the hypotheses 1 and 2.

5. Conclusion and Discussion:

We have investigated the prior knowledge set of the academic inventors. Our first aim was to understand the degree to which inventors rely on their prior scientific knowledge set, and or as opposed to relying on their prior technical knowledge set in their inventive activities. Our results suggested that about one in four inventions issued to academics are in areas in which they have scientific expertize, as witnessed by publications in scientific outlets and less than one in five are in areas of technical expertize, as witnessed by prior patents. Virtually no patents are related to both prior scientific knowledge and prior technical knowledge. It is however important to stress that our measure of relatedness is all but perfect and the estimate may be affected by measurement biases in important ways. To mention just one potential source of biases, it is known that patent documents are edited by the attorneys. Therefore the wording used in patents may not necessarily reflect the wording used in scientific literature and this would cause our software to underestimate the degree of relatedness in document content. Further analysis would be advisable in the area, possibly with the use of surveys. If the result were to be confirmed, this would mean that a sizable, but largely minoritarian share of patents relates or draws substantially from one's scientific production.

Our second aim was to study the correlation between prior knowledge sets and the relative importance of patents. We showed that patents based on prior scientific knowledge sets receive on average 49% more citations than patents non-related to prior scientific knowledge. Conversely, inventions based on areas where the inventor had prior technical knowledge are likely less relevant (receive fewer citations) than unrelated to the inventor's prior technical knowledge. This result is consistent to what emerged in prior studies (Conti, Gambardella and Mariani 2010, Gambardella, Verspagen and Harhoff 2011). Overall, the results seem to suggest that there are high returns from re-using scientific knowledge, but decreasing returns from using technical knowledge.

We use this information Using a novel methodology, we employ a software that identifies semantic similarity to assess whether or not the content of a certain invention (as described in patent document) is related to the prior scientific knowledge set of the academic inventor (as described in scientific publications). We further assess whether or not a certain invention is related to the prior technical knowledge set of the academic inventor, using the standard methodology of Fleming (2001), which infers technical relatedness from IPC patent classes.

We show that the majority of patents are not strictly related to a specific knowledge of the scientist. About one in six patents are related to prior technical knowledge and about one in four patents are related to prior scientific knowledge of the inventor. Virtually no patents are related to both.

Our third aim was to propose and apply a novel methodology that looks at document relatedness by means of content analysis, rather than on temporal contiguity. The methodology that we used is embedded in a commercial software and can be easily replicated in future studies. Improvements in the methodology may certainly be advisable in terms of calibration (threshold of similarity) and in terms of the breadth of text analyzed. In this paper we screened only the title and abstract of the documents, which simplifies computational time and offers the advantage of using documents of comparable lenght. A more refined text analysis could be obtained if longer portions of the texts are used.

These results have several potentially relevant implications for scholars and practitioners. First, over the last 20 years, there has been an increasing interest toward academic patenting among researchers in different fields of science. A large body of literature has investigated the impact of a steady growth in patenting by university researchers on their scientific output. In general, the literature on academic patents has documented that academic patents tend to be more valuable and general than non-academic ones. Furthermore, productive scientists are more likely to become inventors than their less productive colleagues (Stephan, et al., 2007; Fabrizio & Di Minin, 2008; Breschi, et al., 2007) and to produce a flurry of scholarly publications before patenting (Azoulay, et al., 2007; Calderini, et al., 2007). This evidence seems to indicate that inventions are developed as a by-product of a prolific research activity that the scientist has conducted for scientific purposes. In practice, however, the correlation observed in prior works is strictly numerical, meaning that no prior study has established a clear link between the content of the flurry of publications that precede an invention and the invention itself. Our findings advance this literature as they show that patents based on prior scientific research are on average more important (more cited) than those not based on prior scientific research. When, by way of analogy, we tested weather prior technical knowledge visible in patents is also a predictor of patent quality, we see no clear correlation. The findings support the hypotheses that scientific knowledge is a valuable input to the patent process, at least for academic patents.

Second, the metholody that we adopt in this paper draws on the semantic similarity between a patent and a stream of scientific articles, rather than just on numerical occurrences. Applications of this methodology are replicable by means of a simple, yet effective, family of commercial software. Our work provides an example of such application and invites more work in the area.

Third, our results are potentially of interest for agents involved in the technology transfer process at several levels. At the level of university administrators, our findings confirm that scientific research is a needed input to achieve success in technology transfer. Our results are also potentially valuable to technology managers called to assess and triage disclosures, amid limited resources and strong market uncertainty. Our findings indicate that inventions backed by scientific research, all other things being equal, are among the most promising and should therefore be evaluated with special care.

