



OECD Science, Technology and Industry Working Papers
2021/07

Quantitative indicators
for high-risk/high-reward
research

Diogo Machado

<https://dx.doi.org/10.1787/675cbef6-en>

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to the Directorate of Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/STP/GSF/NESTI(2021)1/FINAL

© OECD 2021

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

Foreword

This paper describes the key characteristics of “high-risk/high-reward” research (HRHR) along with examples of quantitative indicators and scientific discoveries representing them. A key element of the paper is the computation and analysis of a novelty indicator at the scientific article and country level. This indicator can help policymakers and the research funding community to benchmark and monitor countries’ performance in supporting the creation of knowledge with HRHR characteristics. The indicator was computed using SCOPUS Custom Data, Elsevier, Version 5, 2019.

The project that forms the basis for this paper was initiated by the OECD Global Science Forum (GSF) in 2019. It culminated in the OECD report “Effective Policies to foster High-Risk/High-Reward research”, which analyses research funding mechanisms designed to foster HRHR research. The preliminary results in this working paper were presented in the GSF international workshop organised in April 2020 — one of the activities of the main project —that explored the major issues and challenges faced by funders in supporting HRHR research.

The author of the paper is Diogo Machado, acting as a consultant to the OECD. During the course of this work, the author acted as Senior Economist at Technopolis-Group, a policy consultancy firm. The start of the work predates the author’s affiliation with Technopolis-Group and the paper is independent of his role with Technopolis. The author wishes to thank OECD colleagues for all the comments and suggestions, especially Brigitte Van Beuzekom, Carthage Smith, Fernando Galindo-Rueda and Frédéric Sgard, as well as Kei Koizumi, the members of the GSF international Expert Group and the participants in the virtual workshop on “Effective Policies to foster High-Risk/High-Reward research”.

Table of contents

Foreword.....	3
Executive summary.....	7
1. Introduction	9
2. Rationale for high risk/high reward research funding: the sources of “risk” and “reward”	11
Knowledge spillovers	12
Long temporal dimension	12
Uncertainty.....	13
Characteristics of HRHR research outputs	13
Basicness	13
Generality	14
Extraordinary levels of novelty	16
3. Aggregate implementation of a combination-based novelty indicator	18
Data sources and computation details	19
Data exploration and visualisation.....	20
Article-level analysis	20
Country-level analysis.....	25
Year-level analysis.....	29
Cross-country/year econometric analysis.....	30
4. Conclusion.....	33
Endnotes.....	35
References.....	3
Annex A.	6
Annex B.	7

Tables

Table 1. Generalised poisson regression with total citations as dependent variable	25
Table 2. Regression table: Share of articles among the top 10% top novelty	31

Table A B.1. Mean Absolute Error Test	12
---------------------------------------	----

Figures

Figure 1. HRHR knowledge characteristics and associate market failures	11
Figure 2. Distribution of the novelty indicator	21
Figure 3. Boxplots of citation performance per novelty category for 2011	22
Figure 4. Novelty indicator vs interdisciplinarity for 2011	23

Figure 5. Generalised poisson distribution fit of forward citations for 2011	24
Figure 6. Share of publications among the top 10% most novel, 2005-17 averages, OECD countries plus China	26
Figure 7. Share of publications among the top 10% most novel vs top 10% most cited, 2007-17 averages, OECD countries plus China	27
Figure 8. Share of publications among the top 10% most novel vs share of publications part of international collaboration, 2005-17 average, OECD countries plus China	28
Figure 9. Average novelty per year, OECD average	29
Figure 10. Average novelty vs HERD growth rate per year, OECD average	30
Figure A B.1. Poisson distribution fit of forward citations for 2011	7
Figure A B.2. Zero inflated poisson distribution fit of forward citations for 2011	8
Figure A B.3. Zero inflated gen. poisson distribution fit of forward citations for 2011	9
Figure A B.4. Negative binomial distribution fit of forward citations for 2011	10
Figure A B.5. Zero inflated negative binomial distribution fit of forward citations for 2011	11

Boxes

Box 1. Examples of quantitative indicators measuring basic research outcomes	14
Box 2. Examples of quantitative indicators measuring scientific generality	15
Box 3. Examples of quantitative indicators measuring the level of novelty	17

This paper describes the key characteristics of high-risk/high-reward research (HRHR), which has gained considerable interest from policy makers as a way to promote the development of new, ‘out-of-the-box’ ideas. It identifies three dimensions that are accentuated in HRHR research: higher levels of basicness, generality and novelty. These knowledge characteristics are commonly associated with market failure and research that requires public investment because it has large spill-overs, long time horizons and high levels of uncertainty. This is illustrated with examples of specific discoveries embedding each knowledge characteristic and the application of appropriate quantitative measures. The paper concludes with the computation and demonstration of an indicator of novelty that may be particularly well suited for the monitoring and evaluation of HRHR research policies.

Executive summary

Funding highly novel scientific research at the knowledge frontier is a risky business. This “high-risk/high-reward” (HRHR) element signals the fact that many research projects of this nature will fail to produce ground-breaking results, while a few are likely to generate the most impactful scientific breakthrough discoveries.

The distinct nature of HRHR research and its resulting knowledge characteristics call for complementary metrics to support science funders to better measure and manage risk. Metrics, such as Journal Impact Factors (JIFs) and H-indexes, are routinely used to help select, monitor and evaluate funding programmes and resulting research outputs. However, excessive reliance solely on these traditional metrics has been shown to be a potential source of risk-aversion in science funding.

The identified knowledge characteristics, which are particularly accentuated in HRHR research, are the higher levels of:

- Basicness: experimental or theoretical discovery resulting from research about the foundations of phenomena and observable facts, without any concrete application or use in view.
- Generality: a general discovery is applicable to a wide number of scientific fields.
- Novelty: an extremely novel discovery is highly different compared to the state of the art, representing a potential leap forward.

This paper examines multiple examples of metrics representing these knowledge characteristics and presents an indicator of scientific novelty. Scientific novelty is a particularly relevant dimension for HRHR research, with high-risk being represented by researchers’ attempts to combine untested and distant knowledge fields.

The computed novelty indicator reported in this paper reveals that novelty in scientific articles is highly skewed. Most articles present no or very low levels of risk-taking, while just a very small fraction of articles are associated with high levels of risk. Higher levels of novelty are also associated with a higher citation variance, meaning that the highly novel category contains many articles that are never cited, or receive a very small number of citations, but also the most highly cited articles in the world. Moreover, highly novel articles are shown to have, on average, significantly higher levels of citations, but this superiority is stronger in the long-term whilst less significant in the short-term.

At a country level, the top performing OECD countries in terms of novel science, as detected with the indicator, are the Netherlands, Switzerland, Denmark and the United States. These countries also have specific HRHR policies in place and are the ones with the highest share of top cited scientific publications. Countries with larger shares of international collaborations and higher R&D spending also appear to be associated with higher levels of novelty. However, in a more robust statistical analysis, the share of international collaborations and R&D spending no longer present statistically significant effects, while

scientific impact in terms of citations, seems to be consistently strongly positively related with novelty at 1% statistical significance level. The results from the most robust econometric model also indicate that higher level of business R&D expenditures are also related with higher levels of scientific novelty.

Fostering risk-taking is a systemic issue for the scientific community, encompassing multiple actors and processes. These range from science agencies designing funding programmes, scientists proposing new research ideas and the scientists reviewing them, and all the processes within the scholarly publishing ecosystem. This paper presents examples of indicators that can support members of the scientific community to better measure and manage risk-taking. Given the high complexity of HRHR research, relying excessively on traditional indicators is too limiting. In this context, complementing the use of traditional bibliometric indicators with new ones provides important insights for monitoring and evaluation activities of HRHR science policymaking.

1. Introduction

Lack of risk-taking in science and technology research dampens the capacity to develop novel scientific breakthroughs. Breakthrough discoveries are essential to boost productivity and tackle the grand societal challenges we are facing, such as climate change, ageing or global pandemics. The type of research underlying breakthrough discoveries has a higher potential to have a major impact but is also subject to higher uncertainty. Moreover, breakthrough discoveries can take longer to have an impact, either due to scientific/technological prematurity or because they require major system changes, which prompt resistance from actors operating within incumbent paradigms. The “high-risk/high-reward” (HRHR) nature of novel scientific research makes it particularly appropriate for public support.

Some members of the scientific community worry that science funding agencies are becoming increasingly risk-averse. As expressed by a Nobel Laureate in Chemistry Roger Kornberg “(today) if the work that you propose to do is not virtually certain of success, then it will not be funded” (Fang & Casadevall, 2016). According to James Rothman, a Nobel Laureate in Physiology or Medicine, “I had five years of failure, really, before I had the first initial sign of success. And I would like to think that that kind of support existed today, but I think there is less of it. And it is actually becoming a pressing national issue, if not an international issue” (Harris, 2013). Evidence of the potential underinvestment by the public sector in “high-risk/high-gain” research in the scientific peer-review system includes analyses by Boudreau et al. (2016) and Wang, Veugelers and Stephan (2017). Making matters worse, the private sector is also showing less appetite for risky science-based R&D, influenced by higher short-term pressure from shareholders and incentives to “outsource” research to public research organisations. Evidence about the decreasing appetite for higher risk science-based R&D in the private sector include findings by Budish et al. (2015) and Arora et al. (2018).

Bibliometric indicators are useful to evaluate a large number of scientific research proposals or articles. However, relying excessively on traditional bibliometric indicators for research evaluation has been shown to be a source of risk-aversion. Indicators such as Journal Impact Factors (JIF) or H-indexes are based on ex-post evaluations of peer “citation impact” [see e.g. Ahuja and Morris Lampert (2001); Leifer et al. (2001); Schoenmakers and Duysters (2010)], which fail to capture (ex-ante) scientific and technological potential of research projects. Citation-based indicators require a long-time horizon, because breakthroughs require a long time to be recognised as such.

Bibliometrics based on JIFs were developed to assess journals, not individual articles, and H-indexes measure past performance, not the quality of research proposals or recently published articles. The forward-looking element of these indicators aim to act as a quality filter — if a scholar succeeds to publish in a journal with a high impact factor or has a high H-index, that can say something about the expected citation “worth” of her work. However, these indicators, if used on a short-term basis, can be biased against risky and highly novel work, as shown by Wang, Veugelers and Stephan (2017). This is one reason why numerous funding organisations, research institutes and individuals signed the San Francisco Declaration

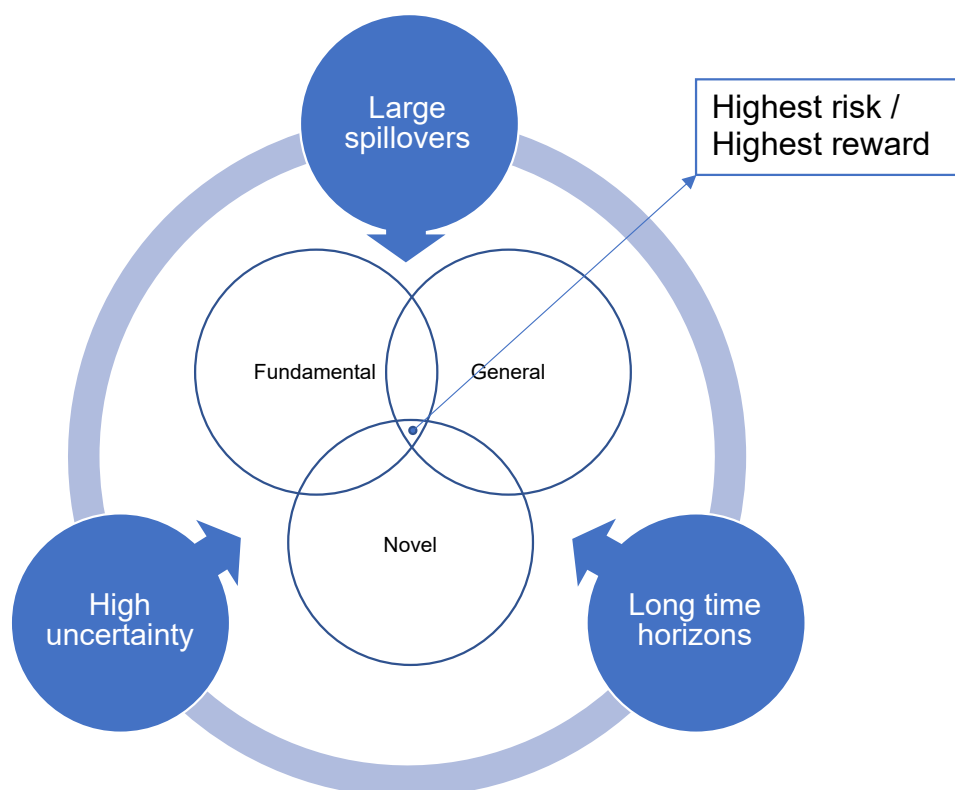
on Research Assessment (DORA) , committing “not to use journal-based metrics, such as JIF, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions”. Similar principles are also proposed in the Leiden Manifesto for Research Metrics. Despite this bias, reviewers regularly justify panel evaluations on the basis of the journals where applicants publish. Scientific panels evaluate proposals based on applicants’ Google Scholar metrics using past citation counts. Some funding agencies directly ask applicants to list JIFs or H-indexes alongside their publications. Even in cases where candidates are not required to report such indicators, they often still report them selectively, otherwise reviewers retrieve such indicators themselves. This practice prevails even at agencies that instruct reviewers not to use bibliometric measures (Stephan, Veugelers, & Wang, 2017). In response, some funding agencies have interpreted the DORA as a call not to use quantitative bibliometric indicators at all. In some cases, funding agencies are refusing to use indicators such as H-indexes, even in the evaluation of research programmes that have been in place for several years.

