Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture?

Sam Arts · Francesco Paolo Appio · Bart Van Looy

Abstract Since Schumpeter's (The theory of economic development, 1934) seminal work on economic development, innovation is considered as one of the main drivers of firm performance and economic growth. At the same time, technological innovations vary considerably in terms of impact with only a minority of new inventions contributing significantly to technological progress and economic growth. More recently a number of indicators derived from patent documents have been advanced to capture the nature and impact of technological inventions. In this paper, we compare and validate these indicators within the field of biotechnology. An extensive analysis of the recent history of biotechnology allows us to identify the most important inventions ($n = 214$) that shaped the field of biotechnology in the time period 1976–2001. A considerable number of these inventions have been patented between 1976 and 2001 ($n = 117$, 55 %). For all USPTO biotech patents filed between 1976 and 2001 ($n = 84,119$), relevant indicators have been

S. Arts (✉) · B. Van Looy
Managerial Economics, Strategy and Innovation, Faculty of Business and Economics,
KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
e-mail: sam.arts@kuleuven.be

S. Arts
FWO, Brussels, Belgium

F. P. Appio
Scuola Superiore Sant'Anna, Istituto di Management, Pisa, Italy
e-mail: F.Appio@sssup.it

B. Van Looy
Expertisecentrum O&O Monitoring (ECOOM) and Research Division INCENTIM,
KU Leuven, Leuven, Belgium

B. Van Looy (✉)
School of Management and Governance, University of Twente, Enschede, The Netherlands
e-mail: bart.vanlooy@econ.kuleuven.be

calculated. In a subsequent step, we assess which indicators allow us to distinguish between the most important patented inventions and their less influential counterparts by means of logistic regression models. Our findings show that the use of multiple, complementary indicators provides the most comprehensive picture. In addition, it is clear that ex-post indicators reflecting impact and value outperform ex-ante indicators reflecting the nature and novelty of the invention in terms of precision and recall.

Introduction: some technological inventions are more influential than others

While the importance of technological innovation is widely acknowledged in terms of value creation on the level of firms and economies as a whole, the nature and impact of technological inventions vary widely. Some new technologies imply a relatively small extension of prior art (i.e. prior technological inventions), while others are highly novel and disrupt or reshape the technological landscape. A large number of new technologies never reach the commercialization phase whereas others allow companies to grow at impressive rates or even stimulate the creation of new industries. For instance, Scherer and Harhoff (2000) found, in the case of eight samples of company and university owned patents, that 10 % of the patents in the sample generated 48–93 % of the total returns. A variety of concepts have been advanced to delineate important technological inventions ranging from radical, revolutionary and breakthrough to discontinuous and disruptive. Technological breakthroughs or radical inventions introduce new concepts that depart significantly from past practices, have the potential to disrupt existing markets, generate new markets, and elicit follow-up innovations. Thus, they can be seen as critical building blocks of a company's or a nation's creative destruction capacity and as a key determinant of long-term economic growth.

In general, the definitions used in the *management literature* tend to characterize the differential nature of inventions both in technological and in economic/financial terms. Adopting a technological perspective, radical inventions rely on a different set of science and engineering principles than previously existing technologies (Henderson and Clark 1990), and/or incorporate substantially different core technologies (Chandy and Tellis 2000). Incremental inventions, in contrast, improve and extend existing technology. Henderson and Clark (1990) introduced the notion of architectural innovation in which core components remain unchanged but are linked differently in a new architecture. Radical innovations, according to their classification, are those where not only the concepts are linked together differently but the core concepts themselves are overturned. Along the economic and financial dimension, technological breakthroughs are listed as adding significant new value to the marketplace or through their impact on competitive dynamics. For example, Tushman and Anderson (1986) defined a technological breakthrough as an order-of-magnitude improvement in the maximum achievable price-versus-performance frontier of an industry. Finally, breakthroughs have been defined in terms of the profound impact they have on firms, industries and markets. Utterback (1994) defined radical

innovations or discontinuous change as "change that sweeps away much of a firm's existing investments in technical skills and knowledge, designs, production technique, plant and equipment," and Henderson (1993) described an innovation as being radical when it renders a firm's information filters and organizational procedures (partially) obsolete. In addition, a number of concepts closely related to radical innovations are popular in the management literature. Tushman and Anderson (1986) classified technological breakthroughs as either competence-enhancing or competence-destroying, depending on whether they either reinforce or destroy established firms' existing competencies, skills, and knowledge. Technological breakthroughs are also described as inventions that serve as the basis for many subsequent technological developments (Fleming 2001; Ahuja and Lampert 2001) and, as such, shape the development of fields and related industries. Christensen (2003) focuses on disruptive technologies and their implications for established firms in an industry. A disruptive technology will have features that initially only a fringe market segment will value. It redefines the performance trajectory (e.g. in the case of the disk drive industry, shrinking the size of disks). These disruptive technologies need not be radical in nature: in fact, Christensen notes that, in general, disruptive innovations are technologically straightforward.

In the *evolutionary economics* tradition, radical innovation is commonly evoked in typologies that attempt to characterize the degree of innovativeness of a product or a process (Dosi 1982). Freeman (1992) proposed a taxonomy for technological innovation involving four levels of change: incremental innovation, radical innovation, changes of technical systems, and changes of techno-economic paradigms. Radical innovations, according to Freeman, are discontinuous as they introduce far-reaching changes in technology and affect different parts of the economy, ultimately leading to entirely new sectors.

In the last two decades, we have not only witnessed the introduction of a variety of definitions, but a number of patent-based indicators have been advanced to assess the nature and value of patented technological inventions. Patents contain detailed information on the nature of the technology and leave a trail of patent citations, backward citations (i.e. the citations made to prior patents) and forward citations (i.e. the citations received from future patents). This information allows us to trace elements of the origin of technologies as well as their influence on future generations of technologies (when they directly or indirectly serve as prior art). In addition, patent citations provide indications of the economic value of patents (Griliches 1984; Jaffe and Trajtenberg 2002). Our contribution builds on the most notable patent-based indicators used in the literature to assess the nature and value of patents, and aims to assess which indicators allow us to identify the most important inventions that shape the development of a technological field. In order to do so, we identified major contributions within the field of biotechnology (time period: 1976–2001). In a subsequent step, the different indicators used in the literature are calculated for all granted USPTO biotech patents (time period 1976–2001). Finally, we rely on logistic regression models to assess which indicators are able to identify the most influential patented technologies. Our findings reveal that combining available indicators results in recall rates exceeding 68 % while precision amounts to 84 %. Ex-post indicators measuring technological impact and economic value clearly outperform ex-ante indicators reflecting the nature and novelty of an invention.

The remainder of the paper is outlined as follows. First, we introduce the data and indicators used in this analysis. Next, we discuss the descriptive statistics and results from multivariate analysis. We conclude with a discussion of the implications as well as the directions for further research.