The paper has several limitations that could possibly be addressed in future research. First, our results are based on a limited sample of academic inventors working in the field of physics. It has been shown that the relationship between scientific productivity and patenting varies across fields (Stephan, et al. 2007) and even across sub-fields of the same area (Calderini, et al., 2009). Second, the inventors in our sample are all "Fellows" of the American Pysical Society which implies that they are exceptionally brilliant scientist within their field. It is possible that further analyses conducted on different areas or more representative samples will lead to different results.

Finally, in this paper we used a sample of academic inventions. It is important to acknowledge that the results should not be generalized outside of the realm of academe. On a more representative sample, the PatVal survey has shown that inventors consider prior "patent literature" and prior "scientific literature" among the top three sources of knowledge used in the invention process for relative importance, although no external source is rated as very relevant in absolute terms (Giuri et al., 2007). Future analysis could be directed to investigate the extent to which scientific and technical knowledge sets are relevant in firms' innovation processes.

6. Bibliography

1. AMBOS, T. C., MAKELA, K., BIRKINSHAW, J. & D'ESTE, P., 2008. When Does University Research Get Commercialized? Creating Ambidexterity in Research Institutions Source. *Journal of Management Studies,* 45(8), pp. 1424-1447.
2. AUDIA, P.G. and GONCALO, J.A., 2007. Past success and creativity over time: A study of inventors in the hard disk drive industry. *Management Science,* 53(1), pp. 1-15.
3. AUTM, 2011. *U.S. Licensing Activity Survey: FY2010,* s.l.: Association of University Technology Managers.
4. AZOULAY, P., DING, W. and STUART, T., 2007. The determinants of faculty patenting behavior: Demographics or opportunities? *Journal of Economic Behavior and Organization,* 63(4), pp. 599-623.
5. BACCHIOCCHI,E. and MONTOBBIO,F., forthcoming. International knowledge diffusion and home-bias effect. Do USPTO & EPO patent citations tell the same story? *Scandinavian Journal of Economics.*
6. BLUMENTHAL, D., CAMPBELL, E.G., ANDERSON, M.S., CAUSINO, N. and LOUIS, K.S., 1997. Withholding research results in academic life science - Evidence from a national survey of faculty. *Jama-Journal of the American Medical Association,* 277(15), pp. 1224-1228.
7. BRESCHI, S., LISSONI, F. and MONTOBBIO, F., 2007. The Scientific Productivity Of Academic Inventors: New Evidence From Italian Data. *Economics of Innovation and New Technology,* 16(2), pp. 101-118.
8. CALDERINI, M., FRANZONI, C. and VEZZULLI, A., 2009. The Unequal Benefits of Academic Patenting for Science and Engineering Research. *IEEE Transactions on Engineering Management,* 56(1), pp. 16-30.
9. CALDERINI, M., FRANZONI, C. and VEZZULLI, A., 2007. If star scientists do not patent: The effect of productivity, basicness and impact on the decision to patent in the academic world. *Research Policy,* 36(3), pp. 303-319.
10. CAMPBELL, E., CLARRIDGE, B., GOKHALE, N., BIRENBAUM, L., HILGARTNER, S., HOLTZMAN, N. and BLUMENTHAL, D., 2002. Data withholding in academic genetics - Evidence from a national survey. *Jama-Journal of the American Medical Association,* 287(4), pp. 473-480.
11. CONTI, R., GAMBARDELLA, A. and MARIANI, M., 2010. *Learning To Be Edison? Individual Inventive Experience and Breakthrough Inventions.* Working Paper. Available at: http://www2.druid.dk/conferences/viewpaper.php?id=501970&cf=43.
12. CRESPI, G., D'ESTE, P., FONTANA, R. and GEUNA, A., 2011. The impact of academic patenting on university research and its transfer. *Research Policy,* 40(1), pp. 55-68.
13. CZARNITZKI, D., HUSSINGER, K. and SCHNEIDER, C., 2011. Commercializing academic research: the quality of faculty patenting. *Industrial and Corporate Change,* 20(5), pp. 1403-1437.
14. EC, 2006. Study on evaluating the knowledge economy what are patents actually worth? The value of patents for today’s economy and society*,* Bruxelles: European Commission.

