Research Policy 51 (2022) 104376

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Breakthrough innovations and where to find them

Giovanna Capponi [a,*], Arianna Martinelli [b], Alessandro Nuvolari [b]

[a] Copernicus Institute of Sustainable Development, Utrecht University, the Netherlands
[b] Scuola Superiore Sant'Anna, Italy

ARTICLE INFO

JEL codes:
O31,O32,O33

Keywords:
Breakthrough Innovations
Queen's Awards
Patent Value

ABSTRACT

Breakthrough innovations significantly depart from common practices and can potentially reshape existing markets, create new markets, and prompt the emergence of new technological trajectories. The crucial role that breakthrough innovations play in technological progress has stimulated a lively line of research investigating methods to identify them in actual empirical contexts. Despite this ongoing effort, the availability of data on breakthrough innovations is still scarce and seems to have prevented, at least so far, an integrated perspective comprising both their technical and economic significance. We address this limitation by developing a method that uses award-winning innovations which have been successfully commercialized to find breakthrough patents on a large scale. For the period 1976–2013, in a sample of 138,467 USPTO patents, we identify 17,176 breakthrough innovations. Relevant robustness checks support the validity of our classification. We then exploit this sample to assess the sources of breakthrough innovations.

1. Introduction

Breakthrough or radical innovations are generally regarded as ruptures along specific technological trajectories, possibly leading to shifts or transformations in the prevailing technological paradigm (Dosi, 1982). Thus, they play a crucial role in the "creative destruction" process that characterizes the long-run dynamics of technological evolution (Ahuja and Lampert, 2001). In contrast, the literature refers to continuous or incremental innovations when the outcome of the innovation process is an improvement of the existing technology (Garcia and Calantone, 2002). While incremental innovations occur, more or less continuously, breakthrough innovations are sporadic. As Fleming puts it, in the technology landscape, "almost all inventions are useless; a few are of moderate value; and only a very, very few are breakthroughs" (Verbatim).

Breakthrough innovations are therefore rare events that are inherently difficult to appraise and characterize. As such, they are better investigated by in-depth qualitative studies rather than by large scale quantitative studies that use noisy proxies of innovation output and innovation quality such as patent data. When studying breakthrough innovations on a large scale, the key challenge becomes the measurement of patent quality or value. In the welter of large patent samples, how can we identify those covering breakthroughs? The increasing appreciation of the high variability in the value of patents has led to the emergence of a lively stream of research investigating whether the information contained in patent documents would be suitable to construct reliable indicators of patent quality (Harhoff et al., 2003). So far, several proxy indicators of patent quality have been tested for their ability to capture the heterogeneity in the quality of the underlying inventions and reproduce the skewed distributions associated with it (see for example Harhoff and Reitzig 2004, Lanjouw and Schankerman 2001, Tong and Frame 1994, Trajtenberg 1990).

The identification of breakthrough innovations is mostly based on the technical merits of an innovation and most commonly carried out using patent citations. Dahlin and Behrens (2005) and Verhoeven et al. (2016) looked at the pattern of patents' backward citations to spot breakthrough innovations based on their ex-ante characteristics. The intuition is that breakthrough patents differ in how they source and recombine existing knowledge compared with previous patents in the same field. Other studies classify inventions as breakthroughs using forward citations, as they will be cited in many follow-up patents (Ahuja and Lampert, 2001; Fleming, 2001). For example, in light of the skewed distribution of patent value and the importance of patent citations in determining patent quality, a commonly adopted criterion defined by Ahuja and Lampert (2001) considers breakthrough inventions the top 1% most cited patents. Even if plausible, it is hard to tell how many genuine breakthroughs this approach would capture.

Despite these attempts, the actual empirical effectiveness of the

* Corresponding author.
E-mail address: g.capponi@uu.nl (G. Capponi).

https://doi.org/10.1016/j.respol.2021.104376
Received 29 October 2019; Received in revised form 4 September 2021; Accepted 14 September 2021
Available online 25 September 2021

proposed metrics remains uncertain. In a recent study, Higham et al. (2020) show how commonly used patent quality measures are generally not in agreement about what makes a "high-quality" patent. They tend to capture different aspects of technological discontinuities. Moreover, their ability to pinpointing the economic significance which emerges during the shift from invention to innovation is still questionable (Bessen, 2008). Information on the use of patented inventions is highly valuable for policymakers to the extent that patenting without commercialization has a limited societal impact (Higham et al., 2020). Extant evidence on this aspect has been collected via surveys (e.g., Giuri and Mariani 2007, Webster and Jensen 2011) or retrieved from virtual patent marking websites (e.g., de Rassenfosse and Jaffe 2018). In general, however, the scarce availability of data on breakthrough innovations seems to have prevented, at least so far, an integrated perspective capable of capturing both technical and economic significance.

This paper seeks to address this gap in three main ways. First, we develop a method that uses award-winning innovations to identify breakthroughs on a large scale. Starting from a sample of patents protecting products that received a UK Queen's Award, we build a classifier that optimizes patent-based indicators' ability to classify award-winning patents correctly. We then use the same parameters to predict the probability of a patent to be a breakthrough on a larger scale. In comparison with existing approaches, our method proposes to shift away from arbitrarily chosen cutoffs and towards a more careful evaluation of the probability of an individual patent to be a breakthrough. Moreover, drawing on a sample of patents that proved to be commercially successful in the "real-world", we ultimately seek to identify breakthrough innovations, as opposed to inventions. While the paper presents an application using the Queen's Awards, the methodology is flexible and replicable in other contexts.

Our second contribution speaks to the scarce availability of data on breakthrough innovations. The output of our method is a dataset comprising 138,467 patents filed between 1976 and 2013, with 17,176 patents meeting our criteria for breakthrough innovations ((Capponi et al., 2021). Relevant robustness checks confirm the validity of our method and the quality of the dataset.

Finally, we exploit this dataset to re-examine the sources of breakthrough innovations. Previous work has shown mixed evidence on the difference in technical merit and commercial success between inventions created by independent inventors and organizations. Dahlin et al. (2004) find that while independent inventors are over-represented among low-impact patents, they also hold the most influential patents. In contrast, focusing on the performance of "lone inventors", Singh and Fleming (2010) find that individuals are less likely to generate breakthroughs and more likely to invent failures. We test the ability of different actors to develop commercially successful breakthroughs and find that both individual inventors and public organizations are less likely than corporations to be the source of breakthroughs.

The remainder of this paper is organized as follows. Section 2 provides an overview of the literature on breakthrough innovation and measurement. Section 3 presents the Queen's Award prize scheme. Section 4 describes the process behind the dataset construction. Section 5 goes through the methodology. Section 6 reports the results and Section 7 shows the robustness checks to validate our procedure. In Section 8, we exploit our dataset to re-examine the sources of breakthrough innovations. Section 9 concludes.

2. Literature review

Breakthrough or radical innovations can be defined along different dimensions and several concepts have been proposed to assess their impact on firms, markets, and the economy at large. Focusing on technological importance, seminal work regards breakthrough innovations as discontinuous, disruptive. Utterback (1994) defines them as "change that sweeps away much of a firm's existing investments in technical skills and knowledge, designs, production technique, plant and equipment". Ahuja and Lampert (2001) describe technological breakthroughs as inventions that shape the development of industries by opening new paths for many subsequent technological developments. Looking at the skewed distribution of inventive value, Fleming (2007) refers to breakthroughs as the "long tail" of innovation. The critical role that breakthrough innovations play in technological progress generated a prolific line of research investigating ways to identify them. Most of the work in this domain relies on patent data and the measurement of patent value.

Patent value heterogeneity reflects the distribution of innovation significance. In the literature, we find two main approaches to its measurement: to estimate patent value directly or indirectly ascertain quality-adjusted measures of inventive output using information contained in the patent document (Bessen, 2008). Early contributions in the first group investigate the value of patents by using data on the stock market evaluation of firms (Griliches, 1981; Hall, 1993; Hall et al., 2007; Pakes, 1985). More recently, Kogan et al. (2017) provided individual patent market evaluation using assignee's abnormal stock return on the date of patent grant. With this approach, they can estimate the economic value of patents issued to US firms between 1926 and 2010.[1] A different line of literature uses renewal decisions to infer the value of patents (Lanjouw et al., 1998; Pakes, 1986). Alternative direct ways to estimate the economic value of patents include indicators based on patent transactions such as the decision to license (Gambardella et al., 2007; Sampat and Ziedonis, 2004) and the level of licensing revenues (Sampat and Ziedonis, 2004). Other studies exploit survey data to collect patents' expected sales value (Gambardella et al., 2008; Harhoff et al., 2003, 1999). An important contribution in this respect is the Pat-Val-EU Survey which collected a direct measure of patent value from inventors' direct ex-post evaluation of the patent for a very large sample of European patents.

Considering the indirect approach, the count of patents' forward citations is by far the most popular indicator of technologically valuable patents (Albert et al., 1991; Carpenter et al., 1981; Trajtenberg, 1990). The number of forward citations relates to technological significance because it signals patents' role in establishing a successful line of innovation. Patents' backward citations have also been exploited to assess patent novelty. In this respect, even if the count of backward citations, in principle, ought to be considered as an indicator of incremental innovation (Ahuja and Lampert, 2001; Squicciarini et al., 2013), they can provide insightful information on the new recombinations of prior knowledge (Verhoeven et al., 2016). Following a similar logic, recent developments in text mining techniques are looking for new recombinations of knowledge within the patent document itself (Arts et al., 2019). Another variable related to patent value is the number of claims, insofar as "An inventor's invention is embodied in his or her claims" (Tong and Frame, 1994, pp. 134). Each patent contains some claims which disentangle the invention in different contributions, which represent built-in inventions themselves. Other studies exploit patent family size which refers to the number of countries in which a patent is filed and enforced. Since the geographical extension of patent coverage is costly, only the inventions with a sufficiently high expected value on the market will be protected abroad (Lanjouw et al., 1998).

While all these metrics have substantially contributed to our understanding of patent value and provided a meaningful ground for its measurement, there are still important pitfalls to address. First, these metrics tend to capture different dimensions of patent value and they can lead to contrasting results. It is becoming clear that a one-size-fits-all approach to using these indicators can be misleading (Higham et al., 2020). The second caveat that predominantly applies to these indirect approaches is the general detachment between the patent and the actual

[1] Note that these data are publicly available and we will used them in Section 7.

value of the invention in the real world. While the value distribution of patents reflects the distribution of inventions' significance, the commercial value of patents remains less explored.

In this paper, we seek to overcome some of these limitations using patent-based indicators and parameters externally validated by a list of patents linked to award-winning innovations. Economic historians proposed prizes as an alternative (or complementary) source of innovation data to address the shortcomings of patent data. For example, analyzing inventions promoted by the Royal Agricultural Society of England from 1839 to 1939, Brunt et al. (2012) found that patents are more likely to be renewed when linked to awarded inventions, indicating a positive relationship between the awards and the value of the patent. Moser (2005, 2011) compiled a dataset of innovations using the catalogues of nineteenth-century industrial exhibitions and highlighted this type of data covers "economically useful innovations" rather than "inventions" (Moser, 2002, pp. 1–2). In this paper, we combine the use of patent-based indicators and prizes to identify breakthrough innovations. While we present an application using the Queen's Awards, the methodology is flexible and replicable in other settings.

