IIMB-WP N0. 701/2024

WORKING PAPER NO: 701

Unpacking novelty: university vs. industry publications and field-effects of industry publishing

Anubha Shokhand
Doctoral Student
Indian Institute of Management Bangalore
anubha.shokhand20@iimb.ac.in

Nilam Kaushik
Assistant Professor
Strategy
Indian Institute of Management Bangalore
nilam.kaushik@iimb.ac.in

Satyam Mukherjee
Associate Professor
School of Management and Entrepreneurship,
Shiv Nadar Institution of Eminence, Noida
satyam.mukherjee@snu.edu.in

Year of Publication – February 2024

Unpacking novelty: university vs. industry publications and field-effects of industry publishing

Anubha Shokhand[a], Nilam Kaushik[a], Satyam Mukherjee[b]

[a]Indian Institute of Management Bangalore, Bangalore, India

[b]School of Management and Entrepreneurship, Shiv Nadar Institution of Eminence, Greater Noida, India

Abstract

Novelty in academia is important for both scientific and technological breakthroughs. While publishing scientific articles is primarily a university-driven activity, industry also regularly indulges in publishing scientific articles in academia as a part of R&D efforts by firms. As an "outsider" to publishing in academia, industry is driven by its own experiences while pursuing scientific research and can bring novelty to academic knowledge. At the same time, the outsider position of industry scientists can also lead to legitimacy problems and discourage the publication of novel knowledge because academics in a field may not accept the new ideas put forth by industry. Further, the industry is often secretive in revealing its knowledge in scientific publications, which can impede novelty in industry-authored publications. In this paper, our initial research question focuses on determining whether the university or the industry generates a larger number of novel publications, and which of these entities' publications exhibit a higher level of novelty. Industry involvement in publishing can impact the novelty of university-authored publications. Further, the extent of industry participation in publishing scientific papers varies across different fields of scientific research. Our second research question explores the relationship between the extent of industry participation in publishing within a specific field and the degree of novelty in the scientific publications emerging from that field. Using 928,787 publications in STEM fields from MAG for scientific articles published in the year 2017, we find that industry-authored publications are less novel than university-authored publications. We also find that the novelty of publications is less in fields with high industry involvement in publishing than in those with low industry involvement in publishing. Our findings remain consistent across various econometric methods, measures, and subsets of data.

Keywords: Novelty; Collaboration; Industry participation; University; STEM; Regression models; Public policy

1 Introduction

Novelty in science is important for scientific (Uzzi et al., 2013; J. Wang et al., 2017) and technological breakthroughs (Ke, 2020), and economic progress (Jaffe, 1989). Scientific articles in academia are primarily published by university (Merton, 1971) yet industry firms also regularly engage in scientific research and publish scientific articles as a part of their R&D activities. However, university and industry firms are distinct institutions with different norms and incentives for scientific knowledge creation (Dasgupta and David, 1994; Sauermann and Stephan, 2013). Industry, being an "outsider" to publishing in academia, is driven by its own experiences while pursuing scientific research and can bring novelty to academic knowledge. At the same time, the outsider position of industry scientists can also lead to legitimacy problems and discourage the publication of novel knowledge because academics in a field may not accept the new ideas put forth by industry (Gassol, 2007). Further, industry is often secretive in revealing its knowledge in scientific publications (Dasgupta and David, 1994), which can hinder novelty in industry-authored publications. Therefore, the first research question posed in this paper examines whether the university or the industry generates a larger number of novel publications, and which of these entities' publications exhibit a higher level of novelty.

Our literature review did not find empirical studies that analyze whether university or industry produces a higher quantity of novel publications. Scant attention has been given to the question of whether publications by university or industry demonstrate a greater degree of novelty (Evans, 2010; Jee and Sohn, 2023). Moreover, prior work has only focused on publications in a particular field of research (Jee and Sohn, 2023) or on a particular topic (Evans, 2010). Large-scale empirical studies from multiple fields of research comparing novelty of knowledge created by scientists from university and industry are missing.

Further, industry firms do not uniformly participate in publishing scientific articles across all research fields (Rosenberg and Nelson, 1994). For example, ∼25% of the total publications in the field of Biomedical research are authored by industry (Arora et al., 2018) as opposed to the negligible participation of industry firms in publications in the area of High Energy Physics. The prevalence of industry-authored publications in a field is indicative of the extent to which knowledge created by industry is accepted in the field community (Merton, 1971). At the same time, the prevalence of industry-authored publications in a field could also indicate the diffusion of industry's norms in knowledge creation by university scientists in the field. Greater industry engagement in publishing could provide an impetus for university scientists to work on more applied research topics (Blumenthal et al., 1986). We argue that a high degree of industry engagement in publishing in a field could also potentially increase secrecy in knowledge sharing in the field. Therefore, our second research question explores the relationship between the extent of industry participation in publishing within a specific field and the degree of novelty in the scientific publications emerging from that field.

Using data of 928,787 scientific publications from Microsoft Academic Graph (MAG) (Sinha et al., 2015) in STEM fields published in the year 2017, we analyze and contrast both the quantity and degree of novelty in publications produced by university and industry. To capture the field-level effects of industry engagement in publishing on novelty, we examine both the quantity and the degree of novelty in publications across fields with varying levels of industry involvement in publishing.

Publishing novel scientific knowledge is risky and difficult (J. Wang et al., 2017). At the same time, novel scientific publications play an important role in advancing future science and technology through knowledge breakthroughs (Ke, 2020; Uzzi et al., 2013; J. Wang et al., 2017). Our first research question aims to understand whether university or industry lead cutting-edge advancements through the publication of novel scientific articles. The second research questions aims to explore the influence of industry on the novelty of publications in academia.

Our research questions also have policy implications. Governments and universities worldwide are emphasizing greater university-industry collaboration (Ankrah and Omar, 2015; Bercovitz and Feldman, 2006). Understanding whether university or industry publish a greater number of novel scientific articles can offer insights into the nature of knowledge that industry contributes to academia. Furthermore, comparing novelty among different fields of study, each with varying degrees of industry-authored publications can further our understanding of the impact of industry publishing on the trajectory of knowledge creation in academia. This can help policymakers in designing incentives and policies for university-industry collaboration.

2 Knowledge creation in academia

Merton (1971) elaborates the process of knowledge creation in academia. New knowledge creation in academia relies on the diffusion of research findings by publishing scientific articles. These scientific articles undergo rigorous scrutiny in the peer-review process prior to publication. This serves as a quality control mechanism and ensures the credibility of publications. In addition, the peer review process operates at the level of the specific field, with reviewers possessing expertise in the subject matter under consideration.

University predominantly leads the publication of scientific articles (Merton, 1971), yet industry also increasingly participates in publishing scientific articles (Arora et al., 2018; Rotolo et al., 2022). There is a rich history of many research-intensive firms, such as IBM and Bell Labs, which regularly published scientific articles and even received the Nobel Prize for their contributions to their respective fields of research (Plantec et al., 2022). However, university and industry are distinct institutions, each operating under unique incentives and logics when publishing scholarly articles in academia (Dasgupta and David, 1994; Sauermann and Stephan, 2013). Further, industry engagement in publishing scientific articles varies across fields of research.

In the subsequent subsections, we discuss the implication of differences in the institutional logics of university and industry on their scientific publications. We also elaborate on the variation of industry participation in publishing scientific articles across different fields.

2.1 Differences in university and industry publishing

Institutional logics are *"the socially constructed, historical pattern of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social reality"* (Thornton and Ocasio, 2008 (p. 101); Friedland, 1991). The institutional logics largely determine the values and behaviour of people in an organization (Friedland, 1991; Thornton and Ocasio, 2008).

The distinct motivations and incentives of university and industry guide the scientific publications of university and industry (Dasgupta and David, 1994). These variations in incentive systems manifest in the conflicting institutional logics of the two entities (Sauermann and Stephan, 2013).

University scientists are motivated primarily by the logic of academia, which rewards university scientists for publishing scientific knowledge by providing recognition and reputation, which is also associated with career promotions and monetary rewards (Merton, 1971; Stephan, 2012). University scientists are also driven by the joy of *puzzle-solving* and enjoy autonomy on their research topics (Sauermann and Stephan, 2013).

Academic publications are a public good (Stephan, 2012). Disclosure of R&D efforts by industry in the form of scientific publications can prevent firms from commercializing knowledge. At the same time, in their comprehensive review of existing literature, Rotolo et al. (Rotolo et al., 2022) highlight various incentives for industry to engage in academic publishing, including accessing external knowledge and resources, attracting and retaining scientists, supporting intellectual property strategies, building the firm's reputation, and facilitating commercialization strategies. However, the primary logic for the industry remains commercial interests (Dasgupta and David, 1994), and unlike university scientists, industry scientists do not necessarily have monetary rewards tied to the publication of scientific findings in academia (Dasgupta and David, 1994; Stephan, 2012). This suggests that disclosure of knowledge by industry can also contribute to the firm's commercial interests. This results in less creative freedom for industry scientists to choose their research topics (Aghion et al., 2008). Because of the public good nature of scientific publications, the industry carefully considers the advantages and disadvantages of publishing scientific articles (Hicks, 1995) and often does not disclose all research findings or delay disclosure of their research findings in the form of publications (Blumenthal et al., 1986; Dasgupta and David, 1994).

