Available online at www.sciencedirect.com
ScienceDirect
Research Policy 36 (2007) 813–831
research policy
www.elsevier.com/locate/respol

# Institutional complementarity and inventive performance in nano science and technology

Andrea Bonaccorsi [a], Grid Thoma [b,*]

[a] School of Engineering, University of Pisa, Pisa, Italy
[b] Laboratory of Economics and Management, Sant'Anna School of Advanced Studies and University of Camerino, Piazza Martiri della Libertà 33, 56127 Pisa, Italy

Available online 9 April 2007

**Abstract**

Academics and policy makers are investigating the relations between science and technology in the emerging field of nano science and technology (NST) and the effectiveness of different institutional regimes. We use multiple indicators to analyze the performance of inventors working in NST. We clustered patents into three groups according to the scientific curricula of the inventors. The first group consists of patents whose inventors are all authors of at least one scientific publication in NST, while the second is made up of patents invented by individuals who have no scientific publication in the field. Thirdly, we isolated those patents that have at least one inventor who is also author of at least one scientific publication in NST. The underlining presumption of this classification is that of a proxy of different institutional complementarities of inventive collective action in NST.


*Keywords:* Science–technology relation; Emerging field; Nanotechnology; Patent quality; Inventive productivity

## 1. Introduction

Nano science and technology (hereafter also NST) regards the understanding and control of matter at the nano scale, which is a billionth of a meter. There is consensus in the scientific community that NST broadly involves: (i) research and technology development at the atomic, molecular or macromolecular levels, in approximately the 1–100 nm range; (ii) creating and using structures, devices and systems that have novel properties and functions because of their small and/or intermediate size; (iii) the ability to control or manipulate on the atomic scale.[1]

There is also consensus among scientists that NST came into being in 1981, when the *Scanning Tunneling Microscope* (US Patent 4343993; hereafter, also STM) was invented by Gerd K. Binnig and Heinrich Rohrer at the IBM Research Laboratory in Zurich. In 1986, they were awarded the Nobel Prize for this discovery. The STM yields atomic-scale images of metal and semiconductor surfaces, something which had not been possible with the so-called *Topografiner*, invented by Russell Young in the late 1960s. The range of materials that can be imaged with a scanning device increased with the invention of the *Atomic Force Microscope* (US Patents 4724318 and RE33387; hereafter, also AFM) by Gerd K. Binnig in 1986.[2]

* Corresponding author. Tel.: +39 0328 3837430; fax: +39 050 883344.
*E-mail address:* thoma@sssup.it (G. Thoma).

[1] The definition is that of the National Nanotechnology Initiative (www.nano.gov). We validated this definition in several interviews conducted with nano scientists in 2004–2005.

[2] The invention was filed in August 1986. A related scientific article had been published 6 months earlier [cf. Gerber, C., Binnig, G., Fuchs, H., Marti, O., Rohrer, H., 1986. Scanning tunneling microscope combined with a scanning electron microscope. Rev. Sci. Instrum. 57 (February), 221–224].

0048-7333/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2007.02.009

These enabling instruments were invented at and with the support of the IBM Corporation, which was interested in scientific advances within the semiconductor industry. They soon realized that the STM and the AFM could be used in a vast array of scientific and technological fields, such as chemistry, biology, biotechnology, telecommunications, and many others.

In this paper, we will try to make sense of and to measure the importance of scientific knowledge in fuelling technological inventions in NST. We contribute to the literature on science–technology interactions and the industrial dynamics of emerging fields by providing recent and large-scale evidence about a young industry and developing new methods for measurement and quantitative analysis. In particular, we develop a methodology to match information on scientific publications and patents, enabling characterization of several *communities of inventors* that implement science–technology interactions in alternative ways. The paper is organized as follows. In the next section, we briefly review the literature, pointing to limitations in existing methodologies and substantive explanations. In Section 3, we introduce descriptive statistics on NST and develop the proposed methodology. In Sections 4 and 5, we investigate the performance implications of alternative ways of arranging science–technology interactions, as evidenced by the matching analysis. We propose some measures of patent quality and test various hypotheses about differences between indicators across patents produced by different communities of inventors. Sections 6 and 7 add new dimensions of performance, analyzing the productivity of individual inventors and the propensity to found a start-up company. Section 8 reviews the evidence and poses new research questions.

## 2. Background literature on science–technology interactions

In the 1990s, the notion that technological developments are increasingly dependent on advancements in science was proposed repeatedly. On one hand, scientometric literature drew attention to the sharp increase in the number and share of non-patent literature citations in patents (Narin and Olivastro, 1992; Narin et al., 1997), suggesting that inventors increasingly make direct use of inputs from published scientific research. Patents may be based not only on the prior art documented in other patents, but in part or fully on new scientific knowledge. Since published scientific research results can be used to illustrate the state of the art against which the application has to be evaluated, patent examiners will then search for relevant references in the scientific literature. The logic of these references is to document the material that is held against the application. Using this metrics, a taxonomy of industries based on the dependence on science can be developed (Grupp, 1992; Heinze and Schmoch, 2004; Tijssen, 2004). More recently, techniques for tracing back networks of patent citations have been developed (Popp, 2005; Verspagen, 2005). Tracing back the full ramification of citations from more recent patents to historical ones can give insights into the underlying dynamics of knowledge.

On the other hand, industry case studies on biotechnology (McKelvey, 1996; Orsenigo, 1990; Owen-Smith et al., 2002; Zucker and Darby, 1996), chemical and electrical engineering (Kenney and Goe, 2004; Mowery and Rosenberg, 1998), semiconductor and laser (Klepper, 2001), and medical instruments (Trajtenberg, 1990) illustrated important examples in which the very definition of industrial applications was only made possible by the discovery of new physical properties of nature. In these fields, the origin of entrepreneurship can often be traced back to scientists from the academic world or to scientists in large and technologically advanced companies.

The importance of this literature can be better understood in relation to the broader theoretical treatment of the relations between science and technology and, more generally, to the conditions for the productive use of knowledge (Dasgupta and David, 1994). In fact, the critique of so-called *linear models* carried out in the 1980s (Kline and Rosenberg, 1986; Rosenberg, 1982) made it clear that technological knowledge is subject to a specific internal dynamics that is relatively independent of scientific advancements. Firms only benefit from science indirectly (Pavitt, 1990), and the use of scientific research for industrial innovation is less about direct collaboration and more about the constitution of human capital (Cohen et al., 1987; Nelson, 1986).

Alongside this institutional treatment, a conceptualization of the nature of technological knowledge has taken place that goes beyond highly stylized representations. Design knowledge started to be characterized as a collection of highly specific rules for problem solving and the selection of acceptable solutions (Klein, 1985; Stankiewicz, 2000; Vincenti, 1990), constituting an autonomous body of knowledge. Design is not applied research and engineering is not applied physics.

Against this background of criticism of the linear model and the articulation of the relative independence of different types of knowledge, the discovery of the sharp increase in the "scientific content" of patents, inventions, and companies still lacks a rigorous theoretical treatment. Is this a sign that the linear model is still valid?

Or is it evidence of a fundamental change in production technology, whereby the flow of knowledge between scientific research and technology is less mediated and more direct? Or is it just a transitory stage in the long-term evolution of industries?

In other words, there is a strong need to go beyond the stylized evidence collected with patent data and industry case studies in the 1990s, and to build up a more general framework for the analysis of the productive use of knowledge.

This task, however, is made difficult by a number of limitations in the existing literature. In substantive terms, the critique of linear models has yet to generate a stream of studies on the specific non-linear ways in which science and technology interact, providing evidence regarding the nature and intensity of feedback loops and iterations. For example, Stankiewicz (1997) made the important distinction between discovery-driven innovation and design-driven innovation. The former depends crucially on an understanding of nature, while the latter is driven by internal technical issues and is mainly influenced by applications and demand considerations. However, these categories of innovation and the possible transition between them in long-term technological evolution have still to be identified and investigated. The micro-mechanisms governing the generation, validation, and transmission of knowledge between science and technology have not yet been explored in great detail. The epistemic foundations of science–technology interactions are still unclear, despite the pioneering analyses offered by Callon et al. (1991).

On the methodological side, several shortcomings in the existing measures should be recognized. First of all, non-patent literature (hereafter also NPL) citations suffer from an important limitation: it is not clear to what extent they are assigned by inventors or by examiners. It is well known that inventors primarily introduce references in the USPTO, while in the European system they are introduced exclusively by the examiners. Breschi and Lissoni (2004) claimed that, at least in the US patent system – since references are assigned by different actors, who quote mainly US references for reasons of availability and for different purposes – there is a severe distortion in the interpretation of data. The full validity of information on cited patents has to be established, given that the motivations for a patent to cite another patent are rather intricate and raise legal and strategic considerations. Therefore, both measurement and validity issues are involved here.

Second, NPL citations do not convey any information about the degree to which the scientific content has generated valuable innovation. Since we know that the distribution of patents by degree of usefulness is extremely skewed, it is possible that patents with a high number of non-patent references are among those that are never used, and so have limited economic value. One approach to mitigate this limitation is given by a careful analysis of patent quality, using the indicators proposed in the literature initiated by Trajtenberg (1990) and fully developed by Jaffe et al. (1993).[3] There is sufficient evidence in the literature that the economic value of patents is associated with the number and quality of citations received in other patents (Hall et al., 2005; Harhoff et al., 1999; Jaffe and Trajtenberg, 2002). Harhoff et al. (2003) and Lanjouw and Schankerman (2001) have suggested a different metrics, i.e. the existence of litigation for patents, implying that patents for which assignees are willing to pay for defense against infringement have greater economic value.