3. The Queen's Award for enterprise: background

The scheme of the Queen's Awards was initially announced on February the 4th 1965 in the House of Commons by Prime Minister Harold Wilson, following the recommendations of a Committee chaired by HRH The Duke of Edinburgh. In a later speech in the House of Common, Wilson specified that: "The purpose of this new scheme is twofold: to reward and to stimulate. I hope that the award will encourage industry in its efforts to achieve the improvements in exports and the technological advance on which our national future so much depends".[2]

To keep the scheme's parameters up to date, the first draft released in 1965 announced that "the working of the Award Scheme as presented in our report should be reviewed after five years so that any modifications which practical experience of its operation had shown to be desirable could be introduced" (Mountbatten-Windsor, 1965). In practice, apart from the first presentation draft, they made only three other official reports: in 1970, 1975 and 1999. In each review, the Committee in charge collected feedback from the applicants to assess whether and to which extent to amend the scheme. Over its lifetime, the nature of the prize remained largely unaffected and, still today, it represents the most prestigious award for innovation for individuals and businesses in the UK (Groom, 2015). Every year, major British newspapers such as the *Financial Times* and *The Guardian* publish the list of winners and specific articles devoted to comment a selection of them.

At present, what is known as "The Queen's Awards for Enterprise" recognizes achievements in three separate fields: innovation, international trade and sustainability (Department of Trade and Industry, 1999). There are no restrictions regarding sectors or predetermined patterns of regional allocation: all large, medium or small organizations that regularly operate as a "business unit" in the UK are eligible to apply. Through the years, another issue commonly discussed was whether the grant should be associated with tangible rewards, for example, special tax reliefs. However, these suggestions have never been endorsed as "their inclusion would detract from the dignity of the Honour" (McFadzean, 1970). Instead, "The Award signifies recognition by the Sovereign of striking achievement and we recommend that it should be represented by an emblem which the holder of the Award should be authorial to display in a wide variety of ways [...] it may be displayed on flags, plaques, note-paper, packaging, and goods themselves" (Mountbatten-Windsor, 1965). To encourage further improvement, the founding Committee limited the currency of the award to 5 years, during which winners have the right to display the emblem together with the year of victory (Mountbatten-Windsor, 1965).

In this study, we only focus on the Queen's Award for Innovation (hereafter QAI). In the following part, we provide some details on the past and current eligibility criteria to give a precise idea of the parameters used to classify innovations as successful.

3.1. Eligibility, selection and relevance of the QAI

At the very beginning, what was known as the Queen's Award for Technological Achievement was rewarding "A significant advance, leading to increased efficiency, in the application of an advanced technology to a production or development process in British industry or the production for sale of goods which incorporate new and advanced technological qualities". Already in the first report in 1970, the Scheme Review Committee underlined the applied connotation of the definition by removing the word "advanced" before "technology", and emphasizing that "the timely application of established technology, as against advanced technology, may well be equally important and deserving of recognition, particularly in the less sophisticated sectors of industry" (McFadzean, 1970).

In the same spirit, in 1999 the Review Committee in charge drafted the final and most substantial reformulation of the criteria for the award. Above all, the name changed from "Queen's Award for Technological Achievement" to "Queen's Award for Innovation", to broaden the eligible set of subject matters to innovation in services and, in general, innovations which are not technology-driven. There are two main groups of eligibility criteria:

- *Outstanding innovation, resulting in substantial improvement in business performance and commercial success, sustained over not less than two years, to levels which are outstanding for the goods or services concerned and for the size of the applicant's operations, and arising in the fields listed below. Or:*
- *Continuous innovation and development, resulting in substantial improvement in business performance and commercial success, sustained over not less than five years, to levels which are outstanding for the goods or services concerned and for the size of the applicant's operations, and arising in the fields listed below.*

Achievements under either criterion may be assessed in any of the following fields: the invention, design, production (in respect of goods), performance (in respect of services, including advice), marketing, distribution, after-sale support, of goods or services (Department for Business Energy and Industrial Strategy, 2013, pp. 6).

The application procedure requires a detailed description of the innovation and evidence of its commercial success covering a period of two to five years before the submission. Specifically, candidates should support their application by providing the patterns of growth of earnings associated with the innovation, providing information on indicators such as profitability, market shares, and others. Because of the meritocratic nature of the award, the Committee refrained from fixing a number of prizes to be given each year (McFadzean, 1970). Initially, the body in charge of selecting winners was a mix between members from within and outside the government, forming an Advisory Committee which was assisting the Prime Minister by making recommendations about the winners. This group was also supported by two subordinate committees under the guidance of the Ministry of Technology (Mountbatten-Windsor, 1965). In 1999, the subordinate committees were replaced by appointed Panels of ad-hoc Judges with the relevant

[2] Available as a transcript of a House of Common sitting at: https://api.parliament.uk/historic-hansard/commons/1965/aug/03/the-queens-award-to-industry

expertise, one for each Award category. The Panel of Judges examining the Queen's Award for Innovation is chaired by the Permanent Secretary of the Department for Business, Innovation and Skills. Applicants are initially screened by a group of contracted technical assessors, who provide the first round of recommendations (Department for Business Energy and Industrial Strategy, 2013). In September 2018, we had the chance to speak with two contracted technical assessors who provided more details on the selection process pipeline[3] and the criteria adopted to establish innovation quality[4] (personal communications with two contracted technical assessors, 28/09/2018)

Despite the rigorous procedure, the evaluation is made on a pool of self-selected applicants. To mitigate potential bias problems, we cross-validate the relevance of the Queen Award-winning innovations by comparing them with the well-known SPRU Innovations Database (Pavitt, 1983). The latter collects over 4000 technological breakthroughs introduced in Britain between 1945 and 1983, selected by experts' recommendations. The two dataset overlap for 18 years, in which the SPRU Innovations Database identifies 2,595 technological breakthroughs, while 424 innovations receive a QAI. Exploiting the names of the companies and the time references reported in both dataset, we considered a match to be successful if the innovator's name was the same and the time difference between the introduction of the innovation in the SPRU Database and the QAI recognition was at most five years. We found 144 matching observations, suggesting that 34% of the prizes awarded between 1966 and 1983 were awarded to innovations listed in the SPRU Database. In our view, this provides a very significant corroboration to the reliability of the QAI selection process in pinpointing breakthrough innovations.[5]

Given the recognized significance of this accolade, some studies dealing with breakthrough innovations have used QAI winners as a source of data, relying mostly on information collected via questionnaires. The first book using the QAI as a source of data was the book *Wealth from Knowledge. Studies of Innovation in Industry* (Langrish et al., 1972) This work contains 84 case studies based on firms winning in 1966 and 1967. In particular, the authors look at "success stories" to capture the relation between technological innovation and organizational change (see Langrish et al., 1972, page 4). Other studies used the Queen's Awards to assess the relevance of informal external linkages in the innovation process (Conway, 1995), analyze the actual contribution of SMEs to innovation in Britain (Romijn and Albaladejo, 2002; Tether, 1998, 1996) and investigated the reasons behind the tendency of innovative firms to congregate in metropolitan regions (Simmie, 1998). To the best of our knowledge, this paper represents the first attempt to systematically gather the QAI data over the entire lifetime of the scheme, match the winning inventions to patents and use QAI as a starting point for detecting breakthrough innovations.

4. Dataset construction

In this section, we first explain how we matched patents to Queen's Award winning innovations. We then construct an "in-sample" of QAI patents and controls that we will use to test which model better predicts the probability of a patent to be a QAI patent. The parameters providing the best estimate "in-sample" will then be used to identify breakthrough patents in the "full sample", a dataset that combines the in-sample patents with all the patents filed at the USPTO by a GB applicant over the same period.

4.1. Patent matching

Every year the full list of the QAI winners is published with the name of the company, the location and a brief description of the innovation, for a total of 1,475 prizes awarded between 1966 and 2015.[6] For example, a typical entry taken from the 1993 list of winners looks as follows:

BT Laboratories (BTL) Optical Research Division, Ipswich, Suffolk, Metal-organic vapor phase epitaxy for making semiconductor devices for the optoelectronics industry.

Private, public and not-for-profit entities are eligible for a QAI provided that the innovation contributes to industrial efficiency. In the case of joint development, the award is given to both entities. It was not before 1975 that the Review Committee in charge required sales figures in the QAI application to provide evidence of the economic value of innovation (Mountbatten-Windsor, 1975). For this reason, we consider only prizes awarded from 1976 till 2015, for a total of 1,234 innovations. Starting from this listing, we first matched the names of the winning firms to the unique code identifying them in AMADEUS Bureau van Dijk (BvD), a database containing comprehensive information on 21 million European companies.[7] We matched 983 innovators (80% of our sample) for which we could collect the relevant firm-level information and the list of patents associated with each firm. We then looked for

[3] The applications are allocated to the assessors depending on their area of expertise and the number of assessors depends on the number of applications. Assessors are qualified in science and technology, most of them are PhD graduates working in different industries. Assessors review the applications and produce a short list of recommended applicants for the Panel of Judges. Every application is appraised by two people independently. Their decisions have then to be discussed and approved by the lead assessor. After the first round of recommendations, shortlisted firms have to produce audits to confirm their figures. Afterwards, the selected applications are forwarded to a Panel of Judges. Panelists are appointed by consultations with stakeholders, both within government and externally. The panel is reviewed every three years, with no fixed term on membership. Candidates' due diligence and CSR practices are the main concerns at this stage and the retention rate tends to be high. The finalist has to be approved by the Prime Minister and its Advisory Committee.

[4] Technical assessors evaluate innovativeness by looking at competitors and current market offerings. They use the description of the nature of the product reported in the application, the narrative content explaining the innovation process and the challenges faced in developing new solutions. Patents are regarded as supporting evidence rather than a determining signal of novelty. Financial criteria are also important: there should be consistency between the numbers applicants claim, the statements, and the firms' size.

[5] It is worth noting that Queen's Awards are given yearly and as such recognize "contemporary" innovations; the SPRU approach is more longitudinal and recognizes innovations "ex-post". Focusing on the two decades of overlapping (1966-1975, 1976-1985), the proportion of Queen's Award innovations which are listed in both dataset is relatively higher in the first decade than in the second. When considering all the Queen's Award given between 1966 and 1985, 56% of the prizes are given in the first decade (1966-1975) and 44% in the second (1976-1985). When considering only the Queen's Awards which are also in the SPRU dataset, the share of prizes given in the first decade becomes 70%, with only 30% in the second decade. This evidence suggests that innovations awarded in the second decade might not have been established enough to be included in the SPRU dataset. Arguably, innovations awarded in the second decade might have been overlooked because experts' could benefit from hindsight to a lesser extent.