Patent indicators that assess the nature and value of technological inventions: an overview

Different indicators relying on patent data have been used in the literature to assess the nature and impact of patented inventions.

Patent-indicators to assess the nature and novelty of technological inventions

To assess the *nature* of an invention, patents can be compared in terms of backward citations, technology classes or both. Patents without backward citations to technical prior art have been labeled 'pioneering' (Ahuja and Lampert 2001) while dissimilar patents have been defined as having backward citations that are different compared to prior patents in the same field (Dahlin and Behrens 2005). The originality of a patent can be identified through a patent's backward citations with original patents relying on prior art from a broad range of technology fields (Trajtenberg et al. 1997). Finally, more creative inventions have been identified as displaying novel pairwise combinations of technology subclasses or components at the patent level (Fleming et al. 2007).

Dissimilar and unique backward citations

A first method of identifying technologically radical inventions was developed by Dahlin and Behrens (2005) using backward patent citations to other patents. By calculating the overlap scores between the backward citations of each patent P granted in year t with all other granted patents[1] in the same field, and averaging these overlap scores within each year relative to the grant year t, one can identify which patents have a dissimilar citation structure with respect to prior art and a unique citation structure with respect to patents granted in year t. Those patents that have low overlapping scores compared to prior art in the field are considered more inventive and unusual. Note that patents without backward citations have the lowest possible overlap score and, as such, are considered more revolutionary or pioneering (Ahuja and Lampert 2001).

New pairwise combination of technology subclasses

New technological inventions originate from recombining and extending pre-existing technological inventions (Nelson and Winter 1982; Basalla 1988). Fleming (2001) conceptualizes technological invention as a recombinant search process across the technology landscape in which inventors experiment with the recombination of technological components. He argues that a patent's technology subclasses capture the different components used to develop the technology. Using patent data, Fleming (2001) empirically shows that breakthroughs, i.e. patents with the highest variability in forward citations (i.e. the citations received by a patent from future patents), most likely originate from the recombination of familiar technological subclasses, i.e. from subclasses with relatively more prior patents. Nevertheless, the findings show that patents re-using the same combination of subclasses as prior patents are less likely to be breakthroughs. Thus, breakthroughs most likely materialize from recombining disconnected but pre-existing technology subclasses. To

---

[1] To calculate the overlap score between patents A and B, the number of overlapping backward citations of A and B is divided by the total number of backward citations of A and/or B. Only citations to patents granted in a year before the minimum grant year of A and B are taken into account.

identify particularly original contributions with a potentially high impact on future technology development, Fleming et al. (2007) sought to look at patents that were the first in history to recombine at least two previously disconnected technology subclasses.[2]

Originality

Trajtenberg et al. (1997) developed a backward-looking measure of the 'basicness' of an invention. Originality captures the extent to which the nature of the research underlying the patent is based on technical prior art from a broad range of technology fields.

Patent indicators to assess the impact and value of technological inventions

To assess the *impact* and *value* of the patent, scholars have primarily used forward citations, backward citations and technology classes. The number of citations a patent receives reflects its direct impact on future technological inventions as well as its private and social value (Gambardella et al. 2008). A similar backward citation structure between a patent and later patents reflects adoption by future generations of technological inventions (Dahlin and Behrens 2005). Patents that are cited by patents from different technology fields are considered to have a more general purpose or impact (Trajtenberg et al. 1997). Finally, technologies with a novel combination of technology subclasses are adopted by future generations of inventions in case the same combination of subclasses is frequently used by future patents (Fleming et al. 2007).

Indicators relying on the count of forward citations

The most popular indicator of patent impact or value is the number of forward citations received from future patents. The number of forward citations that a patent receives is related to its technological importance (Albert et al. 1991; Carpenter and Narin 1993; Jaffe et al. 2000) as well as its social (Trajtenberg 1990) and private value (Harhoff et al. 1999; Hall et al. 2005; Gambardella et al. 2008). The distribution of forward citations is very skewed, with a large share of patents receiving no citations and a small minority of patents obtaining a large number of forward citations. This pattern resembles the distribution of the actual value of inventions. Hence, it is likely that outliers in the distribution of forward citations pertain to more important inventions. Prior research has typically identified breakthrough patents as the top 1 or 5 % in terms of citations received compared to patents with the same application year and technology class (Ahuja and Lampert 2001; Singh and Fleming 2010).

Adoption of backward citations

Besides having a dissimilar and unique backward citation structure, technologically radical patents should also have a backward citation structure that is adopted by future patents in the same field (Dahlin and Behrens 2005). The more similar the backward citations of a patent and future patents in the field, the more influence the patent has on future technological progress.

---

[2] A patent with three technology subclasses A, B and C has three pairwise subclass combinations AB, AC and BC. A novel pairwise subclass combination is identified as the first patent in history with the particular pairwise combination of technology subclasses.

Adoption of a novel pairwise combination of technology subclasses

To assess the diffusion or adoption of a patented invention that recombines two discon-
nected technology subclasses, Fleming et al. (2007) look at the number of future patents
that use the same pairwise combination of technology subclasses. The larger the number of
future patents re-using the same combination of subclasses, the greater impact the patent
has on future technological progress.

Generality

Trajtenberg et al. (1997) develop a measure of generality, capturing the extent to which the
patented invention serves as prior art for a broad range of technology fields. So, while
originality measures the broadness of the prior art of the invention (based on backward
citations), generality captures the extent to which an invention directly serves as prior art
for different technological fields (based on forward citations, i.e. citations received).

Biotechnology

Definition and short history

According to Bud (1993), the term biotechnology was coined as long ago as 1917, the year
of the Russian revolution. Today, the best known definition is perhaps the one spelled out
by the Organization for Economic Co-operation and Development (OECD 2005): "Bio-
technology is the application of scientific and engineering principles to the processing of
materials by biological agents to provide goods and services." Biotechnology is a field that
emerged from agriculture and animal husbandry in ancient times through the empirical use
of plants and animals that could be used as food or dyes (McGloughlin and Re 2010).
Moreover, contrary to its name, biotechnology is not a single technology. Rather, it is a
group of technologies that share two characteristics: working with living cells and their
molecules, and having a wide range of practical uses that can improve our lives (Keener
et al. 2012).
    According to Buchholz and Collins (2010), four periods can be discerned in the history
of biotechnology (before 1850, 1850–1890, 1890–1950, and the period from 1950
onwards). This paper focuses on the later period, more particularly, the period from 1976 to
2001. By the 1950s, large-scale production of, for example, beer, cheese, citric acid,
pharmaceuticals and other products of social and economic relevance such as antibiotics
had become well established. During that time, biotechnology benefited from major public
funding and made an increasing economic impact. Major technological progress was
achieved during the late 1970s and 1980s, most notably due to genetic research and
recombinant technologies. A milestone was the model of DNA providing the molecular
basis of heredity derived by Watson and Crick with the aid of data provided by Rosalind
Franklin who worked in Maurice Wilkin's X-ray crystallography laboratory in 1953
(Watson and Crick 1953). However, the DNA revolution, as Hotchkiss (1979) termed it,
penetrated slowly into technology, initially having little effect on traditional processes and
products. A significant change was triggered by the introduction of recombinant DNA
(Cohen et al. 1972; Cohen and Boyer 1979/1980). The emergence of molecular biology
and biochemical engineering coincided with a growing industrial interest and the range of
products expanded significantly. The field's progress is reflected in the exponential rise in

the number of journals devoted to biotech established in the late 1970s and early 1980s (Buchholz and Collins 2010).