There is another important limitation to non-patent literature. Patent examiners typically start by checking existing patents to look for prior art that might limit a patent's claim. They turn to NPL once these searches are exhausted. Since only one reference to prior art is needed to limit a claim, NPL will be most prevalent when there is little prior patent art. Thus, over time, as more patents are granted in a given field, we would expect NPL citations to fall, though this need not be a signal of falling quality.[4]

More fundamentally, existing methodologies identify science–technology interactions using documents, not individuals. A relation is said to be in place if and only if a paper trail can be identified. This largely ignores the variety of motivations that may lead to citations.

Therefore, a new approach is needed to capture the complexity of interactions between science and technology. In this paper we develop a new methodology for tracing and measuring these interactions, based on matching algorithms of names of individuals between different datasets. We try to match the names of individuals between datasets, i.e. publications (individuals as authors), patents (individuals as inventors) and companies (individuals as entrepreneurs or partners). For all individuals in our final dataset, we know whether and when he/she has published, invented, and created a company, and we know all the associated characteristics (and related indicators) for his/her papers, patents, and companies. This approach does not take the patent as the unit of analysis but rather the individual, i.e. the author, the inventor, or the entrepreneur. The analysis

[3] For a survey of the literature, see Jaffe and Trajtenberg (2002).

[4] We would like to thank one of the two anonymous reviewers for this argument.

focuses on building up time series datasets and on the mobility and carriers of the individual across different institutional settings (Bozeman et al., 2001; Gambardella et al., 2005; Shane, 2004). In combining different roles around the same individual, we will be able to build up the full profile of inventors and to put forward various conjectures about the intrinsic dynamics of science and technology.

Assuming individuals, rather than patents or papers, as the unit of observation has several advantages. First, it helps in identifying heterogeneities in the performance of inventors. Simple but robust classifications such as the one we propose in this paper are powerful enough to throw light on several paths to innovation followed by scientists and inventors in the course of their professional lives. In turn, this contributes to the identification of the dynamic nature of science–technology relations.

Second, the value of scientific production and of invention can be addressed directly. The literature on patent quality, emphasizing the relation between the value of the invention and successive citations in other patents, uses indicators that are internal to the patent system. Studies such as Shane (2004) try to trace forward the history of patents from academia through to the creation of startup firms, but do so with a limited focus (a large US university). Using our dataset we are able to disentangle the full path leading from scientific publications to patents (or vice versa) or from scientific publications and patents to the creation of a company, as well as any other relevant path. Different performance measures – scientific, technological, and economic – are regressed over their initial characteristics and the career paths, as in studies of the industrial dynamics of firms.

Finally, focusing on inventors allows us to carry out a longitudinal analysis at an individual level, helping to identify and follow the steps of careers that lead to scientific activity, invention, and entrepreneurship. In this sense our analysis is a first preparatory work that might be followed by extensive examination of scientists' resumes, in order to trace the evolution of careers and the patterns of mobility.

## 3. Search regimes in nano science and technology

NST is an extremely interesting case in which the micro-mechanisms of science–technology interactions and the origins of entrepreneurship can be detected with great precision, due to the novelty of the field and the relative wealth of available documentation. We characterize this new field along three dimensions. First of all, the rate of growth in the production of scientific results: scientific fields that exhibit exponential growth (or grow at significantly greater rates than average) have completely different properties with respect to regimes that grow linearly. Second, the degree of diversity of directions of research: in some areas all research programmes converge on a few areas, usually associated with crucial experiments based on a commonly held body of theory, while in other areas the agreement on general theories generates a proliferation of (weakly or strongly) competing hypotheses and research programmes, following a divergent dynamics. In NST we expect a proliferation pattern of research programmes, driven by the specific combination of deeper understanding of the properties of matter at low levels of resolution and design objectives. Third, the importance and nature of complementarities in knowledge: while in big science the most important complementarities take place with large experimental facilities, in new emerging fields they are most likely to take the form of human capital and institutional complementarities. In particular, diversified knowledge bases are brought to the frontier of science, while both discovery and invention require a structured interdependence between institutions characterized by different goals (e.g. industry, academia, hospitals). Based on these dimensions, a number of disciplines can be identified, including life sciences after the molecular biology revolution, computer science, materials science, and nano science. These broad disciplines share the following properties: they have been growing exponentially or much more than average for a long period, they follow a dynamic process of divergent research, and they are based on institutional and human capital complementarity. We will use these dimensions to characterize the emerging field of NST, following the label of search regimes proposed in the explanatory work of Bonaccorsi (2005).

### 3.1. Data on nano patents and publications

The search strategy for nanotechnology patents mainly had to be based on keywords, since the specific IPC-subclass B82B for this field was introduced in the year 2000 and does not cover the period up until then. Therefore, it only contributes to identifying a very small portion of all documents.[5]

We used a keyword search strategy suggested by the Fraunhofer-ISI Institute in Karlsruhe, which we found

[5] By retrieving the nanotechnological IPC classes B81B, B81C and B82B, we found the following number of patents: in USPTO there were, respectively, 37, 27 and 14, compared to 20, 12 and 1 in EPO.

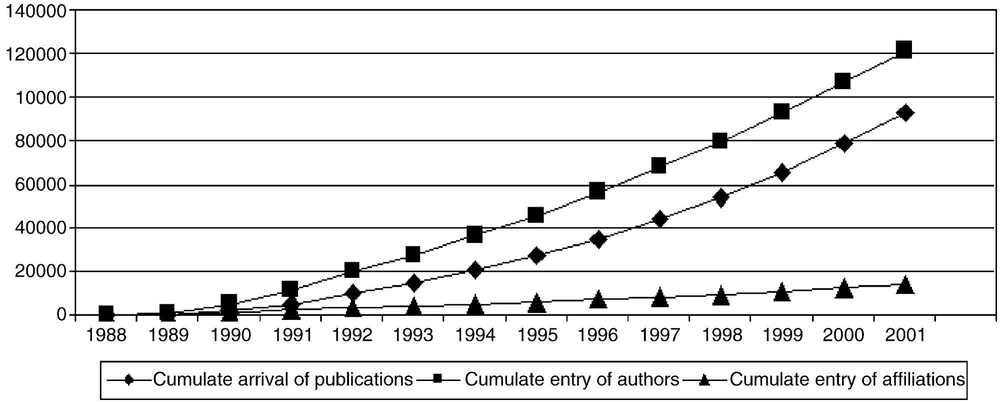


Fig. 1. Cumulate arrivals in nano science. *Source*: Our elaboration from WoS-ISI.

to be the most complete and extensively validated static keyword methodology (Fraunhofer-ISI, 2002).[6]

In the following analysis, we selected data from USPTO since 1971. Given the important role of the United States as a locus of technical change in recent decades, we believe that limiting ourselves to U.S. patenting activity does not constitute a serious drawback for a preliminary investigation of this kind.[7] We executed the ISI strategy with regard to the title and the abstract of a patent.[8]

Nanotechnology proved to be a significant phenomenon. We obtained a sample of 4828 patents granted before May 2004, classified by 1192 examiners in 331 three-digit U.S. technological classes. The collective action in NST involves more than 8000 inventors located in over 3000 cities in 52 countries. The patents are assigned to more than 1900 assignees, located in over 800 cities in 37 countries.

We performed the same keyword search strategy for publications, cleaning it from references to technological classes. Follow-up interviews with scientists in the field validated this strategy for publications (Beltram, 2005). We searched both in the title and keywords of a publication. The data source is the SCI and SSCI of the ISI database for the years 1988–2001. We obtained a pool of 93,149 publications, authored by 119,640 individuals, affiliated to 13,752 institutes.[9]

More generally, our dataset considers only a "seed", from which a more complete dataset might be generated by applying iteration techniques; so, for example, citation links may be used to include not only papers that show one of the keywords but also papers cited in the seed that might be part of an emerging field.[10]

### 3.2. Rate of growth of nano science and nanotechnology

In the case of nano science, it is clear that there has been impressive growth not only in individual fields (such as carbon nanotubes, nanocoatings or nanobiotechnology) but in the discipline as a whole. In less than 10 years, an army of almost 120,000 scientists worldwide has mobilized around the new discipline. Several thousand new institutions worldwide have entered the field. The scientific output of such collective action amounts to about 100,000 publications (see Fig. 1).[11]

[6] In order to circumvent the problem of an accidental selection of keywords given by experts, they listed all terms in the patent database beginning with "nano". An expert in NST assessed each term to ascertain whether it is used in the context of nanotechnology and whether it indicates an unambiguous relation to this field. Forty keyword queries were obtained, identifying singularly a field. See Fraunhofer-ISI (2002), available at www.cwts.nl/ec-coe.

[7] We ran the query procedure also on the EPO database for the same periods, obtaining lass than 20% of the patents identified at USPTO. It is generally recognized that important inventions are patented across patent offices, and are usually found at USPTO.

[8] The source of data is constituted by the Delphion patent database (DPD), which is an on-line proprietary database, accessible from www.delphion.com. It includes data from different national Patents Offices. In particular, it offers a complete text and images of all patents issued by the US Patent and Trademark Office (USPTO) since 1971. It offers the possibility to query in a very intuitive manner the remote database.