2.2 Fields of research

Scientific knowledge creation is influenced by scientific communities organized around shared interests, referred to as "fields" of research (Merton, 1971). These communities of shared interests or research fields have similar norms of language and type of knowledge created (Merton, 1971). Scientific publications in a field of research build on and converse with prior work in the field (Kuhn, 1962). Scientists also identify themselves as researchers in a (small set of) field(s). Scientists in a field also generally attend similar conferences and publish in similar journals. The peer-review process in science is also governed by the peer scientists (Merton, 1971) within the field. Therefore, the forces that drive the creation of scientific knowledge act primarily at the level of the field of research.

2.3 Industry publishing across fields of research

The industry is not uniformly interested in publishing across all fields. Since the industry publishes to meet commercial ends eventually (Dasgupta and David, 1994; Rotolo et al., 2022), it suggests that disclosing knowledge in the form of scientific publications by firms is a strategic decision (Polidoro Jr and Theeke, 2012). Firms carefully analyze the advantages and disadvantages of publishing scientific articles (Hicks, 1995) and decide the fields where they participate in publishing. Industry participation is significant in fields belonging to Pasteur's quadrant (Stokes, 1997), where scientific research directly benefits commercialization (Rosenberg, 1990). For example, industry-authored publications dominate approximately a quarter of the total publications in the field of Artificial Intelligence (Hartmann and

Henkel, 2020), which belongs to the Pasteur's quadrant (Bianchini et al., 2022; Trajtenberg, 2018). Firms are more likely to publish in applied fields than in basic research fields (Blumenthal et al., 1986; Sauermann and Stephan, 2013). Further, firms pursue research in fields that are closer to the technologies of their patent portfolio to broaden their patent scope (Baker and Mezzetti, 2005; Rotolo et al., 2022) and to gain legitimacy around the commercialization of their patents in regulated industries such as Pharmaceuticals (Gambardella, 1992; Polidoro Jr and Theeke, 2012). Scientific research in fields related to the technological elements in an uncertain technological landscape provides a theoretical understanding of the technological components and the interactions between the elements, which can help navigate the uncertain technological landscape (Fleming and Sorenson, 2004). For example, in the Pharmaceutical sector, the trial-and-error method of tweaking existing drug molecules has limited success. Research on how specific body parts function by firms provided them with a more informed approach, uncovering precise drug targets for effective drug discovery (Gambardella, 1992).

Further, the degree of industry involvement in publishing in a field signifies the extent to which knowledge created by industry is accepted in the field community (Merton, 1971). The extent of industry participation in publishing in a field also indicates a high degree of industry interaction with university scientists. Industry scientists publishing in a field engage with university scientists in the field by engaging with prior literature in their scientific articles, attending conferences, and informal interactions between scientists. These interactions can motivate acceptance of industry knowledge creation norms by the community of scientists in the field, which could impact novelty publications in the field.

3 Novelty in Science

Kuhn (1962) describes Science as *puzzle-solving*. The scientific knowledge creation process places importance on publishing novel and original work (Merton, 1971). Emphasis on novelty is also reflected in the peer-review norms of numerous top-ranked scientific journals, such as Nature (Nature, n.d.), which emphasize novelty as a criterion for acceptance of work for publication. Novelty is also important for scientific advances (Uzzi et al., 2013; J. Wang et al., 2017) technological breakthroughs (Ke, 2020), and economic progress (Jaffe, 1989). However, novel scientific publications are inherently risky as judging their future impact is challenging, and assessing the scientific impact of novelty often requires time (Rafols et al., 2012; Stephan et al., 2017; J. Wang et al., 2017).

Novelty is defined as an unusual recombination of elements of existing knowledge (Fleming, 2001; Henderson and Clark, 1990; Nelson, 1982; Schumpeter et al., 1939). Novelty of publications has been operationalized in various ways in prior work. Boudreau et al. (2016) treat MeSH (Medical Subject Headings) as a representation of existing knowledge and operationalize novelty of a research proposal by calculating the ratio pairs of MeSH which occur for the first time in the vast expanse of literature present in PubMed. Foster et al. (2015) categorize publications in the field of Biochemistry into five types (the first three as novel), jump (new chemicals), new bridge (new chemical connections between clusters), new consolidation (new connections between chemicals in the same cluster), repeat bridge (existing chemical connections across different clusters) and repeat consolidation (existing chemical connections within the same cluster).

A set of papers has considered journals as streams of existing knowledge. J. Wang et al. (2017) adjudge the novelty of scientific publications by evaluating whether a referenced journal pair occurs for the first time in a scientific publication and takes into account the level of difficulty in creating these new journal combinations, which is determined by the number of "common friends" that the paired journals had in terms of co-citations. Uzzi et al. (2013) operationalize novelty by evaluating whether referenced article pairs occur less frequently than expected by chance. They do so by comparing referenced article pairs with the ones from Monte Carlo simulations and converting the expectation into z-scores for each referenced journal pair. Boyack and Klavans (2014) employed the K50 statistics method to measure whether a referenced journal pair is novel.

3.1 Novelty of university and industry authored publications

To the best of our knowledge, prior work has not analyzed whether university or industry produces a higher quantity of novel publications. Further, only a few studies have explored which entity - university of industry- produces publications demonstrating a greater degree of novelty (Evans, 2010; Jee and Sohn, 2023; Perkmann and Walsh, 2009). Evans (2010) analyze 18,359 scientific publications related to a plant model *Arabidopsis thaliana* and argue that industry scientists favour theoretically unanticipated, novel research compared to their university counterparts. On a sample of 56,981 publications in the field of Artificial Intelligence Jee and Sohn (2023) show that publications coauthored by industry and university scientists are more novel than those by university scientists. Perkmann and Walsh (2009) interviewed 43 university scientists in Engineering involved in projects with industry and concluded that applied projects in collaboration with industry encourage exploration and lead to new ideas. However, these works only focus on publications in a particular field of research (Jee and Sohn, 2023; Perkmann and Walsh, 2009) or on a particular topic (Evans, 2010). Large-scale empirical studies from multiple research fields comparing the novelty of publications by scientists from university and industry are missing.

The impetus for scientific article publications comes from university as academic knowledge is primarily created by university scientists (Florida and Cohen, 1999; Merton, 1971). This places industry scientists in an advantageous “outsider" position for publishing novel scientific articles. Industry scientists are more likely to work on highly practical problems having a closer impact on innovation than on fundamental problems (Aghion et al., 2008; Evans, 2010). Industry scientists are highly motivated by their commercial interests and industry experience for scientific research (Aghion et al., 2008; Ahmadpoor and Jones, 2017; Evans, 2010) even when working on fundamental science problems (Ahmadpoor and Jones, 2017). For example, Pasteur’s germ theory of disease (that germs, as opposed to non-living things like dust, cause diseases) (Stokes, 1997) was a result of his observation of yeast proliferation during the fermentation process (Ahmadpoor and Jones, 2017). Such experiences of industry scientists bring novelty to academic knowledge because they are more likely not a part of the current research paradigm. These novel experiences motivate connections between theories or paradigms that were unanticipated by the scientific community (Evans, 2010), leading to atypical knowledge recombination.

Further, industry scientists also have more resources than university scientists. Greater resources broaden the horizon of problems that can be solved. For example, OpenAI’s GPT3, a model for text summarizing, translating, generating, and question answering, published as an academic research paper, is a massive language model with 175 Billion parameters and costs $\sim$4.5 million USD just for a single training run (Li, 2020). Therefore, industry scientists’ increased reach of problems is unavailable to university scientists.

However, industry scientists’ advantageous access to problems might not directly translate to novel knowledge creation because the industry’s incentive structure favours lesser creative freedom and limited disclosure of results (Sauermann and Stephan, 2013). While university scientists have the freedom to choose their research fields and problems following their interests (Aghion et al., 2008; Sauermann and Stephan, 2013), research fields and problems of industry scientists are influenced by the firm’s business objectives, which may not align with the preferences of a scientist (Aghion et al., 2008; Blumenthal et al., 1986; Sauermann and Stephan, 2013). Further, university scientists also have a *taste for science* and pay a premium to remain in academia (Roach and Sauermann, 2010). Therefore, university scientists have the incentive to create novel work, greater freedom to engage in any research project, and enjoy pursuing the creative work of research more than industry scientists.

Further, while the academic logic of the university favours publishing scientific knowledge (Merton, 1971), the industry leans towards secrecy and utilizes patents to safeguard its commercial interests (Dasgupta and David, 1994; Sauermann and Stephan, 2013). This can prevent novel knowledge created by industry scientists from reaching academic publications.

New scientific knowledge undergoes rigorous peer scrutiny (Merton, 1971). Being an outsider, the industry is likely to face legitimacy problems for peers (Barber, 1963; Cattani et al., 2014) as new ideas from industry scientists may face a backlash from the community of peers because of their novelty (Boudreau et al., 2016; J. Wang et al., 2017).

Therefore, while industry has access to new ideas, the institutional logic of industry may not allow the freedom to work on these new ideas or to publish the same in scientific articles. Further, the new ideas from the industry may also face backlash from peers in the review process.

3.2 Industry involvement and publication novelty in a field?

University and industry have conflicting institutional logics of scientific knowledge creation. While industry follows commercial logic, the university follows the logic of academia in publishing scientific articles. Unlike university scientists, industry researchers work on topics of commercial interest, have less creative freedom, and are more secretive in sharing knowledge. Nevertheless, industry can infuse new ideas into academia due to their "outsider" position. Consequently, the inquiry arises as to what extent the novel ideas from industry and its logic of knowledge production diffuse to university scientists.