The objective of this paper is to provide an overview of quantitative indicators that are more directly relevant for assessing HRHR research. These indicators can be used to complement, not to substitute, citation-based indicators such as JIFs and H-indexes. The proposed indicators reflect the ex-ante potential of publications, their level of risk, or measure characteristics that go beyond purely citation-based impact. Following this brief introduction, the rest of the paper is organised as follows: section 2. Reviews the market failures that constrain HRHR research - justifying public intervention - and systematically describes the main characteristics of the outputs resulting from HRHR. These elements are illustrated with specific scientific discoveries, and examples of indicators used in the literature of Economics of Science and Innovation based on bibliometrics and patent data. Section 3. provides details about the computation of an indicator of scientific novelty of particular relevance for HRHR research. The indicator is computed for all the articles in journals indexed in the SCOPUS database (reviews and conference proceedings were not included) for the 2005-2017 period and aggregated at a country level. Finally, the novelty indicator is analysed in relation to other relevant policy variables. Section 4. concludes with key takeaway messages.

2. Rationale for high risk/high reward research funding: the sources of “risk” and “reward”

High risk/high reward (HRHR) research has specific characteristics that provide justification for public support. These characteristics are common rationales for the public funding of science and innovation activities of different natures, but the combination and degree of intensity of these characteristics are particularly accentuated in HRHR research. The central argument aligns with classic market-failure reasoning: in the absence of public support, activities liable to market-failures are underprovided because businesses are not well equipped or motivated to invest. Markets do not provide sufficient incentives for investment in scientific research, especially so for scientific research of HRHR nature. The market-failure characteristics of scientific research can include large spillovers, long temporal horizons and high uncertainty. These characteristics are further accentuated in scientific research of HRHR nature.

Figure 1. HRHR knowledge characteristics and associate market failures



Knowledge spillovers

Knowledge spillovers accrue from knowledge being non-rivalrous and only partially non-excludable. After it has been created and shared, others may learn and use knowledge without necessarily paying for it, potentially undermining researchers' incentives to create it in first place (Arrow, 1962). Moreover, knowledge is not depleted when shared, so its use by one individual does not prevent its use by others. The incremental cost of an additional user is virtually zero and not only is the stock of knowledge not diminished by extensive use, but it is also often enlarged (Stephan P. , 1996). Knowledge is also an experience good, as it is often difficult for putative users to properly value a new idea. If researchers reveal enough of the knowledge for a proper valuation, they may reveal the idea itself and further undercut users' incentives to pay for it.

The “reward” dimension of HRHR research is closely linked with the concept of high spillovers. HRHR knowledge should be of higher usage and impact than conventional research, leading, for example, to follow-up discoveries. The larger the spillovers the more difficult it is for a limited group of actors to appropriate all the benefits from investments in research. The absence of public support can result in an underinvestment in research with high spillovers or even no investment at all. The invention of the BLAST is a good example of how knowledge spillovers operate. BLAST (basic local alignment search tool) is an algorithm for comparing primary biological information such as DNA sequences. It was developed by Stephen Altschul and colleagues with two grants from the US National Institutes of Health (NIH) and published in the *Journal of Molecular Biology*. It became a scientific breakthrough — the article describing the invention was published in 1990 and up to 2020 counted more than 68 000 forward citations. Even more interesting with regard to the concept of knowledge spillovers is its commercial application. It is one of the scientific publications with the largest number of citations in patents. In 2020, the article already counted more than 7 000 citations in patent applications from multiple and competing organisations, such as Du Pont, Novartis, or Henkel, who can use the knowledge contained in the article published by Stephen Altschul and colleagues for free. Additionally, these citing patents are not concentrated in a single field of application. The BLAST paper is cited by patents from 8 different Cooperative Patent Classifications (CPC) sections, ranging from Human necessities and Chemistry, to even Physics and Electricity. No single firm, which typically specialise in a specific field, would have been able to develop BLAST and apply it in such diverse set of domains.

Long temporal dimension

Research of HRHR nature typically requires a longer time horizon to achieve results and for these to become widely recognised. Conducting high-risk research with potential high impact in the short-term is not impossible — some of the scientific efforts to find a novel COVID19 vaccine reflect an attempt to do so. However, most often, scientific breakthroughs take years, if not decades, before suddenly emerging as an “overnight success”. This is partly because the research process for HRHR is typically long and because breakthrough results stemming from HRHR research may be confronted with resistance/incomprehension if they threaten accepted paradigms. Findings may also be too early for their time, providing new solutions for problems that have yet to come. An example of discovery with long temporal dimension is the Cryo electron microscopy. As stated by one of the developers “the practical use is immense but there is always a long time between the results of fundamental research to make their way into general knowledge and the practice of medicine.” (Chang, 2017). The research leading to its development started in the 1970s and the three authors were awarded the Nobel prize in Chemistry in 2017. An example of its impact has been the replacement of X-ray technology for determining protein structure in infectious agents, such as the Zika virus. Another topical example of discovery with a long temporal dimension is the method of “backpropagation” - a way to optimise neural networks that is now

broadly used in Artificial Intelligence (AI). It was invented in 1986 by David Rumelhart and colleagues (based on previous fundamental research), but its potential and use was not fully recognised before the late 2000s as it requires computer power that was not available at affordable cost before then. Until the rise of AI applications, “backpropagation” was a solution looking for a problem. During the 90’s and early 2000’s its annual citation rates remained stable, but since 2010 citations have massively increased. In 2019 alone this paper, published in 1986, received more than 600 citations.

Uncertainty

High levels of uncertainty about the final success of research activities is a major obstacle to HRHR research funding. With risk-averse agents, the larger the uncertainty the lower are the incentives to invest in HRHR research. Without bold public funding there would almost certainly be an underinvestment in research with high-risk. An example of long-term research infrastructure investment project with high uncertainty is the Laser Interferometer Gravitational-Wave Observatory (LIGO), which detected four gravitational waves on September 14, 2015, validating Einstein’s prediction from 1916. The project started around 1975 and took 40 years to achieve results. Rainer Weiss, one of the creators of LIGO, affirmed that when they first proposed the project for funding “everybody thought we were out of our minds” (Overbie Dennis, 2017), describing the high uncertainty associated with the proposal. This type of HRHR research could only be financed publicly, with the National Science Foundation (NSF) in the US having invested USD 1 billion in LIGO – the largest project ever funded by the NSF. The discovery led to a Nobel Prize and opened up a whole new field of astro- physics.

Characteristics of HRHR research outputs

The research output that is typically associated with HRHR scientific research is knowledge that is more basic, generalisable and novel. Each of these elements are detailed in the current section. A definition is provided along with specific examples of discoveries containing each element. Additionally, different quantitative measurements for each element are reviewed based on the literature of Economics of Science and Innovation.

Basicness

In its purest form, basic research consists of describing properties, structures and relationships to formulate and test hypotheses, theories or laws. In the Frascati Manual, basic research is defined as experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view (OECD, 2015). Max Planck’s or Albert Einstein’s contributions to quantum mechanics in 1900 and 1905 respectively, are examples of knowledge outcomes that expanded further the fundamental theory of nature. It took several decades, and large investments in applied and experimental follow-up research, to start generating direct applications of economic value that rely on the discovery of quantum mechanics, such as the transistor used for personal computers or lasers used for CT scans (Kakalios, 2010).

Basic research tends to have higher levels of risk. It is associated with long temporal dimensions, high levels of technological uncertainty and information asymmetries between scientists and funders. The more basic a given research process is, the more accentuated are these elements. Firms have limited incentives to invest in research that is far from commercial applications, intensifying the importance of public funding of basic research. For example, firms are reluctant to invest in early-stage research outputs because competing businesses can easily benefit from the knowledge spillovers and produce competing follow-up inventions.

Box 1. Examples of quantitative indicators measuring basic research outcomes

Text-mining heuristics identifying words related with basic or applied research

- **Definition:** Use of specific words or topics that are more common in basic research outputs. For example, in health-related research, words associated with basic biological phenomena are conceptually further away from concrete application and commercialisation in comparison with words related with the use of humans in clinical trials. Therefore, words can be classified in terms of the degree to which they are more related with applied research (e.g. words depicting tools for experiments, animals or clinical trials), or basic research (e.g. biological phenomena and theories). The number and frequency of these words can indicate whether an outcome is more likely to stem from basic or applied research.
- **Rationale:** The type of text used in knowledge outcomes of basic research can often be different from the type of text from more applied research. Part of the difference can stem from distinct stylistic ways of writing between research communities, but also from different types of topics. Text-mining tools can be used to identify and disentangle such differences.
- **Examples of use in the literature:** Papers using this measure include Li, Azoulay and Sampat (The applied value of public investments in biomedical research, 2017) who invented it.

Journal classifications depending on the level of commercial applicability

- **Definition:** The CHI journal classification in the Science Citation Index database of Clarivate Analytics, assigns journals to one of four research levels, from very applied, targeted research to basic research.
- **Rationale:** Taxonomy build by the Science Citation Index database of Clarivate Analytics identifying levels basicness/applicability.
- **Examples of use in the literature:** Malva et al (Basic science as a prescription for breakthrough inventions in the pharmaceutical industry, 2015) used the CHI journal classification to identify publications on biomedical basic research.

Generality

A general discovery is applicable to a wide number of scientific fields. An example of general discovery is Maxwell's equations, which formulate how electric and magnetic fields propagate (Nelson, 1959). Its generality is reflected by the broad scope for application in diverse areas, ranging from optics, to planetary motion or electromagnetics (Lowther & Freeman, 2008). General discoveries are often a result of basic research, so these elements co-occur multiple times. But it is possible to have general discoveries stemming from very applied research, as well as narrow/specific applications resulting from basic research. Therefore, generality is an element on its own, being a characteristic that is often associated with HRHR type of research. An example of applied research resulting in general knowledge outputs is Louis Pasteur's research aiming to solve practical problems concerning fermentation and putrefaction in the French wine industry. His findings ended up establishing the modern science of bacteriology, with important implications to multiple research areas, such as chemistry, biology or immunology. In the case of basic research leading to very applied results, the research of the geneticist George Harrison Shull provides a good illustration. Shull conducted research on the genetic composition of corn plants that was driven by his own academic curiosity and interest with no concrete application in mind. However, his research resulted in the discovery of hybrid corn, which is a type of corn that has a radically higher yield. Hybrid corn has radically changed

the corn industry and is a good example of a specific commercial application stemming from basic research.

Generalisable knowledge has the potential to unleash follow-up usage and discoveries in multiple fields. It requires more public support because firms typically focus on specific fields of activity, not having incentives to produce knowledge outside their area of specialisation. A good illustrative example is the invention of the atomic clock. The atomic clock has led to many scientific and technological advances in multiple sectors, such as global and regional navigation satellite systems, TV transmissions, or applications in the Internet that depend critically on frequency and time standards. Atomic clocks are also used in many scientific disciplines, such as for long-baseline interferometry in radioastronomy (McCarthy & Seidelmann, 2009) and have increasing applications in biotechnology and geological surveys (Norton Quinn, 2007). With applications in several areas and potentially fostering further developments in these and other areas, general knowledge can have much larger spillovers than knowledge that influences only one field.

Box 2. Examples of quantitative indicators measuring scientific generality

Concentration of forward citations in terms of scientific fields

- Definition: This measurement consists in computing the Herfindal index, which is a measure of concentration, of the different fields of forward citations.
- Rationale: The lower the concentration of citing articles' scientific fields, the wider the applicability of the focal article to multiple fields.
- Examples of use in the literature: Azoulay, Graff Zivin and Manso (2011)

Scientific distance of citing articles in terms of scientific fields

- Definition: Scientific fields are typically categorised by different levels of 1, 2, 3, or more digit aggregations, where larger number of digits represent more specific scientific categories. This means that for a given focal article, a citing article with the same 3-digit scientific category is scientifically closer than another citing article only with the same 2 digit category.
- Rationale: If a citing article has a low number of common scientific category digits (or none), the larger is its scientific distance with regard to the focal article, and thus the higher the generality of the focal article.
- Examples of use in the literature: This indicator was used in the context of generality in patents in Trajtenberg, Henderson and Jaffe (1997) and Henderson, Jaffe and Trajtenberg (1998).

Extraordinary levels of novelty

An extremely novel discovery is highly different compared to the state of the art, representing a potential leap forward. As specified in the Frascati Manual, any R&D activity is by definition novel. What is more, all academic publications are also novel by definition, as articles should only be accepted for publication if they contain something that is considered new to the prior art (among other factors). However, not all scientific discoveries are equally novel. Novelty can be seen as a scale, where some articles have extraordinary levels of novelty while others have lower levels of novelty. The same reasoning applies to the other characteristics of described in this study.

Scientific breakthroughs significantly advancing the knowledge frontier are often associated with extraordinary levels of novelty. Extremely novel findings have the potential to lead to major paradigm shifts with large knowledge spillovers. While research that explores novel ideas or uses novel approaches has the potential to be of high impact, it also faces higher levels of uncertainty. The pursuit of extremely high levels of novelty is associated with higher uncertainty because researchers have to take large risks exploring untested processes. In addition, it may take longer for novel knowledge to have an impact due to scientific prematurity, delayed recognition, or resistance from incumbent paradigms. A high degree of novelty is intimately linked with the nature of HRHR research.

Little is known about the processes that lead to highly novel knowledge. One approach has been proposed by Arthur (2009) who describes scientific research combining knowledge pieces. The same concept is used for example by Paul Romer and by Ricardo Hausmann to explain countries' economic development paths based on the combination of knowledge from their economic sectors. For Brian Arthur, using knowledge pieces in well-understood ways corresponds to "exploitative search processes", while using knowledge pieces in new ways and complex ways corresponds to an "exploratory search process", which will generate knowledge outputs with higher levels of novelty. Following this combinatorial perspective of scientific research, a discovery with an extraordinary level of novelty can be defined as new knowledge that recombines pre-existing knowledge components in an unprecedented fashion. An example of discovery that used existing pieces of knowledge in new ways and generated an extremely novel outcome is the "oncomouse" (also known as the "Harvard mouse"). The mouse was genetically modified to carry an activated oncogene that makes it highly susceptible to cancer. It has become a very valuable model for cancer research. The "oncomouse" recombined pre-existing knowledge components in an unprecedented fashion, being the first application of animal genetic modification to cancer research (Arts, Appio, & Van Looy, 2013).