[9] The unit of analysis is constituted by the parent institute. Data on individuals have been cleaned by correcting for classical distortions. Data on affiliations are raw data and should be treated carefully. The cleaning of affiliations is underway.

[10] In a later stage of the *PRIME Nanodistricts* project, a new dataset, based on similar procedures, was developed by Zitt (2005). We plan to investigate this type of database in future research.

[11] The database mentioned in footnote 10 consists of approximately 180,000 publications in the period 1991–2004.

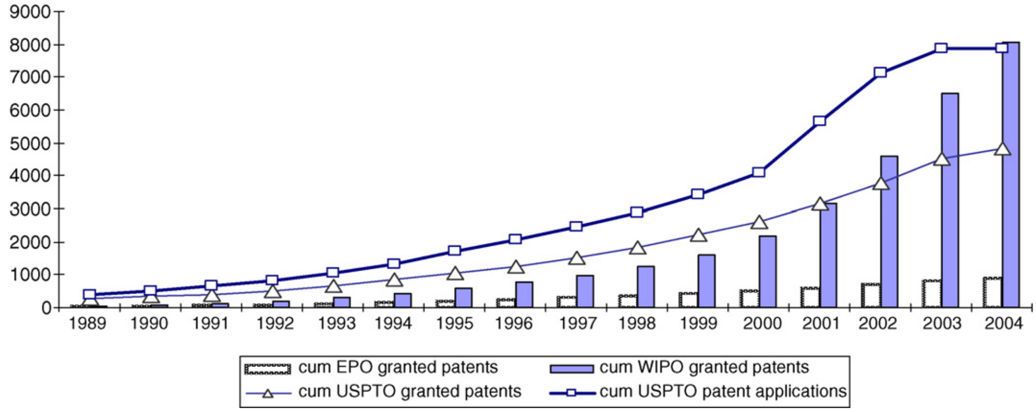


Fig. 2. Cumulate arrivals in nanotechnology, May 2004. *Source*: Our elaboration. *Note*: The patents for 2004 include only those granted before May 2004.

In the case of nanotechnology, we can discern a less stable dynamic of inventive output in terms of growth rates. There has been an impressive growth in the production of patents, especially in recent years (1996–2002). The USPTO has patented several thousand inventions in nanotechnology, with around 6600 files at the end of 2005. In Fig. 2, the data on USPTO patents (used throughout this paper) have been integrated with data on EPO patents and WIPO PCT patents, using the same keyword structure, in order to briefly observe other sources. While European patents are at very low levels, the number of patents that fall within the patent cooperation treaty, signed also by Japan, are very high. In particular, Japanese patents exceed 7000 in the period 1976–2004 (7469 up to December 2004).

What can be observed is a dynamic process characterized by an average growth rate which is far larger than average in science and engineering for all the years in the period. To give an order of magnitude, total publications in SCI grew annually by 3% in the period 1990–1993 and around 1% in 1998–2001. The peak rate was 14% in 2003, following a drop of 2% in 2002 (our elaboration on WoS-ISI data).

However, an exponential growth in publications was not observed, while the growth of applications in patents is exponential until 2002. This is in contrast with recent interesting results from Zucker and Darby (2003), who find exponential growth in publications in the period 1980–2000. We interpret this difference with respect to methodology and substance. Zucker and Darby's dataset was built using the generic word "nano*" while our dataset follows a static combination of keywords. Therefore, the two datasets have different statistical properties. Our dataset may underestimate the production of papers that use only completely new keywords, as it is based on a static keyword list. On the other hand, Zucker and Darby's dataset may include false positives – for example, nano-seconds, nanoplankton, nanoflagelate, and so forth – and is subject to manipulations from authors that include the word "nano" merely out of fashion.[12] In addition, Zucker and Darby's data refer to the period 1980–2004, while our sample is limited to the period 1988–2001. As a matter of fact, the share of nano-articles per 1000 science articles started to grow significantly only after 1990 (Zucker and Darby, 2003, p. 55; Fig. 1).

From a substantive point of view, there may have been a catching-up effect in recent years, due to the legitimating of keywords included in the sample (and hence a decrease in growth rates). Comparing publications with patents it is clear that the impressive growth in publications took place 5–7 years before the surge in patenting. This is a relatively short period for the real economic effects.

### 3.3. Degree of diversity

As argued previously, there are strong reasons to expect divergent dynamics in NST. Within the overall

[12] The use of a single keyword search strategy, in general, might have many drawbacks. For example, Zucker and Darby identify around 2000 USPTO-granted patents in the period 1981–2000 using the word "nano" either in the title or in the description. Following the ISI Fraunhofer methodology, on the contrary, we identified no less than 2700 USPTPO-granted patents in the same period.
When we tried to replicate their approach, we obtained different results. On one hand, we executed a search based only on the word "nano" in the USPTO archive (www.uspto.gov): (TTL/nano$ OR SPEC/nano$) AND ISD/1/1/1980 → 1/1/2001. We found more than 50,000 patents over the 1980–2000 period, clearly an unrealistic amount. More interestingly, use of the word "nano" does not turn up the US04343993 patent granted in 1982 to Binnig and Rohrer for the invention of the scanning tunnelling microscope; we verified directly that the suffix "nano" does not appear in the 11-page document of that patent.

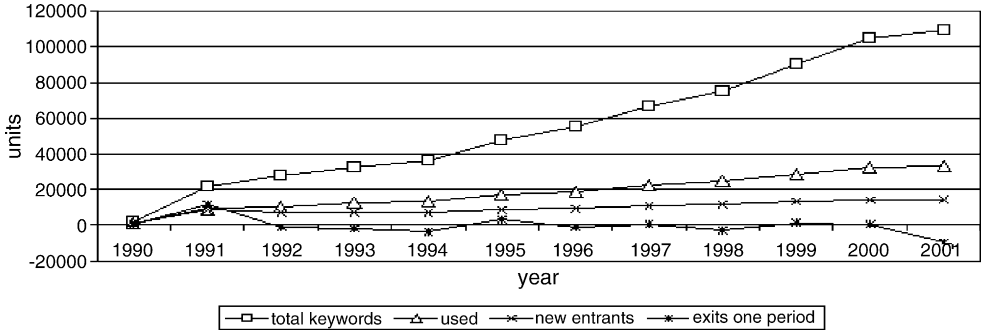


Fig. 3. Stocks, use, entry, and exit of keywords with respect to the overall population by year. *Source*: Our elaboration.

emerging field of NST, there are several well-identified sub-fields. The Fraunhofer-ISI methodology is based on a list of 40 subqueries, each of which is a combination of keywords identified by 40 experts in the field. We ranked all subfields in terms of number of publications and patents and identified the most important.

The identified subfields at the top of the distribution correspond very closely to large scientific areas identified in the literature. In particular, we labelled query number 33 in the Fraunhofer-ISI query list (including keywords such as ionic channels, molecular motors, nanospray) as nanotechnology, query 13 (e.g. scanning probe microscopy) as nano-instrumentation, query 37 (e.g. nanolayer) as nano fabrication, query 5 as nano electronics materials, query 11 (e.g. nanotube, nanowire, nanowhisk) as nanomaterials.

It is certainly possible that these subfields cover articles and patents retrieved under other queries proposed by the Fraunhofer-ISI and found in lower positions in the ranking. What is striking is the extremely high level of concentration of the distribution: for publications these five fields account for around 60% of the total, while for patents they account for slightly more than 50% of the total.

This evidence is consistent with published reports on the main areas of nano science and technology. Basically, applications of nano structures to life sciences, electronics, and new materials absorb the majority of published research, while publications on the development of nano-instrumentation cut across all subfields. The distribution of subfields in patents differs slightly due to the relative immaturity of patentable devices in the nano-biotechnology field.

An important dimension of diversity is dynamic: is diversity increasing or decreasing over time? Are research programmes moving away from others, or are they getting closer as they proceed? Or, put in other terms, can we observe a pattern of divergence or a pattern of convergence among research programmes?

Following the literature on field delineation in bibliometrics and scientometrics, we would say that research programmes might be characterized by keywords and combinations of keywords. A full appreciation of the dynamics of search would require observation of the way in which keywords originate and cluster together over a long period. While this will be the object of future research, we offer here a preliminary exploration of this issue by looking at the industrial dynamics of keywords.

If a discipline is subject to a divergent search regime, there will be many new keywords appearing per unit of time. As can be seen in Fig. 3, there is constant and linear growth in the number of entries of new keywords. We defined new keywords (new entrants) as those that appear for the first time in the dataset at any point in time. If a keyword has been used at least once in any year before an observation, it is not labelled as new. Consequently, each year, the total number of keywords used includes new keywords and old ones, drawn from the set of keywords that appeared for the first time in any previous period.

The use of new keywords as an indicator of emerging areas of research requires some qualification. In particular, it is sometimes observed that scientists relabel their research in order to draw increased attention from governments, funding agencies, and public opinion. In the field of NST, this may have been done by physicists and chemists in the USA after the significant boost in funding for biomedical research during the time of the Clinton Administration.[13] More generally, new words might be used to describe existing objects rather than new objects. However, it is hard to believe that the thousands of new words appearing year after year in the literature are simply the result of strategic relabelling or routine scientific description. After all, proposing new words to

[13] We would like to thank Paula Stephan and participants at the Grenoble Workshop on Nanodistricts (March 2006) for raising our attention to this issue.