[6] Between 1966 and 1975, 55 combined Awards in Export and Technological achievement were assigned. From 1976 onward, combined Awards discontinued, instead, entrants started having the possibility of competing in more categories in the same year (Department for Business Energy and Industrial Strategy, 2013).

[7] AMADEUS covers all the companies having more than ten employees over a moving window of ten years, leaving out those which ran out of business at each update. Also, it keeps track of M&A history and it is possible to find an enterprise even after changes in structure and ownership. Furthermore, the matching between innovators and AMADEUS entries was checked using companies' addresses and their main sector of activity.

Table 1
Patents matched with QAI winning innovations by filing authority.

Authority	No. of families	Share
GB	1,179	0.80
USPTO	720	0.49
EPO	558	0.38
JPO	465	0.32
USPTO & EPO & JPO	267	0.18

Table 2.
Number of observations by sample type.

Dataset	In-sample composition	No. observations	% of total
In-sample (i.e. USPTO patents filed by QAI winners)	QAI patents (5.44%)	524	0.40%
	Controls (94.56%)	9,114	6.60%
Out-of-sample (i.e. universe of USPTO patents filed by GB applicants)		128,829	93.00%
Full sample		138,467	100.00%

possible matches between the innovations winning a QAI and one or more patent family.[8] We did not impose any restrictions on the application authorities, however, we limited the search to a certain time interval which varies depending on the rules governing the prize scheme in every period.[9] Using this approach, we estimated that 32% of the innovations winning a QAI between 1976 and 2015 were protected by at least one patent at the time of the award. Specifically, we singled out 1, 468 patent families filed between 1976 and 2013 protecting 401 innovations.[10] Table 1 reports the number of matched patent families having at least one application filed in a major patent office and the corresponding share calculated on all the matched patent families. We can observe that almost half of them have a filing at the USPTO and 38% at the EPO.

4.2. In-sample and out-of-sample composition

To implement our method and identify breakthrough innovations, we only focus on the applications filed at the USPTO. There are two main reasons behind this choice: first, since the winning companies are all based in the UK, by selecting a foreign authority we are implicitly raising the quality level of the group of valuable patents, possibly capturing the upper tail of the QAI patented innovations. Typically, companies file applications in foreign patent offices for inventions that are planning to economically exploit abroad. Second, we can easily retrieve the scores of the patent-based indicators from the OECD Patent Quality Indicators database, which provides details on roughly 7 million USPTO patents, with filing dates going as back as 1976 (Squicciarini et al., 2013).

We first exploit patent-based indicators to identify the QAI patents within a larger pool of control patents. We build the control group using all the patent applications filed at the USPTO by companies having at least one QAI patent. This sampling choice allows us to compare the awarded inventions with the other innovations developed within the same group of innovators, i.e. leveraging on the same set of resources and capabilities.[11] This dataset consists of 9,638 patents, of which 524 QAI patents (5,4%) and 9,114 controls, we refer to this selection as the "in-sample".[12]

To find commercially successful innovations on a large scale, we construct an "out-of-sample" as a comprehensive set of USPTO patents filed by GB applicants from 1976 to 2013. Starting from the full set of USPTO patent applications listed in the OECD Patent Quality Indicators database (around 7 million patent applications), we single out the patent applications associated with a GB applicant in the EPO's Worldwide Patent Statistical Database, henceforth PATSTAT. Excluding those reporting missing values on at least one of the indicators of interest, we end up with an out-of-sample of 128,829 patents. Table 2 reports the composition of the full sample.

Our method seeks to first identify the model which best classifies QAI patent in-sample, and then use the same parameters to estimate the predicted probability of being a breakthrough for all the patents in the full sample.

5. Methodology: how to identify breakthrough innovations

We start by comparing the ability of patent-based indicators to identify QAI patents in the in-sample. We estimate logit models specified as follows:

$$ln[p / (1-p)] = b_0 + b_1 X + Year\ dummies + Technology\ dummies \qquad (1)$$

Where p=prob(QAI patent=1). *QAI patent* is a direct measure of patent value indicating whether a patent protects a QAI winning innovation or not, X is a vector of patent-based indicators including renewals,[13] number of forward citations, patents' family size and number of claims. We have selected indicators that are ((1) widely adopted as measures of patent quality and (2) standardized and easily retrievable

[8] The OECD Patent Statistics Manual defines patent families as "the set of patents (or applications) filed in several countries which are related to each other by one or several common priority filings" (OECD, 2009, page 71). Our analysis is based on DOCDB families, which are expert-validated families. In principle, the technical content of patents protecting the same invention should be the same. The DOCDB definition builds on experts' assessment on whether the technical content of a patent matches an existing family or not (Martínez, 2011).

[9] A detailed explanation of the patent matching procedure is provided in Appendix A.

[10] However rigorous, the method we followed looks into a specific and limited time window to find matching patents. Thus, it is possible that a few cases were not captured by our search.

[11] The Queen's Award for Innovation strictly recognizes innovations developed in the UK. The innovator can be either a British firm or the UK branch of a multinational company.

[12] The construction of the in-sample requires access to the complete list of the patents associated with a firm, therefore we only considered the winning companies that we successfully matched on AMADEUS. An additional reduction to the sample size resulted from matching the patent applications to the OECD Quality database, which covers patent applications filed between 1976 and 2015 at the USPTO. To have consistent estimates across observations, we exclude patents having missing values on at least one of the indicators of interest.

[13] Note that while the OECD Quality Database Squicciarini et al., 2013) includes a yearly measure of renewals, USPTO patents are renewed after 3.5 years and renewals fees are only available since 1981. From an e-mail exchange with the authors, we learnt that the renewal indicator for the USPTO patents in the OECD Quality Database reflects the lifetime of a patent and counts the number of years between the patent application and the latest date during which any of the following events occurred – payment of the renewal fees (not applicable before 1981), withdrawal, refusal, etc. To better capture the value associated with the payment of renewal fees, we define a variable *Renewal USPTO* that counts the number of renewal fees paid instead (0 if the OECD indicator is <4 years, 1 if [4, 8), 2 if [8, 12), 3 if >=12 years).In our dataset (138,467 observations) we also have 13,350 patents that have been filed before 1981 when the payment of maintenance fees came into force. While this makes the attribution of a *Renewal_USPTO* value problematic, the OECD indicator shows that 95% of the patents in this group were kept alive for less the 4 years anyway. In these cases, *Renewal_USPTO* can take value zero without loss of generality. As for the remaining 5% (696 patents), we assign the *Renewal USPTO* assuming that renewal fees would have been paid throughout the indicated patent lifetime.

Table 3.
Testing patent-based indicators.

DV: QAI patent	M1 Coef/(se)	M2 Coef/(se)	M3 Coef/(se)	M4 Coef/(se)	M5 Coef/(se)	M6 Coef/(se)
Renewals USPTO		0.36***				0.33***
		(0.08)				(0.08)
Forward citations (ln)			0.26***			0.20***
			(0.05)			(0.06)
Family size (ln)				0.26**		0.17*
				(0.08)		(0.08)
Claims (ln)					0.30***	0.26**
					(0.08)	(0.07)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Technology dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−3.00***	−3.04***	−3.21***	−3.42***	−3.68***	−4.08***
	(0.42)	(0.42)	(0.42)	(0.44)	(0.45)	(0.47)
No. observations	9,638	9,638	9,638	9,638	9,638	9,638
Pseudo R^2	0.10	0.11	0.11	0.11	0.11	0.12
Optimal threshold using the ROC curve approach	0.0542	0.0678	0.0578	0.0531	0.0629	0.0776
Sensitivity	70%	62%	69%	72%	65%	60%
Specificity	69%	76%	71%	68%	74%	81%
Accuracy	69%	76%	71%	68%	74%	80%

Note: logit regressions with robust standard errors. Sensitivity indicates the% of true positive case classified as positive. Specificity indicates the% of true negative cases classified as negative. Technology dummies are based on the technical fields in Schmoch (2009). Legend: *** (**, *) indicate a significance level of 0.1% (1%, 5%).

from major databases. We also include sets of dummies controlling for patents' filing year and technology class.[14] Table B.1 in Appendix B provides the list of variables and sources.

Working with binary models, the key issue we face is the definition of a threshold probability which best classifies patents as QAI patents or controls. We first evaluate different specifications using a Receiver Operating Characteristic (ROC) curve, which selects the best model based on its accuracy, i.e. the percentage of correctly classified patents at a certain threshold.[15] The ROC curve estimates the optimal probability threshold as the one maximizing sensitivity and specificity with equal weight.[16] While giving an estimate of the models' accuracy, this approach does not incorporate information on the consequences of misclassification and it fails to tell whether a specification is worth using at all. This concern is particularly prominent in epidemiology where a false negative is much more harmful than a false positive. In these cases, a model with a high level of accuracy but a level of specificity which is slightly higher than sensitivity would be a poor choice. Similarly, since we aim to identify breakthroughs out-of-sample, we would prefer a model which limits the false positive rate rather than a model which maximizes accuracy.

For these reasons, we select the best model adopting a Decision Curve Analysis (DCA), a comparative approach that incorporates information on the consequences of a misclassification (Vickers and Elkin, 2006). This method assesses the effectiveness of the prediction models by looking at the theoretical relationship between the threshold probability p_t of being a breakthrough and the relative value of false positive and false negative. To explain the concept, suppose that a surgeon needs to intervene on a patient depending on a marker prediction on the patient probability to have cancer. On the one hand, an unnecessary intervention can have severe side effects, on the other hand, failure to intervene on a true positive case may lead to the patient's death. Assume that the surgeon would definitely proceed if the probability to have cancer is 30%, but he would not if the probability was only 1%. However, if the risk was 10%, he would be uncertain. To fix p_t at 10% means that the surgeon considers the failure to intervene in a positive case to be 9 times worse than unnecessary surgery (Vickers and Elkin, 2006). This relationship is defined by the net benefit formula (first introduced by Peirce 1884):

$$NetBenefit = (TruePositiveCount/n) - (FalsePositiveCount/n) \times (p_t/(1-p_t)) \tag{2}$$

Where p_t is the threshold probability. In this equation, $(p_t/(1-p_t))$ tells the relative importance of false negative and false positive results. The basic mechanism to follow would then be to (1) choose a meaningful range [0,1] for p_t, (2) look at the net benefit associated with each point in the interval for the models we are comparing, (3) choose the optimal threshold and the best model accordingly. Finally, using the parameters estimated in-sample with the selected model, we predict the probability of a patent to be a commercially successful breakthrough on the full sample. The threshold probability indicated by the DCA will then be used to classify the patents as breakthroughs or not.