The indicator of scientific novelty explored in Wang, Veugelers and Stephan (2017) was shown to capture well the essence of "high-risk" and "high-reward" research. They found that while many highly novel articles had a low number of citations, a small fraction of them became scientific breakthroughs with the highest levels of citations. Researchers take high risks when trying to explore highly novel scientific ideas and many of those ideas tend to fail. However, some will be successful and become major hits with potential to completely revolutionise the scientific paradigm. HRHR research outputs are likely to be associated with multiple knowledge characteristics, but novelty is particularly well-suited to capture the essence of HRHR.

Box 3. Examples of quantitative indicators measuring the level of novelty

Uncommon or new pairwise citations of journals

- Definition: This measure consists in grouping all the possible pairwise combinations of journals cited by each focal article, and counting which combinations were never cited together before and are more unexpected.
- Rationale: New or uncommon pairwise combinations of citations in a given article (measured e.g. through cosine similarity) represent new knowledge combinations, which represent higher levels of novelty.
- Examples of use in the literature: Measurements of this nature are used in e.g. Uzzi et al (2013), Lee, Walsh and Wang (2015), Veugelers and Wang (2016), Verhoeven, Bakker and Veugelers (2016), or Wang, Veugelers and Stephan (2017)

New corpus words or average age of keywords in abstracts

- Definition: This measure relies on using text-mining heuristics to count new corpus words or average age of keywords (as defined in scientific thesaurus).
- Rationale: The larger the number of new words, or the younger, the more novel a piece of research is.
- Examples of use in the literature: Examples of authors using these measurements are e.g. Azoulay, Graff Zivin and Manso (2011), Boudreau et al. (2016), Magerman, Van Looy and Debackere (2015), Packalen and Bhattacharya (2017), or Li, Azoulay and Sampat (2017)

Number of backward citations and mean-citation lags

- Definition: Number of citing articles and the average of their citation lags. The citation lag is the difference between the publication year of the focal article and the publication year a citing article.
- Rationale: This measure assesses the extent to which a discovery builds on newer or older ideas. As argued by the users of this indicator, if an invention draws from fewer and more recent articles, the more novel it is likely to be.
- Examples of use in the literature: Uses of these measurements using patents can be seen in Hall, Jaffe and Trajtenberg (2001) and Nerkar (2003).

3. Aggregate implementation of a combination-based novelty indicator

In this section, the novelty indicator used in Wang, Veugelers and Stephan (2017) is reproduced and aggregated per country, year and scientific discipline using SCOPUS. The goal is to test the utility of the indicator as a tool to monitor countries' efforts to support HRHR research, in this case measured by the levels of novelty of the articles published in each country. The same type of analysis can be done at the level of an individual researcher, university or science funding agency with potential to meet different needs for science policy monitoring and evaluation.

The indicator is computed at the level of scientific articles — each article has a unique novelty score based on how it is combining different knowledge fields. Knowledge fields are represented in terms of scientific journals. Scientific journals are often field-specific, representing specific knowledge spaces, and new journals tend to represent findings from emerging or converging fields. It is important to note that journals are not a perfect representation of knowledge spaces, thus using them as such is only an approximation. For example, some journals — including some of those with the highest impact factors — are multidisciplinary in essence and therefore do not represent a specific knowledge space. Moreover, new journals do not always represent the findings of new knowledge spaces. For example, new journals can result from publishers' strategic decisions, or from editors of existing journals deciding to launch new, closely related, journals. Comparisons over time can be particularly influenced by this factor, therefore they must be interpreted with caution. Recently, the growth in scientific production has been largely represented by the indexation of new journals, which explains 60% of the rise in the number of publications (OECD and SCImago Research Group (CSIC), 2016). Nevertheless, and despite these caveats and exceptions, journals are typically related to particular knowledge spaces, and are often used in the Economics of Science literature as approximations of specific research fields [see e.g. Uzzi et al (2013), Lee, Walsh and Wang (2015), Veugelers and Wang (Novel science for industry?, 2016), or Wang, Veugelers and Stephan (2017)].

Following the same methodology as Wang, Veugelers and Stephan (2017), the patterns of journal citations contained in articles' references can be explored to measure the extent to which articles are combining knowledge fields in more exploratory/risky ways. The computation comprises two steps: First, new journal combinations present in articles' references are identified and counted at the level of each individual article. When an article is the first to combine two journals in its reference list, then that article counts one new combination. Similarly, if two pairs of journals are cited together for the first time in an article, then the article first citing them counts two new combinations. Articles that do not have new journals combinations in their reference list have a novelty indicator of zero. Following the approximation/assumption that journals represent knowledge spaces, this indicator identifies the articles combining new knowledge spaces for the first time.

Second, new journal combinations are weighted according to how far apart their knowledge spaces are. Arguably, if two knowledge spaces are further apart then they are more difficult to combine, and the

corresponding novelty indicator receives a higher weight. The objective of this step is to differentiate between new journal combinations that are expected, or easier to combine as they are in closer fields, against new journals combinations that are more difficult to combine because they represent very dissimilar knowledge domains. Higher knowledge distance between the fields being combined can be indicative of higher uncertainty and thus risk-taking from researchers.

Data sources and computation details

The novelty indicator was computed using SCOPUS Custom Data, Elsevier, Version 5.2019. In the data exploration only publications corresponding to the years from 2005 to 2017 were analysed, but in order to compute the novelty indicator for these years the whole dataset was necessary in order to identify the first time each possible pair of journals was cited together. Therefore, all the articles in journals indexed in the SCOPUS database were used (reviews and conference proceedings were not included), the first publication year being 1996. Each article has a novelty score which consists of 1) counting the number of new journal combinations cited for the first time in the article's references (reflecting that two knowledge pieces are being combined for the first time), and 2) each new combination is weighted with a weight between 0 and 1 according to how unlikely that new combination is.

The weighting element of the computation is operationalised with the metric of cosine distance. For each year of publication, a matrix of journal co-citations was computed (therefore 12 matrices were computed for the 2005-2017 period). The co-citation matrices count the number of times each possible pair of journals are cited together in articles. For each matrix, the novel combinations are assessed with respect to a shifting window — only the journals that were cited between the year of publication and a 4-year window before are considered. For example, the matrix corresponding to 2015 considered all the journals cited in articles, which were published between 2011 and 2015. The weight of new journal combinations consists of the cosine distance between the two journal-vectors of the co-citation matrices. The metric captures how similar/distant the two journal-vectors are. Journal vectors with several common co-citations will have lower cosine distances, meaning that they represent knowledge spaces that are less distant. Journals with less common co-citations will have higher cosine distances, meaning that their knowledge fields are further apart. The cosine distance metric has the following formula:

$$COS\ distance_{i,k} = 1 - \frac{J_i \cdot J_k}{||J_i|| ||J_k||}$$

Where J_i and J_k are two journals. The cosine distance stems from the inner product between two journal-vectors of the journal co-citation matrix. It always results in a value between zero and one, where zero represents complete similarity and one complete dissimilarity. A co-citation matrix would look like the following:

	J1	J2	J3	J4	...Jn
J1	—	4	1	0	...
J2	4	—	9	0	...
J3	1	9	—	5	...
J4	0	0	5	—	...
... Jn	—

The co-citation matrix is symmetric, where each row/column represents a journal. For example, in the matrix above journal 1 and journal 2 are cited together in 4 articles, while journal 1 and journal 3 are cited together in 1 article. The journal-vector of journal 1 is [4,1,0,...] and journal 2 is [4,9,0,...]. The novelty indicator consists of counting the number new journal combinations per articles weighted by their cosine distances. Therefore, the indicator is given by:

$$Novelty_a = \sum_{new\ journal\ pair_{i-k}} COS\ distance_{i,k}$$

The novelty metric for article *a* corresponds to the sum of the cosine distances of the journal-vector pairs being cited together for the first time in article *a*, being an absolute measure without normalisation. The aggregation of the indicator at the country level does not use fractional counting, therefore articles published in international collaboration are fully accounted for on each country. Moreover, the classification of articles among the top 10 % most novel is scientific-field specific, computed at the level of 2 digit All Science Journal Classification Codes (ASJC). Annex A gives a summary of the steps taken to compute the novelty indicator.

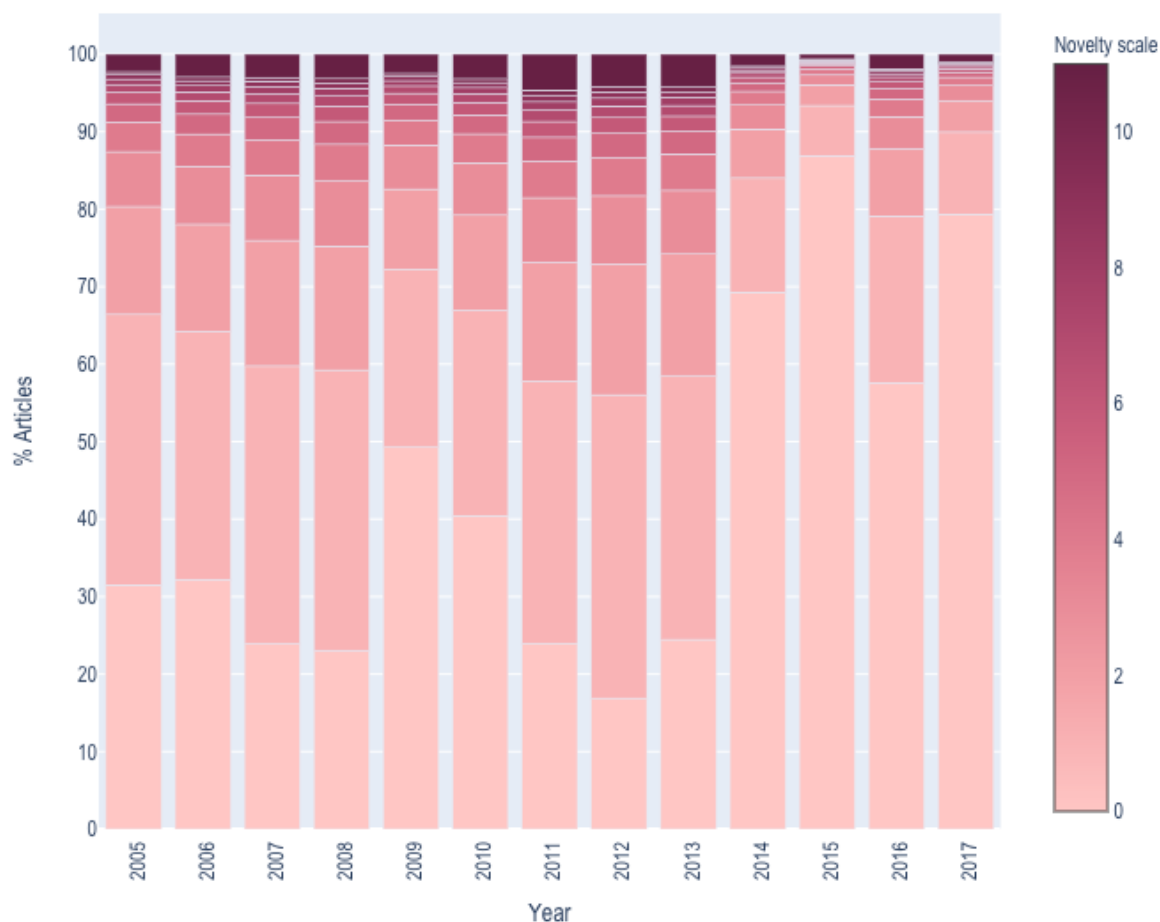
Data exploration and visualisation

Article-level analysis

The novelty indicator is analysed in this section with the objective of better understanding its nature and potential for assessing countries' performance in HRHR research. It is mapped against other variables explore its relationship with other relevant policy dimensions. As explained earlier, articles that represent major scientific breakthroughs have been shown to score very high according to this novelty indicator. However, these represent just a very small minority — most articles with high levels of novelty tend to fail (receiving no or very few citations). Moreover, as shown in Figure 2, novelty itself is already highly skewed — most articles do not attempt to explore knowledge fields that are difficult to combine.

Figure 2 presents the distribution of novelty between the years 2005 and 2017. The indicator ranges from no or very low levels of novelty (ranking zero/one) to high and very high levels of novelty (scoring 10 and higher). For all years, more than 50% of all articles score zero or one in terms of novelty (the two lowest categories), reflecting the fact that most articles do not attempt to make novel combinations of knowledge fields. Just a very small minority, the ones in the darkest colour in Fig 2, score very high on the novelty indicator. This figure illustrates the highly skewed nature of the novelty indicator.

Figure 2. Distribution of the novelty indicator



Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019.

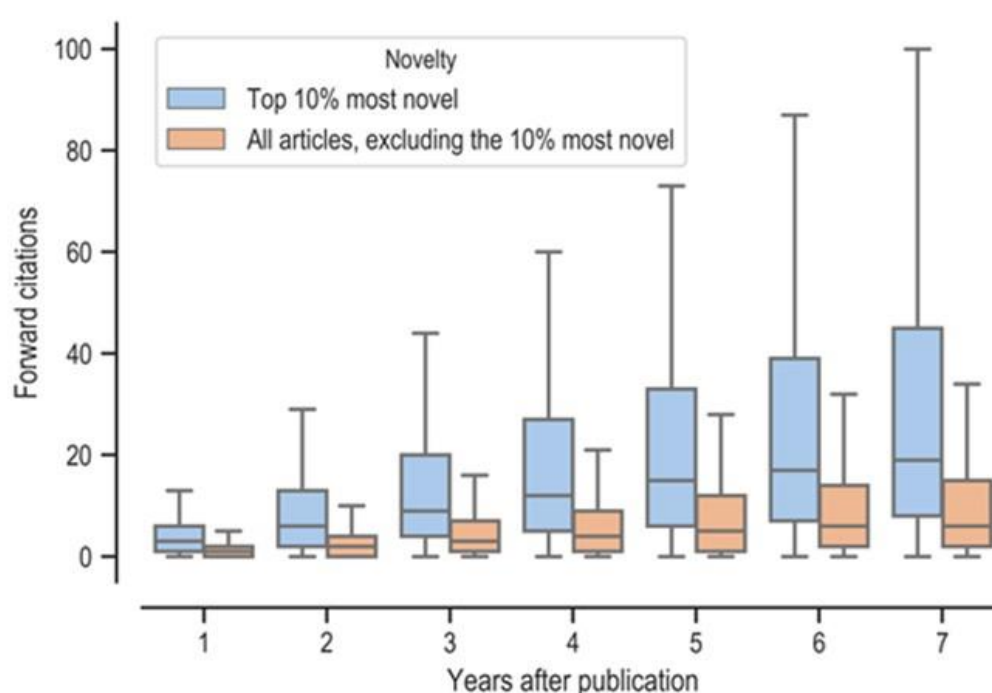
The main reason why novelty is relevant for HRHR research is because highly novel articles are associated with a higher citation performance (the reward part) but also with a higher citation variance (the risk part). To test this, the citation mean and variance were computed for the top 10% most novel articles and compared with the remaining population of articles. The analysis was limited to the year of 2011 in order to have a common and long enough window of data for post publication years.¹ The analysis of the citation mean and variance for different levels of novelty confirms the original hypothesis. In terms of reward, the overall mean of the top 10% most novel articles is 52.86 citations, while for the remaining articles the mean is 13.5 citations. In terms of risk, the standard deviation of the top 10% most novel articles is of 247.29, while for the remaining articles it is only 28.62. Citations are highly influenced by outliers but even when removing these the higher citation mean and variance of highly novel articles prevails. By removing articles with a citation z-score higher than 1, the mean of top novel articles is 19.8, while for the remaining articles the mean is of 10.37. Likewise, the standard deviation of the top novel is of 14.87, while for the remaining it is of 11.83.