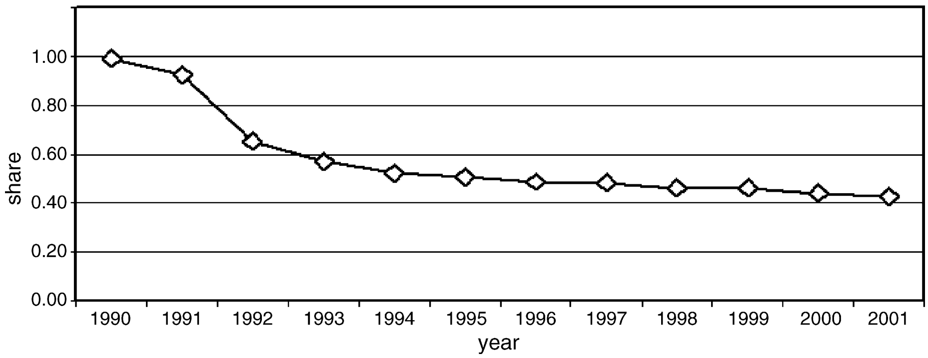


Fig. 4. Ratio between number of new keywords entered and total number of keywords used by year. *Source*: Our elaboration.

the scientific community requires justification if scientists are to preserve their credibility. Therefore, although the crude number of new words at aggregate level admittedly inflates the degree of diversity, the persistence of a massive entry of new words is a clear signal of a proliferation pattern. Future research is needed in order to disentangle this important issue.

The ratio between the number of new keywords and the total number of keywords used each year is an indicator of turnover. As is shown in Fig. 4, this ratio starts at very high levels at the beginning, given that, by construction all the keywords are new in the first year and the set of old keywords starts small. However, it is worth noting that the ratio stands at more than 40% even at the end of the period. Each year, the nano science community is able to generate more than 10,000 new words to describe their production, i.e. 40% of used keywords. We view this ratio as a remarkable indicator of turbulence, deriving from a divergent dynamics of proliferation of new research programmes.

As shown in Bonaccorsi (2005), similar levels of turnover can be found in computer science, where the ratio levelled out at 40% at the end of the observation period. However, in that case the publications of the top 1000 scientists worldwide were examined, and not those of the whole community, and one might argue than top scientists have a better than average ability to generate new research topics on a continuous basis. We can conclude, therefore, that nano science complies fully with the second requirement to qualify as leading science, i.e. divergent dynamics.

### *3.4. Level of complementarities: interface between science and technology*

In leading sciences institutional and human capital complementarities are crucial to the development of research. Institutional complementarities arise because the generation of discovery and invention requires researchers with different perspectives on a given object of research. Due to their different professional backgrounds, these researchers are usually affiliated to different institutional actors (e.g. public research, industry, hospital, public administration, regulatory bodies), bringing to the search process peculiar cognitive attitudes and operational practices. Human capital complementarities arise because the epistemic nature of discovery requires the deployment of several disciplinary competencies, even within the same team and/or the same institution.

Institutional and human capital complementarities are the fundamental mechanism for realizing effective science–technology interactions. In practice, however, there are several possible ways to implement these types of complementarities, for example, in terms of intensity of interaction between researchers, flows of communication, and pattern of mobility. Given the previously discussed relation between discovery and design, in NST the fundamental complementarity is between industry and academia.

To investigate the relation between nano science and technology, we clustered patents in three groups according to the scientific curricula of the inventors. The first group consists of patents whose inventors are all authors of at least one scientific publication in NST (only-authors). By contrast, the second comprises patents whose inventors have no scientific publication in that field (only-inventors). In the third, we isolated those patents that have at least one inventor, who is also author of at least one scientific publication in the field of NST.

This taxonomy is an initial contribution to the development of a new metrics of science and technology relations based on individuals and communities rather than on paper trails. Individual-based indicators can only

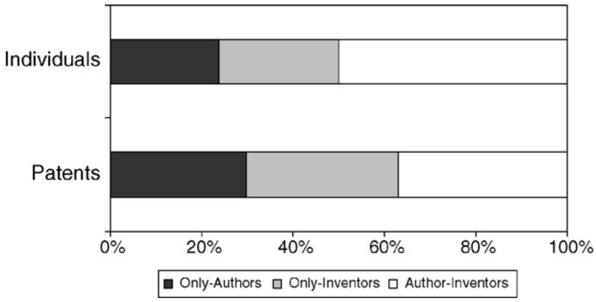


Fig. 5. Composition of communities of inventors in nanotechnology. *Source*: Our elaboration.

be used after appropriate matching procedures between different datasets have been carried out.[14]

In this paper, we present only a crude taxonomy based on the extreme discrete values and very sharp thresholds. Basically, we ask that all/no members of the group have a zero level of a selected variable. If all patent inventors have zero publications they are labelled "only-inventors"; if none have zero publications they are labelled "only-authors"; if none of the two extreme cases is true, they fall into a residual group called "authors–inventors".[15]

The results of the matching procedure of the names of the individuals are illustrated in Fig. 5.

Summing up the only-author and author–inventor groups, it emerges that a significant majority of patents (over 66%) have at least one inventor that is also an active scientist. Therefore, we find evidence of a highly interconnected knowledge system, in which the transformation of scientific achievements into patentable results and of both into commercial ventures is very rapid, and takes place through the multiple roles played by scientists themselves. Interestingly, the author–inventor group is growing more rapidly than the two other groups (see Fig. 6).

Surprisingly, this simple classification has strong validity and good predictive power. Although future research might develop more fine-grained taxonomies, this is a promising start. We interpret the three groups as approximations of different forms of institutional complementarities.[16]

Following our taxonomy, only-inventor patents originate from inventors that have never published in the field. In all likelihood, these inventors are industrial researchers for whom at least one of the following propositions holds true: (i) they have not pursued an academic career; (ii) they are not allowed to publish in the open literature; (iii) they work on applications that cannot be published in the scientific literature; or (iv) they work in institutions where publishing is encouraged but have no original results to submit, or publish in fields other than those of NST, or publish in non-ISI journals, or are technical staff (Fig. 7).

The literature on industrial research has noted that companies that allow their researchers to publish in the open literature do so because they want to gain access, visibility, and a reputation in the scientific community. Industrial researchers that do not have a track record of published articles have more difficulty in getting access to critical external knowledge and in building up absorptive capacity (Rosenberg, 1990). Similarly, companies that want to access external knowledge may hire people with a track record in public research (Gittelman and Kogut, 2003). Assuming that case sub (iv) is less important, we might conclude that this group identifies fairly precisely industrial R&D that has no or very weak institutional complementarity with public research.

On the contrary, in the other two groups all or at least some of the inventors have published at least once in the literature. The only-authors group is interesting, because it is made up of individuals for whom at least one of the following propositions holds true: (i) they all work in public research organizations and carry out both publishing and patenting; (ii) they work in companies, but all of them are encouraged to publish in the open literature; (iii) they work in companies, but all of them have a previous public research career behind them; (iv) a combination

[14] These types of matching cannot avoid the problem of synonymy or homonymy. In the past, the literature has dealt with the problem by collecting CV data and the matching is usually done using ad hoc and tedious procedures. (For a set of contributions, see Bozeman and Mangematin, 2004.)
Given the dimensions of our dataset, this methodology is not really feasible. In future works, we aim to develop automatic name-matching algorithms based on multiple indicator score methods.

[15] We executed a simple matching procedure as follows:

- Inventor versus Inventor: Drexler K Eric = Drexler K$ Eric;
- Author versus Author: Drexler K Eric = Drexler-KE;
- Author versus Inventor: Drexler-KE = Drexler K$ Eric.

Eric K. Drexler is regarded as one of the "founding fathers" of nano science and technology. His most famous book, *Engines of Creation: The Coming Era of Nanotechnology* (New York: Anchor Press/Doubleday), was published in 1986.

[16] It should be borne in mind that co-invention relations are stronger than, for example, co-authorships or co-occurrence of keywords or co-citations in documents. Patents are expensive, create exclusive rights, and may give rise to streams of revenues. Being recognized as an inventor is a right that has full legal implications. Therefore, if more individuals are recorded as co-inventors of a patent, their mutual relations are likely to be institutionalized.

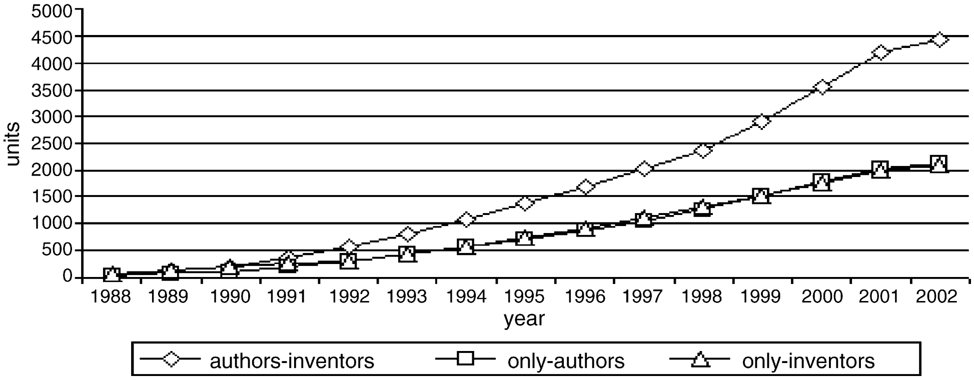


Fig. 6. Entry of individuals by community. *Source*: Our elaboration.