Due to their unique motivations and goals driven by business interests, industry can be a source of new ideas and research topics in academia. (Evans, 2010). A high degree of industry involvement in publishing in a field suggests that industry scientists interact with the community of scientists in the field by conversing with prior work in their publications through the peer review process and formal and informal interactions with scientists at conferences and labs. Therefore, ideas in fields with high industry involvement in publishing are more likely to diffuse to university scientists than those with low industry involvement in publishing. New ideas from industry may also offer novel insights to university scientists by providing a larger pool of knowledge for recombination.

However, the infusion of new ideas from industry does not necessarily equate to diverse ideas. Industry publishes and encourages research on topics of industry interest instead of following the Mertonian way of puzzle-solving (1971). A stream of literature on the *skewing hypothesis* argues that industry participation in publishing encourages university scientists to work on applied topics (Blumenthal et al., 1986; Gulbrandsen and Smeby, 2005; Van Looy et al., 2004). Prior work also shows that collaboration with scientists from industry motivates university scientists to specialize in the field of study of the collaboration project (Bikard et al., 2019).

In addition, a narrow focus on only specific topics of industry research could have decreasing returns on the scale of these ideas (Banal-Estañol et al., 2015). Ideas from industry could also divert attention from topics that require deep research (Banal-Estañol et al., 2015). For example, the thematic diversity of scientific publications by industry in the field of Artificial Intelligence has become stagnant over time (Mateos-Garcia and Klinger, 2023). In addition, all ideas from the industry might not be scientifically important and may not gain any traction by the community of researchers in the field (Banal-Estañol et al., 2015). Consequently, in fields where industry is involved in publishing, introducing new ideas from industry may not enhance novelty within that field.

The academic logic of the university favours the publication of scientific research (Merton, 1971). On the contrary, industry is more secretive in sharing knowledge to safeguard commercial interests (Blumenthal et al., 1986; Dasgupta and David, 1994). University scientists who collaborate with industry scientists delay publishing in academia (Dasgupta and David, 1994) and are less open in sharing their knowledge with other university scientists (Blumenthal et al., 1986; Dasgupta and David, 1994; Welsh et al., 2008). A survey of 1200 faculty members over 40 universities in the US suggested that the work of faculty members with industry funding was more likely to result in trade secrets (Blumenthal et al., 1986). Disclosing knowledge in the form of patents could prevent follow-on research on the patented knowledge, also known as the problem of *anti-commons* (Heller and Eisenberg, 1998; Murray and Stern, 2007). Fields with high industry participation in publishing are more likely to face the problem of *anti-commons* than fields with low industry participation. In such fields, university scientists can use patents as a defensive strategy (Murray, 2010). For example, in the community of scientists working with transgenics mice, university scientists started patenting their transgenics mice instead of sharing information on developing such mice in the form of publications to safeguard against firms like DuPont, which had exclusive licensing rights for the Oncomouse, regulating the research on the topic through their patents (Murray, 2004). Further, this could also result in university scientists becoming more secretive in sharing their knowledge with other scientists (Blumenthal et al., 1986; Walsh and Huang, 2014). This suggests that the scientific community's commitment towards openness could decrease in fields with high industry involvement in publishing. Therefore, in fields with high industry participation in publishing, scientists may create novel knowledge but are less likely to publish novel scientific articles.

Significant industry involvement in publishing has the potential to infuse novel ideas into a field, consequently providing more options for novel recombinations for university scientists. However, significant industry participation in a field may also foster a culture of secrecy among scientists, hindering the free flow of knowledge exchange and impeding the collective progress of the field. Consequently, the influx of ideas from industry may exhibit a limited range of diversity, which, in turn, could lead to the diversion of attention from important research ideas. Therefore, we expect fields with higher industry involvement in publishing to have a lower quantity of novel publications and a lesser degree of novelty than those with lower industry involvement in publishing.

4 Empirical setting

We use the dataset of all scientific publications indexed in Microsoft Academic Graph (MAG)[1] (Sinha et al., 2015), July 11, 2020 version [2] published in the year 2017 across all fields. MAG also provides data on the affiliation of authors of publications by connecting with the Global Research Identifier Database (GRID) (Digital-science, 2017), a public database of organizations indulged in research along with their location and type [3][4]. We use the GRID organization type to classify authors' affiliations as 'university' or 'industry.' We have described the methodology for classifying GRID profiles into 'university' or 'industry' in detail in the Appendix (Section C.2). We consider a cross-sectional sample of the year 2017. We use a cross-sectional sample because the MAG database includes only the authors' latest affiliations. The MAG database closed in 2018, and our sample includes scientific articles published in 2017 to ensure the latest and complete data for the publications in the year.

MAG provides "research of study" codes for papers at four levels. 19 level 0 field codes indicate broader research areas such as Physics and Chemistry. There are 292 level 1 codes indicating narrower fields of research like Artificial Intelligence, Zoology, etc. Following van der Wouden and Youn (2023), we consider level 0 and level 1 codes together to indicate a field of research. For example, work on Zoology in the broader field of Biology refers to a field of research. Similarly, research on Artificial Intelligence in the field of Medicine is also considered a field of research. Level 2 codes (for example, "In silico" and "Rocket Turbine Engine") and level 3 codes (for example, "Magnetization transfer"

[1]MAG is one of the most extensive and widely used databases for bibliometric studies (Martín-Martín et al., 2021; Visser et al., 2021; K. Wang et al., 2019).
[2]Data source url: https://zenodo.org/records/3936556
[3]We use the Version 8, posted on November 21, 2018, in our study. Data source URL: https://doi.org/10.6084/m9.figshare.c.3812929.v8
[4]GRID classifies an organization into one of the following types: 'Government,' 'Company,' 'Bank,' 'Healthcare,' 'Archive,' 'Education,' 'Facility,' 'Nonprofit,' 'Other.'

and "Luciocephalus") are too narrow to classify a field of study. There are 4,226 level 0-level 1 pairs in the data. Since many level 0-level 1 pairs have very few publications categorized under them, we consider a level 0-level 1 pair to indicate a field of study if the number of publications in the field is greater than 10, the median number of publications in all fields. This categorization yields 2,166 fields of research, out of which 1,530 are STEM fields.[5]

There are 2,081,681 journal and conference publications indexed in the MAG database, with at least two references published in 2017. We include articles published in both journals and conferences to include fields such as Computer Science, wherein conferences are the primary medium for publishing findings. [6] We drop 67,677 publications from less cited journals (J. Wang et al., 2017). We consider journals and conferences whose total citation count till 2017 is less than the median citation count of all the journals and conferences indexed in MAG. This allows us to include only well-established journals and conferences, excluding conferences from fields where conferences are not the primary medium of conversing with prior work. Hereafter, both journals and conferences are referred to as journals for brevity in notation. Dropping publications for which information on the field of study is not present yields 1,534,877 publications, of which 1,349,592 publications belong to STEM fields. Further, we drop publications with missing author affiliation data. The final sample of publications in STEM fields for 2017 consists of 928,787 publications.

Dependent variable:

Novelty of a publication

We use a continuous and an ordinal scale variable to measure the novelty of a publication. For the continuous scale measure, use the measure of recombinant novelty as described in J. Wang et al. (2017), which considers journals as bodies of knowledge and a novel combination of journals in a publication indicates a novel recombination of existing knowledge. We identify journal pairs appearing for the first time (*new journal pair*) in the references of a publication[7]. For each of these new journal pairs, we calculate cosine similarity from the common journal friends of the new journal pairs [8]. The cosine similarity is a proxy for the ease of recombination of journal pairs. We subtract the cosine similarity

[5]We categorize STEM fields if the level 1 belongs to either of 'medicine,' 'geology,' 'biology,' 'environmental science,' 'physics,' 'chemistry,' 'materials_science,' 'computer science,' 'engineering' or 'mathematics.'

[6]We exclude 164 publications with more than 500 references.

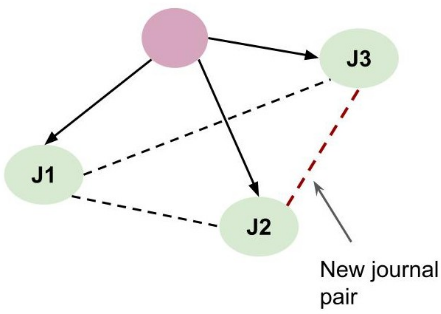
[7]To assess whether a journal pair appears for the first time in a publication, we compare journal pairs for publications in 2017 with the set of journal pairs in all publications across all fields in the past 20 years. We choose the cut-off as 20 years so that our estimates of novelty are comparable with those in J. Wang et al. (2017).

[8]Following J. Wang et al. (2017), we calculate common journal friends for a new journal pair from all scientific articles published in the past three years, 2014-16.