In addition to the higher citation mean and variance, it is assumed that the citation superiority of highly novel articles takes time to consolidate. To test and visualise this, Figure 3 presents boxplots for the top 10% most novel articles and for the remaining population of articles for each year after publication. A total

of 7 years were included in order to enable enough forward citations to accumulate, and the boxplots were computed removing outliers with the interquartile range (IQR) method. The figure presents evidence confirming the original assumption. The median and variance of highly novel articles are systematically higher than the remaining population of articles. However, the difference grows over time — for initial years the difference is not substantial, while for later years it becomes much more pronounced.

The analysis of the medians and dispersion of citation performances shown in Figure 3 raises an important caveat in relation to how risk is interpreted or defined. The citation dispersion for the top 10% most novel articles is always larger than the remaining population and the difference increases over time. However, since the median is also systematically larger, the extent to which the larger citation variance can be interpreted as an indicator of risk is debatable.

Figure 3. Boxplots of citation performance per novelty category for 2011



Source: Author's calculations based on the full SCOPUS Custom Data, Elsevier, Version 5.2019.

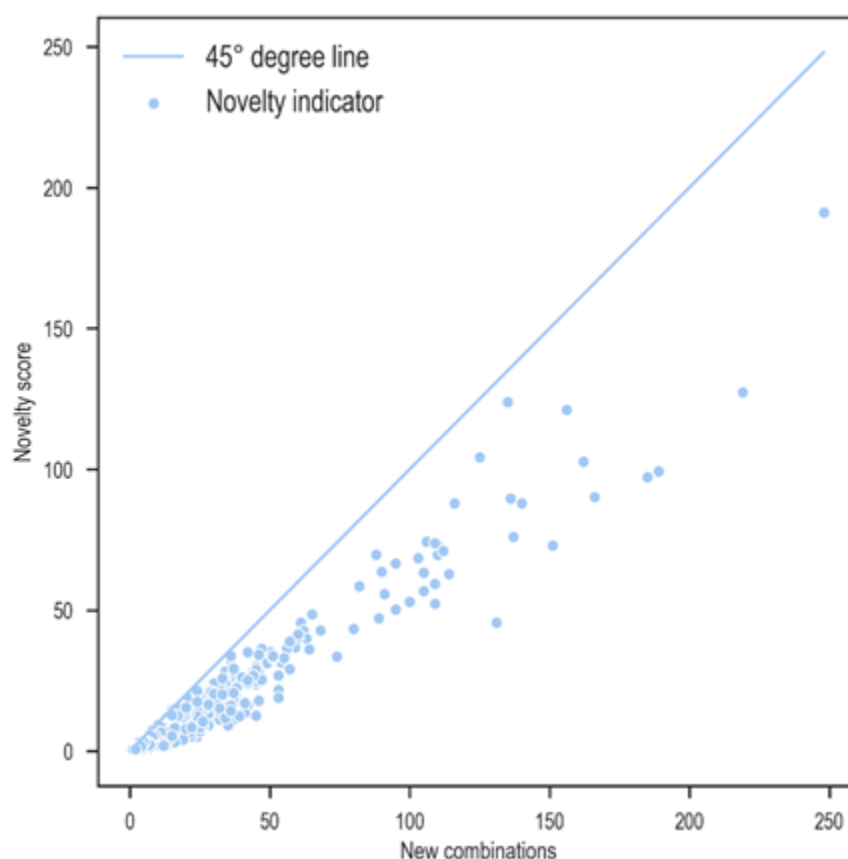
Note: Number of forward citations are not normalised by field.

The novelty indicator measures how difficult or unexpected are the new knowledge combinations made in articles. Therefore, the indicator has two dimensions — the new interdisciplinarity dimension and the uncertainty associated with that new interdisciplinarity. The new interdisciplinarity dimension is estimated by counting the number of times that journal-pairs get cited together for the first time, regardless of how difficult those combinations are. Therefore, this measure is an unweighted version of the novelty indicator. For each article a , the interdisciplinarity dimension is a count model of the number of times a pair of journals ($i-k$) are cited together for the first time:

$$\text{New interdisciplinarity}_a = \sum_{\text{new journal pair } i-k} 1$$

The two dimensions, novelty and new interdisciplinarity, are plotted in Figure 4, also focusing on articles published in 2011. The dots in the figure represent individual articles plotted in terms of the new combinations they make and the respective novelty scores. If the indicator was only based on new combinations (the pure new interdisciplinarity dimension), all the dots would be on the 45° line. The weighting based on the cosine distance method introduces a large amount of variability, enabling one to distinguish between articles with a similar number of new combinations. As shown in a regression framework (fig 5), the weighting dimension is a relevant addition to explain citation performance.

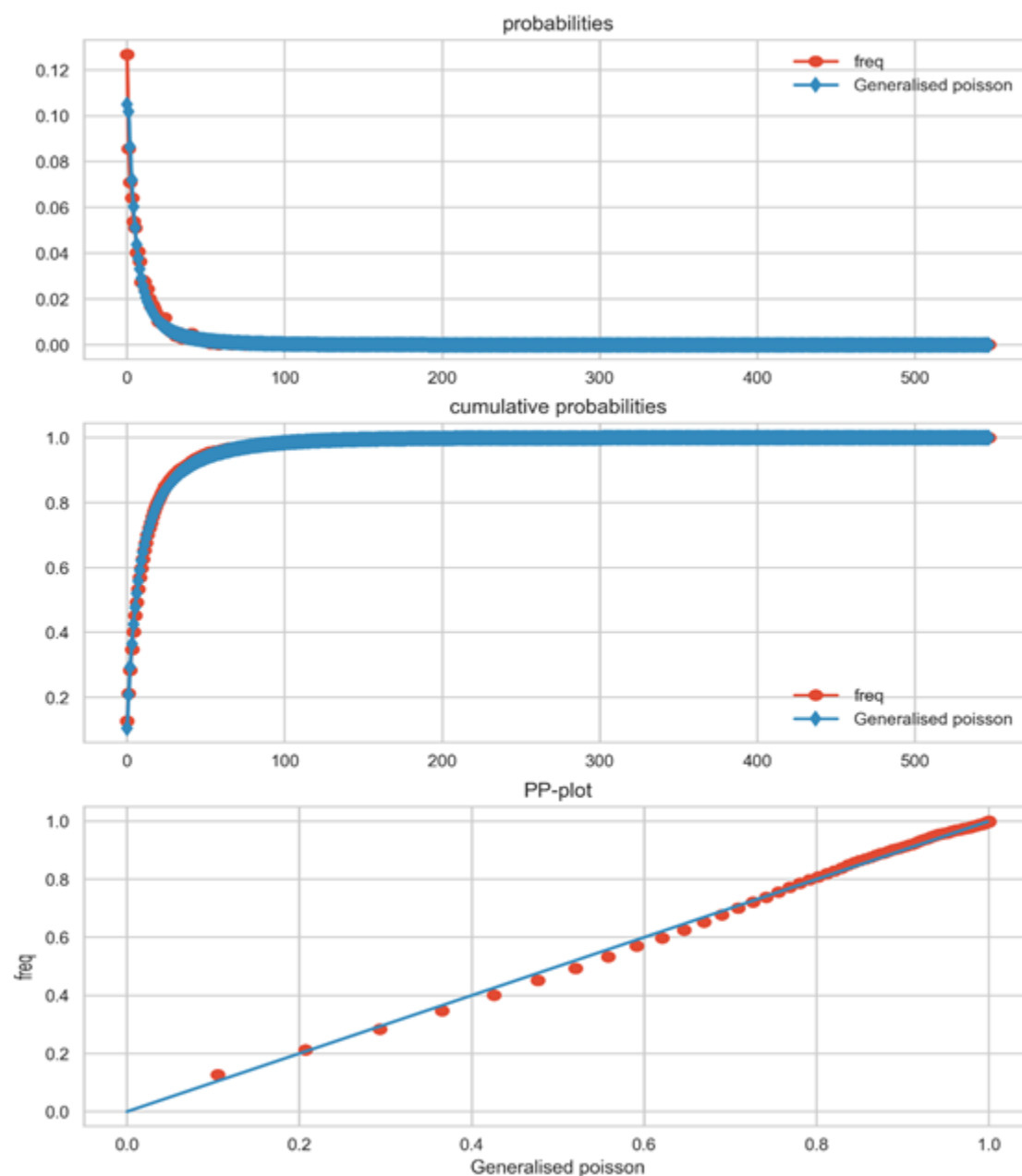
Figure 4. Novelty indicator vs interdisciplinarity for 2011



Source: Author's calculations based on the full SCOPUS Custom Data, Elsevier, Version 5.2019.

Analysing citation performance beyond the sample mean requires identifying the type of distribution underlying citations. Citations are discrete events and are extremely skewed, therefore, test statistics and regression models that assume an underlying normal distribution are not appropriate. Multiple discrete distributions were fitted against the citation series in order to identify which one best portrays the data. The distributions that were tested include the Poisson, Zero inflated poisson, Generalised poisson, Zero inflated generalised poisson, Negative binomial and Zero inflated negative binomial. The distribution that provided the best fit is the Generalised poisson (Figure 5)p. This was confirmed using a Mean Absolute Error (MAE) test, in which this distribution presented the lowest MAE. The remaining distribution fitting figures and test statistics are included in Annex B.

Figure 5. Generalised poisson distribution fit of forward citations for 2011



Source: Author's calculations based on the full SCOPUS Custom Data, Elsevier, Version 5.2019.

Based on the findings of distribution fitting test, a Generalised poisson regression model was implemented. The dependent variables are the total number of forward citations and multiple independent variables were included individually and combined. The results of the econometric model are presented in Table 1. All the regression models include country and scientific field fixed effects, the latter measured by 2-digit All Science Journal Classification Codes (ASJC). In column (1), new combinations are used as a regressor, representing the interdisciplinary element of the novelty indicator before the cosine weight is included. The model predicts that one additional new combination is associated with 0.002 more citations on average. In comparison, in columns (2), an increase by 1 unit of the novelty indicator (after including the cosine weight)

is associated with 0.003 more citations on average. In column (3), both measures are included simultaneously. In this framework, the model predicts that, for the same level of new combinations, an increase by one unit of the novelty indicator is associated with a 0.01 increase in forward citations. That said, for the same level of new interdisciplinarity measure, when the knowledge fields being combined are further apart, the resulting number of forward citations tends to be larger. In contrast, when holding the level of novelty fixed, an additional new journal combination is associated with less than 0.007 forward citations. This represents a situation where the new knowledge fields being combined are very closely related and thus not novel. The AIC and BIC criteria balance good fit with parsimony for each model, the lower their values, the better is the fit/parsimony. By comparing models 1 and 2, both AIC and BIC indicate that the novelty indicator is a relevant addition to explain citation performance. Columns 4-5 use dummy variables as regressors. Column 4 shows that articles that rank among the top 10% most novel have on average 1.3 more citations than articles in the bottom 90% least novel, and column 5 shows that articles that rank among the top 1% most novel have on average 1.7 more citations than articles in the bottom 99% least novel. Both coefficients are also statistically significant at the 1% level.