6. Results

As expected, the descriptive statistics within the in-sample show that forward citations, family size and claims present a right-skewed distribution, we therefore follow previous papers (see for example Singh and Fleming 2010) and use their logarithmic transformation in the analysis.[17] The number of renewals is not affected by outliers as it only takes four possible values. Looking at the distribution separately for QAI patents and controls, we notice that the means for the QAI patents are higher for all indicators except for the family size (see Table B.2 in Appendix B). To better interpret this outcome, it should be pointed out that we set an explicit requirement on the breadth of control patents'

[14] Based on (Schmoch, 2009) classification, which is reported in the OECD Quality database (Squicciarini et al., 2013)

[15] The ROC curve is defined on a space having sensitivity as y coordinate and 1-specificity (fall out rate) as x coordinate. The ROC curve of a diagnostic test leading to perfect discrimination would be a vertical line from (0,0) to (0,1) joined with a line from (0,1) to (1,1), while the curve of a poor performing test would resemble the 45 lines, which is the chance level (Hajian-Tilaki, 2013). In this study, we adopt the empirical (non-parametric) ROC estimator, which constructs the curve by connecting the (1-specificity, sensitivity) points obtained at all possible cut-off values (Park et al., 2004).

[16] To obtain the optimal threshold for each model we adopt the Youden Index, which selects the point on the curve having the maximum vertical distance from the chance line, maximizing both sensitivity and specificity (Youden, 1950).

[17] In doing so, we first added one to the number of forward citations because they can take a value of 0.

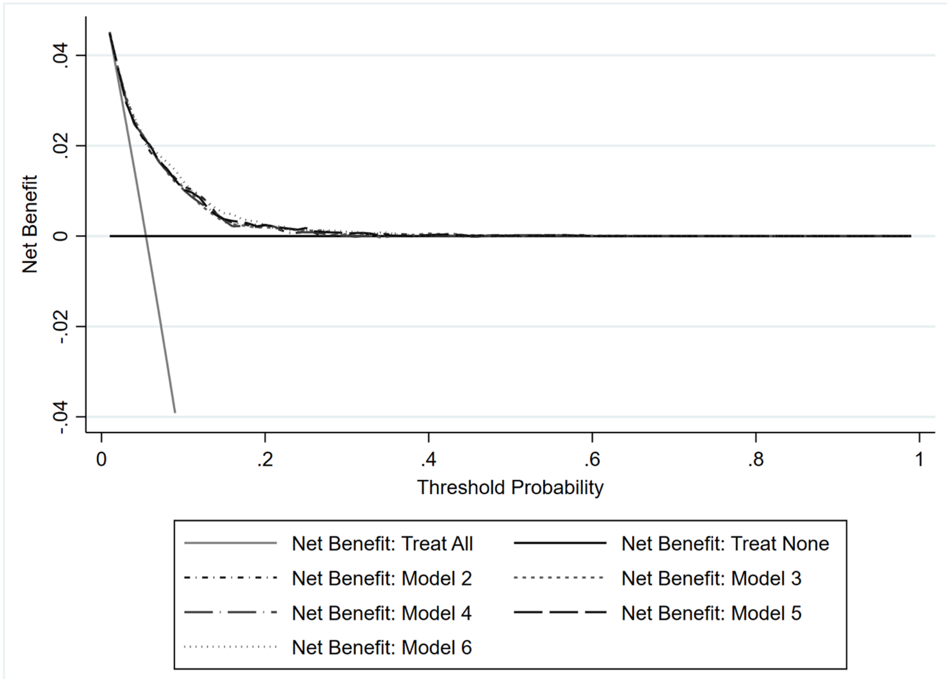

Fig. 1. Decision curves.

families by considering only USPTO filed applications. Also, jurisdictions vary in terms of market size and they should not be considered equally relevant (Van Zeebroeck, 2011).

We run the logit models to test the effect of each patent-based indicator on the probability of a patent to be a QAI patent.

The coefficients are positive and significant for all patent indicators in the six specifications we defined (see Table 3). However, the models' discriminatory power varies: a ROC curve approach would select Model 6 as the most accurate model, with 80% of the patents correctly classified. This evidence is in line with Higham et al. (2020) who describe patent quality as a multidimensional concept that cannot be captured by a single metric. Maximizing specificity and sensitivity with equal weight, Model 6 classifies a patent as 1 if its probability of being a QAI patent is above 7.8%. This cut off is rather low because of the strong presence of control patents in the sample (more than 90%). Indeed, the level of accuracy is heavily driven by the ability of the model to classify the zeros. Accordingly, Model 6 shows a relatively high level of specificity at the optimal threshold. For these reasons, the model may perform poorly when it comes to the identification of breakthrough innovations in the full sample.

Given the very small share of QAI patents in-sample (5.5%), suppose that we would classify a patent as 1 if its probability of being a QAI patent is at least 15%. Fig. 1 shows the decision curves for each model estimated in Table 3, the x-axis is p_t while the y-axis is the net benefit, which indicates the percentage of true positive cases that a model captures at different probability thresholds (see Eq. (2)).

At each threshold probability, the model leading to the highest net benefit is the best. Looking at the y-axis, we can observe that the highest possible net benefit is the true prevalence of QAI patents in the sample. For example, identifying breakthroughs using a model that at a certain p_t achieves a net benefit of 2% (0.02), is equivalent to a strategy that correctly picks 2,000 breakthroughs out of 100,000 patents. Note that this graph does not give any information about the number of false negatives and false positives.[18]

In all models, as the threshold probability rises, we assign 1 to fewer patents, and the share of true positives that we capture tails off to zero. Instead, if we set the threshold probability to zero, the net benefit reached by the models approaches the prevalence. In general, the models we are testing achieve a higher or equal net benefit at each probability threshold, compared to the baseline strategies. At this point, we should identify a suitable range for p_t and select the best model. Fig. 1 shows that for $p_t > 0.2$ the net benefit is very low and we risk missing out on too many breakthroughs. Instead, for $p_t < 0.1$ the net benefit increases but the risk of including a high number of false positives increases too. Fig. 2 zooms on this interval.

We can observe that Model 6 performs better than the other models for probability thresholds ranging from 13% to 19%. To have a better sense of the outcome for this model, we calculate the prevalence of false positive and false negative for p_t=0.15, p_t=0.16, p_t=17, p_t=0.18 and p_t=0.19 based on the estimates obtained for Model 6 (in Table 3).

[18] The decision curve analysis enables us to compare the performance of the models vis-a-vis the two baseline strategies of "pick all", or "pick none". The horizontal line (pick none) corresponds to a strategy in which we assume that there are no breakthroughs in the sample. In this case, the net benefit would be zero, regardless of p_t. Conversely, a "pick all" strategy assumes that all the innovation in the dataset are breakthroughs. This approach has a performance which is comparable with the models only if we set p_t=0, that is if we do not attach any harm to the inclusion of false positive, while we are 100% committed to avoid false negative. In this case, the intercept point on the y axis is 5.5%, i.e. the QAI patents' prevalence. If we assign 1 to every patent in the dataset, than 5.5% will be true 1s, hence the net benefit, with 94.5% false positive rate.

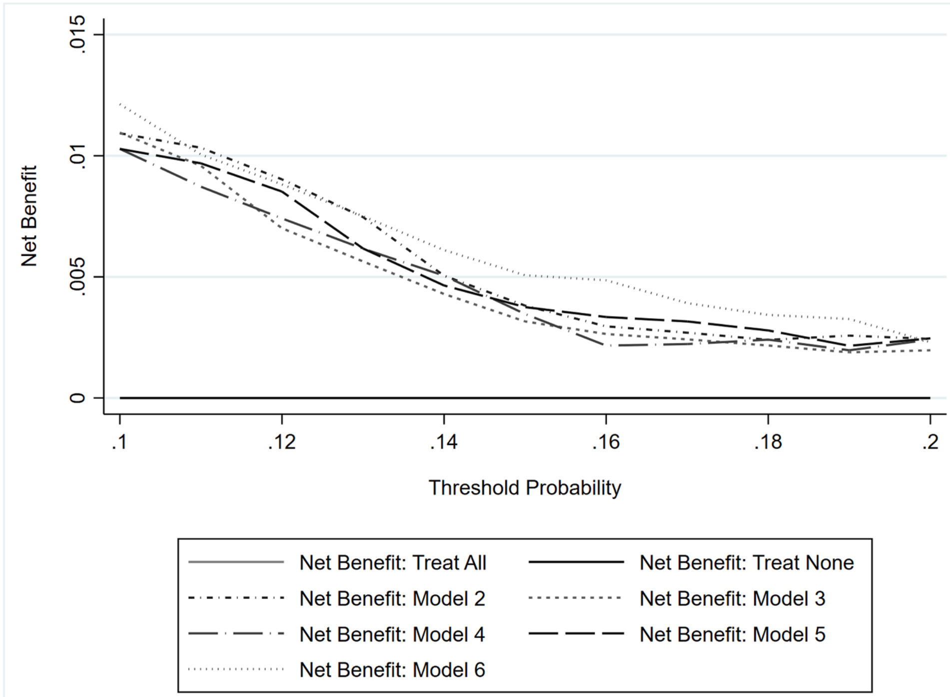

Fig. 2. Decision curves for $0.1 \leq p_t \leq 0.2$.

Table 4.
Prevalence of false positives and false negatives at selected p_t using the ROC and DCA approach.

	ROC	DCA				
p_t	0.0776	0.15	0.16	0.17	0.18	0.19
True positives	312	142	132	116	102	95
True negatives	7,357	8,586	8,667	8,732	8,800	8,843
False positives	1,757	528	447	382	314	271
False negatives	212	382	392	408	422	429
Total	9,638	9,638	9,638	9,638	9,638	9,638
% of true positives in the group classified as +	15%	21%	23%	23%	25%	26%
% of false positives in the group classified as +	85%	79%	77%	77%	75%	74%

Note: The results are derived using the result estimation of Model 6 in Table 3.

Table 4 shows the results; as p_t increases, the group of patents classified as QAI patents becomes smaller and the share of false positive decreases.

Opting for a conservative approach, we estimate the predicted probability in the full sample using Model 6 and setting p_t=0.19. We then define a new variable *predicted breakthrough* which takes value 1 if the predicted probability calculated in the full sample is p>=0.19, 0 otherwise.[19] This method classifies 17,176 of the patents in the full sample as breakthroughs (12%). In the next section, we check the validity of our method testing both the economic significance and the technical merits of our predicted breakthroughs.

7. Robustness checks

7.1. Relation with independent estimates of patent value

We first examine the correlation between our predicted breakthroughs and patent value. We use the estimates developed by Kogan et al. (2017) to measure the monetary value of patents.[20] By matching their data with ours, we could assign a monetary value to 2, 887 patents of our in-sample (29.9%) and 32,221 patents of our full sample (23.2%).[21]

The regressions in Table 5 examine the correlation between QAI patents, our predicted breakthroughs, and the log of patent value estimated by Kogan et al. (2017). In all regressions, we include year and technology dummies as controls. In some specifications, we also include the number of forward citations of the patent, to see whether the QAI patents or the predicted breakthrough variable remain significant while including one of the most used indicators of patent value.