15. ELFENBEIN, D.W., 2007. Publications, patents, and the market for university inventions. *Journal of Economic Behavior & Organization,* 63(4), pp. 688-715.
16. FABRIZIO, K.R. and DI MININ, A., 2008. Commercializing the laboratory: Faculty patenting and the open science environment. *Research Policy,* 37(5), pp. 914-931.
17. FARRELL, M., 2008. *Universities that turn research into revenue.* [Online] Available at: http://www.forbes.com/2008/09/12/google-general-electric-ent techcx_mf_0912universitypatent.html [Accessed 19 July 2012].
18. FLEMING, L., 2001. Recombinant uncertainty in technological search. *Management Science,* 47(1), pp. 117-132.
19. FLEMING, L. and SORENSON, O., 2004. Science as a map in technological search. *Strategic Management Journal,* 25(8-9), pp. 909-928.
20. FRANZONI, C., 2007. Opportunity recognition in technology transfer organizations five case studies from UK and Italy. *International Entrepreneurship and Management Journal,* 3(1), pp. 51-67.
21. FRANZONI, C., 2009. Do scientists get fundamental research ideas by solving practical problems?. *Industrial and Corporate Change,* 18(4), pp. 671-699.
22. FRANZONI, C. and SCELLATO, G., 2010. The grace period in international patent law and its effect on the timing of disclosure. *Research Policy,* 39(2), pp. 200-213.
23. GALLINI, N.T. and WRIGHT, B.D., 1990. Technology-Transfer Under Asymmetric Information. *Rand Journal of Economics,* 21(1), pp. 147-160.
24. GAMBARDELLA, A., HARHOFF, D. and VERSPAGEN, B., 2011. *The Determinants of The Private Value Of Patented Inventions.* Working Paper edn. Available at: http://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_econ_ge_2_11/wipo_ip_econ_ge_2_11_determinants.pdf.
25. GITTELMAN, M. and KOGUT, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science,* 49(4), pp. 366-382.
26. GIURI, P., MARIANI, M., BRUSONI, S., CRESPI, G., FRANCOZ, D., GAMBARDELLA, A., GARCIA-FONTES, W., GEUNA, A., GONZALES, R., HARHOFF, D., HOISL, K., LE BAS, C., LUZZI, A., MAGAZZINI, L., NESTA, L., NOMALERI, O., PALOMERAS, N., PATEL, P., ROMANELLI, M. and VERSPAGEN, B., 2007. Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research Policy,* 36(8), pp. 1107-1127.
27. GÖKTEPE-HULTEN, D. and MAHAGAONKAR, P., 2010. Inventing and patenting activities of scientists: in the expectation of money or reputation? *Journal of Technology Transfer,* 35(4), pp. 401-423.
28. GRILICHES, Z., 1981. Market value, R&D, and patents. *Economics Letters,* 7(2), pp. 183-187.
29. HAEUSSLER, C., JIANG, L., THURSBY, J. & THURSBY, M., 2009. Specific and general information sharing among academic scientists. *NBER Working paper,* Issue 15315.
30. HALL, B.H., JAFFE, A. and TRAJTENBERG, M., 2005. Market value and patent citations. *Rand Journal of Economics,* 36(1), pp. 16-38.