Table 1. Generalised poisson regression with total citations as dependent variable

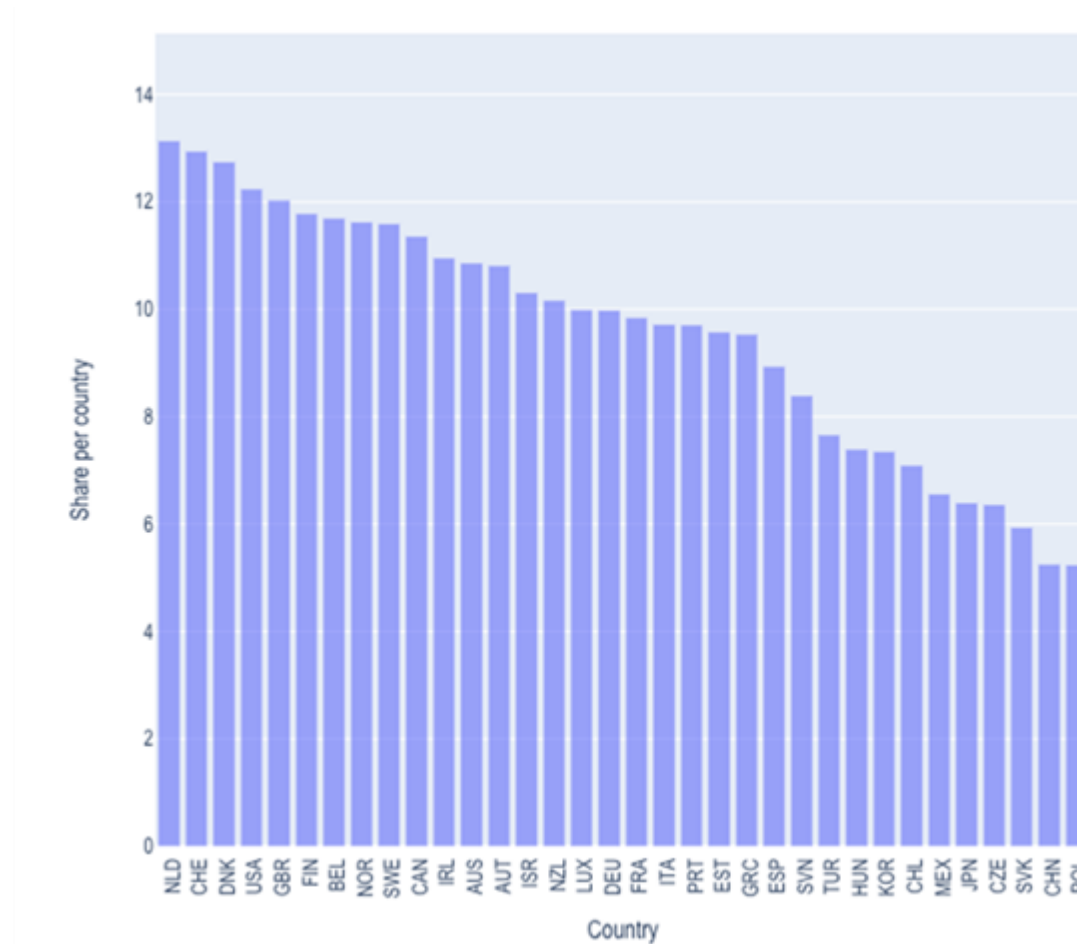
	1	2	3	4	5
New combinations	0.00248***	—	-0.00685***	—	—
Novelty indicator	—	0.00397***	0.0147***	—	—
Top10% most novel dummy	—	—	—	1.299***	—
Top1% most novel dummy	—	—	—	—	1.685***
Constant	Yes	Yes	Yes	Yes	Yes
2-digit ASJC FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
AIC	31.82925	31.81086	31.78241	30.61309	30.54498
BIC	6.49E+07	6.48e+07	6.46e+07	5.86e+07	5.83e+07
Observations	5126995	5126995	5126995	5126995	5126995

Note: Statistical significance is denoted by: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. In order to account for country and scientific field FE, the number of observations is pumped up by cases of international collaborations and multidisciplinary at a 2-digit ASJC.

Country-level analysis

The indicator can be aggregated at a country-year level in order to benchmark countries' performance in supporting novel science. With the indicator being highly skewed, the exploration of extraordinary levels of novelty requires focusing in the articles ranking in the top novelty categories. To illustrate that, Figure 6 shows the share of articles per country that score among the top 10% most novel in the world.

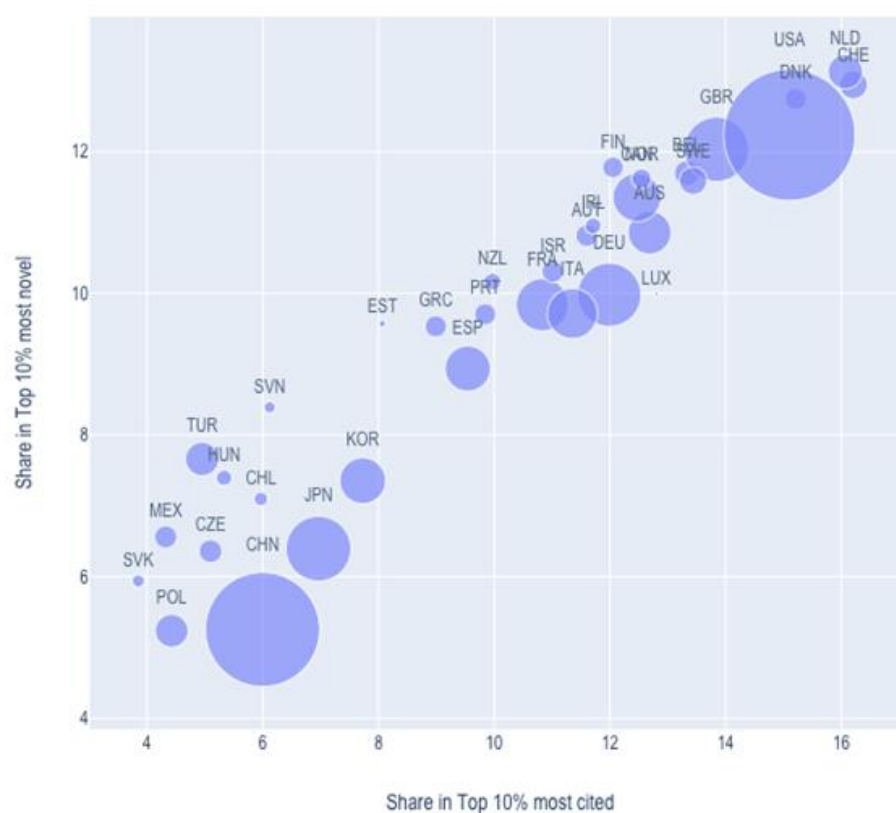
Figure 6. Share of publications among the top 10% most novel, 2005-17 averages, OECD countries plus China



Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Novelty percentiles are computed at the level of scientific disciplines following the All Science Journal Classification Codes (ASJC) used by SCOPUS at a 2-digit level, and are later aggregated at a country level. On average, for the period of 2005-2017, the country with the highest share of articles scoring among the top 10% most novel in the world was the Netherlands, where 13% of all articles with a Dutch author ranked among the top 10%. The Netherlands is followed by Switzerland, Denmark, the United States and the United Kingdom. These are also countries that tend to score very high in terms of scientific impact measured through citations. Thus, in order to further investigate the relationship between novelty and impact at the country level, Figure 7 presents the share of articles scoring among the 10% highest novelty against the share of articles scoring among the 10% top cited. The top 10% top cited documents indicator stems from OECD calculations database based on SCOPUS Custom Data, Elsevier, Version 5.2019.

Figure 7. Share of publications among the top 10% most novel vs top 10% most cited, 2007-17 averages, OECD countries plus China



Source: Author's calculations based on OECD and SCOPUS Custom Data, Elsevier, Version 5.2019

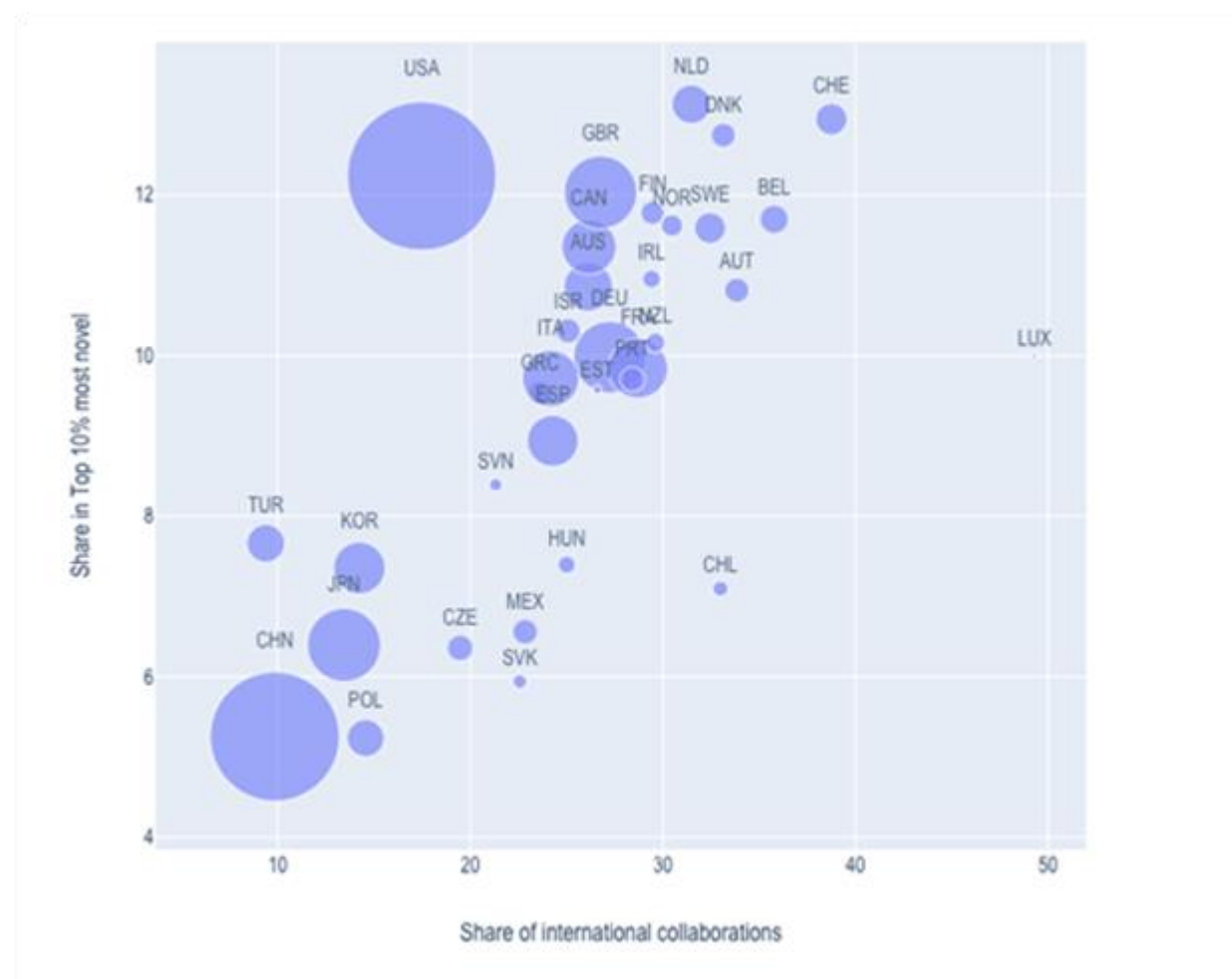
Note: bubble size represents the total number of scientific publications per country.

Making novel combinations of knowledge fields seems to be associated with high-reward. The relationship between novelty and impact, measured with forward citations, seems positive and strong. Figure 7 presents both indicators for the period 2005-2017 for all fields of science. The three small bubbles at the top right end of the figure represent the Netherlands, Switzerland and Denmark and the large bubble also at the top right end represents the United States. The size of the bubbles represents the total number of publications (not only the top 10% most cited) in fractional counts using OECD calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019.

Citations remain the most commonly used quantitative measure of scientific impact — the reward part of HRHR. Hence, finding this positive relationship between novelty and impact at a country level is an important outcome. The potential danger with the use of citations is that these are by nature a long-term indicator. It takes a long time for articles to be cited, especially so for the so called “sleeping beauties” (van Raan, 2004; Ke, Ferrara, Radicchi, & Flammini, 2015), scientific discoveries that take longer to be recognised and cited due to being extremely novel and of difficult acceptance by the existing scientific establishment (a discovery “ahead of its time”). Given the long-term nature of citations using them excessively, or solely, for short-term evaluation or policy decision-making can be problematic. It can distort recruitment practices by universities aiming to artificially boost their citation performance in the short-term. It can also potentially undermine scientific integrity with research results being manipulated or exaggerated

to attract citations. Most importantly it can discourage risky research projects as these are more uncertain and likely to deliver less short-term outputs. This said, while the novelty metric is not a substitute to citation-based indicators, nor immune to potential manipulations, it is a valuable complement to these traditional bibliometric measures. Citations should not be used in the short-term because articles need a long-time horizon to accumulate citations, but this novelty indicator is of short-term nature — the indicator can be computed at the time an article is published. The indicator could even potentially be computed at the research proposal stage if the key references are available, although this is was beyond the scope of the current analysis due to data availability. When taking a long-term perspective, the novelty indicator is positively related with citations — publishing highly novel research is a good measure of likely future impact both at a micro and country level. Thus, combining the two indicators can be helpful for managing risk and reward in portfolios of research projects, making sure that enough risky projects get funded (highly novel) and that in the long-term the portfolio delivers the expected level of reward/research impact (citation performance).

Figure 8. Share of publications among the top 10% most novel vs share of publications part of international collaboration, 2005-17 average, OECD countries plus China



Source: Author's calculations based on OECD and SCOPUS Custom Data, Elsevier, Version 5.2019.

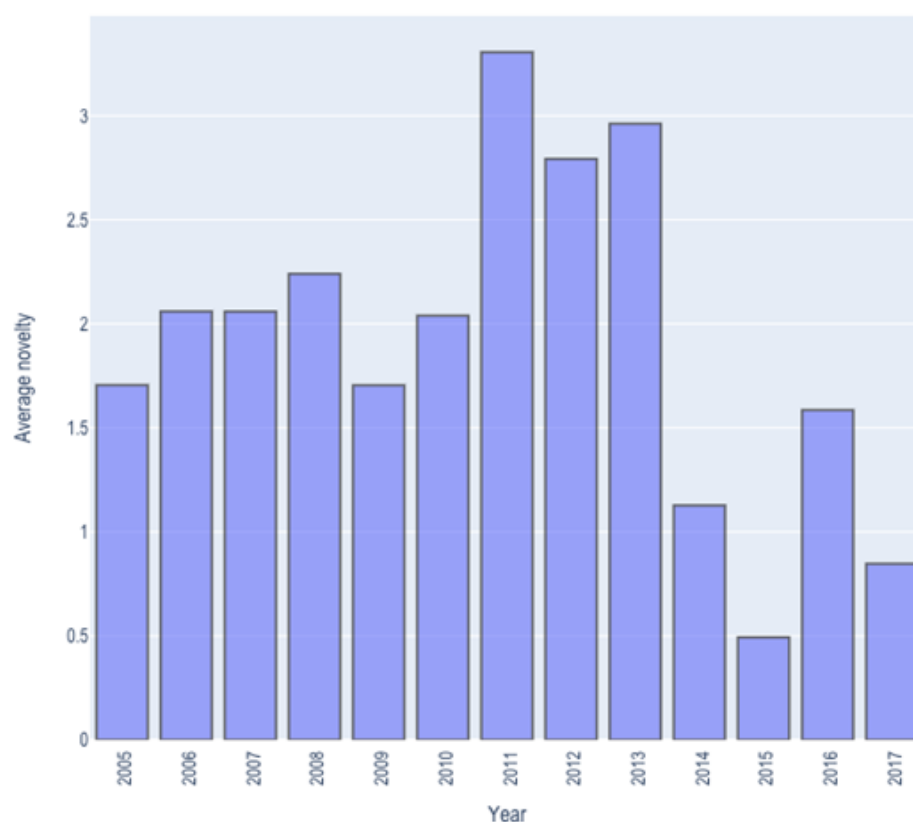
Note: bubble size represents the total number of scientific publications per country.