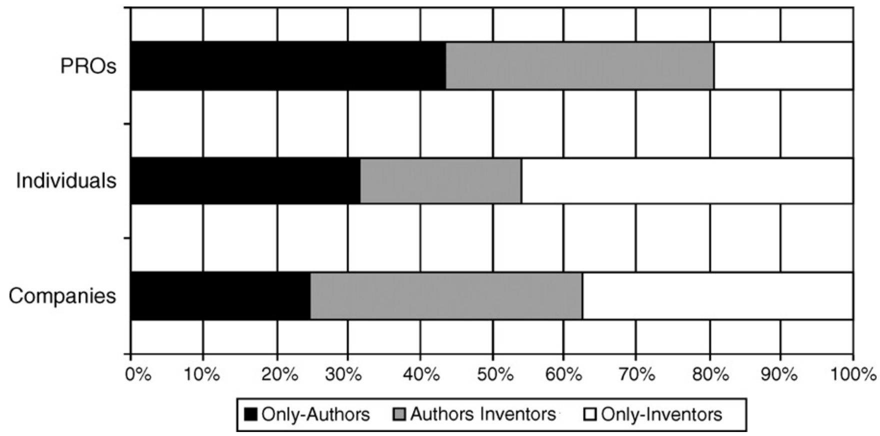


Fig. 7. Distribution of patents by community in relation to assignee type. *Source*: Our elaboration.

of proposition (ii) and (iii); (v) some of them work in companies and are permitted to publish/have published in their early career, some of them work in public institutions, and they collaborate and patent together.

Basically, these three communities represent different ways of organizing the complementarity between industry and academia. Only-inventors access scientific knowledge through codified channels such as publications and conferences. They do not involve professional scientists in their inventor community, nor do they have a track record of scientific publications in their past. Only-authors have full access to scientific knowledge, but have no structured and permanent access to knowledge of potential industrial applications, i.e. have difficulties in combining knowledge about physical structures (discovery) with knowledge about design. Since design-related knowledge is structurally more idiosyncratic and less codified than scientific knowledge, only-authors that come from academia implement the minimum level of complementarity. Finally, the community of authors–inventors gets access to both discovery and design knowledge, both in the codified and in the embodied form, on a permanent and organized way. They realize most of the complementarity. We anticipate that the more intense the complementarity, the more effective the invention.

To explore the different institutional search regimes, we classified the dataset of our patent assignees into three groups, according to whether they are private companies, public research organizations (hereafter also, PROs) or individuals.[17] We found that around 68% of the patents are assigned to private companies while PROs own 26% of them.

The results of Fig. 8 allow us to conclude that there is a higher probability that inventors-only are employed primarily in companies.

[17] PROs include higher education institutions (both public and private) and other non-university public research performing organizations such as government labs, large national research centres and public hospitals.

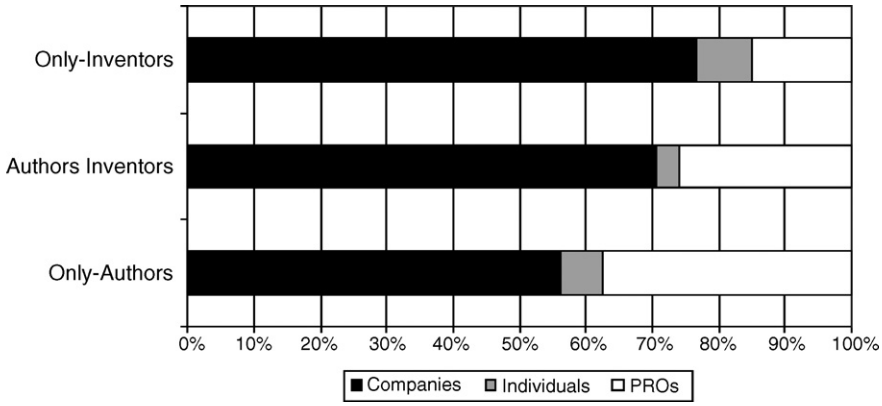


Fig. 8. Distribution of patents by assignee type in relation to community. *Source*: Our elaboration.

At least one active scientist is involved in the invention of around 80 and 60% of the patents assigned respectively to PROs and private companies.

## 4. The impact of complementarity on inventive performance

As we have seen from the matching exercise, there is considerable overlap between the different social roles of researchers in the production of public knowledge and of appropriable results. Two-thirds of patents have been produced by a team of inventors in which at least one inventor has published in the field.

We propose that the specific pattern of interaction between different types of inventors has an impact on the quality of patents, as measured in the relevant literature (see below).

Patents in the only-author group originate from two different patterns: researchers in the public sector that publish and patent, or industrial researchers in companies that maintain close relations with academia. In future research we will disentangle the two communities, looking at individual data on patents and publications and tracing careers across various affiliations. For the time being, we predict that this community will produce patents with a lower quality than the author–inventor community.

The latter will have the best performance, because here all groups of inventors benefit from at least one individual who has experience in publishing (currently or in the early career), while the others may well be pure industrial researchers. In this sense, the author–inventor community is characterized by the highest degree of institutional complementarity. In fact, in the only-author group it is required that all inventors are also active in publishing. Therefore, if the patent is obtained within a company, only industrial researchers that are all exposed to the scientific community will be involved, while if it is obtained in collaboration between industry and academia, public researchers collaborate only with those industrial researchers that also publish. In a word, in the only-author group there is low distance (cognitive, social, and institutional) between co-inventors. In practice, these patents do not involve industrial researchers or technicians who, while not personally involved in publishing, may have deep knowledge of the technology. Therefore, we expect that the author–inventor community, based on interaction patterns that materialize high levels of institutional complementarity, will exhibit better quality in the inventive activity.

The relation between the two groups of only-author and only-inventor is a bit more complex. On one hand, patents originated by groups of inventors in which no one has published (only-inventors) are characterized by access to scientific knowledge through codified channels, leading to a prediction of poor quality. On the other hand, the community of only-authors may be heavily influenced by academic inventors that do not have access to embodied design knowledge, also leading to a prediction of poor quality. Balancing these two effects is an empirical matter. We do not therefore advance any specific proposition.

The proposition regarding the superiority of the author–inventor community extends to the overall quality of patents, and to the upper tail of the distribution of inventive productivity. We rank inventors by number of patents produced and investigate whether the author–inventor community is more than proportionally present in the upper part of this distribution. This qualification is important, given the skewness of the distribution of inventive productivity.

Finally, we extend this proposition to one of the most important consequences of the quality of patents, namely the probability that an inventor becomes a founder of a new company.

We therefore put forward the following testable propositions:

**Proposition 1.**

(a) *The quality of patents of the inventor-only community will be lower than the quality of patents of the author–inventor community.*
(b) *The quality of patents of the author-only community will be lower than the quality of patents of the author–inventor community.*

**Proposition 2.** *The productivity of authors–inventors' inventive activity will be higher if counted by the number of patents produced in the top percentiles of the distribution of patents/inventors.*

**Proposition 3.** *Given the higher technological permanence of authors–inventors, we will observe more of them as founders of companies.*

## 5. Patent quality across different communities

In this section, we suggest some indicators that can be used in measuring the performance of inventive activity. Patent indicators have been widely used as a proxy of the intensity and quality of innovations in different fields of social sciences. It is not our intention to review that literature here, but a broad survey can be found in Jaffe and Trajtenberg (2002). We will follow a multiple indicators approach as suggested by Hall and Trajtenberg (2004), Henderson et al. (1998), and Lanjouw and Schankerman (2004).

One of the most widely used patent indicators are patent references or citations. In patents, citations serve an important legal function, since they delimit the scope of the property rights awarded by the patent. Thus, if patent B cites patent A, it implies that patent A represents a piece of previously existing knowledge upon which patent B builds, and over which B cannot have a claim. The applicant has a legal duty to disclose any knowledge of prior art, but the decisions regarding which patents to cite ultimately rests with the patent examiner, who is supposed to be an expert in the area and hence able to identify any relevant prior art the applicant may miss or conceal. The presumption is thus that citations are informative of links between patented innovations. Hence, citations made (or backward citations) may constitute a paper trail for spillovers, i.e. the fact that patent B cites patent A may indicate knowledge flowing from A to B.

Some scholars have suggested that a large number of citations to others implies that a particular innovation is likely to be more derivative in nature (Lanjouw and Schankerman, 2004). This is more evident when citations are within the same technological field. Hence, dispersion index measures that take into account the distribution across technological classes have been elaborated, for instance the originality index, which is 1 minus the Herfindal index of backward citation across technological classes (Henderson et al., 1998).

As citations are a paper trail for spillovers, received (or forward) citations may be indicative of the importance of the cited patent (Trajtenberg, 1990). Received citations over the long term indicate an innovation that has contributed to future research. Citations received soon after patent application suggests rapid recognition of its importance as well as the presence of others working in a similar area, and thus the expectation of a valuable technological area. Due to the problem of citation lag, we cannot trace forward citations over the long term in NST. We suggest a count indicator (FWcit5) of forward citations over the short term, in particular 5 years between the publication date of the cited and the application date of the citing. We expect FWcit5 to be positively correlated to the patent value.