31. HARGADON, A. and SUTTON, R.I., 1997. Technology brokering and innovation in a product development firm. *Administrative Science Quarterly,* 42(4), pp. 716-749.
32. HENDERSON, R., JAFFE, A.B. and TRAJTENBERG, M., 1998. Universities as a source of commercial technology: A detailed analysis of university patenting, 1965-1988. *Review of Economics and Statistics,* 80(1), pp. 119-127.
33. JAFFE, A.B., TRAJTENBERG, M. and HENDERSON, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics,* 108(3), pp. 577-598.
34. JENSEN, R. and THURSBY, M., 2001. Proofs and prototypes for sale: The licensing of university inventions. *American Economic Review,* 91(1), pp. 240-259.
35. JENSEN, R. A., THURSBY, J. G. & THURSBY, M. C., 2003. The Disclosure and Licensing of University Inventions*.* National Bureau of Economic Research, Inc., *NBER Working Papers 9734* .
36. KOGUT, B. and ZANDER, U., 1992. Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science,* 3(3), pp. 383-397.
37. Lissoni, F., Montobbio, F. & Seri, R., 2010. Ownership and impact of European university patents. *working paper.*
38. MOWERY, D.C., NELSON, R.R., SAMPAT, B.N. and ZIEDONIS, A.A., 2001. The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh-Dole act of 1980. *Research Policy,* 30(1), pp. 99-119.
39. MURRAY, F. and STERN, S., 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization,* 63(4), pp. 648-687.
40. NELSON, R. and WINTER, S., 1982. *An Evolutionary Theory of Economic Change.* Cambridge, MA.: Belknap.
41. OWEN-SMITH, J. and POWELL, W.W., 2001. To Patent or Not: Faculty Decisions and Institutional Success in Academic Patenting. *Journal of Technology Transfer,* 26(1), pp. 99-114.
42. SAMPAT, B.N., MOWERY, D.C. and ZIEDONIS, A.A., 2003. Changes in university patent quality after the Bayh-Dole act: a re-examination. *International Journal of Industrial Organization,* 21(9), pp. 1371-1390.
43. SCOTCHMER, S., 2004. *Innovation and Incentives.* Cambridge, MA: MIT Press.
44. SHANE, S., 2002. Selling university technology: Patterns from MIT. *Management Science,* 48(1), pp. 122-137.
45. STEPHAN, P.E., GURMU, S., SUMELL, A.J. and BLACK, G., 2007. Who's Patenting In The University? Evidence From The Survey Of Doctorate Recipients. *Economics of Innovation and New Technology,* 16(2), pp. 71-99.
46. STRUMSKY, D, LOBO, J and VAN DER LEEUW, S, 2012: Using patent technology codes to study technological change, *Economics of Innovation and New Technology*, 21(3), pp. 267-286
47. SWAMIDASS, P. M. and VULASA, V., 2008. Why university inventions rarely produce income? Bottlenecks in university technology transfer. *Journal of Technology Transfer,* 34, pp. 343–363.

48. THURSBY, J., FULLER, A.W. and THURSBY, M., 2009. US faculty patenting: Inside and outside the university. *Research Policy,* 38(1), pp. 172-187.
49. TRAJTENBERG, M., 1990. A Penny for Your Quotes - Patent Citations and the Value of Innovations. *Rand Journal of Economics,* 21(1), pp. 172-187.
50. TRAJTENBERG, M., HENDERSON, R. & JAFFE, A., 1997. University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology,* 5(1), pp. 19-50.
51. WALSH, J. P., CHO, C. & COHEN, W. M., 2005. View from the bench: Patents and material transfers. *Science,* 309(5743), pp. 2002-2003.
52. WEITZMAN, M.L., 1998. Recombinant growth. *Quarterly Journal of Economics,* 113(2), pp. 331-360.
53. WOOLDRIDGE, J.M., 2002. *Econometric Analysis of Cross-Section and Panel Data.* Cambridge: MIT Press.
54. ZUCKER, L.G., DARBY, M.R. and ARMSTRONG, J.S., 2002. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management Science,* 48(1), pp. 138-153.

7. TABLES

Table 1- Variables construction

Variable	Description
Forward citation	Number of forward citation after 6 years.
Prior scientific knowledge	Binary variable being 1 if the patent content is resonant with at least one prior publications of the scientists; 0 otherwise.
Prior technical knowledge	Number of prior patent at the same IPC (whole IPC) as current patent relative to total prior patents.
Inventor technical productivity (log+1)	The logarithm of prior patents plus one.
Inventor scientific productivity (log+1)	The logarithm of papers in the year of priority and two previous years plus one.
Number of claims (log)	The logarithm of Number of claims.
Number of inventors (log)	The logarithm of Number of Inventors in the same patent.
Count of cited patents (log+1)	The logarithm of patents that have been cited by a patent (backward citations) plus one.
Count of scientific articles cited (log+1)	The logarithm of publications that have been cited in the patent document (backward citations) plus one
First patent	Binary variable being 1 if the inventor is at his/her first patent; 0 otherwise.
Gender (Male)	Binary variable being 1 if the inventor is a man; 0 if a woman.
Career-age and Career-age squared	Years since PhD was awarded at the patent priority year.
Assignee type	Set of 3 binary variables based on different types of assignees: firm, university or others; 0 otherwise.
Assignee's Patenting experience	Categorical variable based on 6 intervals of prior patents count (0; 1-10; 11-100; 101-1000; 10001-10000 and over 10000).
Patent scope	Number of IPC classes assigned to the patent.
Patenting Trend	Number of patent granted by USPTO after 5 years.
Spatial Distribution	Number of patents granted by USPTO in the U.S state of assignee.
IPC classes dummies	Set of 22 dummy variables based on 2-digits main IPC class.
Year dummies	Set of 14 dummy variables based on the priority year of the patent.