The integration of applied microbiology, biochemical engineering and molecular biology led to the creation of biotechnology as a scientific discipline in its own right, with a common paradigm at the level of molecular research. Sub-disciplines such as genomics, transcriptomics, proteomics, metabolic flux analysis with quantitative analysis of complex metabolic pathways and, finally, biochemical engineering and bioinformatics have merged to create bio-systems engineering (Sinskey 1999; Stephanopoulos 1999; Reuss 2001).

Identification of major contributions that shape the field of biotechnology

In order to identify the most important technological developments that have shaped the evolution of biotechnology, we relied on secondary sources including books, journal articles, websites of inventors, academics, companies and research institutes, and expert reports. Amongst those sources, we principally relied on scientific books providing a consistent and exhaustive overview of major technological accomplishments in biotechnology or a particular subfield of biotechnology. An exhaustive overview of the major sources used in this respect can be found on-line (see Supplementary Material). We verified multiple secondary sources to strengthen the overall consistency of our list of important inventions, since any account may well be conditioned by the personal interests or values of the authors. We concentrated initially on events labeled as discontinuous, pioneering, important, breakthrough, revolutionary, radical, drastic, cutting edge, fundamental, groundbreaking, dramatic, leap forward and original, among others. In particular, we searched for those inventions that were described as contributing to the evolution of biotechnology, highlighting fundamental leaps on certain key research trajectories or establishing new ones that were clearly stated by authors.

Relating important contributions to patents

After carrying out a comprehensive screening and assessment of technological inventions that shaped the field of biotechnology, we systematically searched for patents and publications associated with those inventions. Using information on the description and timing of the invention, the associated researchers, institutions and/or companies, we searched for corresponding patents and publications in the USPTO patent database and the ISI Web of Science (WOS), respectively. In some cases, we found more than one corresponding patent and/or more than one corresponding publication, whereas we were unable to find publications or patents for a minority of important inventions. Of the 214 important inventions identified, 117 (55 %) were found in the USPTO patent database while 153 (71 %) were found in the WOS database as scientific publications. For 37 (17 %) of the events, we found at least one corresponding patent but no publication, for 72 (34 %), we found at least one publication but no patent while for 80 (37 %), we found patent-paper pairs. For eight events, we found multiple corresponding patents. For 25 (12 %) contributions, neither a patent document nor a scientific publication (present in the Web of Science) has been identified. A detailed list of all major inventions considered in this analysis as well as the related USPTO patents can be found on-line (see Supplementary Material). Notice that a number of inventions resulted in multiple USPTO patent documents. All patents that fall within the relevant time window (filed between 1976 and 2001 and granted before 2004) and that are situated within the biotechnology domain (see infra) have been included in the subsequent analysis ($n = 122$).

Data and findings

Sample selection

To identify all USPTO biotechnology patents, we made use of the OECD classification scheme that relies on IPC codes (OECD 2005). Data have been extracted from the Patstat patent database (version October 2011) and include all patents filed at the USPTO between 1976 and 2001 and granted before 2004, which fall into at least one of the IPC classes. The final sample used for analysis consists of 84,119 patents. From the 84,119 patents, 122 have been identified as relating to 117 important inventions that shaped the field of biotechnology. For the calculation of citation-related indicators, we employed the updated NBER patent database.[3]

Variables

Dissimilar, unique and adopted backward citations

We follow the methodology of Dahlin and Behrens (2005) and calculate, for each patent P granted in year t, the average annual overlap scores between the backward citations of P with, respectively, all other patents filed in the same field (3-digit US technology class) within a time window of 5 years before and 5 years after the grant year of P (i.e. patents granted between $t - n$ and $t + n$ with $0 \leq n \leq 5$). We extend their methodology by comparing a patent to all other US granted patents with, at least, one similar technology class. So, for each of a patent's 3-digit technology classes, we follow the methodology outlined in Dahlin and Behrens (2005). First, we label P as dissimilar compared to prior art when the average annual overlap score is 0 or the average standardized annual overlap score is smaller than or equal to the 10th percentile of all patents for each year $t–n$ with $n > 0$ and $n \leq 5$. Due to truncation, not all patents have a time window of 5 years before (and after) grant. For patents that we can only observe three or 4 years before grant, we require the patent's average overlap score to be 0 or its standardized average annual overlap score to be equal to or smaller than the 10th percentile threshold for each of the observed years before grant. Patents that cannot be observed at least 3 years before and after the grant are not taken into account during the analysis. Following this methodology, 47 % of the patents in our sample have a dissimilar citation structure. Second, a patent is labeled 'unique' when the average standardized overlap score in the year of grant is below or equal to the 10th percentile threshold. We find 67 % of the patents to pass the uniqueness criteria. Third, a patent is labeled 'adopted' when the annual overlap score passes the 90th percentile threshold for each year after grant. 8 % of the patents in our sample pass the adoption criteria. Finally, 2.2 % of the patents pass all three criteria.

New and adopted pairwise combination of technology subclasses

To identify patents that recombine two technology subclasses for the first time in history, we use the 2008 US technology subclass concordance to investigate all technology subclass assignments of all US-granted patents in order to identify all first pairwise subclass combinations. For each new subclass combination, we count the number of future patents re-using the same pairwise combination. In our sample of biotech patents, we find 45 % of

---

all patents displaying a new combination of technology subclasses with, on average, 49 future patents re-using the same combination.

Originality and generality

In line with Trajtenberg et al. (1997) and Hall et al. (2001), we calculate originality and generality using a measure reflecting the concentration of backward and forward citations, respectively, within technology classes. Originality is calculated as $1 -$ bias-corrected[4] Herfindahl index of the technological classes (main) of all cited patents. Generality is calculated as $1 -$ bias-corrected Herfindahl index of technological classes (main) of all citing patents. The average originality score of the biotech patents in our sample is 0.52 while the average generality score is 0.51.