Models 1 and 2 examine the correlation between receiving a Queen's Award and the patent monetary value in the original in-sample set of patents. We find that QAI patents have a positive and significant higher economic value as compared to similar patents owned by the same set of innovators. Models 3 and 4 report the correlation between *predicted breakthrough* and the patent monetary value in our full-sample set of patents. We find a significant positive correlation between our estimated variable and patent monetary value in Model 3. When including forward citations, *predicted breakthrough* is still significant, albeit at the 10% level (Model 4).

Finally, Models 5 and 6 show the results of an estimation similar to the one presented in Model 1 and 2 but performed on the full sample of patents. In this case, the correlation between receiving a Queen's Award and the patent monetary sample is negative but insignificant. This is a surprising, but interesting result. The regressions of Models 5 and 6 suggest that when studied against the background of a very large set of patents, the QAI patents are probably not characterized by monetary values that set themselves out from the "average" patent. In other words, the extension of the sample from Models 1 and 2 to Models 5 and 6 brings in several additional patents of relatively high monetary value, so

[19] The distribution of the predicted probability on the full sample is very skewed (mean=0,09, median=0,06), thus we can reasonably expect to capture high quality patents setting $p_t = .19$

[20] The data are freely available at: https://iu.app.box.com/v/patents.

[21] Note that Kogan et al. (2017) provide the patent value only for patents assigned to listed companies.

Table 5.
Patent value (Kogan et al., 2017), QAI patents and predicted breakthroughs.

	M1 *In-sample* Coef/(se)	M2 *In-sample* Coef/(se)	M3 *Full sample* Coef/(se)	M4 *Full sample* Coef/(se)	M5 *Full sample* Coef/(se)	M6 *Full sample* Coef/(se)
QAI patents	0.89***	0.93***			−0.16	−0.18
	(0.18)	(0.18)			(0.19)	(0.19)
Forward citations (ln)		−0.09***		0.10***		0.10***
		(0.02)		(0.01)		(0.01)
Predicted breakthrough			0.10**	0.06		
			(0.03)	(0.03)		
Constant	0.48***	0.63***	0.49***	0.28***	0.49***	0.28***
	(0.06)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Technology dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. observations	2,887	2,887	32,221	32,221	32,221	32,221
R^2	0.34	0.34	0.20	0.20	0.20	0.20

Note: OLS regressions with robust standard errors. In-sample refers to the USPTO patents filed by QAI winners (i.e. it includes the QAI patents and the controls). Full sample refers to all the patents in the in-sample and the patents in the out-of-sample (i.e. universe of USPTO patents filed by GB applicants). Technology dummies are defined at the IPC-class level. Legend: *** (**, *) indicate a significance level of 0.1% (1%, 5%).

Table 6.
Breakthroughs impact on IPC subgroup.

	DV: Number of patents filed					
	M1 *Shock at (t)* Coef/(se)	M2 *Shock at (t)* Coef/(se)	M3 *Shock at (t)* Coef/(se)	M4 *Shock at (t-5)* Coef/(se)	M5 *Shock at (t-5)* Coef/(se)	M6 *Shock at (t-5)* Coef/(se)
Post predicted breakthrough	30.07***	2.90*	2.88*	30.89***	0.94	1.13
	(2.22)	(1.29)	(1.29)	(1.76)	(0.99)	(1.00)
Year trend		2.88***	2.10***		3.14***	1.09**
		(0.27)	(0.40)		(0.22)	(0.35)
Year trend squared			0.01**			0.03***
			(0.00)			(0.00)
Constant	13.53***	−75.40***	−62.88***	8.10***	−91.84***	−57.54***
	(1.22)	(9.46)	(10.40)	(0.95)	(7.83)	(8.48)
Technology dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. observations	66,350	66,350	66,350	65,655	65,655	65,655
R^2	0.47	0.47	0.47	0.42	0.43	0.43

Note: OLS regressions with robust standard errors clustered at the WIPO IPC subgroup level. Technology dummies are defined at the IPC subgroup level. Legend: *** (**, *) indicate a significance level of 0.1% (1%, 5%).

that QAI patent becomes insignificant. However, if we employ our predicted breakthrough indicator in the same large sample, we find still find some effect. All in all, we interpret the findings of Table 5 as further corroboration of our approach.

7.2. Impact on innovations in the same technology class

The evolutionary economics literature makes a fundamental distinction between the search and selection of new technological paradigms and the technological progress along a specific trajectory (Dosi, 1982). In this context, breakthrough innovations have the potential to shift the incumbent trajectory and stir the direction of technical progress accordingly. We now investigate whether this property characterizes the breakthroughs identified by our method.

Following previous studies, we consider patents as belonging to the same trajectory if they belong to the same technology class (Andersen, 1999). To be very precise about the scope of the effect of a breakthrough, we define the technological domain at the highest level of disaggregation using the WIPO International Patent Classification up to the subgroup level.[22] As regarding the timing, we examine the effect of each predicted breakthrough patent on its technological domain for the ten years after their filing, as compared to the previous ten. Our dependent variable is the *number of patents filed* in a given year, the independent variable of interest *post predicted breakthrough* is a time-variant dummy that takes value 1 from the year in which our predicted breakthrough patent was filed. We estimate OLS regressions testing the effect of the filing of a predicted breakthrough patent on the number of subsequent patents filed controlling for the year trend, the year trend squared, and the IPC subgroup. To address possible issues of autocorrelation of the residuals, we cluster the standard errors by IPC subgroup (Cameron and Miller, 2015). Table 6 shows that the effect is positive and significant.

To check the robustness and sensitivity of our results, we run the same models in a "placebo" setting, anticipating the year of the occurrence of the predicted breakthrough patent (t) by five years (t-5). Consider for example an IPC subgroup in which the predicted

[22] The WIPO Classification separates the whole body of technical knowledge using the following hierarchical levels: section, class, subclass, group and subgroup. For example, the group "H01S 3/00" corresponds to "Laser", the subgroup "H01S 3/14" corresponds to "Laser characterised by the material used as the active medium" (the complete classification is available at https://www.wipo.int/classifications/ipc/en/). In the dataset we use for this exercise we have 4,043 technical subgroups.

Table 7.
The source of breakthrough innovations.

	DV=Predicted breakthrough				
	M1 Coef/(se)	M2 Coef/(se)	M3 Coef/(se)	M4 Coef/(se)	M5 Coef/(se)
Public assignee	−0.23**	−0.26***	−0.01	−0.26***	0.00
	(0.07)	(0.07)	(0.12)	(0.07)	(0.12)
Independent inventors	−0.43***	−0.37***	−0.37***	−0.36***	−0.35***
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Team		0.32***	0.34***	0.33***	0.34***
		(0.03)	(0.03)	(0.03)	(0.03)
Public assignee *team			−0.35*		−0.35*
			(0.15)		(0.15)
Independent inventors *team				−0.04	−0.05
				(0.12)	(0.12)
Experience, 10 year (ln)	−0.04***	−0.04***	−0.04***	−0.04***	−0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Experience, 10 year, same IPC subclass (ln)	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Year dummies	Yes	Yes	Yes	Yes	Yes
Technology dummies	Yes	Yes	Yes	Yes	Yes
Constant	−8.61***	−8.74***	−8.75***	−8.75***	−8.75***
	(1.08)	(1.08)	(1.08)	(1.08)	(1.08)
No. observations	119,853	119,853	119,853	119,853	119,853
Pseudo R2	0.55	0.55	0.55	0.55	0.55
Prob > chi2	0.00	0.00	0.00	0.00	0.00

Note: logit regressions with robust standard errors. Technology dummies are defined at the IPC subclass level. Legend: *** (**, *) indicate a significance level of 0.1% (1%, 5%).

breakthrough was filed in 2000, the placebo model would replace the shock year (i.e. when *post predicted breakthrough* becomes 1) with 1995 and test the effect of this fictitious shock on the subsequent 10 years. Models 5 and 6 show that when arbitrarily changing the shock year (and controlling for the time trend) the effect is no longer significant. The positive relation between the breakthrough arrival and the development of more inventions in the same technological areas suggests that our method correctly identifies breakthroughs.

8. The sources of breakthrough innovations

From an evolutionary perspective, breakthrough innovations emerge from a process of recombinant search and selection: technical novelty is associated with the ability to find new recombinations of prior knowledge which are then evaluated in the selection phase. This mechanism has been investigated along different dimensions to improve the understanding of the sources of breakthrough innovations. An important debate is whether independent inventors are more or less likely to introduce breakthrough innovations than corporate inventors. On the one hand, being industry "outsiders" with no preconceived notions, independent inventors can be better equipped to come up with novel ideas. On the other hand, financial constraints coupled with the absence of a corporate setting and structured intellectual support may lead to the generation of inventions of little relevance (Dahlin et al., 2004). Singh and Fleming (2010) focus on the "lone inventor" as a possible source of breakthroughs and argue that individuals are less likely than teams to generate breakthroughs because they cannot benefit from the help of collaborators in sorting and identifying promising new ideas.

Within our full sample, 93% of the patents are assigned to at least one organization.[23] To analyze which conditions favor the emergence of breakthroughs, we define three mutually exclusive dummy variables capturing whether at least one assignee is a public organization (6% among universities, government and hospitals), at least one private company and no public organizations (87%) or only individual inventors (7%).[24]

The analysis is carried out at the patent level and we estimate the probability of a patent to be a *predicted breakthrough* depending on the type of assignee and the number of inventors. To account for the prior experience and resources of the assignee(s), we also include the cumulative number of patents filed over the ten previous years and the cumulative number of patents filed over the ten previous years in the same technology class.[25] Finally, we include year and technology dummies.[26]

Table 7 presents the results. Looking at the independent variables of interest, M1 shows that the odds of patents developed by individual inventors and the public sector to be breakthroughs are 35% and 20% lower than the odds for corporate patents. This evidence is expected given that our method seeks to identify inventions that became commercially successful, corporations are generally better equipped in this sense. In line with Singh and Fleming (2010), Model 2 shows that the *team* indicator is positive and significant, however, it does not improve the chances of a patent to be a breakthrough in these cases (Models 3 and 4). Our findings seem to suggest that individual inventors (or "garage inventors") are less likely to contribute to the upper tail of the distribution regardless of whether they work in a team or not. Intriguingly, the presence of a team seems to be detrimental when public organizations are involved. Arguably, in the context of public organizations, single individuals can perform better than teams because they can benefit from a higher level of intellectual freedom than in a corporate setting.[27] At the same time, they can rely on the organizational structure that "lone inventors" miss. This evidence suggests that universities and more in general public institutions can potentially create a productive research environment for single inventors, to the extent that they offer a combination of formal support and intellectual freedom. Finally, it is interesting to notice that organizations' experience over the past 10 years is negatively associated with the probability of a patent to be a breakthrough. This evidence supports the incumbent curse argument whereby incumbents would be less likely than new entrants to come up with breakthrough innovations.