Table 2- Descriptive statistics

Variables	Obs.	Mean	St.dev.	Min.	Max.
Forward citations	295	18.59	35.78	0	421
Prior scientific knowledge (=1 if the patent content is resonant)	295	0.24	0.43	0	1
Prior technical knowledge	295	.06	.20	0	1
Inventor technical productivity	295	9.94	17.95	0	76
Inventor technical productivity (log)	295	1.52	1.23	0	4.34
Inventor scientific productivity	295	22.53	19.09	0	90
Inventor scientific productivity (log)	295	2.77	1.02	0	4.51
Number of claims	295	25.31	19.87	1	176
Number of claims (log)	295	2.97	0.76	0	5.12
Number of inventors	295	3.53	1.94	1	13
Number of inventors (log)	295	1.11	0.55	0	2.56
Count of cited patents	295	15.99	24.65	0	190
Count of cited patents (log)	295	2.27	1.01	0	5.25
Count of scientific articles cited	295	10.39	15.02	0	128
Count of scientific articles cited (log)	295	1.70	1.26	0	4.56
First Patent (=1 if it is first patent of inventor)	295	0.30	0.46	0	1
Gender (=1 if inventor is a man)	295	0.95	0.21	0	1
Career age	295	21.98	11.35	4	55
Career age squared	295	611.60	642.57	16	3025
University (=1 if assignee is a university)	295	0.61	0.49	0	1
Firm (=1 if assignee is a firm)	295	0.29	0.45	0	1
Others (=1 if assignee is individual/ government / both university and firm)	295	0.10	0.30	0	1
Assignee patenting experience	292	3.08	1.30	1	6
Patent scope	295	1.90	1.76	1	15
Patenting Trend	295	760008.7	101443.8	523070	875789
Patenting Trend (log)	295	13.53	0.15	13.17	13.68
Spatial Distribution	263	6949.82	6298.96	55	19692
Spatial Distribution (log)	263	8.34	1.15	4.01	9.89
Priority year	295	1998.45	3.63	1992	2005

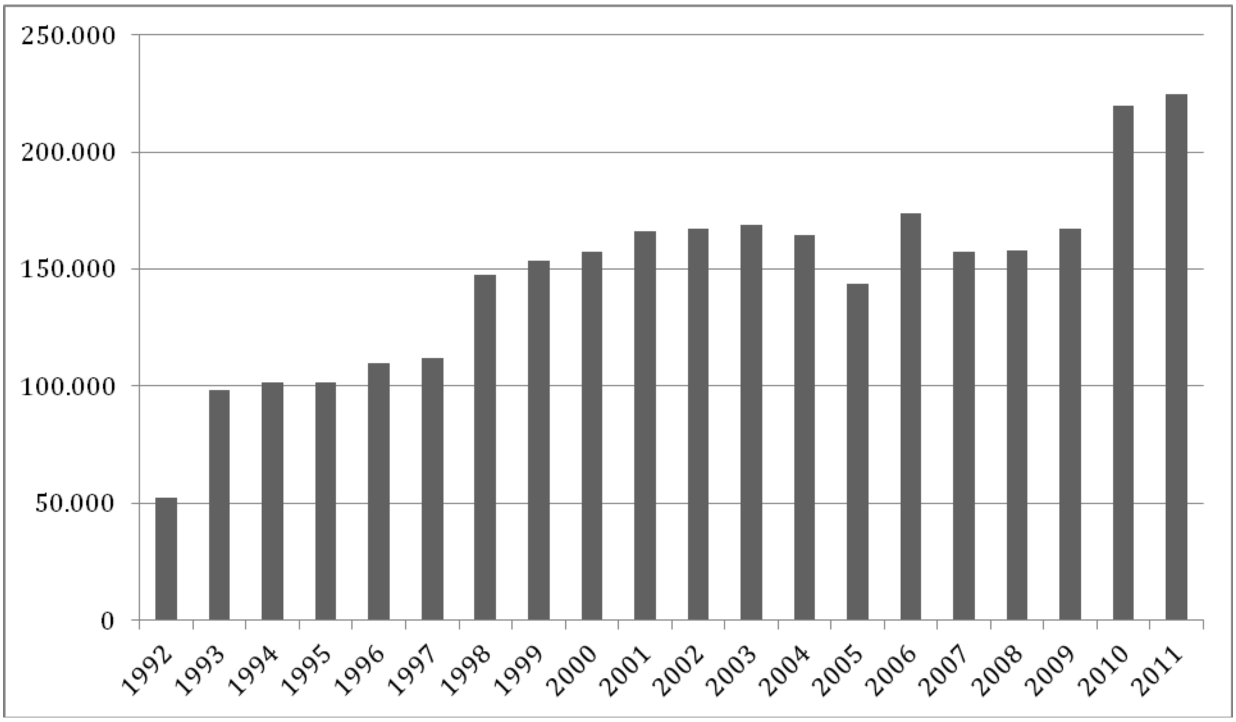
Table 3- Distribution of patents for each inventor (n=95)