Figure 8 shows the relationship between the novelty indicator and the share of scientific articles in a country that are part of an international collaboration. The novelty indicator on the Y axis represents the share of articles per country among the top 10% most novel in the world. The X axis represents the share of articles per country published with co-authors from other countries. The relationship between novelty and international collaboration is positive but is less clear-cut than the relationship between novelty and citations (Fig 7). The countries with the highest levels of international collaborations are Switzerland, Belgium, Austria, Denmark, Sweden and the Netherlands (in order in the figure from right to left). Making novel combinations of knowledge pieces is a risky activity, so engaging in international research collaborations to spread such risk may be a risk-management strategy. Moreover, the interdisciplinarity element of the novel indicator may require putting together a diverse/multidisciplinary research team, which often entails international collaboration.

Year-level analysis

As explained at the start of this section, the temporal dimension of the indicator needs to be interpreted with caution, in particular due to the emergence of new journals. New journals contribute accurately to the yearly evolution of the novelty indicator when they represent new or emerging knowledge fields — new combinations of past knowledge spaces end up creating new knowledge spaces. However, new journals do not always represent this scientific evolution process, and that is not taken account of in the analysis that was conducted. Keeping this caveat in mind, Figure 9 presents the average novelty of all articles for different years between 2005 and 2017 for all OECD countries.

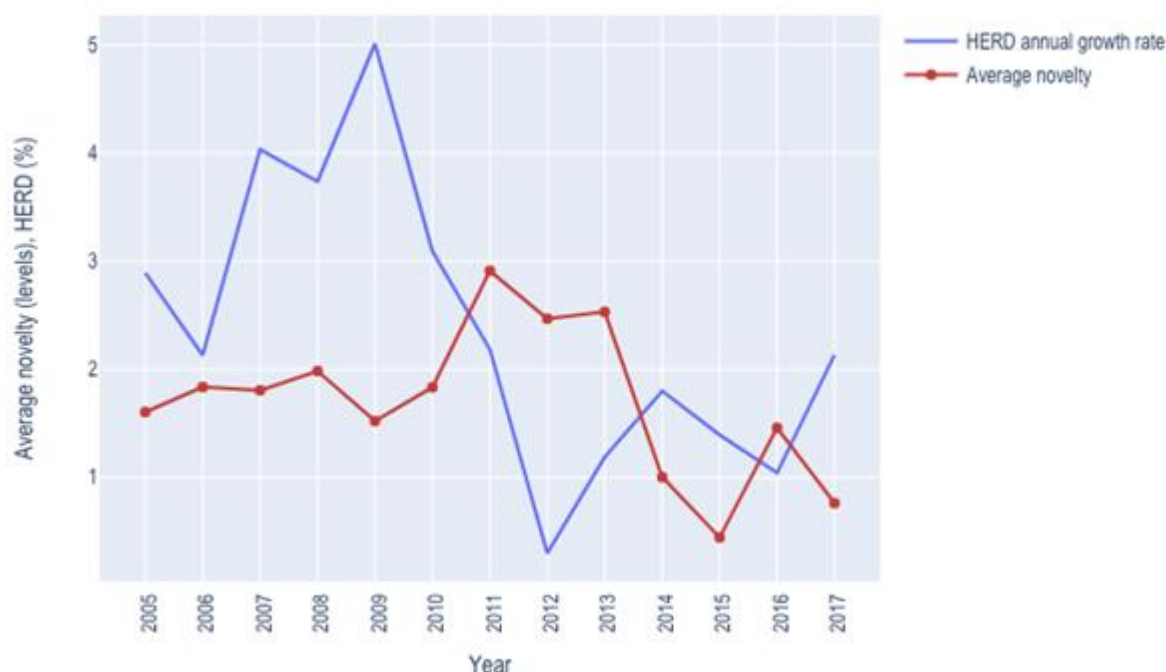
Figure 9. Average novelty per year, OECD average



Source: Author's calculations based on OECD and SCOPUS Custom Data, Elsevier, Version 5.2019.

The average of the novelty indicator decreased and rose substantially from 2009 to 2011. Between 2011 and 2015 novelty decreased drastically and then increased in 2016 before dropping again in 2017. Interestingly, this pattern is common to most countries and science fields, suggesting that the evolution of the indicator is driven by common factors. When including the rest of the world, the novelty trend follows the same pattern as the presented for OECD countries. Moreover, the trend for the new interdisciplinarity indicator average is similar.

Figure 10. Average novelty vs HERD growth rate per year, OECD average



Source: Author's calculations based on OECD and SCOPUS Custom Data, Elsevier, Version 5.2019.

Figure 10 presents two time series, the average novelty of publications from OECD countries and the annual growth rate of Higher Education Expenditures in R&D (HERD). The novelty indicator (red line) is the same as in Figure 9, while the annual growth rate of HERD for OECD countries (blue line) is retrieved from the OECD's Main Science and Technology Indicators database (MSTI). The evolution of the indicators over time is interesting — novelty follows the direction of HERD growth rates with a 3 to 4-year lag. HERD accelerates in 2007, followed by an increase in average novelty in 2010-11, HERD then decelerates slightly in 2008 and accelerates again in 2009, which is mirrored by a slight decrease in novelty in 2012 and a slight increase in 2013. Between 2009 and 2012 HERD decelerates massively reaching close to zero growth (some countries experienced actual decreases in HERD expenditures). This deceleration in HERD expenditures is followed by a significant fall in the levels of novelty between 2013 and 2015. Finally, pattern in HERD over 2012 and 2016, is similar to that of novelty between 2015 and 2017.

Cross-country/year econometric analysis

With the naked eye, the course over time of the novelty indicator and HERD seem related, but that does not necessarily mean that variations in HERD cause variations in novelty. The same reasoning applies to

the previous variables measuring citations and international collaborations. For example, HERD, citations and international collaborations can be related with other observable or even unobservable phenomena that mediate their relationship with novelty. In order to make a more robust assessment of the relationship between these variables, multiple statistical models were tested. The objective was to take into account other observable and unobservable factors that might be influencing novelty in order to have a better understanding about the true relationship between the variables of interest and novelty. The results of the statistical models are summarised on Table 2.

Table 2. Regression table: Share of articles among the top 10% top novelty

	Pooled OLS	First Differences	Time FE	Entity FE	Time and Entity FE
	(1)	(2)	(3)	(4)	(5)
Percent top cited	0.4481***	0.3619***	0.4178***	0.4314***	0.3931***
Percent international	0.0148	0.0401**	0.0326	-0.16	-0.1485**
HERD	2.2431***	2.3548***	2.2545***	-0.6367	-2.4733
BERD	-0.4008**	-0.1520	-0.2870***	0.3775	0.5444**
GOVERD	1.0882	1.6992	0.5466	0.6600	0.6032
Constant	Yes	No	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes
Country effects	No	No	No	Yes	Yes
Time effects	No	No	Yes	No	Yes
Observations	525	483	525	525	525

Notes: The dependent variable is Percent top novelty. Additional controls include GDP, population, value added of industry and total number of publications. Standard errors are clustered at the entity level. All countries with enough available data points in the OECD's MSTI database are included, i.e. all OECD countries plus Argentina, Romania, Russia, Singapore and South Africa.

Statistical significance is denoted by: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

The dependent variable for all the 5 models in Table 2 is the share of scientific articles per country that score among the 10% highest novelty in the world. The level of analysis is at the country-year level, covering the period of 2005-2017 and all the countries from OECD's MSTI database. In all specifications multiple control variables such as GDP or total number of researchers in a country, are taken into account, and the independent variables of interest are: share of articles among the top 10% most cited, share of international collaborations and R&D expenditures. R&D expenditures from the high education sector, business and government R&D expenditures are considered (HERD, BERD and GOVERD, respectively), all measured as a percentage of GDP.

Each specification from (1) to (5) represents a different econometric model, (1) being the least robust and (5) the most robust. Model (1) is a pooled OLS, where all countries and years are pooled together. In model (2) the format of the variables is in 1-year differences instead of in absolute value in order to account for potential autocorrelation (thereby decreasing the number of observations). With all variables in first differences the interpretation of the coefficients is in terms of acceleration (positive coefficients) or deceleration (negative coefficients). Model (3) includes year fixed effects, which means that all possible unobservable factors that are fixed in time are taken into account, while model (4) includes country fixed effects, taking into account unobservable factors that are fixed across countries. Model (5) includes both year and country fixed effects.

In all specifications, the share of articles per country/year that score among the 10% most cited is positive and with high statistical significance at 1% level. This provides strong evidence that high-risk is clearly associated with high-reward. Countries where researchers take higher risks to explore novel knowledge combinations are also producing the most impactful scientific discoveries. The relationship between the share of international publications and novelty is inconclusive. In models (1) to (3) the relation is positive and it is statistically significant under model (2) at the 5% level. However, after accounting for country fixed effects the relation becomes negative and statistically significant at the 5% level under model (5). The coefficients of R&D expenditures provide interesting insights. As might be expected from Figure 10, HERD is positively related with novelty, with large positive coefficients, which are statistically significant at the 1% level under models (1) to (3). However, after taking into account country fixed effects the coefficient becomes statistically insignificant and changes sign. Interestingly, the coefficient of business R&D expenditures (BERD) as a percentage of GDP, becomes positive and statistically significant at the 5% level after taking into account country fixed effects in model (5). The coefficients for government expenditures in R&D (GOVERD) as a percentage of GDP are positive but statistically insignificant in all the models. Perhaps more interesting than each individual coefficient, is the way that country and time fixed effects change the results. This may reflect the relevance of unobserved country-level factors that are constant through time, such as, for example scientific culture or overall risk-taking attitudes.

4. Conclusion

Quantitative bibliometric indicators, when well used and interpreted, are powerful tools for science policy monitoring and evaluation, and for evidence-based policymaking. On the other hand, excessive reliance and misuse of bibliometric indicators, such as JIF and H-indexes, has been shown to have pernicious effects, particularly if they are used as short-term performance measures. The goal of this paper was to show that HRHR research calls for the use of new quantitative indicators to complement traditional ones. New indicators are not substitutes to the rigorous use of traditional indicators and this paper does not claim that citation-based indicators should stop being used. The argument put forward in this paper is that the combined use of multiple bibliometric indicators can help monitor and evaluate the different knowledge characteristics that are associated with HRHR research in order to promote it. Knowledge characteristics that are relevant for HRHR research are the levels of basicness, generality and novelty, as well as levels of spill-overs, the longer time lags and the uncertainty associated with its outcomes. This paper provides examples of discoveries to illustrate each characteristic and reviews relevant quantitative metrics that have been published in the academic literature.

In order to demonstrate the potential of a recently described quantitative indicator to inform HRHR research policy, a metric of novelty was computed and analysed at the article and country level. The indicator is based on the theory that science evolves by combining different knowledge pieces. Combining knowledge pieces (represented by journals) that are closer/easier to combine, corresponds to exploitative research processes, which tend to lead to incremental improvements. Combining knowledge pieces that are further apart corresponds to exploratory research processes, which are more closely linked with HRHR type of research. The results of exploratory research can be more uncertain because combining knowledge pieces that are further apart is harder. However, major scientific discoveries typically stem from exploratory research. Therefore, while highly novel research projects have more unpredictable results, some end up producing the most significant scientific breakthroughs.

The metric computed in this paper shows that novelty captures some of the essence of HRHR research, but not perfectly. Most articles show very low levels of novelty, and highly novel articles have a higher citation performance in the long-term combined with higher citation unpredictability. However, the median citations of highly novel articles are systematically larger than those of less novel articles, both in the shorter and longer-run. Therefore, the higher citation uncertainty of highly novel papers is not a perfect reflection of the high-risk component typically associated with HRHR research. No unique indicator can fully represent the whole essence of HRHR research. This said, the novelty indicator can be used as a useful complement to the other measures reviewed in this paper and more traditional indicators, to more effectively monitor how research portfolios are performing over time with regard to multiple knowledge characteristics. The novelty indicator is useable in the short-term (it can be computed at the time an article is published) and one of its main documented features is its strong correlation with long-term scientific impact (measured with citations in a 7-year window), both at the article and country levels. Citation-based

indicators require a long-time span to be useful from a monitoring and evaluation perspective, thus complementing them with the novelty indicator can be particularly powerful.

The analysis of the novelty indicator indicates that novelty in publications is highly skewed. Most articles show no risk-taking, making no new combinations of knowledge pieces or make new combinations from very similar knowledge fields. Just a very small fraction of articles present high levels of risk-taking combining multiple and distant knowledge pieces. Restricting the analysis to 2011-2013 (to have enough years after publication date in order to accumulate citations) shows that higher levels of novelty are associated with a higher citation variance. Moreover, highly novel articles are shown to have significantly higher levels of citations, but this superiority is less significant in the short-term, becoming more pronounced as time goes by. By aggregating the indicator at a country level, the top performers in terms of novel research publications in the OECD are the Netherlands, Switzerland, Denmark and the United States. These countries have specific HRHR policies in place and are also the ones with the highest share of highly-cited scientific publications. Countries with larger shares of international collaborations and higher R&D spending also appear to be associated with higher levels of novelty. In a more robust statistical analysis, only high scientific impact (reward) in terms of citations seems to be consistently strongly positively related with novelty at 1% statistical significance. The results from the most robust econometric model also indicate that countries with a higher level of business R&D expenditures tend to have higher levels of scientific novelty.