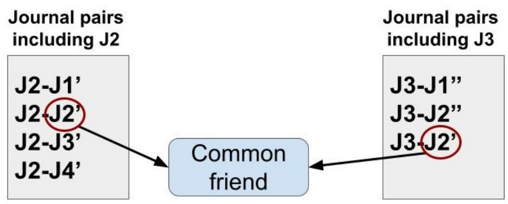
score for a new journal pair from 1 to get the novelty score for the new journal pairs. Finally, we evaluate the novelty score for a publication ($Novelty$) by adding novelty scores for all new journal pairs and scaling by the number of journal pairs in the publication[9].

$$Novelty = \frac{\sum_{ij_{new}}(1 - CosineSimilarity_{ij})}{N_{all\ journal\ pairs}}$$

Figure 2 presents an illustration for calculating the cosine similarity for novel journal pair in a publication. A publication has references which belong to three journals J1, J2 and J3. The corresponding journals pairs for the publication will J1-J2, J2-J3 and J1-J3. Let's assume that the journal pair J2-J3 is a novel journal pair. Next, we calculate the ease of recombining the journals J2 and J3. We list all the journals cited along with J2 in any publication in the past three years: J2-J1', J2-J2', J2-J3' and so on. We repeat the same for the journal J3 and get J3-J1", J2-J2", J2-J2' and so on. Here, J2' is a common journal friend of J2 and J3. We calculate cosine similarity for J2-J3 pair by counting the number of their journal friends (1) and scaling by the number of journals cited along with J2 (4) and the number of journals cited along with J3 (3). Therefore, the cosine similarity of the journal pair J2-J3 is 1/12.



(a)



(b)

Figure 1: (a) An illustration of journal pairs (J1-J2, J2-J3 and J1-J3) referenced in a scientific article. (b) An illustration of depicting common journal friends (J2') of a new journal pair (J2-J3).

We also construct an ordinal variable ($Novelty_{ordinal}$) for the novelty of a publication from the values of the continuous variable of novelty. We assign one of the three values of "Not novel," "Low Novelty," and "High Novelty" to $Novelty_{ordinal}$ for a publication. "Not Novel" is assigned to publications with a zero novelty score on the continuous scale. While "High Novelty" is assumed for publications whose novelty score on the continuous scale falls above the 75th percentile mark of the novel papers and "Low

[9] J. Wang et al. (2017) calculates two other alternatives of the novelty measure. We estimate the regression models in the paper for the two measures of novelty for robustness. The details of the two measures and regression estimates are in Section C.1 of the appendix.

Novelty" for the rest of the publications[10].

Independent variables:

Our model has two independent variables: first, a dummy indicating whether a publication is authored by industry or university ($Industry$), and second, the degree of industry involvement in publishing in a field ($PIndAuth_f$ or $PIndAuth_{max}$).

University vs. industry authored publications

The presence of an industry author on a paper indicates the industry's influence on the publication. Therefore, we refer to publications with all authors affiliated with a university as *university* or *university-authored* publications and publications with at least one industry author as *industry* or *industry-authored* publications. The independent variable $Industry$ is a dummy variable that takes the value 1 if the publication is authored by industry and 0 if the university authors the publication.

Industry involvement in publishing in a field

We calculate the degree of industry involvement in publishing in a field by measuring the percentage of publications with at least one industry author ($PIndAuth_f$).

A publication can belong to multiple fields of study. We observe that 46.69% of total publications in the sample belong to more than one field. Therefore, we use a measure of industry involvement unique to a publication ($PIndAuth_{max}$) by taking the 'maximum' value of $PIndAuth_f$ among the fields to which the paper belongs. We used 'maximum' because a publication classified in two fields of study (say one field with a high degree of industry involvement and the other with low industry involvement) is likely to be under the radar of industry authors from the field with higher industry involvement as well[11].

Controls

We also include control variables that confound the relationship between the independent and the dependent variables. We control for the publication's coauthor team size (T_{size}). Team size has an inverted-U-shaped relationship with the novelty of publications (Lee et al., 2015). University/industry publications and publications in different fields can have different tendencies to cite more or less a

[10] We also calculate $Novelty_{ordinal}$ with other percentiles as cut off for robustness. We plot the coefficient estimates of the variables of interest in Figures 6 and 7.

[11] We also run our analysis by considering 'minimum' value of $PIndAuth_f$ among the fields to which the paper belongs (Table C7).

number of references and journals, which can bias our estimates. Therefore, we also control for the number of references (N_{refs}) and the number of journals cited in the publication ($N_{journals}$). A publication classified under more than one field of research is more likely to draw on diverse and make novel recombinations. We account for this by controlling for the number of fields in the publication (Pub_{fields}) in our model. Some fields of research align themselves better for international collaboration of authors (Coccia and Wang, 2016). International collaboration can also impact the novelty of the publications (Wagner et al., 2019). Therefore, we also control for whether the coauthor team members are from different countries ($International$). The coauthors of a publication can belong to the same or distinct organizations. Some fields could be more aligned for inter-organizational collaboration than others (Abramo et al., 2013). A higher count of organizations of a publication's authors could bring together broader aspects of knowledge, which can impact novelty positively and could also lead to coordination problems, which can impact novelty negatively (Zhang et al., 2023). We account for this by controlling for the count of organizations of authors in a publication (N_{orgs}).

Firms are more likely to publish in more applied fields (Rosenberg, 1990). Therefore, we need to account for whether the field is applied. A good proxy for whether a field is applied is its technological impact. We use a dummy, $Applied_f$, which takes the value 1 (Applied field) if there is at least one follow-on patent on the publication(s) in the field and 0 (Basic field) otherwise.

Variable name	Description
Publication-level	
N_{refs}	Number of references
$N_{journals}$	Number of journals in the references
T_{size}	Number of authors
$International$	International collaboration
Pub_{fields}	Number of fields a paper belongs to
N_{orgs}	Number of organizations of paper authors
$Industry$	Dummy indicating if the publication is a university (0) or an industry publication (1
Novelty measures	
$Novelty$	Novelty
$Novelty_{Ordinal}$	Ordinal measure of novelty with three categories: Non-novel: $Novelty = 0$, Low-novelty: $Novelty \in (0, 75^{th}$ percentile), High-novelty: $Novelty \geq 75^{th}$ percentile
Field-level	
$Applied_f$	Dummy indicating if field is applied (1) or basic (0)
$PIndAuth$	Percentage of industry involvement in publishing in field
Publication-field-level	
$PIndAuth_{max}$	Maximum value of $PIndAuth_f$ among the fields to which the paper belongs

Table 1: Description of variables.

	All publications				Only Novel			
	mean	sd	min	max	mean	sd	min	max
T_{size}	4.802	3.137	1.000	57.000	4.617	2.827	1.000	42.000
N_{refs}	39.514	21.894	2.000	475.000	45.252	21.522	2.000	475.000
Pub_{fields}	1.575	0.686	1.000	7.000	1.589	0.704	1.000	7.000
$International$	0.265	0.442	0.000	1.000	0.270	0.444	0.000	1.000
$N_{journals}$	16.100	8.874	2.000	243.000	20.201	8.677	2.000	243.000
N_{orgs}	1.862	1.196	1.000	21.000	1.843	1.122	1.000	19.000
$Applied_f$	0.867	0.340	0.000	1.000	0.828	0.377	0.000	1.000
$Industry$	0.233	0.422	0.000	1.000	0.193	0.394	0.000	1.000
$PIndAuth_{max}$	24.252	11.064	0.000	64.706	22.840	10.732	0.000	64.706
$Novelty$	0.015	0.057	0.000	7.764	0.080	0.112	0.001	7.764
Observations	928,787				172,675			
Number of fields	1,530				1,518			

Table 2: Summary statistics

Descriptive statistics

Table 2 shows the summary statistics for the variables. As expected, the distribution of novelty is highly skewed. Only 20.54% percent of the total publications in the sample are novel. Of the total 928,787 publications, 21.53% have at least one industry author.

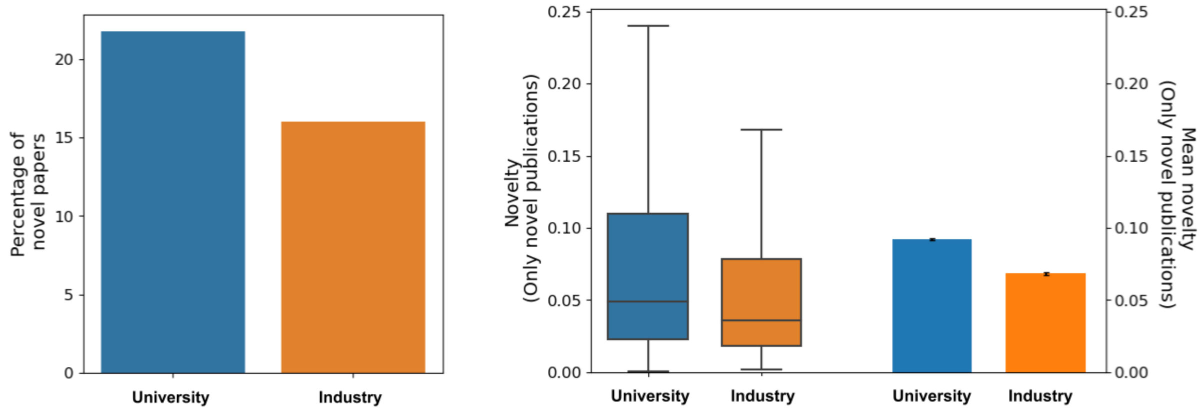
	$Novelty$	T_{size}	N_{refs}	Pub_{fields}	$International$	$N_{journals}$	N_{orgs}	$Applied_f$	$Industry$	$PIndAuth_{max}$
$Novelty$	1.000									
T_{size}	-0.076	1.000								
N_{refs}	0.001	0.261	1.000							
Pub_{fields}	0.006	-0.001	-0.013	1.000						
$International$	-0.012	0.263	0.099	0.000	1.000					
$N_{journals}$	0.131	0.286	0.686	-0.003	0.098	1.000				
N_{orgs}	-0.030	0.540	0.148	0.005	0.538	0.156	1.000			
$Applied_f$	-0.126	0.189	0.094	0.142	0.019	0.089	0.028	1.000		
$Industry$	-0.057	0.261	0.036	0.009	0.179	0.025	0.322	0.067	1.000	
$PIndAuth_{max}$	-0.129	0.328	0.118	0.129	-0.000	0.131	0.150	0.242	0.280	1.000
Observations	1058812									

All correlations are significant ($p < 0.01$) except for N_{refs}-$Novelty$, Pub_{fields}-T_{size}, Pub_{fields}-$International$ and $International$-$PIndAuth_{max}$.