Prior patents	Freq.	Percent	Cum.
0	27	28.42	28.42
1	21	22.11	50.53
2	15	15.79	66.32
3	6	6.32	72.63
4	5	5.26	77.89
5	2	2.11	80
6	2	2.11	82.11
7	2	2.11	84.21
8	1	1.05	85.26
9	4	4.21	89.47
10	3	3.16	92.63
12	2	2.11	94.74
13	1	1.05	95.79
21	1	1.05	96.84
22	2	2.11	98.95
76	1	1.05	100
Total	95	100	

Table 4 - Patents by prior knowledge set of the inventor

	Prior scientific knowledge		
Prior technical knowledge	**0**	**1**	**Total**
0	182 (61.7%)	62 (21.0%)	244 (82.7%)
1	42 (14.2%)	9 (3.1%)	51 (17.3%)
Total	224 (76.0%)	71 (24.0%)	295

Figure 1- Yearly distribution of patents (utility) granted by USPTO[10]



[10] Data extracted from USPTO: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports_stco.htm

Table 5- Correlation matrix

	Forward citation	Prior scientific knowledge	Prior technical knowledge	Inventor technical productivity (log)	Inventor scientific productivity (log)	Number of claims (log)	Number of inventors (log)	Count of cited patents (log)	Count of scientific articles cited (log)	First patent	Gender (Male)	Career-age	Career-age squared	University	Firm	Assignee's Patenting experience	Patent scope	Trend_5year (log)	Spatial dist (log)
Forward citation	1.000																		
Prior scientific knowledge	0.182	1.000																	
Prior technical knowledge	-0.062	-0.075	1.000																
Inventor technical productivity (log)	0.087	-0.182	-0.020	1.000															
Inventor scientific productivity (log)	0.029	0.196	0.103	-0.019	1.000														
Number of claims (log)	0.096	-0.104	-0.050	0.030	-0.021	1.000													
Number of inventors (log)	0.068	-0.027	0.017	-0.112	0.093	0.195	1.000												
Count of cited patents (log)	0.045	-0.135	0.253	0.075	-0.052	0.070	-0.036	1.000											
Count of scientific articles cited (log)	0.156	0.037	0.069	-0.033	0.065	0.070	0.080	0.275	1.000										
First patent	-0.099	-0.028	-0.197	-0.607	-0.143	-0.044	-0.019	-0.045	0.003	1.000									
Gender (Male)	0.045	0.007	0.068	0.104	0.025	0.015	-0.045	-0.053	-0.065	-0.040	1.000								
Career-age	-0.036	-0.269	-0.065	0.639	-0.183	0.028	-0.049	0.015	-0.099	-0.220	0.146	1.000							
Career-age squared	-0.019	-0.254	-0.083	0.687	-0.172	-0.002	-0.096	0.005	-0.093	-0.212	0.138	0.974	1.000						
University	0.098	0.204	-0.200	-0.084	0.040	0.148	0.037	-0.198	0.177	0.007	-0.091	0.033	0.013	1.000					
Firm	-0.087	-0.221	0.227	0.116	-0.048	-0.133	-0.084	0.221	-0.170	-0.024	0.075	0.013	0.029	-0.917	1.000				
Assignee's Patenting experience	0.035	0.021	-0.191	0.068	0.083	-0.094	0.033	-0.206	0.036	-0.015	-0.037	-0.053	-0.010	0.107	-0.138	1.000			
Patent scope	-0.028	-0.102	0.018	-0.015	0.083	-0.006	-0.007	-0.005	-0.017	-0.073	0.092	0.037	0.015	0.082	-0.063	-0.126	1.000		
Trend_5year (log)	0.037	0.064	0.110	0.087	0.156	0.147	0.229	0.052	0.100	-0.363	0.059	0.141	0.069	0.257	-0.259	-0.184	0.055	1.000	
Spatial dist (log)	0.129	-0.084	0.110	0.216	0.096	0.082	-0.009	0.152	0.139	-0.182	-0.131	0.070	0.105	0.088	0.025	-0.028	-0.095	0.214	1.000