The claims in the patent specification delineate the property rights protected by the patent. Principal claims define the essential novel features of the invention and subordinate claims describe detailed features of the innovation. The patentee has an incentive to claim as much as possible in the application but the patent examiner may require that the claims be narrowed before granting. The number of claims could be considered an indication that an innovation is broader and of greater potential profitability (Lanjouw and Schankerman, 2004).

Patent family size, measured as the number of jurisdictions in which a patent grant has been sought, should be directly related to the expected (private) value of protecting an innovation and thus to the value of the innovation, since applying for protection in each country is costly (Lanjouw et al., 1998). Hence, family size (*Family*) may be particularly well suited as an indicator of the economic value of patent rights.[18]

[18] We obtained the family size from Delphion-Derwent, a private vendor. Delphion extends the family data received from INPADOC to create the family unit and considers the set of patents filed with different patenting authorities that refer to the same invention.

Table 1
Descriptive statistics of patent indicators over the period 1989–1999

| | Sample | Min | Max | Median | Mean | Std | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| Overall | | | | | | | | |
| Originality | 2932 | 0.00 | 0.93 | 0.56 | 0.50 | 0.28 | −0.72 | 2.33 |
| FWcit5 | 2932 | 0.00 | 114.00 | 3.00 | 5.15 | 7.18 | 3.92 | 33.39 |
| Family | 2932 | 1.00 | 884.00 | 7.00 | 11.56 | 26.48 | 17.04 | 464.09 |
| Claims | 2932 | 1.00 | 236.00 | 15.00 | 18.80 | 16.11 | 3.79 | 32.64 |
| Authors–inventors | | | | | | | | |
| Originality | 1056 | 0.00 | 0.93 | 0.59 | 0.51 | 0.28 | −0.77 | 2.39 |
| FWcit5 | 1056 | 0.00 | 73.00 | 3.00 | 5.34 | 7.43 | 3.27 | 20.10 |
| Family | 1056 | 1.00 | 422.00 | 7.00 | 12.76 | 24.68 | 8.71 | 108.59 |
| Claims | 1056 | 1.00 | 199.00 | 17.00 | 20.20 | 17.21 | 3.86 | 31.61 |
| Only-authors | | | | | | | | |
| Originality | 919 | 0.00 | 0.90 | 0.53 | 0.48 | 0.29 | −0.61 | 2.08 |
| FWcit5 | 919 | 0.00 | 54.00 | 3.00 | 5.17 | 6.14 | 2.34 | 11.56 |
| Family | 919 | 1.00 | 468.00 | 4.00 | 7.64 | 18.02 | 19.01 | 469.15 |
| Claims | 919 | 1.00 | 155.00 | 14.00 | 17.97 | 14.70 | 2.76 | 16.90 |
| Only-inventors | | | | | | | | |
| Originality | 957 | 0.00 | 0.91 | 0.56 | 0.50 | 0.26 | −0.77 | 2.55 |
| FWcit5 | 957 | 0.00 | 114.00 | 2.00 | 4.92 | 7.80 | 5.17 | 51.39 |
| Family | 957 | 1.00 | 884.00 | 8.00 | 14.01 | 33.82 | 18.41 | 460.48 |
| Claims | 957 | 1.00 | 236.00 | 14.00 | 18.06 | 16.06 | 4.39 | 43.30 |

*Source*: Our elaboration.

### 5.1. Empirical evidence

In the following analysis, we decided to consider only the utility granted patents with an application date in the period 1988–1999. We opted to start from 1988 because the period of our dataset of scientific publications in NST covers 1988–2001. Secondly, our patent dataset is limited to 1999, since the FWcit5 cannot be computed for the following years.[19]

Table 1 reports the descriptive statistics of the suggested indicators. Some of them have a large spectrum of variation (max–min) and are strongly skewed, which fits with previous findings in the literature that most patents turn out to be of very little value, and only a handful have significant importance.

We compared the performance of the inventive activity across the three communities in three different ways.

Firstly, we compared the means of indicators for different groups, noting a number of statistically significant differences (Table 2). The originality index for only-authors is significantly lower than for authors–inventors and only-inventors. In these terms, the only-authors rely less on interdisciplinary previous art.

The higher originality of authors–inventors patents is followed by a larger technological importance and patent scope, with regard to which the other two groups do not differ significantly. Only-inventors patents have received greater protection, which could signal that they have a higher economic impact. Authors–inventors produce patents that are significantly more original, and have wider scope and expected value than only-authors. With respect to only-inventors, patents produced by authors–inventors are also more cited, in addition to having a wider scope. These results fully confirm Proposition 1(a and b), i.e. the quality of patents obtained by the author–inventor community dominates.

On the other hand, patents produced by only-authors are significantly less original and have less expected value than patents produced by only-inventors. We can interpret this finding in two ways. First, since patents produced by only-inventors are usually assigned to companies (see Fig. 8), it may be argued that they receive

[19] There might be some truncation bias in the sample. One possible source of truncation bias is the application lag for patents applied for before 1999 and not granted at the end of the period. It can be argued, however, that this should not be very large given that the average application lag is 2.28 years and we found that less that 1.5% of nano patents have an application lag greater than 5 years. Secondly, there may be a truncation bias relating to the citation lag when we include in the estimates patents applied during 1999, given that we normalized the citation to a 5-year lag window and built up the dataset in the middle of 2004. To overcome this limitation, we ran the regressions with the data up to 1998 and the results hold. In this paper, we decided to report the estimates up to 1999 because of the larger degrees of freedom.

Table 2
Comparison of the means: *t*-test statistics

| | Authors–inventors vs. only-authors | Authors–inventors vs. only-inventors | Only-authors vs. only-inventors |
|---|---|---|---|
| Originality | 2.78*** | 1.18 | −1.69*** |
| FWcit5 | 0.58 | 1.26* | 0.78 |
| Family | 5.31*** | −0.94 | −5.12*** |
| Claims | 3.10*** | 2.88*** | −0.13 |

*Source*: Our elaboration. *Note*: *Statistically significant at 10% level; **statistically significant at 5% level; ***statistically significant at 1% level.

stronger protection, due to potential strategic behaviour, and hence have a larger family size. Alternatively, it may be possible that only-authors are relatively more represented by academic inventors, who focus only on a limited technological area. If this is true, it appears that accessing complementary knowledge is easier when it is formed by codified scientific knowledge than when it refers to idiosyncratic design knowledge.

Secondly, given the strong skewness of the quality of innovation, comparison of the means has well-known limits. It might therefore be more interesting to look at the distribution of patents across different communities at the top percentiles of the distribution of the suggested patent indicators. Table 3 depicts forward citations and shows that on average authors–inventors have more patents in those percentiles than the other two communities.

Finally, we built a quality index with the mean of different patent indicators, as suggested in Appendix A. In particular, we built a common factor patent quality index that is assumed to be an unobserved characteristic of a patent positively influencing four quality indicators: backward citations, forward citations, number of claims, and family size. Estimation of the common quality index q is based on information extrapolated from the covariance matrix of our four observable indicators. By assuming the normality distribution of the common factor, we can estimate by maximum likelihood, which ensures a unique solution. The quality index is distributed normally with zero mean and variance $\sigma^2$.

Table 4 presents the factor loadings for the specified model. Both the restriction of no common factor and on more factors are rejected. It is worth noticing that the sign of the coefficients is positive, as we expected, excluding family size for only-inventors. We interpret this as over-protection of their patents with respect to their value.

Table 5 reports the distribution of the quality index for different communities given the above factor loading. The quality index gives a one-dimensional summary of four different indicators of patent quality, normalized around zero. Again the authors–inventors exhibit the highest performance, while only-authors display a lower quality of inventive activity with respect to the other two communities.

Similar results were obtained even when the factor index was conditioned for the geographic origin of the inventors and the patent application year. Table 6 shows the OLS regression of patent characteristic on the quality index, taking authors–inventors as the baseline. Model 1 includes only the categories of inventors, while the others add control variables and Model 3 also adds the interaction effects between categories of inventors

Table 3
Patents by different community at the top percentiles of the distribution of the patent indicator

| Percentiles | 75p (%) | 90p (%) | 95p (%) | 99p (%) |
|---|---|---|---|---|
| Authors–inventors | 43.88 | 37.41 | 44.90 | 42.86 |
| Only-authors | 30.61 | 31.97 | 12.59 | 28.57 |
| Only-inventors | 25.51 | 30.61 | 42.52 | 28.57 |

*Source*: Our elaboration.

Table 4
Maximum likelihood factor loadings of one latent variable model

| | Only-authors | Only-inventors | Authors–inventors | Overall |
|---|---|---|---|---|
| Originality | 0.39 | 0.21 | 0.33 | 0.29 |
| FWcit5 | 0.18 | 0.43 | 0.22 | 0.25 |
| Family | 0.20 | −0.08 | 0.34 | 0.15 |
| Claims | 0.24 | 0.26 | 0.31 | 0.31 |

*Source*: Our elaboration.

Table 5
Distribution of the quality index across communities

| Percentiles | Only-authors | Only-inventors | Authors–inventors | Overall |
|---|---|---|---|---|
| P25 | −0.39 | −0.32 | −0.28 | −0.32 |
| P50 | −0.02 | −0.01 | 0.05 | 0.01 |
| P75 | 0.30 | 0.31 | 0.40 | 0.34 |
| P90 | 0.52 | 0.56 | 0.69 | 0.60 |
| P99 | 0.97 | 1.03 | 1.16 | 1.06 |

*Source*: Our elaboration.