9. Conclusion

In this paper, we have introduced a method to identify breakthrough innovations on a large scale starting from award-winning innovations. The output of our procedure is a dataset comprising 138,467 patents filed over 37 years, of which 17,176 are classified as breakthroughs. We then exploit this dataset to investigate the sources of breakthrough innovations depending on the type of assignee and the number of inventors.

Our proposed method emphasizes the link between patent data and

[23] In terms of patent filing, the share of individual inventors has decreased dramatically over the past century, from 86% in 1910 to 15% in 1998 (USPTO, 1998).

[24] These information are retrieved from Table TLS206_PERSON in PATSTAT.

[25] Note that 74% of the patents have only one assignee, whenever the number is higher we take the average of the assignees' experience.

[26] Following the set up proposed by Singh and Fleming (2010), we also tested models controlling for the number of claims made by a patent, the number of backward citations and the number of non-patent references. The results (not shown) remain broadly consistent.

[27] In line with Singh and Fleming (2010), we find that working in teams further enhance the probability of a patent to be a breakthrough when the assignee is a private company.

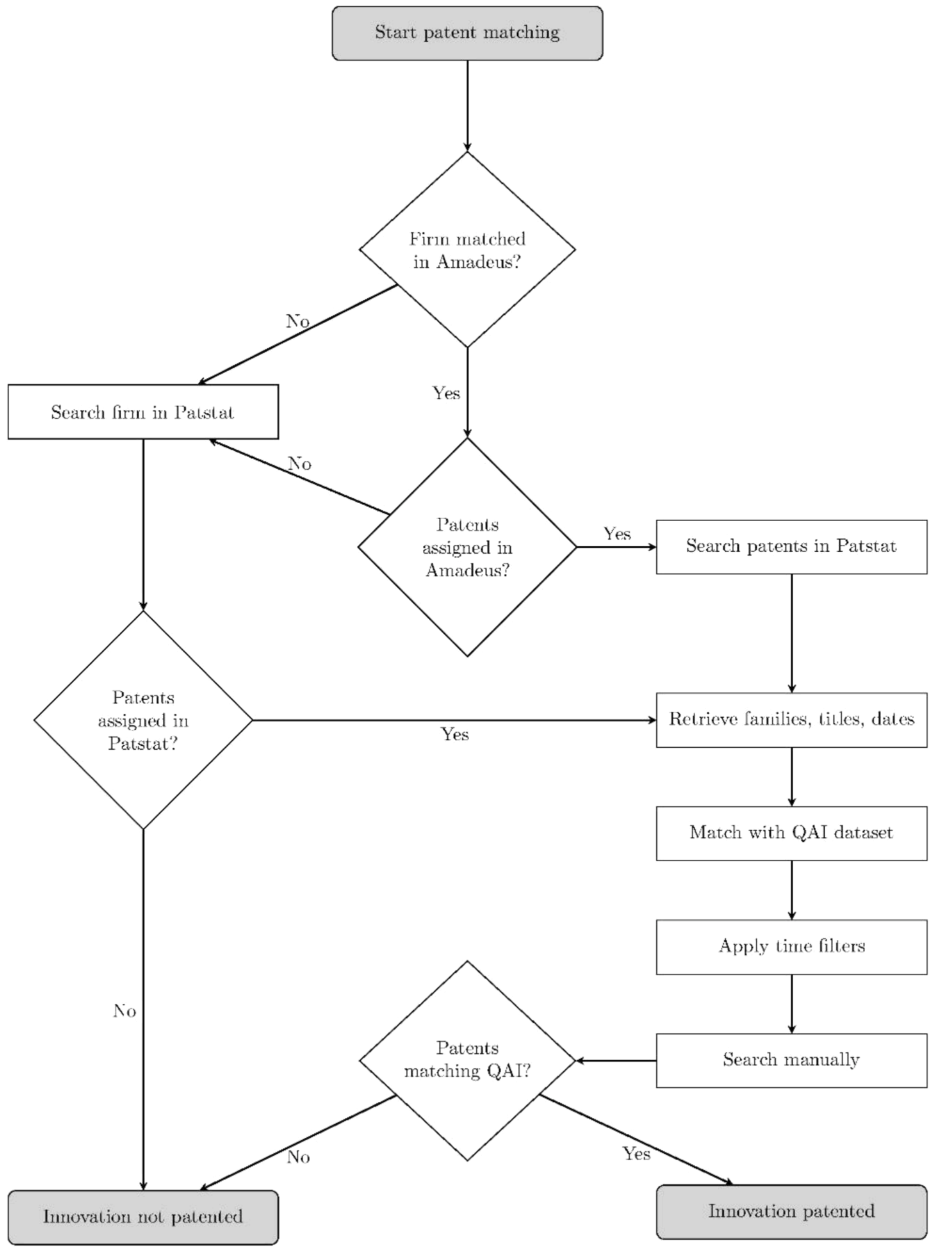

Fig. A.1. Patent matching flowchart.

Table B.1.
Overview of the variables used in Table 3.

Name	Description	Source
QAI	Binary variable equal to 1 if the patent refers to an awarded innovation, 0 otherwise	Authors' elaboration
Renewals USPTO	No. of renewal fees paid	Authors' elaboration based on the OECD Quality Database (Squicciarini et al., 2013)
Forward citations	No. of citations that a patent receives over a period of five years after the publication date	OECD Quality Database (Squicciarini et al., 2013)
Family size	No. of patent offices at which an invention has been protected	
Claims	No. of claims per patent	
Year dummies	38 dummies for the patents' earliest filing year	
Technology dummies	35 dummies for the patents' technical field	

innovations with demonstrated technical and commercial success. Through external validation of patent value, we seek to advance research in this field by addressing the often missing connection between measures of patent value and the actual use and commercial success of such patents. It is also worth stressing that future contributions in this direction could replicate, adapt or extend our approach taking as inputs other prize schemes or different evidence of innovations' economic and technical performance.[28] This would in turn increase the stock of available data and our understanding of fundamental issues on the origin, characteristics and impact of breakthrough innovations. With this goal in mind, we also encourage the use of the dataset constructed in this paper and the further evaluation of potential differences – both conceptual and empirical - between the breakthroughs identified with this procedure and those classified by other

[28] We provide a detailed and generalized sequence of the steps we followed in Table D.1 Appendix D.

Table B.2.
Descriptive satistics in-sample.

Variable	QAI	No. observations	Mean	SD	0	0.25	0.5	0.75	1
Renewals USPTO	1	524	1.88	1.11	0.00	1.00	2.00	3.00	3.00
	0	9,114	1.74	1.14	0.00	1.00	2.00	3.00	3.00
Forward citations (ln)	1	524	1.36	0.95	0.00	0.69	1.39	2.08	4.28
	0	9,114	1.11	0.88	0.00	0.69	1.10	1.61	4.70
Family size (ln)	1	524	1.86	0.74	0.00	1.39	1.79	2.20	3.69
	0	9,114	1.91	0.77	0.00	1.39	1.79	2.48	3.85
Claims (ln)	1	524	2.49	0.76	0.00	2.08	2.48	3.00	5.50
	0	9,114	2.30	0.73	0.00	1.95	2.30	2.77	5.00

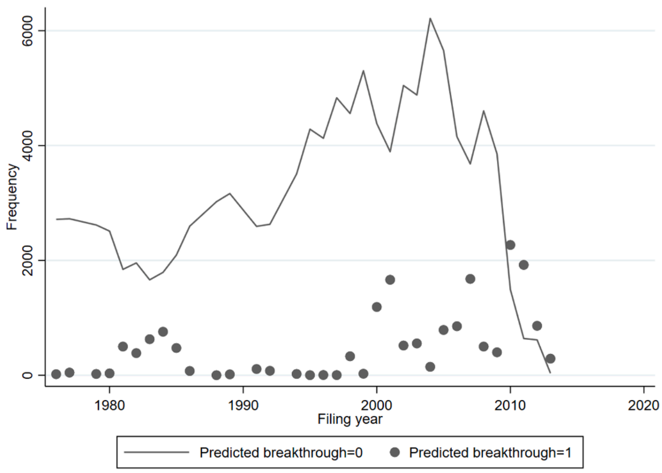

Fig. C.1. Frequency of breakthroughs and other patents over time.

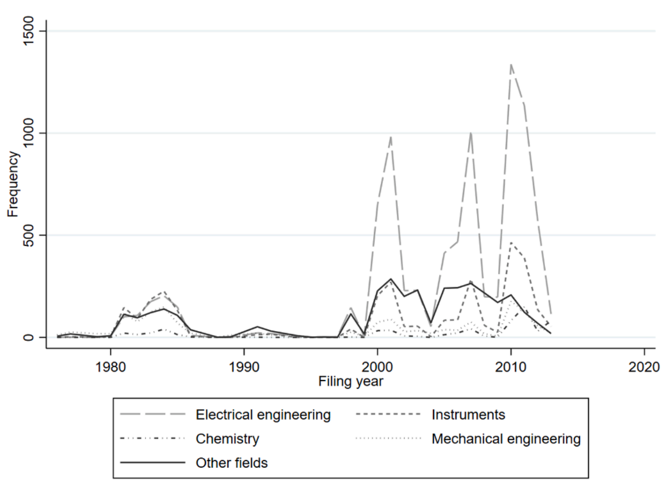

Fig. C.2. Comparing trends in sectors.

Table C.1.
Patents' concentration in technology fields.

	Predicted breakthrough =0	Predicted breakthrough =1
HHI	0.04	0.09
Equivalent number (1/ HHI)	24.24	11.40
C3	0.22	0.41

new methods (see for example Funk and Owen-Smith 2017).

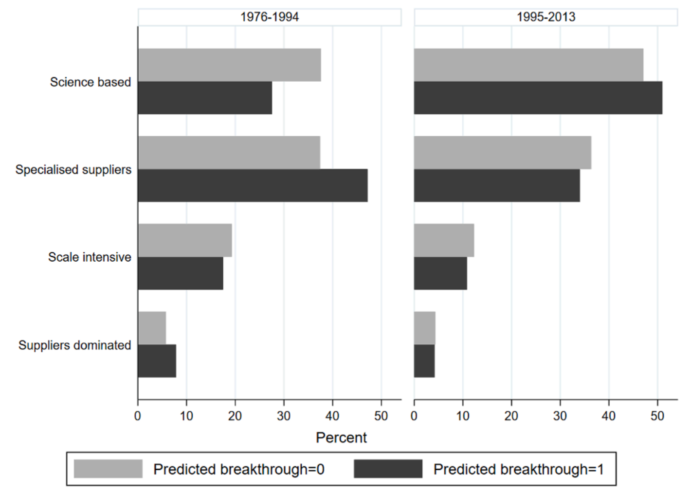

Fig. C.3. Pavitt Taxonomy.

This paper suffers from some limitations. First, while our procedure is flexible and can be applied in different contexts, it is quite labor-intensive. The search for an external measure of patent quality remains a difficult task and the association with patent data can be time-consuming. Also, the adoption of a binary classification oversimplifies the issue of patent value estimation, leading to an extreme classification whereby a patent is either valuable or not. Another word of caution stems from the fact that while we provide validation of economic and technical significance *ex-ante,* we cannot quantify nor disentangle these two aspects at the patent level. Finally, this study suffers from the common pitfall characterizing research that uses archival patent data, above all the fact that many important inventions are never patented.