Indicators relying on the distribution of forward citation counts

To identify patents with the largest impact on future technologies, we calculate, for all granted US patents, the count of forward citations as the number of US patents citing the patent (citations from patents granted until 2006 inclusive) and the truncated count of forward citations as the number of citations received within 5 years of application. Prior research has typically identified breakthrough patents as the top 1 or 5 % in terms of citations received compared to patents within the same application year and technology class (e.g. Ahuja and Lampert 2001; Singh and Fleming 2010). This definition assumes that each technology field has a fixed share of high impact inventions each year and does not compare patents across years. To avoid a definition that forces a fixed proportion of breakthroughs every year into each class while allowing similar patents to be compared across years, we follow the methodology of Arts (2012) and consider the distribution of both forward citations received within five years and the distribution of forward citations received from all future patents. We use the full count of forward citations to compare all patents sharing at least one 3-digit US technology class filed in the same year and the truncated citation count to compare all patents sharing at least one technology class irrespective of time of filing. For each of the distributions, we calculate the mean and standard deviation of (truncated) forward citation counts. A patent is labeled as having a high impact on future generations of inventions when both its truncated and full count of forward citations are larger than the mean plus n times the standard deviation in at least one of its technology classes. So, for each patent's technology class, the patent is compared with two distributions: the distributions of full and truncated forward citation counts. Using a 1, 2, 5 standard deviation rule to identify outliers in the distribution of forward citations, we find that, respectively, 7, 3, 0.5 and 0.1 % of biotechnology patents in our sample are labeled as having a disproportionate impact on future patents.

For all USPTO patent documents under study we calculate the abovementioned indicators. An illustrative example for the inventions related to recombinant DNA and polymerase chain reaction can be found on-line (see Supplementary Material). As a close inspection of both patent documents and the implied indicators reveal, not all indicators signal discriminatory values for these important contributions. Whether this is also the case for the sample as a whole, a more systematic, multivariate, analysis will reveal (see Sect. "Multivariate analyis").

---

[4] A bias correction is necessary because not all patents have the same number of backward and forward citations.

Descriptive statistics

Table 1 gives an overview of descriptive statistics for the set of major technological inventions and for the control patents including a mean-comparison T test between the two groups.

In terms of indicators reflecting the origins and nature of the patented invention, we do not observe significant differences in the proportion of patents without citations to technical prior art. By contrast, most important inventions seem to more frequently cite other patents compared to the control group (ten backward patent citations compared to six on average). Furthermore, major inventions have more dissimilar backward citations (55 % of dissimilar patents compared to 47 % for the control group) and rely on more recent technical prior art with an average backward citation lag of 6.1 versus 7.5 years for the control group. Both groups are not different in terms of originality indicating that important biotech patents do not rely on prior art stemming from a broader range of technology fields. Nevertheless, patent documents associated with important technological inventions contain a larger number of technology main classes and subclasses, so they seem to cover a larger part of the technology landscape (2.5 main classes and 8.0 subclasses on average compared to 2.2 and 6.2 for the control group, respectively). Finally, important inventions display a much larger number of citations to non-patent literature (44 vs. 22 on average), are more likely to have a novel pairwise combination of technology subclasses (66 vs. 45 % on average) and contain a larger number of claims (23 vs. 15 on average). In conclusion, patents associated with major technological inventions cite more patents, cite more recent technical prior art, contain more references to non-patent literature, and have dissimilar backward citations compared to prior art in the same field but do not rely on prior art from a broader range of technology fields. Nonetheless, major patents seem to be more novel and serve a more general purpose by covering more technology fields, subfields and claims, and are more likely to combine previously disconnected technology subfields.

Besides being based on a different set of science and engineering principles and/or incorporating substantially different core technologies, the most important technological contributions are expected to have a higher and broader impact on future technology trajectories. In line with expectations, patents associated with important contributions receive significantly more forward citations on average (146 forward citations and 36 forward citations within 5 years compared to 7 forward citations and 3 forward citations within 5 years for the control group). Accordingly, looking at outliers in the distributions of forward citations we observe that 74 % of the important patents are 1 SD outliers compared to 7 % for non-radical patents, and 29 % are 10 SD outliers compared to only 0.1 % for the control group. Besides serving more extensively as prior art for future generations of inventions, they also tend to remain cited for a longer time with an average forward citation lag of 7.6 years compared to 5.9 for the control group. Furthermore, patents associated with important inventions seem to serve as prior art for a broader range of technology fields, reflected by an average generality score of 0.74 (compared to 0.51 for the control group). Likewise, for patents that make at least one new pairwise combination of previously disconnected technology subfields, the pairwise combination of important contributions is adopted by a much larger number of future patents. On average, 1,526 future patents will use the same component configuration compared to 46 future patents for the control group. Also, the backward citations of major patents are more likely to be adopted by future patents with 36 % of important inventions having adopted backward citations compared to only 8 % for the control group.

Table 1 Descriptive statistics

	Important contribution	Final sample 122 important contributions (0.15 %)			
		n	Mean	SD	Pr(|*T*| > |*t*|)
Forward citation count	0	83997	6.932	14.146	0.0000
	1	122	146.082	218.402	
Forward citation count 5 years	0	83997	2.813	5.461	0.0000
	1	122	36.230	45.135	
Outlier 1SD class and class year	0	83997	0.068	0.252	0.0000
	1	122	0.738	0.442	
Outlier 2SD class and class year	0	83997	0.028	0.166	0.0000
	1	122	0.648	0.480	
Outlier 5SD class and class year	0	83997	0.005	0.069	0.0000
	1	122	0.385	0.489	
Outlier 10SD class and class year	0	83997	0.001	0.025	0.0000
	1	122	0.287	0.454	
No forward citations	0	83997	0.258	0.438	0.0000
	1	122	0.008	0.091	
Forward citation lag	0	62299	5.874	3.542	0.0000
	1	121	7.575	3.666	
Generality	0	83997	0.510	0.276	0.0000
	1	122	0.735	0.066	
Count claims	0	83970	15.421	15.182	0.0000
	1	122	23.336	20.511	
Count main technology classes	0	83997	2.195	1.042	0.0030
	1	122	2.475	1.201	
Count technology subclasses	0	83997	6.248	4.418	0.0000
	1	122	8.016	7.337	
First subclass combi dummy	0	83997	0.452	0.498	0.0000
	1	122	0.656	0.477	
First subclass combi count re-use	0	37976	45.677	248.231	0.0000
	1	80	1525.700	4754.776	
Count backward citations	0	83997	5.879	12.859	0.0006
	1	122	9.861	15.280	
No backward citations	0	83997	0.193	0.395	0.5610
	1	122	0.172	0.379	
Backward citation lag	0	67581	7.506	4.047	0.0003
	1	101	6.063	2.837	
Originality	0	83997	0.516	0.269	0.8924
	1	122	0.513	0.316	
Count non-patent references	0	83997	22.439	39.464	0.0000
	1	122	43.836	47.297	
Dahlin and Behrens dissimilarity (before grant)	0	82969	0.465	0.499	0.0625
	1	122	0.549	0.500	

Table 1 continued

	Important contribution	Final sample 122 important contributions (0.15 %)			
		n	Mean	SD	Pr(|T| > |t|)
Dahlin and Behrens uniqueness (year grant)	0	83997	0.668	0.471	0.0699
	1	122	0.590	0.494	
Dahlin and Behrens adoption (after grant)	0	83997	0.076	0.265	0.0000
	1	122	0.361	0.482	
Dahlin and Behrens composite	0	82969	0.022	0.145	0.0000
	1	122	0.139	0.348	

Finally, Dahlin and Behrens (2005) suggest the use of a composite measure to identify technologically radical inventions. Besides having a backward citation structure that is dissimilar to prior art and becomes adopted by future patents, they add an additional uniqueness criteria, i.e. having backward citations different from patents granted in the same year in the same technology field. We find important contributions have a backward citation structure that is less unique with 59 % of the important patents satisfying the uniqueness criteria compared to 67 % for the control patents. According to the authors, technologically radical inventions should satisfy all three criteria. We find that 14 % of the most important inventions satisfy all three criteria compared to 2 % for the control group.