Table 6
OLS regression of the inventor type on the multidimensional quality index

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Coeff. | S.E. | Sign | Coeff. | S.E. | Sign | Coeff. | S.E. | Sign |
| Only-authors | −0.12 | 0.02 | *** | −0.06 | 0.02 | *** | −0.08 | 0.03 | *** |
| Only-inventors | −0.07 | 0.02 | *** | −0.07 | 0.02 | *** | −0.06 | 0.02 | *** |
| PROs × only-authors | | | | | | | −0.04 | 0.03 | |
| PROs × only-inventors | | | | | | | −0.07 | 0.04 | * |
| Control variables | | | | | | | | | |
| Number of inventors | | | | −0.01 | 0.05 | | −0.01 | 0.05 | |
| Time dummies | | | | × | | | × | | |
| Country dummies | | | | × | | | × | | |
| Constant | | | | −0.04 | 0.44 | | −0.04 | 0.44 | |
| Number of observations | 2802 | | | 2802 | | | 2802 | | |
| Adjusted $R$-squared (%) | 2 | | | 16 | | | 17 | | |

*Notes*: (1) ***1% level significance; **5% level significance; *10% level significance. (2) The likelihood-ratio test rejects the null hypothesis across all models. (3) Models 2 and 3 have been estimated by including time and country dummies.

and the institutional role, namely PROs. The negative and significant coefficient for only-authors and only-inventors in all models suggest that patents produced by authors–inventors have higher quality, as predicted by Proposition 1(a and b). This effect does not depend on country differences or time trends. Moreover the interaction effects between assignee type and PROs are negative, although not significant for only-authors. This is consistent with the above theoretical discussion. On one side the only-inventor group might lack adequate scientific knowledge, while the only-authors might not have structured and permanent access to potential industrial applications.

## 6. Productivity of publishing and patenting for different communities

In the previous section, using multiple patent indicators we found that the inventive activity of only-authors is of lower quality, while that of the authors–inventors is the highest. In this section we explore the relation between patenting and publishing. In particular, the aim was to test if authors–inventors have a higher participation rate in the top percentiles of the distribution of patents and publications across individuals. The distribution is obtained by ranking all authors or inventors by total number of articles or patents produced in the period. The findings show an interesting pattern. First, although authors–inventors do not out-perform only-authors in terms of publishing distribution in the top percentiles, the observed difference is very small (see Table 7). The top 1% of most productive scientists that have patented consists of 87 individuals, a figure that can be broken down almost equally into inventors that cooperate only with other scientists ($n = 46$) and inventors that cooperate also with people without a publication record. This means that even highly productive scientists benefit from strong complementarities with inventors with different backgrounds. The higher quality of patenting is strongly reflected in the top list of inventors by count (Table 8): here, authors–inventors represent 87% of the top 1% of most productive inventors and 77% of the top 5%. These results strongly confirm Proposition 2, i.e. authors–inventors have a larger number of patents in the top percentiles of patenting productivity.

## 7. Entrepreneurial productivity

The list of founders we used is provided by a questionnaire survey done by www.netinvestor.com. They interviewed around 1000 companies that have launched NST-based products. We extracted a list of 425 founders of such companies. It turned out that 67 of them hold at least one patent as defined in our database.

In Table 9, we classified those 67 by community membership according to the suggested taxonomy. As can be seen, 70% of founders belong to the author–inventor community; the result holds even if we weight for the size of the community.

This confirms Proposition 3, i.e. given the higher technological permanence of authors–inventors, we will observe more of them as founders of companies. The post-entry performance at these firms is an interesting research question, which will be addressed in future works.

Table 7
Distribution of inventors across different communities in the top percentiles of the publishing distribution

| Top percentiles | Only-authors | Only-inventors | Authors–inventors | Overall |
|---|---|---|---|---|
| Units | | | | |
| 1p | 46 | Nd | 41 | 87 |
| 5p | 209 | Nd | 226 | 435 |
| 10p | 437 | Nd | 433 | 870 |
| 25p | 1129 | Nd | 1146 | 2275 |
| **Top percentiles** | **Only-authors (%)** | **Only-inventors** | **Authors–inventors (%)** | **Individuals (%)** |
| Shares | | | | |
| 1p | 53 | 0 | 47 | 100 |
| 5p | 48 | 0 | 52 | 100 |
| 10p | 50 | 0 | 50 | 100 |
| 25p | 50 | 0 | 50 | 100 |

*Sources*: Our elaboration.

Table 8
Distribution of inventors across different communities in the top percentiles of the patenting distribution

| Top percentiles | Only-authors | Only-inventors | Authors–inventors | Overall |
|---|---|---|---|---|
| Units | | | | |
| 1p | 5 | 6 | 76 | 87 |
| 5p | 37 | 65 | 333 | 435 |
| 10p | 102 | 151 | 617 | 870 |
| 25p | 412 | 476 | 1387 | 2275 |
| **Top percentiles** | **Only-authors (%)** | **Only-inventors (%)** | **Authors–inventors (%)** | **Individuals (%)** |
| Shares | | | | |
| 1p | 6 | 7 | 87 | 100 |
| 5p | 9 | 15 | 77 | 100 |
| 10p | 12 | 17 | 71 | 100 |
| 25p | 18 | 21 | 61 | 100 |

*Sources*: Our elaboration.

Table 9
Distribution of founders by community

| Founders | Only-inventors | Only-authors | Authors–inventors | Overall |
|---|---|---|---|---|
| Units | 4 | 16 | 47 | 67 |
| Share % | 6 | 24 | 70 | 100 |
| Normalization ratio | 2.10 | 1.93 | 1.00 | |
| Normalized units | 8 | 31 | 47 | 86 |
| Normalized share % | 10 | 36 | 55 | 100 |

*Notes*: The normalization procedure takes into account the fact that the author–inventor community is larger than the other two communities. The normalization ratio adjusts for the size of the community.

## 8. Conclusions and suggestions for further research

The importance of scientific discoveries in fuelling technological inventions has been widely documented in many fields. This is particularly evident in NST where we found that the production of more than two-thirds of the nano patents involves an active scientist. To summarize, we have presented detailed and original evidence on an emergent field in which:

- the production of new knowledge is growing much faster than the average for science and engineering;
- although the application areas seem quite well defined, within each area there is evidence of a divergent pat-

tern of growth, following a process characterized by turbulent entry dynamics of new keywords;
- scientists have a tremendous impact on patenting activity in a variety of forms, and the whole field is characterized by high levels of institutional complementarity between industry and academia.

These features qualify NST as a new leading science, characterized by high growth, divergent dynamics, and new forms of complementarity. Some pointers seem to suggest that in NST these elements are present with greater intensity and speed than in other leading sciences examined in their birth period (for instance, biotechnology), although this proposition will need to be subjected to rigorous testing in future research.

In order to provide evidence on these aspects, we developed a battery of new indicators, namely indicators of entry and turnover of keywords, and individual-centred indicators based on the matching of publication and patent data.

Based on this descriptive and interpretative evidence, we developed some testable propositions that relate the institutional setting of NST research to performance, as measured by a factor model of patent quality. In spite of the evidence of a highly interconnected knowledge production system, the transformation of scientific discoveries into economic welfare is not immediate and direct. A simple taxonomy of inventors revealed evident differences in their technological and economic performance, according to some standard indicators. In particular, communities characterized by the highest levels of institutional complementarity (authors–inventors) perform better in both patenting and entrepreneurial activity, and achieve a remarkable performance in publishing as well.

Further research is needed for wider validation of the results and to combine the suggested framework with geographical and institutional contexts in which the inventors are embedded.

## Acknowledgements

This paper arose out of a PRIME Network of Excellence project on "Technological districts in nanotechnology", where we had intense discussions with Eric Avenel, Philippe Laredo, Vincent Mangematin, Christopher Palmberg, Arie Rip, Rikard Stankiewicz, and Michel Zitt. There have been other interesting discussions with Ian Cockburn, Maryanne Feldman, Mark Granovetter, Stine Grodal, Simcha Jong, Francesco Lissoni, Ed Steinmuller, and Paula Stephan, and other colleagues from the SiVNAP Project Group at the Department of Sociology of Stanford University, and from CESPRI at the Bocconi University. We gratefully acknowledge the contribution of Fabio Beltram, Director of the National Enterprise for Nano Science and Technology at the Scuola Normale Superiore, Pisa. The competent assistance of Donatella Caridi and Francesca Pierotti in data analysis is also kindly acknowledged. All errors are ours, though hopefully they are not too far from the nanoscale.

## Appendix A. A multidimensional measure of patent quality

The construction of the multidimensional measure of patent quality relies on factor analysis. In factor models, each series of data (quality indicator in our case) is broken down into a common component and an idiosyncratic component. The common component is driven by only a few common shocks, denoted by $Q<N$, where $N$ is the number of indicators. In a static factor model, the common shocks affect the indicators only contemporaneously. The basic model is given by $X=UB+E=K+E$, where $X$ is the $(T\times N)$ matrix of observations on $N$ series (indicators) of length $T$. The series are normalized to have mean 0 and variance 1. $U$ is the $(T\times Q)$ matrix of $Q$ common shocks and $B$ is the $(Q\times N)$ matrix of factor loadings, which determines the impact of common shock $q$ on series $n$. The common shocks and the factor loadings together make up the common component $K$. After the influence of common shocks has been removed, only the idiosyncratic component ($E$) remains. To estimate the common component, we have to find a linear combination of the indicators in $X$ that explains as much as possible the total variance of each indicator, minimizing the idiosyncratic component (for a technical discussion of factor models, see Jolliffe, 2002).