Table 3: Correlation matrix

Table 3 presents bivariate correlations with the continuous measure for novelty, $Novelty$, and $PIndAuth_{max}$ measure for field industry involvement in publishing. All correlations are significant ($p < 0.01$) except for N_{refs}-$Novelty$, Pub_{fields}-T_{size}, Pub_{fields}-$International$ and $International$-$PIndAuth_{max}$ which are not significant. The correlations show a negative association between $Novelty$ and $Industry$. We observe a similar relationship between $Novelty$ and $PIndAuth_{max}$.

Figure 2a illustrates that university has a higher proportion of novel publications (21.78%) than industry (16.03%). Figure 2b depicts a similar trend in the median and mean novelty of publications by university and industry. The mean novelty of university publications is significantly higher than industry publications (Single-tailed t-test, $p < 0.01$). Further, the Kolmogorov-Smirnoff test indicates that university-authored and industry-authored publications belong to different populations ($p < 0.01$). Different distributions of novelty for university and industry-authored publications align with our argument that university and industry authors belong to different institutions, which can impact the novelty of their publications.



(a) (b)

Figure 2: Comparison of quantity of novel publications (a) and median and mean novelty (b) between publications by university and industry.

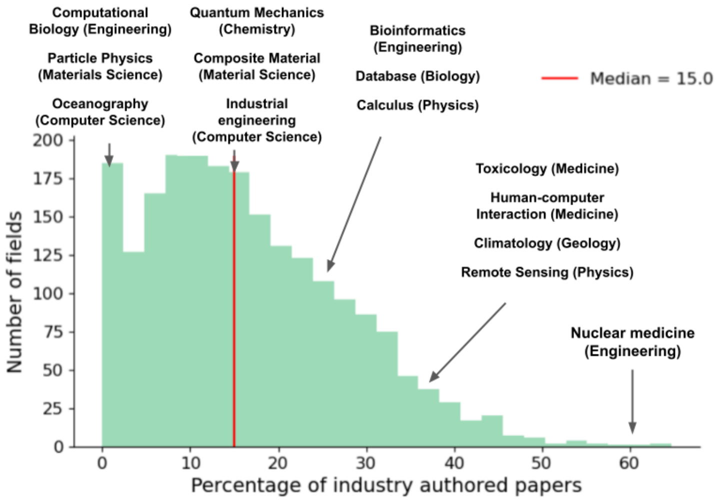
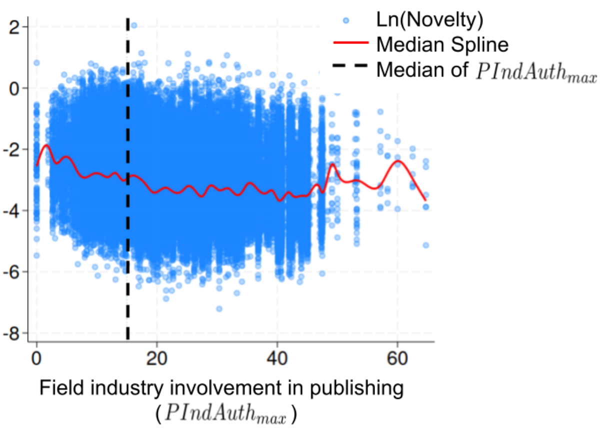


Figure 3: Variation of industry involvement in publishing across different fields.

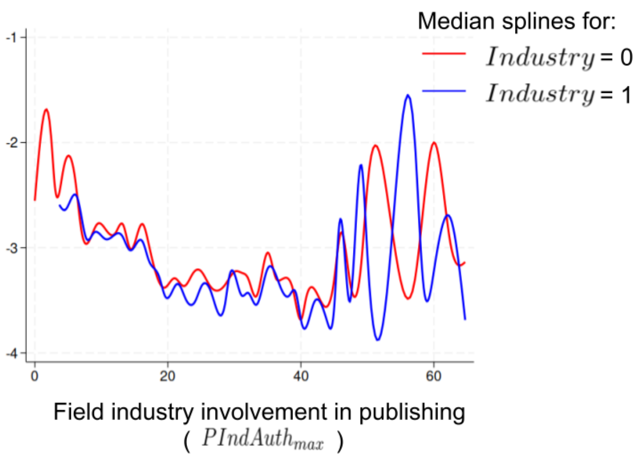
There are 1,530 fields of research in our data. Figure 3 shows a wide variation of industry involvement in publishing across different research fields. Of the 1530 fields, 1471 (96.14%) have at least one industry-authored publication. The median level of industry involvement in a field is 15.0, while the maximum is 64.71%, indicating a right-skewed distribution of industry involvement in publishing. Fields with less than 1% industry involvement in publishing include Particle Physics (Materials Science), Oceanography (Computer Science), Quantum Mechanics (Materials Science), and Computational Biology (Engineering). Fields such as Quantum Mechanics (Chemistry), Composite Material

(Materials Science), Anatomy (Physics), and Industrial Engineering (Computer Science) are around the median level of field industry involvement in publishing. At the higher end of field industry involvement in publishing, we have the fields of Database (Biology) and Bioinformatics (Engineering), with 25% of the publications authored by industry, Toxicology (Medicine), and Human-computer Interaction (Medicine) with 35% of the publications authored by industry and Nuclear Medicine (Engineering) with 60% industry-authored publications.

We present how the novelty of publications varies across different degrees of industry involvement in publishing in Figure 4 through a median spline plot of the novelty of publications. Figure 4 shows that the average novelty of publications decreases as we go from fields with low industry involvement in publishing to those with high industry involvement in publishing. The decrease in the novelty of publication slows after the median level of field industry involvement in publishing. There seems to be an increase in the average novelty of publications for fields with a very high degree of industry involvement in publishing (>45). However, very few fields have very high industry involvement in publishing. Figure 4b shows that the trends of the novelty of publications and degree of industry involvement in publishing in the field are similar across university and industry-authored publications.



(a)



(b)

Figure 4: (a) Scatter plot and median spline plot fitted to the scatter plot of $Log(Novelty)$ of novel publications vs. field's industry involvement in publishing ($PIndAuth_f$). (b) Median spline fitted to the scatter plot of $Log(Novelty)$ vs. $PIndAuth_f$ for university-authored publications ($Industry$=0) and industry-authored publications ($Industry$=1).

Empirical strategy

Around 80% of the publications in our data are non-novel, making the OLS regression model unsuitable. We use Ordered Logistic Regression, with the $Novelty_{ordinal}$ as the dependent variable to capture both the odds of a publication being novel ("Not Novel" vs. "Low Novelty") and the odds of a change in the level of novelty ("Low Novel" vs. "High Novelty") of a publication. For a finer analysis of the change in the degree of novelty with respect to the independent variables, we use the OLS regression model only on the novel publications.

Second, we also employ a Hierarchical Linear regression model with publications at the lower level and field at the higher level for robustness. Our data is cross-classified, wherein publications can belong to more than one field, each with multiple publications.

To incorporate the degree of industry involvement in a field, we use the $PIndAuth_{max}$ in the Ordered Logistic and OLS regression models and $PIndAuth_f$ in the Hierarchical Linear regression model.

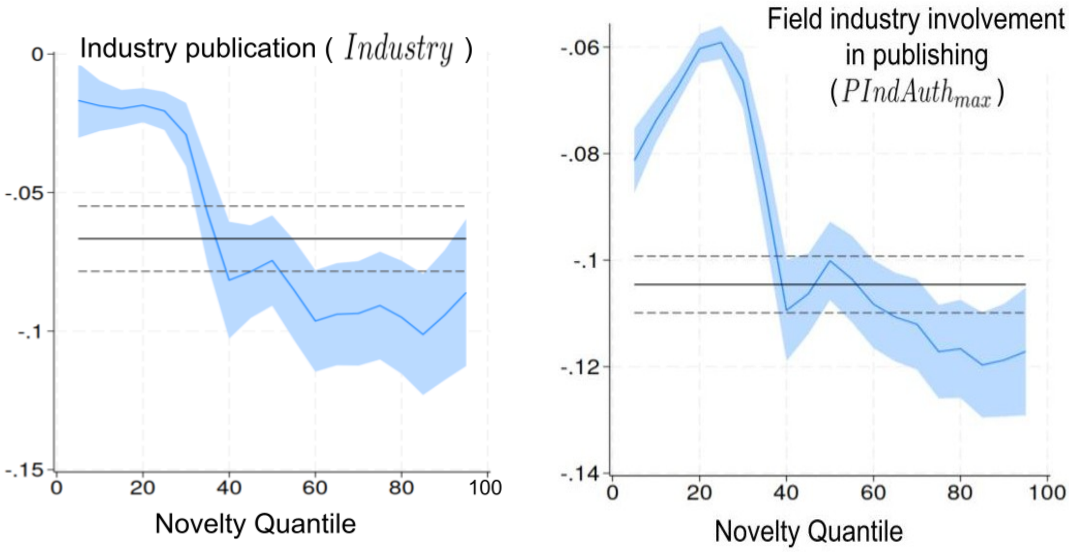
We cluster standard errors at the level of field in all the regression estimates. The standard deviation of novelty of publications is high (see Table 2), which suggests a broader and possibly skewed distribution of novelty of publications. Therefore, we use log transformation of the continuous measure of novelty ($Novelty$) as the dependent variable in the OLS and multilevel Linear regression models.