Table 6- Negative Binomial and QLM Poisson models. Dependent variable: forward citation

	Negative Binomial			QML Poisson		
	Model1	Model2	Model3	Model1	Model2	Model3
Prior scientific knowledge	0.399	0.449	0.496	0.535	0.569	0.599
	(0.194)**	(0.213)**	(0.197)**	(0.251)**	(0.246)**	(0.238)**
Prior technical knowledge	-0.635	-0.558	-1.091	-0.743	-0.745	-0.958
	(0.345)*	(0.354)	(0.318)***	(0.334)**	(0.348)**	(0.337)***
Inventor’s technical productivity (log)	0.265	0.330	0.145	0.187	0.215	0.119
	(0.118)**	(0.137)**	(0.130)	(0.158)	(0.177)	(0.173)
Inventor’s scientific productivity (log)	0.158	0.126	0.169	0.051	0.001	0.032
	(0.072)**	(0.102)	(0.085)**	(0.096)	(0.104)	(0.092)
Number of claims (log)	0.327	0.309	0.241	0.287	0.278	0.264
	(0.094)***	(0.102)***	(0.101)**	(0.102)***	(0.100)***	(0.106)**
Inventor count (log)	0.234	0.263	0.301	0.327	0.350	0.329
	(0.165)	(0.188)	(0.180)*	(0.295)	(0.300)	(0.270)
Count of cited patents (log)	0.130	0.146	0.209	0.163	0.177	0.165
	(0.074)*	(0.078)*	(0.089)**	(0.094)*	(0.109)	(0.087)*
Count of scientific articles cited (log)	0.194	0.199	0.229	0.178	0.173	0.161
	(0.065)***	(0.072)***	(0.065)***	(0.086)**	(0.097)*	(0.093)*
First patent	0.167	0.240	-0.098	-0.257	-0.207	-0.381
	(0.274)	(0.302)	(0.301)	(0.305)	(0.319)	(0.301)
Gender (Male)	0.589	0.533	0.647	1.146	1.293	1.308
	(0.473)	(0.595)	(0.497)	(0.532)**	(0.620)**	(0.536)**
Career age	-0.061	-0.075	-0.076	-0.034	-0.058	-0.020
	(0.034)*	(0.034)**	(0.034)**	(0.030)	(0.032)*	(0.027)
Career age squared	0.001	0.001	0.001	0.001	0.001	0.000
	(0.001)	(0.001)*	(0.001)**	(0.001)	(0.001)	(0.001)
University assignee	0.241	0.405	0.461	0.592	0.742	0.560
	(0.258)	(0.320)	(0.399)	(0.247)**	(0.283)***	(0.370)
Firm assignee	0.057	0.084	0.284	0.368	0.422	0.299
	(0.277)	(0.337)	(0.425)	(0.192)*	(0.227)*	(0.323)
Assignee experience	0.006	-0.043	-0.036	0.098	0.086	0.090
	(0.063)	(0.070)	(0.080)	(0.098)	(0.103)	(0.100)
Patent scope	0.004	0.006	0.028	0.064	0.071	0.106
	(0.037)	(0.036)	(0.036)	(0.057)	(0.060)	(0.051)**
Trend			-3.475			-2.876
			(6.577)			(7.811)
Geographical distribution			0.039			0.228
			(0.081)			(0.096)**
IPC dummy1-21	Yes	Yes	Yes	Yes	Yes	Yes
Years dummy1-14	Yes	Yes	Yes	Yes	Yes	Yes

_cons	-4.406	-4.042	42.719	-5.810	-5.774	31.078
	(1.279)***	(1.510)***	(89.340)	(1.953)***	(2.105)***	(106.397)
Lnalpha	0.124	0.157	0.089			
	(0.089)	(0.098)	(0.089)			
N	295	267	263	295	267	263

Clustered Robust Std. Err. in parentheses * p<0.1; ** p<0.05; *** p<0.01

8. Appendices

Appendix 1-Robustness Checks of functional form. Negative Binomial estimates.