The parallel with least squares estimation is clear from this formulation, but the fact that the common shocks are unobserved complicates the problem. The standard way to extract the common component in the static case is to use principal component analysis. In principal component analysis, the first $Q$ eigenvalues and eigenvectors are calculated from the variance–covariance matrix of the dataset $X$. The common component is then defined as: $K=XVV'$, with $V=[p_1, \ldots, p_Q]$ and where $p_i$ is the eigenvector corresponding to the $i$th largest ($i=1, \ldots, Q$) eigenvalue of the covariance matrix of $X$. This method does not ensure a unique solution. A further problem is that *ex ante* it is not known how many common shocks $Q$ affect the series in $X$. Following the approach suggested by Lanjouw and

Schankerman (2004), we use a multiple-indicator model with an unobserved common factor:

$$y_{ki} = \lambda_k q_i + \beta' X + e_{ki}$$

where $y_{ki}$ indicates the value of the $k$th patent indicator for the $i$th patent, $q$ the common factor with factor loadings $\lambda_k$ and normally distributed, and $X$ is a set of controls. The main underlining assumption is that the variability of each patent indicator in the sample may be generated by the variability of a common factor across all the indicators and an idiosyncratic part $e_k$ not related to the other indicators and distributed $N(0, \sigma_k^2)$.

In our setting, the common factor is the unobserved characteristic of a patent that positively influences four quality indicators: backward citations, forward citations, number of claims, and family size. Estimation of common quality index $q$ is based on information extrapolated from the covariance matrix of our four indicators. By assuming the normality of $q_i$ and $e_k$, we can estimate by maximum likelihood, which ensures a unique solution. Once the estimates of $\lambda_k$ are obtained, the model is inverted to calculate $q$.

## References


Beltram, F., 2005. Interview with the authors on the discovery methods and instrumentation tools in Nano Science and Technology. Available upon request from the authors.

Bonaccorsi, A., 2005. Search regimes and the industrial dynamics of science. Paper presented at the PRIME General Assembly, Manchester: under review. Available at: http://www.prime-noe.org/index.php?project=prime&locale=en&level1=menu1_prime_1b8057d059a36720_1&level2=6&doc=Annual_Conference&page=3.

Bozeman, B., Mangematin, V., 2004. Editor's introduction: building and deploying scientific and technical human capital. Research Policy 33, 565–568.

Bozeman, B., Dietz, J.S., Gaughan, M., 2001. Scientific and technical human capital: an alternative model for research evaluation. International Journal of Technology Management 22, 716–740.

Breschi, S., Lissoni, F., 2004. Knowledge networks from patent data: methodological issues and research targets. In: Glänzel, W., Moed, H., Schmoch, U. (Eds.), Handbook of Quantitative S&T Research. Kluwer Academic Publishers.

Callon, M., Courtial, J.P., Laville, F., 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics 22, 155.

Cohen, W.M., Levin, R.C., Mowery, D.C., 1987. Firm size and R&D intensity: a re-examination. The Journal of Industrial Economics 35, 543–565.

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy 23, 487–521.

Fraunhofer-ISI, 2002. Search methodology for mapping nanotechnology patents. Karlsruhe, Germany: ISI. In: Mapping Excellence in Science and Technology across Europe Nanoscience and Nanotechnology, EU Report. Available at: www.cwts.nl/ec-coe.

Gambardella, A., Giuri, P., Mariani, M., 2005. The Value of European Patents. Evidence from a Survey of European Inventors. Final Report of the PATVAL EU Project.

Gittelman, M., Kogut, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. Management Science 49, 366–382.

Grupp, H., 1992. In: Grupp, H. (Ed.), Dynamics of Science Based Innovation. Springer-Verlag.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND Journal of Economics 36, 16–38.

Hall, B., Trajtenberg, M., 2004. Uncovering GPTs with Patent Data. In: NBER Working Paper No. 10901. Cambridge, MA.

Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Review of Economics and Statistics 81, 511–515.

Harhoff, D., Scherer, F.M., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Research Policy 32, 1343.

Heinze, S., Schmoch, U., 2004. Opening the black box. In: Glänzel, W., Moed, H., Schmoch, U. (Eds.), Handbook of Quantitative S&T Research. Kluwer Academic Publishers.

Henderson, R., Jaffe, A.B., Trajtenberg, M., 1998. Universities as a source of commercial technology: a detailed analysis of University Patenting, 1965–1988. Review of Economics and Statistics 80, 119–127.

Jaffe, A.B., Trajtenberg, M., 2002. Patents, Citations, and Innovations: A Window on the Knowledge Economy. The MIT Press, Cambridge, Massachusetts.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. The Quarterly Journal of Economics 108, 577–598.

Jolliffe, I.T., 2002. Principal Component Analysis, second ed. Springer-Verlag, New York.

Kenney, M., Goe, R.W., 2004. The role of social embeddedness in professorial entrepreneurship: a comparison of electrical engineering and computer science at UC Berkeley and Stanford. Research Policy 33, 691–707.

Klein, S., 1985. What is technology? Bulletin of Science, Technology and Society 1, 215–218.

Klepper, S., 2001. Employee startups in high-tech industries. Industrial and Corporate Change 10, 639–674.

Kline, S.J., Rosenberg, N., 1986. An overview of innovation. In: Landau, R., Rosenberg, N. (Eds.), The Positive Sum Strategy. National Academy Press, Washington, DC.

Lanjouw, J.O., Pakes, A., Putnam, J., 1998. How to count patents and value intellectual property. The uses of patent renewal and application data. Journal of Industrial Economics 46, 405–432.

Lanjouw, J.O., Schankerman, M., 2001. Characteristics of patent litigation: a window on competition. RAND Journal of Economics 32 (1).

Lanjouw, J.O., Schankerman, M., 2004. Patent quality and research productivity: measuring innovation with multiple indicators. Economic Journal 114, 441–465.

McKelvey, M., 1996. Evolutionary Innovations. The Business of Biotechnology. Oxford University Press, Oxford.

Mowery, D., Rosenberg, N., 1998. Paths of Innovation: Technological Change in 20th-Century America. Cambridge University Press, New York.

Narin, F., Olivastro, D., 1992. Status report: linkage between technology and science. Research Policy 21, 237–249.

Narin, F.S., Hamilton, K.S., Olivastro, D., 1997. The increasing linkage between U.S. technology and public science. Research Policy 26, 317–330.

Nelson, R.R., 1986. Institutions supporting technical advance in industry. The American Economic Review 76, 186–189.

Orsenigo, L., 1990. The Emergence of Biotechnology. Cambridge University Press, Cambridge and New York.

Owen-Smith, J., Riccaboni, M., Pammolli, F., Powell, W.W., 2002. A Comparison of U.S. and European University–Industry Relations in the Life Sciences. Institute for Operations Research, Management Science: INFORMS, p. 24.

Pavitt, K., 1990. What we know about the strategic management of technology. California Management Review 32, 17.

Popp, D., 2005. They Don't Invent Them Like They Used To: An Examination of Energy Patent Citations Over Time. In: NBER Working Paper No. 11415.

Rosenberg, N., 1982. Inside the Black Box. Cambridge University Press, Cambridge and New York.

Rosenberg, N., 1990. Why do firms do basic research (with their own money)? Research Policy 19, 165–174.

Shane, S., 2004. Academic Entrepreneurship: University Spinoffs and Wealth Creation, Edward Elgar.

Stankiewicz, R., 2000. On the concept of "Design Space". In: Ziman, J. (Ed.), Technological Innovation as an Evolutionary Process. Cambridge University Press, Cambridge and New York.

Stankiewicz, R., 1997. The development of beta-blockers at Astra-Hässle and the technological system of the Swedish Pharmaceutical Industry. In: Carlsson, B. (Ed.), Technological Systems and Economic Performance. Kluwer Academic Publishers.

Tijssen, R.J.W., 2004. Measuring and evaluating science and technology connections and interactions. In: Glänzel, W., Moed, H., Schmoch, U. (Eds.), Handbook of Quantitative S&T Research. Kluwer Academic Publishers.

Trajtenberg, M., 1990. A penny for your quotes: patent citations and the value of innovations. RAND Journal of Economics 21, 172–187.

Verspagen, B., 2005. Mapping Technological Trajectories as Patent Citation Networks. A Study on the History of Fuel Cell Research, Merit Research Paper, Maastricht, Merit.

Vincenti, W., 1990. What Engineers Know and How They Know It. Johns Hopkins Press, Baltimore.

Zitt, M., 2005. Delineating complex scientific fields by hybrid lexical-citation method: an application to nanoscience (under review). Available on request.

Zucker, L.G., Darby, M.R., 1996. Star scientists and institutional transformation: patterns of invention and innovation in the formation of the biotechnology industry. In: Proceedings of the National Academy of Sciences, pp. 709–712, 716.

Zucker, L.G., Darby, M.R., 2003. Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in Nanotechnology. In: NBER Working Paper No. 9825.