	(1) $Novelty_{ordinal}$	(2) $Novelty_{ordinal}$	(3) $Novelty_{ordinal}$	(4) $Ln(Novelty)$	(5) $Ln(Novelty)$	(6) $Ln(Novelty)$	(7) $Ln(Novelty)$	(8) $Ln(Novelty)$	(9) $Ln(Novelty)$
main									
T_{size}	-0.283***	-0.273***	-0.217***	-1.040***	-1.034***	-0.935***	-0.052***	-0.051***	-0.050***
	(0.018)	(0.017)	(0.016)	(0.048)	(0.048)	(0.046)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.075***	-0.076***	-0.081***	-0.352***	-0.352***	-0.356***	-0.453***	-0.453***	-0.453***
	(0.021)	(0.021)	(0.021)	(0.010)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.061***	0.062***	0.096***	0.077***	0.076***	0.091***	0.012***	0.012***	0.013***
	(0.022)	(0.022)	(0.021)	(0.008)	(0.008)	(0.007)	(0.003)	(0.003)	(0.003)
$International$	0.037***	0.038***	-0.034***	0.048***	0.050***	0.009	-0.012**	-0.010**	-0.011**
	(0.013)	(0.013)	(0.012)	(0.009)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.688***	0.686***	0.729***	0.664***	0.663***	0.658***	0.275***	0.274***	0.274***
	(0.031)	(0.031)	(0.030)	(0.032)	(0.032)	(0.031)	(0.011)	(0.011)	(0.011)
N_{orgs}	0.014	0.042***	0.042***	0.009**	0.022***	0.020***	0.000	0.005	0.006*
	(0.010)	(0.009)	(0.011)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.538***	-0.537***	-0.564***	-0.213***	-0.212***	-0.220***	-0.074***	-0.073***	-0.076***
	(0.078)	(0.077)	(0.079)	(0.026)	(0.026)	(0.021)	(0.011)	(0.011)	(0.011)
$Industry$		-0.243***	-0.115***		-0.125***	-0.055***		-0.052***	-0.050***
		(0.023)	(0.012)		(0.010)	(0.007)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.270***			-0.148***			
			(0.048)			(0.015)			
$PIndAuth$									-0.132***
									(0.008)
IMR				8.234***	8.240***	7.699***			
				(0.441)	(0.440)	(0.430)			
Constant				-4.183***	-4.163***	-3.871***	0.224***	0.233***	0.204***
				(0.232)	(0.232)	(0.227)	(0.011)	(0.011)	(0.011)
Cut1:	1.142***	1.090***	1.099***						
Not Novel|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.887***	2.836***	2.854***						
Low Novelty|High Novelty	(0.065)	(0.065)	(0.061)						
Ln(sd(Constant))							-1.217***	-1.223***	-1.350***
							(0.025)	(0.025)	(0.028)
Ln(sd(Residual))							-0.137***	-0.137***	-0.137***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,445	1,445	1,445	1,518	1,518	1,518
R^2				0.186	0.188	0.206			
AIC	1016481	1015322	1008144	454513	454057	450207	706204	706075	705868
BIC	1016586	1015439	1008273	454603	454158	450318	706309	706191	705995

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) models and Multilevel Linear regression (Models 7-9) models.



(a) (b)

Figure 5: Quantile regression estimates for (a) dummy indicating an industry publication ($Industry$) and (b) degree of industry involvement in publishing in a field ($PIndAuth_{max}$) at different quantiles of $Novelty$ (Blue) in comparison with the respective OLS estimates (Black).

Next, we present estimates from the regression models. Models 1-3 in Table 4 show estimates for the Ordered Logistic regression. Given the substantial size of our sample, we employ the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) tests to assess and compare the goodness of fit among various models (Raftery, 1995).

In Model 1, we begin with the control variables. We control for author team size (T_{size}), number of references (N_{refs}), number of fields a publication belongs to (Pub_{fields}), dummy indicating whether the authors are from different countries ($International$), number of journals in the references ($N_{journals}$), number of organizations of all the authors (N_{orgs}) of the publication and whether the publication belongs to an applied field ($Applied_f$).

For the control variable T_{size}, in line with prior work (Lee et al., 2015), we find that the novelty of publications ($Novelty_{ordinal}$) decreases with an increase in team size ($p < 0.01$). N_{refs} (number of references) and Pub_{fields} (number of fields a publication belongs to) also have statistically significant effects on novelty at $p < 0.01$ level. N_{orgs} and $International$ do not significantly impact the publication's novelty. $N_{journals}$ has a minimal statistical significance on the novelty of publications. This suggests that the number of journals in themselves does not lead to economically significant levels of novel recombinations. We also find that publications in applied fields are less novel than those in basic fields.

In Model 2, we add the first independent variable denoting whether the publication is authored by industry ($Industry$). From Model 1 to 2, the AIC value drops from 1,016,481 to 1,015,322 (> 10), and the BIC value drops from 1,016,586 to 1,015,439 (> 10), indicating a better fit of Model 2. In Model 3, we introduce the second independent variable, field industry involvement in publishing. In Model 3, AIC drops to 1,008,144 (> 10), and BIC drops to 1,008,273 (> 10), which suggests a better fit for Model 3.

The coefficient estimate for $Industry$ remains negative and significant ($p < 0.01$) in both Models 2 and 3. The odds ratio corresponding to the point estimate for $Industry$ in model 3 is 0.838. In other words, with a standard deviation increase in the field's industry participation in publishing (11.06 percentage points), the likelihood of a publication being highly novel with respect to a publication having low novelty (and a publication having low novelty with respect to a non-novel[12]) decrease by 16.2%. Therefore, on average, industry-authored publications are less likely to be novel than university-authored publications. Further, industry-authored publications are likely to have lower novelty than industry-authored publications.

The coefficient estimate for the degree of field industry participation in publishing (-0.116, with an odds ratio of 0.891) is negative and significant. This suggests that ceteris paribus, as field industry involvement increases, publications tend to have low levels of novelty as well as tend to be non-novel.

Next, we analyze how the degree of novelty of publications changes with the independent variables, $Industry$ and $PIndAuth_{max}$ for only novel publications through OLS estimates in models 4-6 of Table 4[13]. We begin with controls in Model 4 of Table 4. We add the independent variable, $Industry$, in Model 5 and the independent variable $PIndAuth_{max}$ in Model 6. From model 4 to 5, the AIC drops from 454513 to 454057 (> 10), and the BIC drops from 454603 to 454158 (> 10), indicating a better fit for model 5. We also see a drop in the AIC from 454057 to 450207 (> 10) and the BIC from 454158 to 450318 (> 10) between Models 5 and 6, suggesting a preference for Model 6. The effect size for $Industry$ indicates that university-authored publications are 5.5% more novel than industry-authored publications. At the field level, with one standard deviation (11.06 increase) in the field industry involvement in publishing ($PIndAuth_{max}$), the novelty of a publication is likely to drop by 14.8%.

[12] Ordered Logistic regression assumes the same odds ratio for all level thresholds.

[13] We include Heckman's correction in the OLS models. The first stage estimates are in Table A1.

Robustness tests

In the Ordered Logistic and OLS regression models, if a publication is classified in more than one field, we assign a unique field of research to the publication. We allow a publication to be classified under more than one field in Hierarchical Linear regression (Models 7-9) models presented in Table 4. The estimates from the Hierarchical Linear regression models are similar to those of OLS regression models presented in Models 4-6 Table 4.

We employ Quantile regression to explore whether highly novel publications drive our findings or if the distinctions are observed across all levels of novelty. In Figure 5a, we plot Quantile regression estimates for industry authorship ($Industry$) at different quantiles of $ln(Novelty)$ and find that coefficient estimates for $Industry$ remain negative and significant, indicating that industry-authored publications are consistently less novel than university-authored publications at all quantiles of novelty. We also see publications by industry are much less novel in comparison with university publications at higher quantiles of novelty ($> 40^{th}$ percentile) than at lower quantiles of novelty.

Figure 5b shows the Quantile regression estimates for the association of the degree of industry participation in publishing in a field with the novelty of publications at different quantiles of novelty. In line with our estimates for other regression models, we find that the coefficient estimates for industry participation in publishing in a field ($PIndAuth_{max}$) are negative and significant for all quantiles for novelty. This indicates that the observed decrease in novelty with an increase in a field's industry involvement in publishing is not driven by only highly novel publications but at all quantiles of novelty. We also find that the decrease in novelty with an increase in industry involvement in publishing in a field is more for higher quantiles of novelty ($> 40_{th}$ percentile) than at lower quantiles of novelty.

In Figure 4b, the novelty of university-authored publications is very high for fields with a low degree of industry participation in publishing (less than 10). This suggests that fields with low industry involvement in publishing could drive high average novelty of university publications. Further, the comparison of novelty between university and industry publications is unclear for fields with very high degrees of industry involvement in publishing (≥ 45) because of only 26 fields in the region. Therefore, we estimate the regression models for fields where industry involvement in publishing exceeds 10 and is less than 45 (Table B2).