CRediT authorship contribution statement

Giovanna Capponi: Methodology, Data curation, Formal analysis, Resources, Investigation, Writing – original draft, Writing – review & editing. **Arianna Martinelli:** Conceptualization, Methodology, Data curation, Validation, Writing – original draft, Writing – review & editing. **Alessandro Nuvolari:** Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table D.1.
Methodology steps.

Step	Description	Additional comments
Data preparation	(1) Identify innovations that proved to be successful	Prize schemes are one way to sample innovations that proved to be successful. Alternative strategies could rely on exhibitions, rankings, innovations appearing in specialized journals or magazines, to name a few.
	(2) Match the innovations to patents	This step is essential to translate the innovations identified in (1) into workable data.
In-sample estimates	(3) Construct the in-sample	The in-sample consists of the "treated" patents matched to (1) and the "control" patents selected within the patent portfolio of the same innovators. Depending on the sampling strategy, additional (quality) selection criteria based on the patent filing office and/or the applicant's nationality can be applied.
	(4) Compare models and select the one with the highest accuracy	We find that using multiple patent indicators improves a model's predictive power (Model 6 in Table 3). We would however recommend testing and comparing different models first.
	(5) Select a threshold to classify patents as breakthroughs	The use of Decision Curves to select an appropriate probability threshold to classify breakthroughs allows us to introduce judgment in this process. Instead of using the threshold which maximizes accuracy – identified in (4) – we explicitly try to limit the share of false positives within the breakthrough group. The calibration of the threshold probability using this criterion (see Table 4) increases the chances that a patent classified as 1 is a genuine breakthrough.
Out-of-sample estimates	(6) Construct the out-of-sample	The out-of-sample consists of the population of patents meeting the same criteria as the patent in (3) in terms of nationality of the applicant and filing office. We include all the USPTO patents filed by GB applicants.
	(7) Predict the probabilities out-of-sample using the model selected at (4)	In this paper, the best performing model in-sample is Model 6 in Table 3. We run this specification on the full sample (in-sample + out-of-sample) to obtain the estimates of the predicted probabilities on the patents out-of-sample as well.
	(8) Construct a binary variable *predicted breakthroughs* applying the threshold selected at (5)	This last step will result in the construction of a binary variable defined on a large-scale sample.

Appendix A. Patent matching

This section reports the detailed patent matching procedure, with a flowchart summary in Fig. A.1. Starting from the BvD codes we matched, we saved all the patent publication numbers assigned to a QAI winning firm within AMADEUS. We then retrieved the patents DOCDB family codes, the applications' titles and earliest filing date from PATSTAT, the EPO Worldwide Patent Statistical Database. We first defined and tested a possible patent selection process on the most recent awarded innovations, for which the Queen's Award Office reports more information. Since 2011, the descriptions often mention whether the innovation was patented or not at the time of the award, allowing us to identify an appropriate time range to look for possible matching patents. For instance, an entry taken from the 2011 report looks as follows:

Checkmate Lifting & Safety LLP, New Road, Sheerness, Kent ME12 1PZ,
Website: www.checkmateuk.com,
Employees: 41,
Managing Director: Mr O Auston,
Ultimate Parent: N/A,
Contact for Press enquiries: Mr Oliver Auston, Tel: 01,795 662,590; 07,770 395,919
(Mobile), E-mail: oa@checkmateuk.com.

An Innovation Award is made to Checkmate Lifting & Safety LLP for the design, development and successful sales of its Xcalibre Fall Arrest Blocks. The device is designed for use when working at height and arrests a worker should they fall and reduces the peak forces to the body and structure it is attached to. The company's patented technology including the SBM (Sealed Brake Module) and injection moulded one-piece drum were both inspired by the automotive industry and use the very latest aerospace materials and automated manufacturing techniques. These unique features allow the device to function in more arduous environments and reduces both the weight and manufacturing costs. In a conservative market the company has forged successful sales at home and overseas.

Acknowledging the outcome of this exploratory search and the eligibility criteria of the QAI, we considered a patent's DOCDB a possible match if the earliest filing date was at most the first year of financial figures to be submitted. For example, an innovation awarded in 1995 with two years of sales required could have been patented in 1993 at the latest, any later filing is considered an improvement. We set the lower bound ten years before the award, assuming that any prior filing would be no longer innovative enough to receive the prize. It is worth noting that for this study we matched patents which are timely associated with the innovations' launch on the market. By doing so, we expect to disentangle patents carrying technological and economic value from those filed for strategic reasons, possibly at a later stage. From this narrower pool of patent families, keeping only those whose title was in English, we manually singled out the matching cases.

As for the companies which are not in AMADEUS, or do not have patents assigned in AMADEUS, we searched directly in PATSTAT by matching the name of the Award-winning firms to the assignees. We then retrieved the patents DOCDB codes, the applications' titles and earliest filing dates using the PERSON_ID code in PATSTAT. The subsequent steps were the same as for the other group.

Appendix B. Variables and in-sample descriptive statistics

Table B.1 reports the description of the main variables used in Table 3 and the sources.

We also report the summary statistics of the patent value indicators within the in-sample separately for the control and the QAI patents (Table B.2).

Appendix C. Breakthrough innovations: a reappraisal of the UK case, 1976–2013

The identified 17,176 breakthroughs can be used to describe the innovative landscape in the UK over the past forty decades. Fig. C.1 compares the frequency of the filing years for our *predicted breakthrough* innovations vis-a-vis the other patents. The line trend depicts the increasing propensity to patent characterizing the past few decades, which is extensively covered in the literature. In contrast, the dotted path shows that breakthrough patents emerged mainly in the 1980s and the 2000s. While they present an increasing trend as well, they show a tendency to cluster in time.

Focusing on breakthrough innovations alone, Fig. C.2 shows the absolute patent frequency by sectors over time.[29] Patents in electrical engineering are by far the most frequent from 2000 to 2010. At a more disaggregated level, we observe that the sub-classes of computer technology, telecommunications and digital innovation prevail. In the most recent years, we find a relatively high share of breakthroughs in instruments (mainly measurement and optics). Also, a growing number of breakthrough patents are filed in what Schmoch (2009) categorized as “other fields”, driven chiefly by innovations in civil engineering. The least represented classes are mechanical engineering, which in the graph is relatively more present in the 1980s, and chemistry. Interestingly, the patents that we do not classify as breakthroughs do not show the same tendency to cluster in time.

Table C.1 reports the scores on three widely adopted concentration indexes calculated on 35 technology fields.[30] Breakthrough patents exhibit a higher level of sectoral concentration than the rest of the sample, with the top three fields covering 41% of the patents.

Finally, we classify the innovations following the Pavitt Taxonomy, a classification that describes the behavior of innovating firms depending on their technological competencies (Pavitt, 1984).[31] The original taxonomy presented four main categories: supplier dominated firms which operate in traditional industries (ex. textile), specialized suppliers of equipment who work in close collaboration with their customers, science-based firms whose main source of knowledge is in-house R&D, and scale intensive firms operating in mass production industries (Archibugi, 2001). It is worth highlighting that Pavitt’s seminal work is based on evidence emerging from the SPRU innovation database, which also includes innovations winning a Queen’s Award. Hence, it seems to be particularly appropriate to analyze our sample using the Pavitt Taxonomy. Fig. C.3 shows the patents’ distribution across the five classes, comparing the composition of breakthroughs and other patents over two time periods.

Between 1976 and 1994, breakthrough innovations are relatively more frequent in specialized suppliers sectors. Pavitt recognizes core specialized suppliers in sectors like machinery and instruments, where innovation is driven by performance-sensitive clients valuing product design (Pavitt, 1984). This evidence is consistent with the significant proportion of successful innovative activities in the complex product systems (CoPS) branches of UK manufacturing during the 1980s and early 1990s (Hobday and Rush, 1999).

In the second period, between 1995 and 2013, we notice the role of science-based sectors as generators of breakthroughs. At a more disaggregated level, it emerges that breakthroughs are driven by innovative activities in the field of electronic engineering while the rest of the sample presents a higher share of patents in chemistry. Arguably, this trend captures the heavy digitalization process we witnessed in the past two decades, which shifted the locus of breakthrough innovations consolidating information and communications technology (ICT) as a new techno-economic paradigm (Freeman and Louçã, 2001).

Of course, this is just a set of exploratory stylized facts emerging from our new dataset of UK breakthrough innovations that could be further analyzed in future work.

Appendix D. Methodology steps

See Table D.1

References

Ahuja, G., Lampert, C.M., 2001. Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. Strateg. Manag. J. 22, 521–543. https://doi.org/10.1002/smj.176.

Albert, M.B., Avery, D., Narin, F., McAllister, P., 1991. Direct validation of citation counts as indicators of industrially important patents. Res. Policy 20, 251–259. https://doi.org/10.1016/0048-7333(91)90055-U.

Andersen, B., 1999. The hunt for S-shaped growth paths in technological innovation: a patent study. J. Evol. Econ. 9, 487–526. https://doi.org/10.1007/s001910050093.

Archibugi, D., 2001. Pavitt'S taxonomy sixteen years on: a review article. Econ. Innov. New Technol. https://doi.org/10.1080/10438590100000016.

Arts, S., Hou, J., Gomez, J.C., 2019. Text mining to measure novelty and diffusion of technological innovation. In: Proceedings of the 17th International Conference on Scientometrics and Informetrics ISSI, pp. 1798–1800. https://doi.org/10.34726/pst2019.2.

Bessen, J., 2008. The value of U.S. patents by owner and patent characteristics. Res. Policy 37, 932–945. https://doi.org/10.1016/j.respol.2008.02.005.

Bogliacino, F., Pianta, M., 2010. Innovation and employment: a reinvestigation using revised pavitt classes. Res. Policy 39, 799–809. https://doi.org/10.1016/j.respol.2010.02.017.

Brunt, L., Lerner, J., Nicholas, T., 2012. Inducement prizes and innovation. J. Ind. Econ. 60, 657–696. https://doi.org/10.1111/joie.12002.

Cameron, A.C., Miller, D.L., 2015. A Practitioner 's guide to cluster-robust inference. J. Hum. Resour. 50, 317–372.

Capponi, G., Martinelli, A., Nuvolari, A., 2021. Breakthrough innovations and where to find them. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E149201V2.

Carpenter, M.P., Narin, F., Woolf, P., 1981. Citation rates to technologically important patents. World Pat. Inf. 3, 160–163. https://doi.org/10.1016/0172-2190(81)90098-3.

Conway, S., 1995. Informal boundary-spanning communication in the innovation process: an empirical study. Technol. Anal. Strateg. Manag. 7, 327–342. https://doi.org/10.1080/09537329508524216.