Table 2 presents the correlation coefficients between the most notable indicators. The dummies representing outliers in the distribution of forward citations correspond most with being an important contribution to the field, particularly 5 and 10 SD outliers. Furthermore, the number of future patents re-using the same pairwise combination of subclasses displays a strong correlation.

In conclusion, the descriptive results suggest that both backward-looking measures reflecting novelty with respect to prior art, as well as forward-looking measures of value and impact, signal important inventions within a field. In particular, measures reflecting an impact on future technological progress seem to reveal discriminatory power.

Multivariate analysis

Given that our dependent variable, indicating whether the patent was identified as being a major contribution to the field, is binary (0/1), we use logit models to assess the discriminatory power of the different indicators. All models include technology dummies for each of a patent's main technology classes (3-digit) as well as a set of additional control variables including the number of assignees, the number of inventors, and patent age. Note that patents in a main technology class without important contributions are dropped from the analysis during estimation. To assess the discriminatory power of the different indicators, we provide a number of statistics below each regression model in Table 3. We are particularly interested in the recall, i.e. the percentage of externally identified major patented inventions that are predicted as such, as well as in the precision, i.e. the percentage of patents that are predicted to be very important and that actually prove to be so.

Table 3 presents the results obtained for the full set of patents. Note that important contributions only represent 0.15 % of the total sample, which seriously hampers the assessment of precision and recall rates of the different models (any model that would

Table 2 Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Important contribution	1												
(2) Outlier 1SD class and class year	0.1005*	1											
(3) Outlier 2SD class and class year	0.1396*	0.6384*	1										
(4) Outlier 5SD class and class year	0.1994*	0.2680*	0.4198*	1									
(5) Outlier 10SD class and class year	0.3409*	0.1174*	0.1840*	0.4382*	1								
(6) First subclass combi dummy	0.0156*	0.0868*	0.0668*	0.0343*	0.0210*	1							
(7) First subclass combi count re-use	0.1640*	0.0640*	0.0652*	0.0603*	0.1010*	0.1068*	1						
(8) D&B dissimilarity (before grant)	0.0065	0.0006	-0.0015	0.0003	0.0007	0.1222*	0.0477*	1					
(9) D&B uniqueness (year grant)	-0.0062	-0.0455*	-0.0357*	-0.0180*	-0.0082	0.1191*	0.0314*	0.5293*	1				
(10) D&B adoption (after grant)	0.0408*	0.1136*	0.0959*	0.0542*	0.0328*	-0.0095*	0.0432*	-0.0895*	-0.1543*	1			
(11) D&B composite	0.0310*	0.0540*	0.0448*	0.0185*	0.0158*	0.0634*	0.0804*	0.1597*	0.1054*	0.5145*	1		
(12) Generality	0.0310*	0.1651*	0.1120*	0.0549*	0.0299*	0.1078*	0.0410*	0.0507*	-0.0473*	0.0725*	0.0431*	1	
(13) Originality	-0.0005	0.0727*	0.0549*	0.0294*	0.0179*	0.0375*	-0.0257*	-0.2769*	-0.2256*	0.0923*	-0.0168*	0.1345*	1

* $p < 0.01$

predict all patents as not important would classify over 99 % of all patents correctly). Therefore, we present parallel results for a reduced sample of matched patents in Table 4. We generate a more balanced sample of patents by only retaining control patents with exactly the same combination of technology main classes (3-digit), application year, and grant year as, at least, one of the patents associated with a major technological invention. Treatment and control patents for which no proper matches are found are excluded from the analysis. For each of the 92 remaining major patents, we randomly sample 4 control patents among those that match and rerun the model on the reduced sample.

Marginal effects are calculated as the percentage change with respect to the average likelihood of being an important invention, i.e. 0.15 %. In terms of the indicators capturing the nature and novelty of the patented invention, we find recombining previously disconnected technology subclasses makes a patent 47 % more likely to be a major invention (Table 3, Column 3). Column 4 (Table 3) presents the findings for the different measures suggested by Dahlin and Behrens (2005). We find that patents dissimilar to prior art in terms of backward citations are 67 % more likely to be a major contribution to the field. Surprisingly, the uniqueness dummy has a negative impact, suggesting that important contributions have more similar backward citations to patents filed in the same year and field. Patents with unique backward patent citations are 73 % less likely to be associated with major inventions. This might indicate that similar and parallel inventions building on the same set of technical prior art are conducted during the same time period. Furthermore, we find major patents are not more original, i.e. they do not rely on technical prior art from a broad range of technology fields. In fact, in Column 8 of Table 3, originality is negative and significant. A standard deviation increase in originality reduces the likelihood of a major invention with 23 % compared to the average likelihood.

However, in line with the descriptive statistics, the number of citations to non-patent literature and the number of claims have a positive and significant impact. Surprisingly, the number of main technology classes has a negative and significant effect while the number of backward patent citations is insignificant.