We present regression estimates for alternative measures for novelty and author affiliation (see Appendix: Sections C.1 and C.2 for details) in Tables C1-B6. Our theoretical foundation primarily draws on works focused on industry engagement in academia. However, most research in this area has been carried out in the context of Biotechnology. This raises the possibility that our findings could be mainly driven by publications in the field of Biotechnology. To account for this, we estimate the regression models for the subsample excluding the fields of Biotechnology (see Table B1). Our results remain consistent across all the above robustness tests.

An alternative explanation is that the industry publishes in fields that are past the exploration stage and are in the exploitation stage. We account for this by including whether a field is applied as a control in the estimates.

We also estimate the regression models for alternative measures of novelty, author classifications, and field industry involvement in publishing in Section C, and the estimates remain similar.

Post-hoc analysis

The focus of this paper is publications in STEM fields. We conducted a post-hoc analysis on a broader set of publications, including STEM and non-STEM fields, and found that our findings remain consistent (see Table B3).

We also explore how the relationship between $Industry$ and the novelty of publications changes with the increase in the field's industry involvement in publishing. This involves including an interaction of the independent variables, $Industry$ and $PIndAuhth_{max}$ (or $PInAuth$) in the regression models (Model 2 in Tables D4-D6). We find that the odds ratio corresponding to the coefficient estimate for the interaction between $Industry$ and $PIndAuhth_{max}$ in Ordered Logistic regression (Model 2 in Table D4), 1.08, is statistically significant and greater than one, indicating a positive effect. However, the coefficient estimate for the interaction term in the OLS model (Model 2 in D5) is not statistically significant. In the hierarchical model (Model 2 in Table D6), we find a small statistically significant negative effect of the interaction between $Industry$ and $PIndAuth_{max}$ and novelty.

The discrepancy in the direction of the interaction effect between the Ordered Logistic regression and other models could be because the positive interaction effect could be driven by differences in "Not Novel" and "Low Novelty" publications rather than by both "Not Novel" - "Low Novelty" and "Low Novelty" - "High Novelty" thresholds. We account for this by using separate Logistic regression models to analyze whether industry-authored publications are more or less likely to be novel between fields with high and low industry involvement in publishing (Models 2 in Table D7) and whether among novel publications, industry-authored publications, are more or less likely to shift from "High Novelty" to "Low Novelty" category (Models 4 in Table D7).

For the Logistic regression model comparing non-novel and novel publications, we make the ordinal novelty measure binary by assigning the value 1 to publications in the "Low Novelty" and "High Novelty" categories and 0 otherwise. We use only novel publications to analyze the "Low Novelty" (0) - "High Novelty" (1) threshold by the Logistic regression model. In the Logistic regression model analyzing the "Non novel"-"Novel" threshold, we find that the odds ratio corresponding to the interaction between $Industry$ and $PIndAuth_{max}$ is statistically significant and greater than one (1.08) (Model 2 in Table D7). This indicates that with one standard deviation increase in field industry involvement (11.06), the odds of an industry-authored publication being novel increase by 8%. In the Logistic regression model analyzing the "Low Novelty"-"High Novelty" threshold, the coefficient estimate for the interaction between $Industry$ and $PIndAuth_{max}$ is statistically insignificant (Model 4 of Table D7), which is in line with the OLS estimates. Therefore, industry-authored publications are more likely to be novel in fields with high industry involvement in publishing than in those with low industry involvement in publishing. However, we do not find a consistent statistically significant difference in the novelty of industry-authored publications between fields with high industry involvement in publishing and those with low industry involvement in publishing.

We also investigate the boundary conditions for our findings by incorporating the interaction between different factors moderating the relationship between the independent variables, $Industry$ and $PIndAuhth_{max}$ (or $PIndAuth$), and novelty of publications (Tables D1-D6). We test for the moderation of $Industry$ with $PIndAuth_{max}$ (or $PIndAuth$) with $First\ Author\ Industry$, dummy indicating whether the first author of the industry publications is an industry author (1) or a university author (0). We also include gender diversity of the coauthor teams through a dummy indicating the presence (1) or absence (0) of a female coauthor on a publication ($Female$). We assign the authors' gender using Wiki-Gendersort, a gender database trained on names from Wikipedia (Bérubé et al., 2020)[14]. Wiki-Gendersort classifies a first name into one of the five categories: male, female, unknown, unisex, and initials. We only consider publications with names that can be classified as male or female for the regression estimates. We also test for the moderating effect of country diversity of the coauthor teams through a dummy indicating whether a coauthor team is national (0) or international (1)[15]. We also include *Repeated collaboration* whether a team collaboration is a repeated (1) or a first-time (0) collaboration by analyzing all collaborations in the past three years.

[14] Source URL: https://github.com/nicolasberube/Wiki-Gendersort

[15] We use the nationalities of the authors provided in the GRID database.

Models including the moderating effect of the variables *First Author Industry*, and *International* on *Industry* and $PIndAuth_{max}$ (or $PIndAuth$), do not have a better fit (as indicated by the AIC and BIC value), suggesting no significant moderating effect (Models 2 and 6 in Tables D1-D3, and Models 3 and 6 in Tables D4-D6).

We do not find a significant direct association of a female coauthor with the novelty of the publication in the OLS and Hierarchical linear regression estimates. The odds ratio corresponding to the coefficient estimate for *Female* in Ordered Logistic regression is statistically significant but close to 1 (1.06). We do not find a statistically significant moderation effect of *Female* with *Industry* or $PIndAuth_{max}$ (or $PIndAuth$).

The model, including the direct impact of *Repeated Collaborations* on novelty, has a better fit for only Ordered Logistic regression as indicated by the AIC and BIC values (Model7 in Tables D1-D3). The effect size for *Repeated Collaborations* in Ordered Logistic regression estimates shows that repeated collaborators only have a slight increase of 2% in their odds of publishing novel scientific articles compared to first-time collaborators. We do not find a statistically significant moderating effect of *Repeated Collaborations* on the relationship between *Industry* and novelty of publications. However, we find a small positive and significant interaction effect between *Repeated Collaboration* and $PIndAuth_{max}$ (or $PIndAuth$) for the OLS and Hierarchical linear regression models (Tables D5, D6) In the OLS model, with one standard deviation increase in field industry involvement (11.06), there is a slight increase of 0.8% in the novelty of publications by teams who have collaborated before.

5 Discussion

Our findings show that university not only publishes a greater number of novel publications compared to industry but also that the degree of novelty in university-authored publications surpasses that of industry-authored ones. However, in fields where industry publishing is high, the novelty of publications tends to be lower than in fields with lower industry publishing.

Our post-hoc analysis shows that the industry is more likely to publish novel scientific articles in fields with high industry involvement in publishing than those with low industry involvement in publishing. This suggests that the industry has an advantage in creating and publishing novel knowledge in fields of their interest.

We also explore the boundary conditions of the relationship between industry-authored publications and the novelty of publications, and the degree of a field's industry involvement in publishing and the novelty of publications in the field. We study the moderation effects of various factors, including whether the publication's first author is affiliated with the industry, the publication is female-authored, the team is international, and the team is a first-time collaboration. We do not find any economically significant moderation effect consistent across the regression models.

Publishing novel scientific knowledge is risky and difficult (J. Wang et al., 2017). At the same time, novel scientific publications are important for advancing future science and technology through knowledge breakthroughs (**ke**; Uzzi et al., 2013; J. Wang et al., 2017). Our findings show that universities play a pivotal role in generating novel knowledge. This insight reaffirms the critical function of academic institutions in pushing the boundaries of knowledge.

University-industry collaboration is a dynamic system of knowledge exchange (Perkmann et al., 2021). Our paper sheds light on the nature of academic knowledge that emanates from university and industry and how industry participation in publishing can shape knowledge creation in academia.

Policymakers around the world are pushing for greater university-industry collaboration. Our findings add to the debate on the extent to which university-industry collaboration is beneficial for academia. Our findings suggest that we may be missing out on the novelty of publications in fields with industry participation in publishing. However, this does not suggest that higher industry participation is not beneficial. Higher industry participation in publishing could be helpful in faster translation of science to technology. Therefore, policymakers can focus on hybrid models where rather than providing blanket incentives for all universities for greater collaboration with industry, policymakers could choose not to incentivize some universities for industry collaboration by providing them more government funding. More work is required to clarify science policy implications in this direction further. Future work can also focus on unfolding the mechanisms responsible for low novelty in fields with high industry involvement.

Our work also extends the literature on the role of outsiders in creating new knowledge. While prior work highlights the tension between new ideas vs. legitimacy for outsiders (Cattani et al., 2017), it overlooks the incentives of outsiders to create novel knowledge. Outsiders may have either similar or different objectives when it comes to generating novel knowledge. In domains like music or innovation, newcomers might align their goals with those of established entities, aspiring, for instance, to produce highly influential music or groundbreaking innovations. Conversely, the objectives of outsiders may deviate from those of incumbents. In the context of industry-driven scientific article publications, the firm's primary goal remains commercial interests.

Our work also contributes to hybridization of institutional logics (Murray, 2010; Pache and Santos, 2013). Prior work shows that conflicting logics university and industry can give rise spaces with hybrid logics (Murray, 2010). Our work adds to this stream of literature by potentially capturing outsiders' combined direct and indirect impact on the whole field. This is achieved through large-scale evidence across many fields, contributing to a broader understanding of the hybridization of institutional logics.