	Model1	Model 2	Model 3	Model 4	Model 5
Prior scientific knowledge	0.404	0.399	0.348	0.348	0.458
	(0.195)**	(0.196)**	(0.208)*	(0.204)*	(0.181)**
Prior technical knowledge	-0.991		-0.636	-0.811	-0.958
	(1.598)		(0.285)**	(0.370)**	(0.306)***
Prior technical knowledge squared	0.401	-0.628			
	(1.677)	(0.355)*			
Inventor's technical productivity (log)	0.271	0.261	0.275	0.300	0.179
	(0.121)**	(0.119)**	(0.128)**	(0.126)**	(0.122)
Inventor's scientific productivity (log)	0.156	0.159	0.140	0.147	0.113
	(0.072)**	(0.072)**	(0.079)*	(0.079)*	(0.079)
Number of claims (log)	0.325	0.333	0.304		0.315
	(0.094)***	(0.094)***	(0.096)***		(0.094)***
Inventor Count	0.233	0.235	0.265	0.276	0.243
	(0.165)	(0.166)	(0.156)*	(0.165)*	(0.157)
Count of cited patent (log)	0.129	0.132		0.128	0.187
	(0.073)*	(0.074)*		(0.077)*	(0.082)**
Count of scientific articles cited	0.193	0.195		0.199	0.226
	(0.065)***	(0.065)***		(0.067)***	(0.069)***
First patent	0.171	0.176	0.175	0.089	-0.057
	(0.277)	(0.276)	(0.291)	(0.282)	(0.256)
Gender (Man)	0.586	0.592	0.551	0.682	0.562
	(0.473)	(0.473)	(0.484)	(0.416)	(0.512)
Career age	-0.060	-0.062	-0.051	-0.041	-0.074
	(0.035)*	(0.034)*	(0.033)	(0.034)	(0.031)**
Career age squared	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)**
University assignee	0.242	0.242	0.430	0.263	0.350
	(0.259)	(0.258)	(0.278)	(0.267)	(0.264)
Firm assignee	0.061	0.050	0.366	0.047	0.123
	(0.279)	(0.277)	(0.279)	(0.287)	(0.247)
Assignee exprience	0.008	0.006	-0.035	0.001	-0.055
	(0.063)	(0.063)	(0.065)	(0.066)	(0.069)

Patent Scope	0.005	0.003	0.009	-0.028	-0.033
	(0.038)	(0.037)	(0.038)	(0.035)	(0.031)
	(0.388)***	(0.375)***	(0.358)***	(0.372)***	
IPC dummy 1-21	Yes	Yes	Yes	Yes	Yes
Years dummy 1-14	Yes	Yes	Yes	Yes	No
Share patent cited/Total citations			-0.581		
			(0.265)**		
Policy change					-0.697
					(0.219)***
_cons	-4.397	-4.445	-3.387	-2.886	-1.833
	(1.275)***	(1.283)***	(1.294)***	(1.222)**	(1.120)
lnalpha	0.124	0.125	0.162	0.154	0.211
	(0.089)	(0.089)	(0.087)*	(0.090)*	(0.087)**
N	295	295	288	295	295

Clustered Robust Std. Err. in parentheses * p<0.1; ** p<0.05; *** p<0.01

Appendix 2- Robustness check by considering all patents belonging to patent family. Negative Binomial estimates.

	Negative Binomial	
	Model1	Model 2
Prior scientific knowledge	0.386	0.410
	(0.195)**	(0.212)*
Prior technical knowledge	-0.381	-0.400
	(0.216)*	(0.225)*
Inventor's technical productivity (log)	0.210	0.230
	(0.131)	(0.149)
Inventor's scientific productivity (log)	0.098	0.060
	(0.080)	(0.109)
Number of claims (log)	0.348	0.345
	(0.093)***	(0.099)***
Inventor count (log)	0.266	0.297
	(0.137)*	(0.151)**
Count of cited patents (log)	0.013	0.015
	(0.076)	(0.077)
Count of scientific articles cited (log)	0.157	0.160
	(0.056)***	(0.061)***
First patent	0.144	0.142
	(0.265)	(0.281)
Gender (Male)	0.603	0.585
	(0.425)	(0.524)
Career age	-0.068	-0.089
	(0.039)*	(0.041)**
Career age squared	0.001	0.001
	(0.001)	(0.001)*
University assignee	0.207	0.253
	(0.205)	(0.260)
Firm assignee	0.148	0.117
	(0.216)	(0.253)
Assignee experience	-0.036	-0.081
	(0.079)	(0.086)
Patent scope	-0.020	-0.016
	(0.034)	(0.035)
IPC dummy 1-21	Yes	Yes
Years dummy 1-14	Yes	Yes
_cons	-3.442	-2.936
	(1.293)***	(1.537)*
Lnalpha	0.210	0.239
	(0.086)**	(0.093)***

N	373	348

Clustered Robust Std. Err. in parentheses * p<0.1; ** p<0.05; *** p<0.01