For the indicators reflecting impact and value, we find that the dummies indicative of outliers in the distribution of forward citations have the most predictive power of all indicators. In Column 2 of Table 3, we find patents that are 1 SD outliers in the distribution of forward citations are 173 % more likely to be of major importance for the evolution of biotechnology, 2 SD outliers are 313 % more likely while 10 SD outliers are 693 % more likely. The stricter the criteria of being an outlier, the better the discriminative performance.[5] While a new pairwise subclass combination makes a patent 47 % more likely to be a major invention, the number of future patents adopting the same pairwise subclass combination is also positive and significant (Table 3, Column 3). A standard deviation increase in the number of future patents adopting the same pairwise combination is associated with a predicted increase of 9 % in the likelihood of being a major invention. Also, important contributions have backward patent citations that become adopted by future patents. This effect is strong; patents whose backward citations strongly overlap with future patents are predicted to be 146 % more likely to become breakthroughs, whilst we find no support for a significant impact of combining dissimilarity, uniqueness and adoption (Table 3, Column 4). Column 5 of Table 3 presents results for the originality and generality measures. While important contributions clearly serve a general purpose as technical prior art, they themselves do not seem to rely on technical prior art from a broad

---

[5] Note that more conservatively defined outliers are also less conservatively defined outliers. For instance, a patent which is a 5 SD outlier is also a 1 and 2 SD outlier.

range of fields. A standard deviation increase in generality is associated with an expected increase of 303 % in terms of the likelihood of being a very important invention. Finally, we present the results for all ex-ante indicators reflecting dissimilarity or novelty with respect to prior art (Table 3, Column 6), ex-post indicators reflecting impact (Table 3, Column 7) and both combined (Table 3, Column 8). The results clearly signal the superior performance of ex-post indicators. Measures reflecting direct use as prior art through forward citations, indirectly reflecting adoption through subclass combination and backward citations, display considerably more predictive power. The full model (Table 3, Column 8) combining all indicators is able to identify 25 % of the major patents (recall) while 70 % of the patents predicted to be important are indeed important (precision). Note that ex-post indicators account for the recall and precision. The model with only the ex-ante indicators has a recall of 0 % (Table 3, Column 6).

Table 4 presents the regression results for the matched sample of patents. Note that being a 10 SD outlier in the distribution of forward citations is a perfect predictor of importance; not a single control patent belongs to this category. Therefore, we dropped it from the regression analysis. Thus, the calculated recall and precision are an underestimation. The obtained results are in line with the results on the full sample of patents in Table 3. In contrast to the results for the full sample, having dissimilar or unique backward citations is no longer significant. Moreover, the adoption of backward citations by future patents becomes insignificant in Columns 7 and 8. The full model (Table 4, Column 8) combining all indicators is able to identify 69 % of the important patents (recall) while 84 % of the patents predicted to be of high importance are indeed important (precision). Note again that ex-post indicators account for the lion's share of both recall and precision.

Discussion and conclusion

Technological innovation is an important constituent of economic growth. At the same time, technological inventions vary widely in terms of nature and impact. While there has been a great interest in the development of new technologies and their commercialization, only a minority of these new technologies will contribute significantly to private and social welfare in the future. Consequently, analyzing and understanding both the discovery and the exploitation phases of technological inventions, including their differentiated nature and impact, is important for companies and countries alike. While there has been great interest in the competitive dynamics following the commercialization of inventions, large-scale empirical research on the actual discovery of important technological contributions is scarce. As noted by Dahlin and Behrens (2005), this is mainly due to the lack of reliable indicators that allow such a large-scale quantitative assessment.

In this contribution, we rely on secondary sources to identify the most important technological inventions in the field of biotechnology and relate these to US patent data. Thus, it becomes feasible to examine whether, and to what extent, patent-based indicators advanced in the literature are able to identify these distinctive technological inventions. Indicators advanced so far can be labeled 'ex ante' to the extent that the indicator can be calculated as soon as the invention appears as a patent publication (e.g. dissimilarity of backward citation patterns, new technology subclass combination). Available ex-ante indicators reflect—at least partly—the nature and novelty of the patented invention. At the same time, a number of indicators can only be assessed ex post, i.e. after an invention's impact becomes visible (e.g. the number of received citations, the number of patents displaying similar citation patterns afterwards). Available ex-post indicators reflect—at

Table 3 Identifying important contributions: full sample USPTO biotechnology patents

Variables	(1) BT Logit	(2) BT Logit	(3) BT Logit	(4) BT Logit	(5) BT Logit Ex ante	(6) BT Logit Ex post	(7) BT Logit	(8) BT Logit
Model								
Outlier FC 1SD		2.1788***					1.7515***	1.7598***
		[0.375]					[0.370]	[0.371]
Outlier FC 2SD		1.7081***					1.5563***	1.5491***
		[0.359]					[0.362]	[0.361]
Outlier FC 5SD		0.9558***					0.8302**	0.8339**
		[0.370]					[0.364]	[0.372]
Outlier FC 10SD		3.6892***					3.5345***	3.5716***
		[0.461]					[0.458]	[0.471]
First subclass combi dummy			0.4561**			0.4531**		0.2251
			[0.220]			[0.211]		[0.259]
First subclass combi count re-use			0.0004**				0.0003***	0.0003***
			[0.000]				[0.000]	[0.000]
D&B dissimilarity				0.6188**		0.4843**		0.2160
				[0.243]		[0.216]		[0.337]
D&B uniqueness				−0.6838***		−0.8888***		−0.2029
				[0.252]		[0.237]		[0.319]
D&B adoption				1.3450***			0.5778**	0.6334*
				[0.300]			[0.253]	[0.355]
D&B composite				0.0990				0.1313
				[0.425]				[0.559]
Generality					10.7871***		7.2033***	7.7490***
					[1.000]		[1.177]	[1.261]
Originality					−0.4447	−0.0723		−1.0977***
					[0.349]	[0.382]		[0.423]

Table 3 continued

Variables / Model	(1) BT Logit	(2) BT Logit	(3) BT Logit	(4) BT Logit	(5) BT Logit Ex ante	(6) BT Logit Ex post	(7) BT Logit	(8) BT Logit
Count assignees	0.0181	0.0929	0.0172	0.0353	0.0165	0.0320	0.0846	0.0997
	[0.141]	[0.136]	[0.138]	[0.132]	[0.150]	[0.137]	[0.137]	[0.136]
Count inventors	−0.0103	−0.0482	−0.0077	−0.0092	−0.0140	−0.0112	−0.0576	−0.0468
	[0.045]	[0.057]	[0.046]	[0.045]	[0.045]	[0.046]	[0.059]	[0.058]
Count PRS	0.0009	0.0054	0.0010	−0.0052	−0.0011	−0.0013	0.0039	0.0063
	[0.003]	[0.004]	[0.003]	[0.004]	[0.003]	[0.004]	[0.005]	[0.004]
Count NPRS	0.0046***	0.0019*	0.0045***	0.0041***	0.0042***	0.0045***	0.0013	0.0014
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Count claims	0.0099***	0.0083***	0.0101***	0.0099***	0.0090***	0.0099***	0.0088***	0.0092***
	[0.002]	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]
Count tech. classes	−2.4708**	−3.3732**	−2.5014**	−2.3964**	−2.6907***	−2.4373**	−3.3615***	−3.5808***
	[1.000]	[1.341]	[1.027]	[1.006]	[1.020]	[1.010]	[1.254]	[1.386]
Count tech. subclasses	0.0454**	0.0170	0.0059	0.0437**	0.0409**	0.0324	0.0018	−0.0032
	[0.020]	[0.025]	[0.025]	[0.020]	[0.020]	[0.021]	[0.024]	[0.024]
Patent age	−0.1142***	−0.1110***	−0.0864***	−0.1167***	−0.1025***	−0.1125***	−0.0958***	−0.0841***
	[0.016]	[0.020]	[0.015]	[0.016]	[0.016]	[0.016]	[0.021]	[0.023]
Technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	−796.08	−525.67	−774.40	−771.60	−732.18	−783.82	−495.08	−489.51
Pseudo R^2	0.1255	0.4225	0.1493	0.1512	0.1957	0.1377	0.4562	0.4615
Exp. Pr. ≥ 0.5 as cut off								
Recall Pr(+ | BT)	0.00 %	22.95 %	2.46 %	0.00 %	0.00 %	0.00 %	25.41 %	24.59 %
Specificity Pr(− | NBT)	100.00 %	99.98 %	100.00 %	100.00 %	100.00 %	100.00 %	99.99 %	99.98 %
PrecisionPr(BT | +)	100.00 %	70.00 %	50.00 %	100.00 %	100.00 %	100.00 %	73.81 %	69.77 %
Neg. pred. value Pr(NBT | −)	99.84 %	99.88 %	99.85 %	99.84 %	99.84 %	99.84 %	99.88 %	99.88 %