The OECD report “Effective Policies to foster High-Risk/High-Reward research” provides several examples of science programmes supporting HRHR research from the top performing countries as measured with the novelty indicator. These include, for example, The Netherlands’ *Off Road programme* and the Swiss National Science Foundation’s *Sinergia programme*, which explicitly emphasise the “out-of-the-box” character and originality of research proposals at project selection stage. In Denmark, the Lundbeck Foundation sets aside 10% of its research funding for alternatives to classical peer review evaluation, in part to select research projects deemed too risky in classical peer review processes. The US National Science Foundation (NSF) has introduced small programmes deviating from classical peer review evaluation processes, such as the *Early-concept Grants for Exploratory Research (EAGER) programme* that relies on programme-manager evaluation in making funding awards. These specific programmes and initiatives do not necessarily explain why these countries are ‘top performers’ when it comes to novelty. However, they help to illustrate how funding organisations are experimenting and adapting the way that research projects are selected in order to avoid biases against highly novel and risky research proposals.

More broadly, the OECD report makes the argument that accepting or incentivising risk-taking is a systemic issue for the scientific community, encompassing multiple actors and processes. In this context, the availability and effective usage of new indicators is a critical element for the monitoring and evaluation of policies to foster HRHR research. Such indicators can play an important role in research portfolio management. One of the lessons from the OECD Global Science Forum (GSF) work is that it is important to take a portfolio approach to supporting HRHR research, since risk and reward are better managed at the level of the research portfolio rather than at the level of each individual project.

Complementing the use of traditional bibliometric indicators with new ones can provide important insights for science funding and science policy evaluation. Given the high complexity of HRHR research, relying excessively, or solely, on indicators that are purely based on long-term performance to make decisions in the short-term is too limited. In the same way that traditional indicators are currently used, new indicators can be computed at a national or subnational level and for specific regions, science agencies, funding programmes, research organisations, or even individual researchers. Multiple indicators have been proposed, tested and validated in the academic literature but their use in science policy remains limited. The time has come to complement existing science monitoring and evaluation tools with new indicators, in order to have a better understanding and inform efforts to support HRHR research.

Endnotes

¹ Using instead the top 15%, 5% or 1% most novel articles leads to similar conclusions. Likewise, the same analysis was conducted for the years of 2012, 2013 and 2014 with a lower post publication window but producing similar results.

References

- Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, 22(6-7), 521-543. doi:10.1002/smj.176
- Arora, A., Belenzon, S., & Pataconi, A. (2018). The decline of science in corporate R&D. *Strategic Management Journal*, 39(1), 3-32. doi:10.1002/smj.2693
- Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The rate and direction of inventive activity: Economic and social factors* (pp. 609 - 626). Princeton University Press.
- Arthur, W. (2009). *The nature of technology: what it is and how it evolves*. Free Press.
- Arts, S., Appio, F., & Van Looy, B. (2013). Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? *Scientometrics*, 97(2), 397-419. doi:10.1007/s11192-013-1045-1
- Azoulay, P., Graff Zivin, J., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *Source: The RAND Journal of Economics*, 42(3), 527-554. doi:10.1111/j.1756-2171.2011.00140.x
- Boudreau, K., Guinan, E., Lakhani, K., & Riedl, C. (2016). Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Management Science*. doi:10.1287/mnsc.2015.2285
- Budish, E., Roin, B., & Williams, H. (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. *The American economic review*, 105(7), 2044-2085. doi:10.1257/aer.20131176
- Chang, K. (2017). Nobel Prize in Chemistry Awarded for 3D Views of Life's Biological Machinery - The New York Times. *The New York Times*. Retrieved 07 05, 2019, from <https://www.nytimes.com/2017/10/04/science/nobel-prize-chemistry.html>
- Fang, F. C., & Casadevall, A. (2016). Research Funding: the Case for a Modified Lottery. *Mbio*, 7(2). doi:e00422-16
- Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *NBER Working Paper*(8498). doi:10.3386/w8498
- Harris, R. (2013). *Scientists Win Nobel For Work On How Cells Communicate* : NPR. NPR. Retrieved 07 05, 2019, from <https://www.npr.org/templates/story/story.php?storyId=230192033>
- Henderson, R., Jaffe, A., & Trajtenberg, M. (1998). Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988. *Review of Economics and Statistics*, 80(1), 119-127. doi:10.1162/003465398557221
- Kakalios, J. (2010). *The amazing story of quantum mechanics : a math-free exploration of the science that made our world*. Gotham Books.

- Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, 112(24), 7426-7431. doi:<https://doi.org/10.1073/pnas.1424329112>
- Lee, Y.-N., Walsh, J., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy*, 44(3), 684-697. doi:10.1016/j.respol.2014.10.007
- Leifer, R., O'Connor, G., & Rice, M. (2001). Implementing radical innovation in mature firms: The role of hubs. *Academy of Management Perspectives*, 15(3), 102-113. doi:10.5465/ame.2001.5229646
- Li, D., Azoulay, P., & Sampat, B. (2017). The applied value of public investments in biomedical research. *Science*, 356(6333), 78-81. doi:10.1126/science.aal0010
- Lowther, D., & Freeman, E. (2008). The application of the research work of James Clerk Maxwell in electromagnetics to industrial frequency problems. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 366(1871), 1807-20. doi:10.1098/rsta.2007.2188
- Magerman, T., Van Looy, B., & Debackere, K. (2015). Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. *Research Policy*, 44(9), 1702-1713. doi:10.1016/J.RESPOL.2015.06.005
- Malva, A., Kelchtermans, S., Leten, B., & Veugelers, R. (2015). Basic science as a prescription for breakthrough inventions in the pharmaceutical industry. *The Journal of Technology Transfer*, 40(4), 670-695. doi:10.1007/s10961-014-9362-y
- Mccarthy, D., & Seidelmann, P. (2009). *TIME – From Earth Rotation to Atomic Physics*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA. Retrieved 10 27, 2017
- Nelson, R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, 67(3), 297-306. doi:10.1086/258177
- Nerkar, A. (2003). Old Is Gold? The Value of Temporal Exploration in the Creation of New Knowledge. *Management Science*, 49(2), 211-229. doi:10.1287/mnsc.49.2.211.12747
- Norton Quinn. (2007). *How Super-Precise Atomic Clocks Will Change the World in a Decade* | WIRED. Retrieved 10 27, 2017, from <https://www.wired.com/2007/12/time-nist/?currentPage=2>
- OECD. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*. Paris: OECD Publishing. doi:<https://doi.org/10.1787/9789264239012-en>
- OECD and SCImago Research Group (CSIC). (2016). *Compendium of Bibliometric Science*. Paris: OECD publishing. Retrieved from <http://oe.cd/scientometrics>
- Overbie Dennis. (2017). 2017 Nobel Prize in Physics Awarded to LIGO Black Hole Researchers - The New York Times. *New York Times*. Retrieved 10 27, 2017, from <https://www.nytimes.com/2017/10/03/science/nobel-prize-physics.html>
- Packalen, M., & Bhattacharya, J. (2017). Neophilia ranking of scientific journals. *Scientometrics*, 1(1), 43-64. doi:10.1007/s11192-016-2157-1
- Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, 39(8), 1051-1059. doi:10.1016/J.RESPOL.2010.05.013
- Stephan, P. (1996). The Economics of Science. *Journal of Economic Literature*, 34(3), 1199-1235. doi:10.2307/2729500

- Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*, 544(7651), 411-412. doi:10.1038/544411a
- Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology*, 5(1), 19-50. doi:10.1080/10438599700000006
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science*, 342(6157), 468-472. doi:10.1126/science.1240474
- van Raan, A. F. (2004). Sleeping Beauties in science. *Scientometrics*, 59(3), 467-472. doi:https://doi.org/10.1023/b:scie.0000018543.82441.f1
- Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, 45(3), 707-723. doi:10.1016/J.RESPOL.2015.11.010
- Veugelers, R., & Wang, J. (2016). *Novel science for industry?* IEEE. doi:10.1109/ICMIT.2016.7605046
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*. doi:10.1016/j.respol.2017.06.006

Annex A.

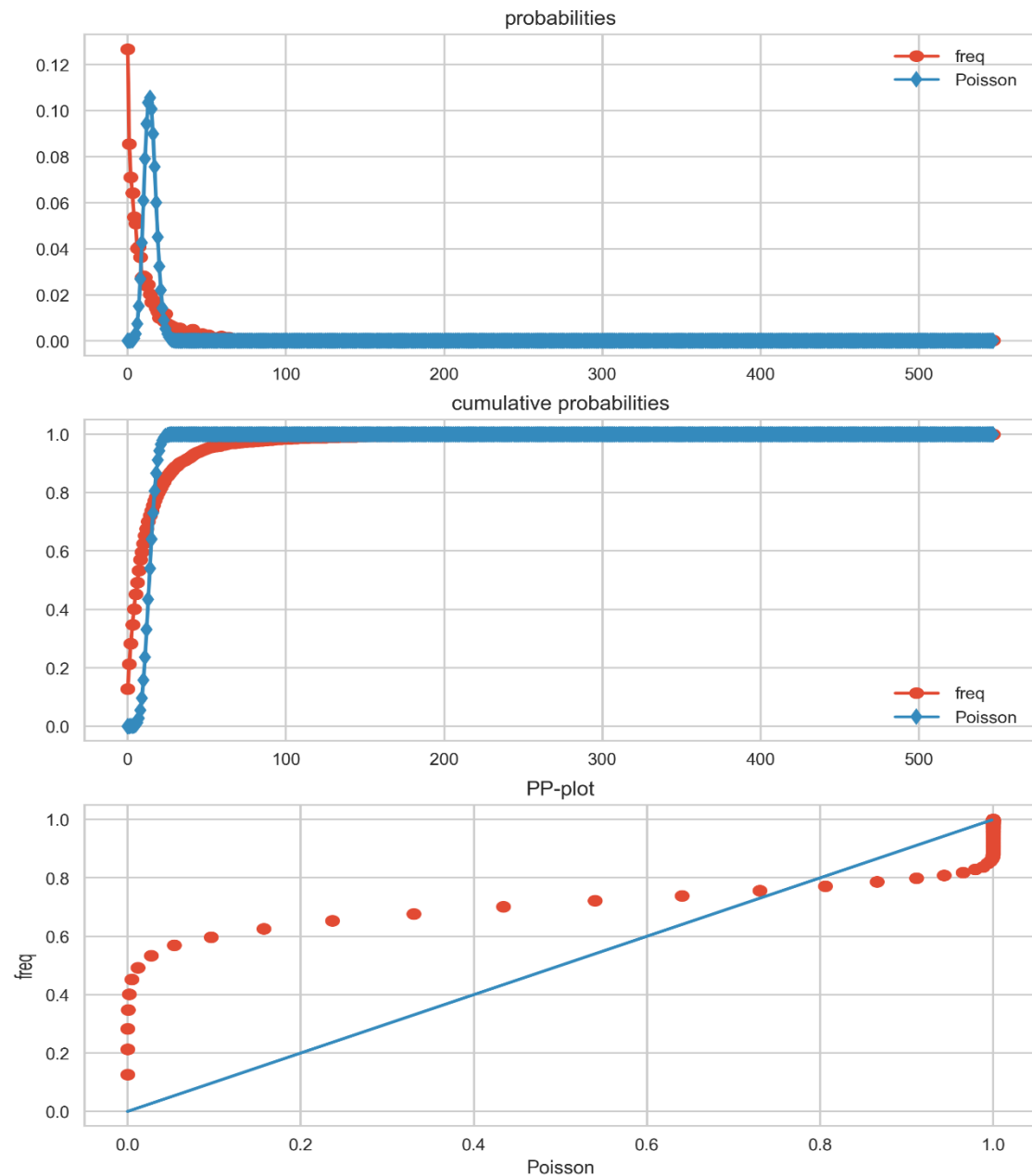
Novelty indicator computation steps using Scopus:

Prepare list of pairwise combinations of journals cited in each focal article reference list

1. Identify the combinations of journals being cited together for first time since 1996 (first available year on Scopus)
2. Link each new combination to the corresponding article citing the two journals
3. Prepare a journal co-citation matrix for each year containing all the times journals are cited together between the focal year and 4-year lag (or more if computationally feasible)
4. Use the citation matrix to compute the cosine distance between the journal-vectors corresponding to each new pair of journal combination
5. Sum the cosine distances of each new journal pair at the level of the focal article combining the two journals for the first time
6. In this paper, the classification of articles among the top X% most novel is scientific-field specific, computed at the level of 2 digit All Science Journal Classification Codes (ASJC)

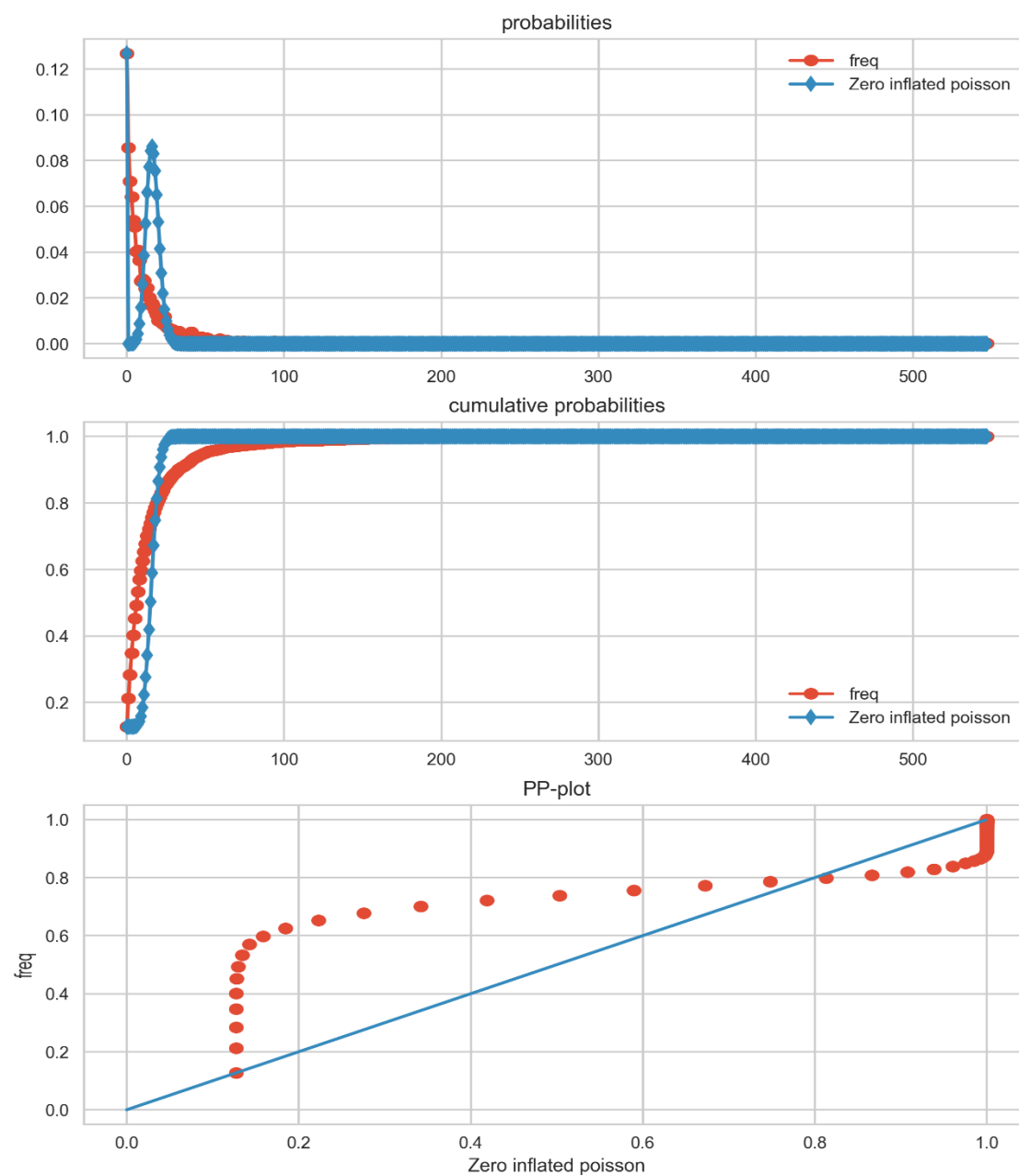
Annex B.