Dahlin, K., Taylor, M., Fichman, M., 2004. Today’s Edisons or weekend hobbyists: technical merit and success of inventions by independent inventors. Res. Policy 33, 1167–1183. https://doi.org/10.1016/j.respol.2004.06.003.

Dahlin, K.B., Behrens, D.M., 2005. When is an invention really radical?: defining and measuring technological radicalness. Res. Policy 34, 717–737. https://doi.org/10.1016/j.respol.2005.03.009.

de Rassenfosse, G., Jaffe, A.B., 2018. Are patent fees effective at weeding out low-quality patents? J. Econ. Manag. Strateg. 27, 134–148. https://doi.org/10.1111/jems.12219.

Department for Business Energy and Industrial Strategy, 2013. The Queen’s Awards for Enterprise: general background notes for business awards. UK Government, London.

Department of Trade and Industry, 1999. The Queen’s Awards for Enterprise: report of the 1999 review of the Queen’s Awards undertaken by a Committee under the chairmanship of His Royal Highness the Prince of Wales. UK Government, London.

Dosi, G., 1982. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Res. Policy 11, 147–162.

Fleming, L., 2007. Breakthroughs and the “long tail” of innovation. MIT Sloan Manag. Rev. 49.

Fleming, L., 2001. Recombinant uncertainty in technological search. Manage. Sci. 47, 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671.

Freeman, C., Louçã, F., 2001. As Time Goes by: from the Industrial Revolutions to the Information Revolution. Oxford University Press.

[29] As defined by Schmoch (2009)

[30] To correct for the upward bias in concentration resulting from the count nature of the data, we adopt the adjusted version of the HHI index proposed by Hall (2005). The unbiased index is as follows: $\widehat{\eta} = (NxHHI - 1)/(N - 1)$ (4)The HHI index ranges from a minimum concentration value of 1/35=0.03 to a maximum score of 1.We also include the equivalent indicator (1/HHI) which provides the number of equal size industries corresponding to the HHI score. The C3 concentration ratio is calculated as the sum of the shares of the three classes with the highest frequency of patents.

[31] In this exercise we first matched NACE Rev. 2 sector codes to patent applications using PATSTAT and sum up the weights of each NACE classification for the two patent groups. Then, we used the correspondence table elaborated in Bogliacino and Pianta, (2010) to link the sector codes to the taxonomy’s categories.

Funk, R.J., Owen-Smith, J., 2017. A dynamic network measure of technological change. Manage. Sci. 63, 791–817. https://doi.org/10.1287/mnsc.2015.2366.
Gambardella, A., Giuri, P., Luzzi, A., 2007. The market for patents in. Europe. Res. Policy 36, 1163–1183. https://doi.org/10.1016/j.respol.2007.07.006.
Gambardella, A., Harhoff, D., Verspagen, B., 2008. The value of European patents. Eur. Manag. Rev. 5, 69–84. https://doi.org/10.1057/emr.2008.10.
Garcia, R., Calantone, R., 2002. A critical look at technological innovation typology and innovativeness terminology: a literature review. J. Prod. Innov. Manag. 19, 110–132. https://doi.org/10.1111/1540-5885.1920110.
Giuri, P., Mariani, M., 2007. Inventors and invention processes in Europe. Results from the PatVal-EU survey. Res. Policy. https://doi.org/10.1016/j.respol.2007.07.007.
Griliches, Z., 1981. Market value, R&D, and patents. Econ. Lett. 7, 183–187. https://doi.org/10.1016/0165-1765(87)90114-5.
Groom, B., 2015. Half a century of the Queen's Awards. Financ. Times.
Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp. J. Intern. Med.
Hall, B., 1993. The stock market's valuation of R&D investment during the 1980's. Am. Econ. Rev. 83, 259–264. https://doi.org/10.2307/2117674.
Hall, B.H., 2005. A note on the bias in herfindahl-type measures based on count data. Rev. d'économie Ind. 110, 149–156. https://doi.org/10.3406/rei.2005.3076.
Hall, B.H., Thoma, G., Torrisi, S., 2007. The market value of patents and R&D: evidence from european firms. In: Proceedings of the Academy of Management Annual Meeting: Doing Well by Doing Good, AOM Academy of Management. https://doi.org/10.5465/ambpp.2007.26530853.
Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Rev. Econ. Stat. 81, 511–515. https://doi.org/10.1162/003465399558265.
Harhoff, D., Reitzig, M., 2004. Determinants of opposition against EPO patent grants-the case of biotechnology and pharmaceuticals. Int. J. Ind. Organ. 22, 443–480. https://doi.org/10.1016/j.ijindorg.2004.01.001.
Harhoff, D., Scherer, F.M., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Res. Policy 32, 1343–1363. https://doi.org/10.1016/S0048-7333(02)00124-5.
Higham, K., De Rassenfosse, G., Jaffe, A., 2020. Patent Quality: Towards a Systematic Framework for Analysis and Measurement. NBER Work, Pap.
Hobday, M., Rush, H., 1999. Technology management in complex product systems (CoPS)-ten questions answered. Int. J. Technol. Manag. 17, 618–638. https://doi.org/10.1504/IJTM.1999.002739.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132, 665–712. https://doi.org/10.1093/qje/qjw040.
Langrish, J., Gibbons, M., Evans, W.G., Jevons, F.R., 1972. *Wealth from knowledge: Studies of innovation in industry*. Springer.
Lanjouw, J.O., Pakes, A., Putnam, J., 1998. How to count patents and value intellectual property: the uses of patent renewal and application data. J. Ind. Econ. 46, 405–432. https://doi.org/10.1111/1467-6451.00081.
Lanjouw, J.O., Schankerman, M., 2001. Characteristics of Patent Litigation: a Window on Competition. RAND J. Econ. 32, 129. https://doi.org/10.2307/2696401.
Martínez, C., 2011. Patent families: When do different definitions really matter? Scientometrics 86 (1), 39–63. https://doi.org/10.1007/s11192-010-0251-3.
McFadzean, W., 1970. The Queen's Award to Industry. Report of the 1970 Review Committee Under the Chairmanship of the Right Hon. The Lord Mcfadzean, London.
Moser, P., 2011. Innovation without patents: evidence from world's fairs. J. Law Econ. 55, 43–74. https://doi.org/10.1086/663631.
Moser, P., 2005. How do patent laws influence innovation? Evidence from nineteenth-century world's fairs. Am. Econ. Rev. https://doi.org/10.1257/0002828054825501.
Moser, P., 2002. The determinants of innovation: new evidence from nineteenth-century world fairs. University of California, Berkeley.
Mountbatten-Windsor, P., 1975. The Queen's Award to Industry. Report by a committee under the chairmanship of His Royal Highness the Duke of Edinburgh. Her Majesty Startionary Office, London.
Mountbatten-Windsor, P., 1965. The Queen's Award to Industry. Report by a committee under the chairmanship of His Royal Highness the Duke of Edinburgh. Her Majesty Stationary Office, London.
Pakes, A., 1986. Patents as options: some estimates of the value of holding european patent stocks. Econometrica 54, 755. https://doi.org/10.2307/1912835.
OECD, 2009. OECD Patent Statistics Manual. OECD Publishing, Paris.
Pakes, A., 1985. On patents, R & D, and the stock market rate of return. J. Polit. Econ. 93, 390–409. https://doi.org/10.1086/261305.
Park, S.H., Goo, J.M., Jo, C.H., 2004. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J. Radiol. https://doi.org/10.3348/kjr.2004.5.1.11.
Pavitt, K., 1984. Sectoral patterns of technical change: towards a taxonomy and a theory. Res. Policy 13, 343–373. https://doi.org/10.1016/0048-7333(84)90018-0.
Pavitt, K., 1983. Characteristics of innovative activities in British industry. Omega (Westport) 11, 113–130. https://doi.org/10.1016/0305-0483(83)90001-4.
Peirce, C.S., 1884. The numerical measure of the success of predictions. Science (80). https://doi.org/10.1126/science.ns-4.93.453-a.
Romijn, H., Albaladejo, M., 2002. Determinants of innovation capability in small electronics and software firms in southeast England. Res. Policy 31, 1053–1067. https://doi.org/10.1016/S0048-7333(01)00176-7.
Sampat, B.N., Ziedonis, A.A., 2004. Patent citations and the economic value of patents. Handbook of Quantitative Science and Technology Research. Springer, Netherlands, pp. 277–298. https://doi.org/10.1007/1-4020-2755-9_13.
Schmoch, U., 2009. Document IPC/CE/41/5, Annex, Conception of a Technology Classification for Country Comparisons. WIPO, 41th session of the IPC Committee of Experts.
Simmie, J., 1998. Reasons for the development of "islands of innovation": evidence from Hertfordshire. Urban Stud. 35, 1261–1289. https://doi.org/10.1080/0042098984358.
Singh, J., Fleming, L., 2010. Lone inventors as sources of breakthroughs: myth or reality? Manage. Sci. 56, 41–56. https://doi.org/10.1287/mnsc.1090.1072.
Squicciarini, M., Dernis, H., Criscuolo, C., 2013. Measuring patent quality: indicators of technological and economic value. OECD Sci. Technol. Ind. Work. Pap. 70 https://doi.org/10.1787/5k4522wkw1r8-en.
Tether, B.S., 1998. Small and large firms: sources of unequal innovations? Res. Policy 27, 725–745. https://doi.org/10.1016/S0048-7333(98)00079-1.
Tether, B.S., 1996. Virtual Panacea & Actual reality : Small firms, Innovation & Employment creation; Evidence from Britain during the 1980s. University of Sussex.
Tong, X., Frame, J.D., 1994. Measuring national technological performance with patent claims data. Res. Policy 23, 133–141. https://doi.org/10.1016/0048-7333(94)90050-7.
Trajtenberg, M., 1990. A Penny for Your Quotes: patent Citations and the Value of Innovations. RAND J. Econ. 21, 172. https://doi.org/10.2307/2555502.
USPTO, 1998. Ideas that Become Valuable Innovations: a Patent and Trademark Office Review. Washington DC.
Utterback, J., 1994. Mastering the dynamics of innovation: How companies can seize opportunities in the face of technological change. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.
Van Zeebroeck, N., 2011. Economics of Innovation and New Technology The puzzle of patent value indicators The puzzle of patent value indicators. Econ. Innov. New Technol. 20, 33–62. https://doi.org/10.1080/10438590903038256.
Verhoeven, D., Bakker, J., Veugelers, R., 2016. Measuring technological novelty with patent-based indicators. Res. Policy 45, 707–723. https://doi.org/10.1016/j.respol.2015.11.010.
Vickers, A.J., Elkin, E.B., 2006. Decision curve analysis: a novel method for evaluating prediction models. Med. Decis. Mak. 26, 565–574. https://doi.org/10.1177/0272989X06295361.
Webster, E., Jensen, P.H., 2011. Do patents matter for commercialization? J. Law Econ. 54, 431–453. https://doi.org/10.1086/658487.
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35, 10.1002/1097-0142(1950)3:1<32::AID−CNCR2820030106>3.0.CO;2-3.