Table 3 continued

Variables	(1) BT Logit	(2) BT Logit	(3) BT Logit	(4) BT Logit	(5) BT Logit Ex ante	(6) BT Logit Ex post	(7) BT Logit	(8) BT Logit
Model								
False+ for NBT Pr(+ │ NBT)	0.00 %	0.02 %	0.00 %	0.00 %	0.00 %	0.00 %	0.01 %	0.02 %
False− for BT Pr(− │ BT)	100.00 %	77.05 %	97.54 %	100.00 %	100.00 %	100.00 %	74.59 %	75.41 %
False+ for BT Pr(NBT │ +)		30.00 %	50.00 %				26.19 %	30.23 %
False− for NBT Pr(BT │ −)	0.16 %	0.12 %	0.15 %	0.16 %	0.16 %	0.16 %	0.12 %	0.12 %
Correctly classified	99.84 %	99.86 %	99.84 %	99.84 %	99.84 %	99.84 %	99.87 %	99.86 %

Robust standard errors in brackets

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4 Identifying important contributions: matched sample USPTO biotechnology patents

Variables	(1) BT Logit	(2) BT Logit	(3) BT Logit	(4) BT Logit	(5) BT Logit	(6) BT Logit Ex ante	(7) BT Logit Ex post	(8) BT Logit
Outlier FC 1SD		1.5898*** [0.500]					0.9133 [0.606]	0.9783* [0.591]
Outlier FC 2SD		1.6466*** [0.591]					1.9823*** [0.691]	1.9991*** [0.731]
Outlier FC 5SD		2.6926*** [1.000]					1.8784** [0.846]	2.1092*** [0.813]
Outlier FC 10SD								
First subclass combi dummy			0.7056** [0.336]			0.8372** [0.332]		0.6515 [0.465]
First subclass combi count re-use			0.0004* [0.000]				0.0004* [0.000]	0.0003* [0.000]
D&B dissimilarity				0.4999 [0.392]		0.3600 [0.390]		0.0668 [0.548]
D&B uniqueness				−0.5751 [0.393]		−0.9180** [0.407]		−0.6018 [0.555]
D&B adoption				1.1302** [0.456]			0.4001 [0.454]	0.5749 [0.566]
D&B composite				−0.3595 [0.820]				−0.6684 [0.915]
Generality					12.4757*** [1.913]		9.0847*** [2.152]	8.8964*** [2.181]
Originality					−0.5293 [0.573]	−0.4069 [0.537]		−1.6120* [0.891]

Table 4 continued

Variables	(1) BT Logit	(2) BT Logit	(3) BT Logit	(4) BT Logit	(5) BT Logit	(6) BT Logit Ex ante	(7) BT Logit Ex post	(8) BT Logit
Model								
Count assignees	0.4870	0.4101	0.5020	0.5789	0.2076	0.5494	0.2253	0.3385
	[0.396]	[0.365]	[0.397]	[0.371]	[0.433]	[0.411]	[0.423]	[0.390]
Count inventors	−0.0373	−0.1682	−0.0620	−0.0380	−0.0448	−0.0422	−0.1881	−0.1568
	[0.077]	[0.107]	[0.081]	[0.078]	[0.094]	[0.077]	[0.120]	[0.113]
Count PRS	−0.0166	0.0149	−0.0161	−0.0247*	−0.0033	−0.0193*	0.0112	0.0116
	[0.011]	[0.014]	[0.011]	[0.013]	[0.011]	[0.012]	[0.016]	[0.016]
Count NPRS	0.0150***	0.0080*	0.0158***	0.0141***	0.0130***	0.0165***	0.0076	0.0080
	[0.005]	[0.004]	[0.005]	[0.005]	[0.004]	[0.005]	[0.005]	[0.005]
Count claims	0.0490***	0.0388***	0.0495***	0.0481***	0.0518***	0.0516***	0.0399***	0.0459***
	[0.010]	[0.012]	[0.010]	[0.010]	[0.011]	[0.010]	[0.014]	[0.014]
Count tech. classes	0.4422	0.0834	0.4705	−0.0792	0.4491	0.6230	−0.1230	0.3086
	[0.766]	[1.025]	[0.790]	[1.000]	[0.797]	[0.766]	[1.073]	[1.196]
Count tech. subclasses	−0.0678	−0.1199**	−0.1253**	−0.0777	−0.0988*	−0.1256**	−0.1569**	−0.2045***
	[0.048]	[0.053]	[0.055]	[0.051]	[0.056]	[0.052]	[0.063]	[0.063]
Patent age	−0.0579**	−0.0278	−0.0189	−0.0565**	−0.0497	−0.0342	0.0027	0.0350
	[0.027]	[0.037]	[0.029]	[0.027]	[0.035]	[0.029]	[0.039]	[0.044]
Technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	−187.10	−124.55	−182.46	−181.67	−150.02	−181.88	−109.12	−105.75
Pseudo R^2	0.1237	0.4167	0.1455	0.1492	0.2974	0.1482	0.4890	0.5047
Exp. Pr. ≥0.5 as cut off								
Recall Pr(+ | BT)	17.98 %	62.92 %	22.47 %	22.47 %	35.96 %	21.35 %	67.42 %	68.54 %
Specificity Pr(− | NBT)	96.21 %	95.90 %	96.21 %	96.85 %	94.32 %	96.21 %	94.95 %	96.21 %
Precision Pr(BT | +)	57.14 %	81.16 %	62.50 %	66.67 %	64.00 %	61.29 %	78.95 %	83.56 %
Neg. pred. value Pr(NBT | −)	80.69 %	90.21 %	81.55 %	81.65 %	83.99 %	81.33 %	91.21 %	91.59 %