Figure A B.1. Poisson distribution fit of forward citations for 2011



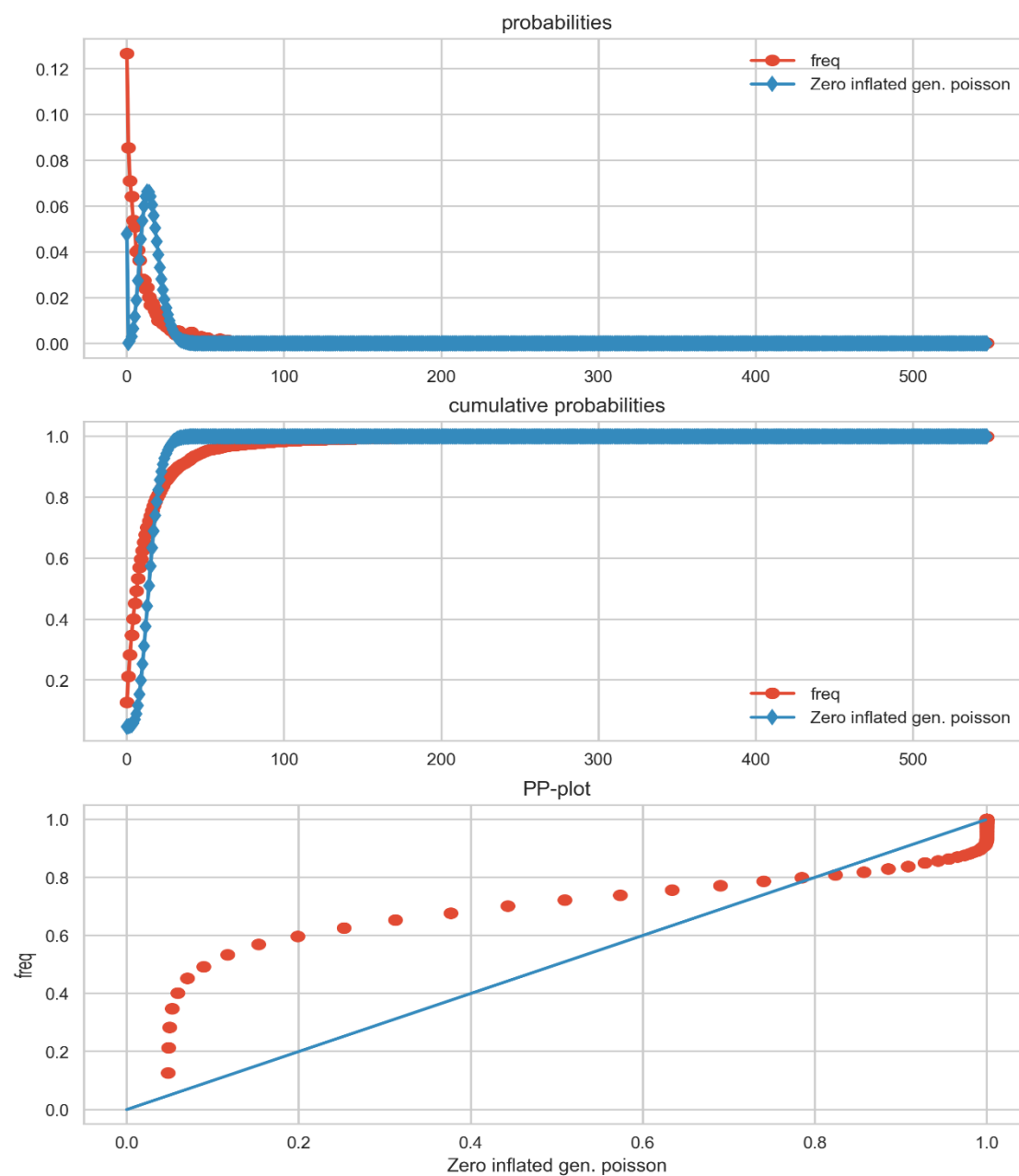
Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Figure A B.2. Zero inflated poisson distribution fit of forward citations for 2011



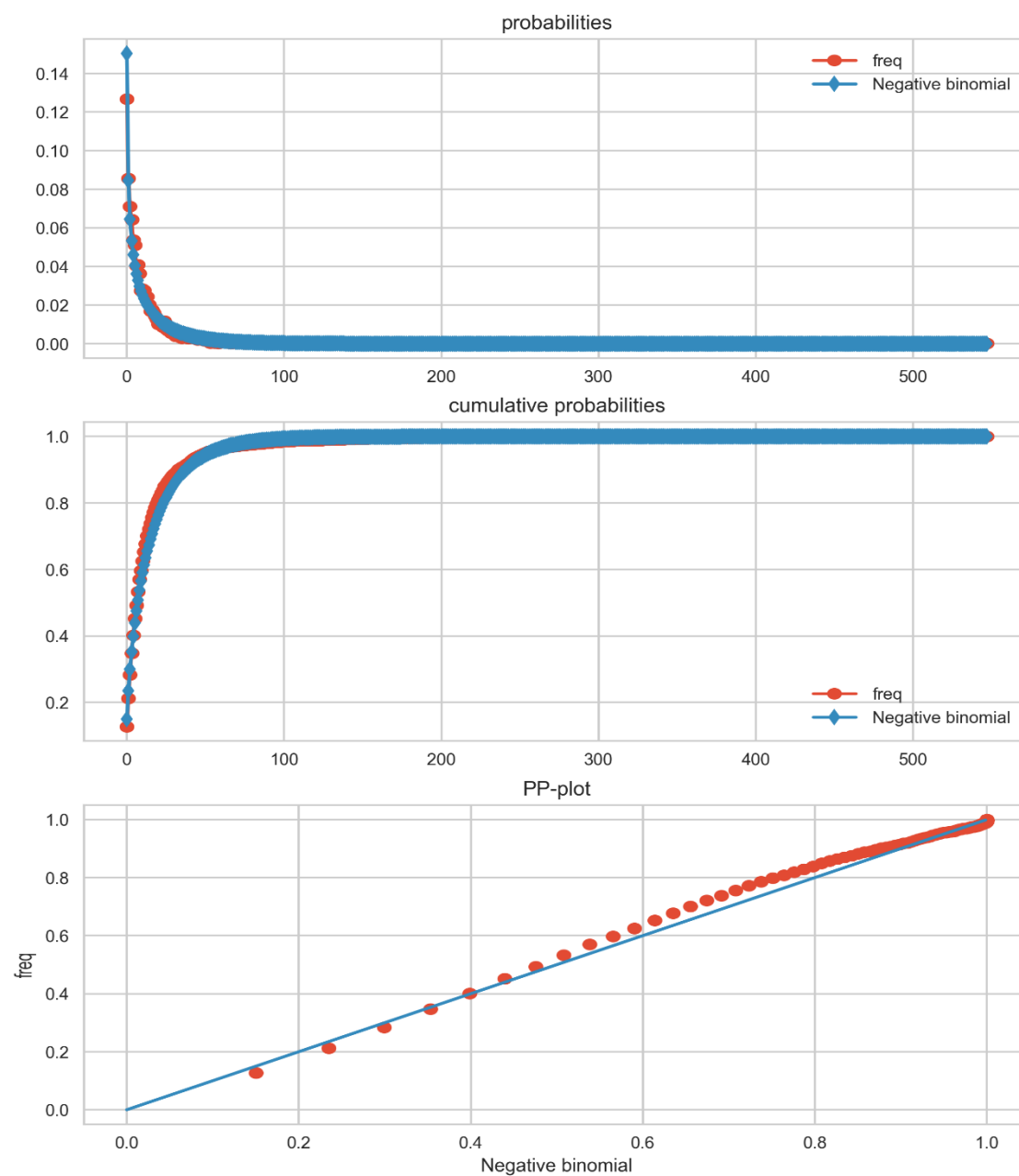
Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Figure A B.3. Zero inflated gen. poisson distribution fit of forward citations for 2011



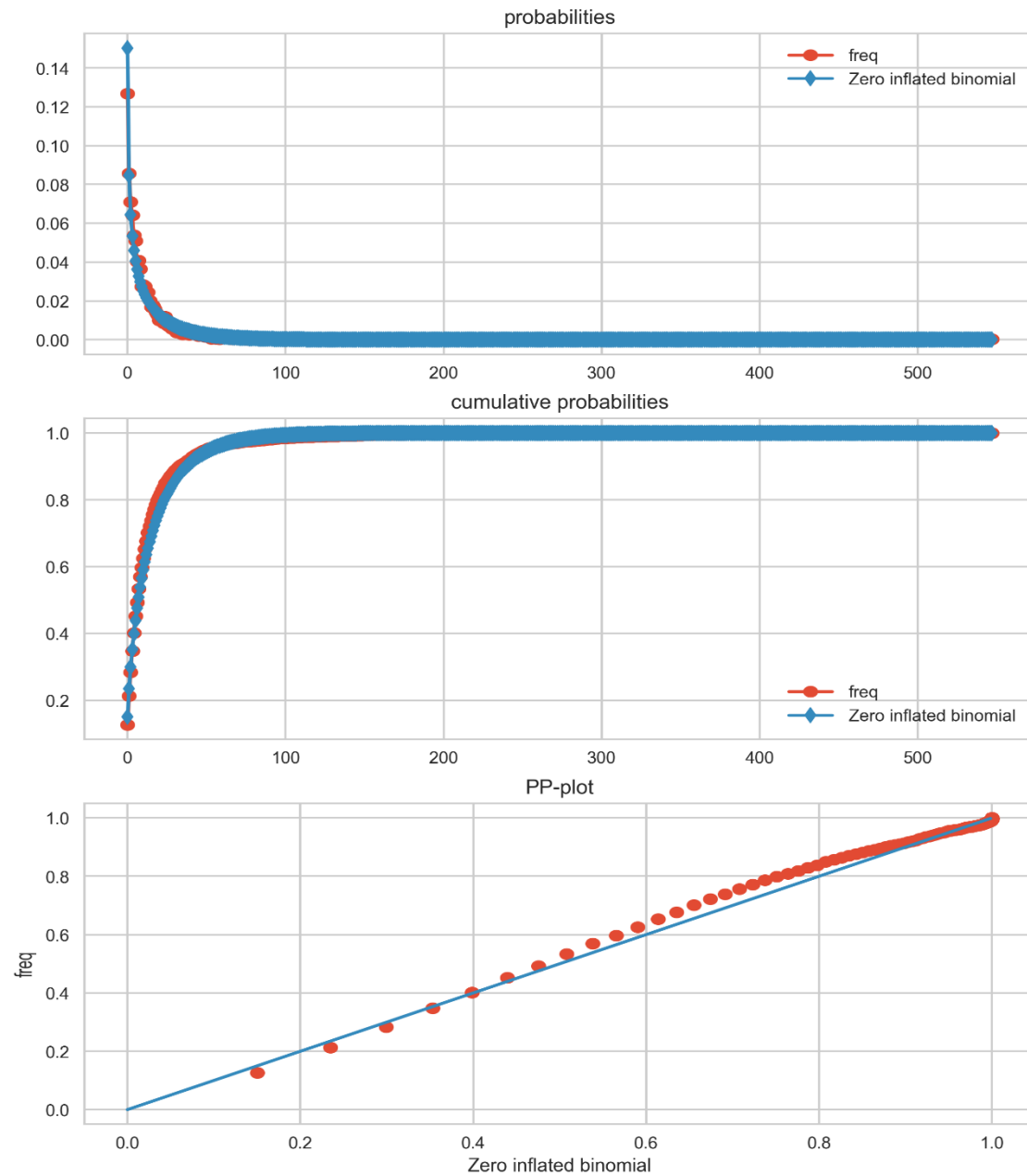
Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Figure A B.4. Negative binomial distribution fit of forward citations for 2011



Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Figure A B.5. Zero inflated negative binomial distribution fit of forward citations for 2011



Source: Author's calculations based on SCOPUS Custom Data, Elsevier, Version 5.2019

Table A B.1. Mean Absolute Error Test

Discrete distribution	Mean Absolute Error
Poisson	0.00243
Generalised poisson	0.00030
Inflated poisson	0.00208
Zero inflated gen. poisson	0.00188
Negative binomial	0.00033
Zero inflated negative binomial	0.00033