Table 4 continued

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | BT | BT | BT | BT | BT | BT | BT | BT |
Model	Logit	Logit	Logit	Logit	Logit	Logit Ex ante	Logit Ex post	Logit
False+ for NBT $Pr(+ \mid NBT)$	3.79 %	4.10 %	3.79 %	3.15 %	5.68 %	3.79 %	5.05 %	3.79 %
False—for BT $Pr(- \mid BT)$	82.02 %	37.08 %	77.53 %	77.53 %	64.04 %	78.65 %	32.58 %	31.46 %
False+ for BT $Pr(NBT \mid +)$	42.86 %	18.84 %	37.50 %	33.33 %	36.00 %	38.71 %	21.05 %	16.44 %
False—for NBT $Pr(BT \mid -)$	19.31 %	9.79 %	18.45 %	18.35 %	16.01 %	18.67 %	8.79 %	8.41 %
Correctly classified	79.06 %	88.67 %	80.05 %	80.54 %	81.53 %	79.80 %	88.92 %	90.15 %

Robust standard errors in brackets, outlier FC 10SD is dropped because it predicts important contributions perfectly in the matched sample

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

least in part—the impact and value of inventions. Our results show that relying on ex-post indicators allows us to identify patents associated with the most important inventions on a much larger scale (67 % are correctly identified) and more accurately (79 % of the patents that are predicted to be of major importance are correctly classified) compared to ex-ante indictors (21 and 61 % respectively in the matched sample). The 'ex-post' indictors clearly outperform the 'ex-ante' indicators in terms of precision and recall. In consequence, some of the recent proposed indicators, which rely heavily on novelty, do not qualify as accurate predictors of important contributions to the field. In addition, our findings clearly signal potential for future research seeking to identify the nature of inventions more precisely. As currently available indicators do not directly take into account the technical *content* of the inventions under study (e.g. by engaging in a textual analysis of the abstracts or claims), novel, complementary indicators relying on text mining algorithms may well achieve better results (e.g. Magerman et al. 2010; Kaplan and Vakili 2012). In terms of impact, the deployment of more sophisticated, network-oriented indicators could enhance the prospect of added value. Finally, the role of non-patent references, as well as the number of claims, would appear to deserve further attention. When included as a control variable, they tend to consistently predict important contributions in a positive manner. While precision and recall rates for the control variables only are modest, it seems worthwhile to further investigate the nature of this relationship and to assess the relevance of additional indicators based on claims and non-patent references, respectively. We hope our contributions inspire colleagues to engage in such endeavors.

References

Ahuja, G., & Lampert, C. M. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal, 22*, 521–543.

Albert, M. B., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*(3), 251–259.

Arts, S. (2012). *Path dependency and recombinant exploration: How established firms can outperform in the creation of technological breakthroughs. FEB Research Report MSI_1228.*

Basalla, G. (1988). *The evolution of technology*. Cambridge, MA: Cambridge University Press.

Buchholz, K., & Collins, J. (2010). *Concepts in biotechnology history, science and business*. Weinheim: Wiley, VCH.

Bud, R. (1993). *The uses of life: A history of biotechnology*. Cambridge, MA: Cambridge University Press.

Carpenter, M., & Narin, F. (1993). Validation study: Patent citations as indicators of science and foreign dependence. *World Patent Information, 5*(3), 180–185.

Chandy, R. K., & Tellis, G. J. (2000). The incumbent's curse? Incumbency, size, and radical product innovation. *The Journal of Marketing, 64*(3), 1–17.

Christensen, C. (2003). *The Innovator's Solution*. Cambridge: Harvard Business School Press.

Cohen, S., & Boyer, H. (1979/1980). *Process for producing biologically functional molecular chimeras*. US Patent 4,237,224.

Cohen, S., Chang, A. C. Y., & Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *E. coli* by R-factor DNA. *Proceedings of the National Academy of Sciences of the United States of America, 69*, 2110–2114.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy, 34*(5), 717–737.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy, 11*, 147–162.

Fleming, L. (2001). Recombinant uncertainty in technology search. *Management Science, 47*(1), 117–132.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity and creative success. *Administrative Science Quarterly, 52*(3), 443–475.

Freeman, C. (1992). *The economics of hope*. New York: Pinter Publishers.

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of European patents. *European Management Review, 5*(2), 69–84.

Griliches, Z. (1984). *R&D, patents, and productivity*. Chicago: University of Chicago Press.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498*.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics, 36*(1), 16–38.

Harhoff, D., Narin, F., Scherer, F., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics, 81*(3), 511–515.

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation—evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics, 24*(2), 248–270.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly, 35*(1), 9–30.

Hotchkiss, R. D. (1979). The identification of nucleic acids as genetic determinants. In P. R. Srinivasan, J. S. Fruton, & J. T. Edsall (Eds.), *The origins of modern biochemistry*. New York: New York Academy of Sciences.

Jaffe, A., & Trajtenberg, M. (2002). *Patents, citations and innovations*. Boston, MA: MIT Press.

Jaffe, A., Trajtenberg, M., & Fogarty, M. S. (2000). *The meaning of patent citations: Report on the NBER/ case-western reserve survey of patentees. NBER Working Paper* No. W7631.

Kaplan, S., & Vakili, K. (2012). *Breakthrough innovations: Using topic modeling to distinguish the cognitive from the economic: Academy of Management Best Paper Proceedings*.

Keener, K., Hoban, T., & Balasubramanian, R. (2012). *Biotechnology and its Applications*. Retrieved February 22, 2012 from http://www.ces.ncsu.edu/depts/foodsci/ext/pubs/bioapp.pdf.

Magerman, T., Van Looy, B., & Song, X. (2010). Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics, 82*(2), 289–306.

McGloughlin, M. N., & Re, B. E. (2010). *The evolution of biotechnology. From Natufians to nanotechnology*. Dordrecht: Springer.

Nelson, R., & Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge: Belknap Press.

OECD. (2005). A framework for Biotechnology Statistics.

Reuss, M. (2001). Editorial. *Bioprocess and Biosystems Engineering, 24*(1), 1.

Scherer, F. M., & Harhoff, D. (2000). Policy implications for a world with skew-distributed returns to innovation. *Research Policy, 29*(4), 559–566.

Schumpeter, J. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.

Singh, J., & Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science, 56*(1), 41–56.

Sinskey, A. J. (1999). Foreword. *Metabolic Engineering, 1*(3), 3–5.

Stephanopoulos, G. (1999). Metabolic fluxes and metabolic engineering. *Metabolic Engineering, 1*(1), 1–11.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics, 21*(1), 172–187.

Trajtenberg, M., Jaffe, A., & Henderson, R. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology, 5*(1), 19–50.

Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly, 31*(3), 439–465.

Utterback, J. M. (1994). *Mastering the dynamics of innovation: How companies can seize opportunities in the face of technological change*. Boston: Harvard Business School Press.

Watson, J. D., & Crick, F. H. C. (1953). The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18, 123–131.