Our study is not without limitations. The study is constrained by the coverage of the MAG dataset. Further, the cross-sectional data prevents us from establishing a causal relationship in our findings. We use a bibliometric indicator for novelty. Future works should investigate whether a different quantification of novelty yields contrasting findings.

References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). The collaboration behaviors of scientists in italy: A field level analysis. *Journal of Informetrics*, *7*(2), 442–454.

Aghion, P., Dewatripont, M., & Stein, J. C. (2008). Academic freedom, private-sector focus, and the process of innovation. *The RAND Journal of Economics*, *39*(3), 617–635.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587.

Ankrah, S., & Omar, A.-T. (2015). Universities–industry collaboration: A systematic review. *Scandinavian Journal of Management*, *31*(3), 387–408.

Arora, A., Belenzon, S., & Patacconi, A. (2018). The decline of science in corporate r&d. *Strategic Management Journal*, *39*(1), 3–32.

Baker, S., & Mezzetti, C. (2005). Disclosure as a strategy in the patent race. *The Journal of Law and Economics*, *48*(1), 173–194.

Banal-Estañol, A., Jofre-Bonet, M., & Lawson, C. (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the uk. *Research Policy*, *44*(6), 1160–1175.

Barber, B. (1963). Resistance by scientists to scientific discovery. *American Journal of Clinical Hypnosis*, *5*(4), 326–335.

Bercovitz, J., & Feldman, M. (2006). Entpreprenerial universities and technology transfer: A conceptual framework for understanding knowledge-based economic development. *The journal of technology transfer*, *31*(1), 175–188.

Bérubé, N., Ghiasi, G., Sainte-Marie, M., et al. (2020). Wiki-gendersort: Automatic gender detection using first names in wikipedia.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, *51*(10), 104604.

Bikard, M., Vakili, K., & Teodoridis, F. (2019). When collaboration bridges institutions: The impact of university–industry collaboration on academic productivity. *Organization Science*, *30*(2), 426–445.

Blumenthal, D., Gluck, M., Louis, K. S., Stoto, M. A., & Wise, D. (1986). University-industry research relationships in biotechnology: Implications for the university. *Science*, *232*(4756), 1361–1366.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management science*, *62*(10), 2765–2783.

Boyack, K. W., & Klavans, R. (2014). Atypical combinations are confounded by disciplinary effects. *STI 2014 Leiden*, *64*.

Cattani, G., Ferriani, S., & Allison, P. D. (2014). Insiders, outsiders, and the struggle for consecration in cultural fields: A core-periphery perspective. *American Sociological Review*, *79*(2), 258–281.

Cattani, G., Ferriani, S., & Lanza, A. (2017). Deconstructing the outsider puzzle: The legitimation journey of novelty. *Organization Science*, *28*(6), 965–992.

Coccia, M., & Wang, L. (2016). Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences*, *113*(8), 2057–2061.

Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research policy*, *23*(5), 487–521.

Digital-science. (2017). Data-science. In *Grid. digital science. collection.* https://doi.org/\url{https://doi.org/10.6084/m9.figshare.c.3812929.v19}

Evans, J. A. (2010). Industry induces academic science to know less about more. *American Journal of Sociology*, *116*(2), 389–452.

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

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic management journal*, *25*(8-9), 909–928.

Florida, R., & Cohen, W. (1999). Engine or infrastructure? the university role in economic development. In *Industrializing knowledge. university–industry linkages in japan and the united states* (pp. 589–610). MIT Press, Cambridge MA/London.

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, *80*(5), 875–908.

Friedland, R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. *The new institutionalism in organizational analysis*, 232–263.

Gambardella, A. (1992). Competitive advantages from in-house scientific research: The us pharmaceutical industry in the 1980s. *Research policy*, *21*(5), 391–407.

Gassol, H. J. (2007). The effect of university culture and stakeholders' perceptions on university–business linking activities. *The Journal of Technology Transfer*, *32*, 489–507.

Gulbrandsen, M., & Smeby, J.-C. (2005). Industry funding and university professors' research performance. *Research policy*, *34*(6), 932–950.

Hartmann, P., & Henkel, J. (2020). The rise of corporate science in ai: Data as a strategic resource. *Academy of Management Discoveries*, *6*(3), 359–381.

Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? the anticommons in biomedical research. *Science*, *280*(5364), 698–701.

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

Hicks, D. (1995). Published papers, tacit competencies and corporate management of the public/private character of knowledge. *Industrial and corporate change*, *4*(2), 401–424.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, 957–970.

Jee, S. J., & Sohn, S. Y. (2023). Firms' influence on the evolution of published knowledge when a science-related technology emerges: The case of artificial intelligence. *Journal of Evolutionary Economics*, *33*(1), 209–247.

Ke, Q. (2020). Technological impact of biomedical research: The role of basicness and novelty. *Research Policy*, *49*(7), 104071.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University Press Chicago.

Lee, Y.-N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research policy*, *44*(3), 684–697.

Li, C. (2020, June 3). Openai's gpt-3 language model: A technical overview [Accessed: 2022-07-09].

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google scholar, microsoft academic, scopus, dimensions, web of science, and opencitations' coci: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*(1), 871–906.

Mateos-Garcia, J., & Klinger, J. (2023). Artificial intelligence in science: Is there a narrowing of ai research?

Merton, R. K. (1971). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.

Murray, F. (2004). The role of academic inventors in entrepreneurial firms: Sharing the laboratory life. *Research policy*, *33*(4), 643–659.

Murray, F. (2010). The oncomouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. *American journal of sociology*, *116*(2), 341–388.

Murray, F., & Stern, S. (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization*, *63*(4), 648–687.

Nature. (n.d.). Peer review [Accessed: October 30, 2023].

Nelson, R. R. (1982). *An evolutionary theory of economic change*. harvard university press.

Pache, A.-C., & Santos, F. (2013). Inside the hybrid organization: Selective coupling as a response to competing institutional logics. *Academy of management journal*, *56*(4), 972–1001.

Perkmann, M., Salandra, R., Tartari, V., McKelvey, M., & Hughes, A. (2021). Academic engagement: A review of the literature 2011-2019. *Research policy*, *50*(1), 104114.

Perkmann, M., & Walsh, K. (2009). The two faces of collaboration: Impacts of university-industry relations on public research. *Industrial and Corporate Change*, *18*(6), 1033–1065.

Plantec, Q., Weil, B., & Le Masson, P. (2022). Nobel laureates and the role of the industry in the emergence of new scientific breakthroughs. *Academy of Management Proceedings*, *2022*(1), 14322.

Polidoro Jr, F., & Theeke, M. (2012). Getting competition down to a science: The effects of technological competition on firms' scientific publications. *Organization Science*, *23*(4), 1135–1153.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research policy*, *41*(7), 1262–1282.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 111–163.

Roach, M., & Sauermann, H. (2010). A taste for science? phd scientists' academic orientation and self-selection into research careers in industry. *Research policy*, *39*(3), 422–434.

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

Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research policy*, *23*(3), 323–348.

Rotolo, D., Camerani, R., Grassano, N., & Martin, B. R. (2022). Why do firms publish? a systematic literature review and a conceptual framework. *Research Policy*, *51*(10), 104606.

Sauermann, H., & Stephan, P. (2013). Conflicting logics? a multidimensional view of industrial and academic science. *Organization science*, *24*(3), 889–909.

Schumpeter, J. A., et al. (1939). *Business cycles* (Vol. 1). Mcgraw-hill New York.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of microsoft academic service (mas) and applications. *Proceedings of the 24th international conference on world wide web*, 243–246.

Stephan, P. (2012). *How economics shapes science*. Harvard University Press.

Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*, *544*(7651), 411–412.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Thornton, P. H., & Ocasio, W. (2008). Institutional logics. *The Sage handbook of organizational institutionalism*, *840*(2008), 99–128.

Trajtenberg, M. (2018). Artificial intelligence as the next gpt: A political-economy perspective. In *The economics of artificial intelligence: An agenda* (pp. 175–186). University of Chicago Press.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.

van der Wouden, F., & Youn, H. (2023). The impact of geographical distance on learning through collaboration. *Research Policy*, *52*(2), 104698.

Van Looy, B., Ranga, M., Callaert, J., Debackere, K., & Zimmermann, E. (2004). Combining entrepreneurial and scientific performance in academia: Towards a compounded and reciprocal matthew-effect? *Research Policy*, *33*(3), 425–441.

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. *Quantitative science studies*, *2*(1), 20–41.

Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy*, *48*(5), 1260–1270.

Walsh, J. P., & Huang, H. (2014). Local context, academic entrepreneurship and open science: Publication secrecy and commercial activity among japanese and us scientists. *Research Policy*, *43*(2), 245–260.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., & Rogahn, R. (2019). A review of microsoft academic services for science of science studies. *Frontiers in Big Data*, *2*, 45.

Welsh, R., Glenna, L., Lacy, W., & Biscotti, D. (2008). Close enough but not too far: Assessing the effects of university–industry research relationships and the rise of academic capitalism. *Research policy*, *37*(10), 1854–1864.

Zhang, N., Cheng, L., Sun, C., Callaert, J., & van Looy, B. (2023). The role of inter-and intra-organisational networks in innovation: Towards requisite variety. *Scientometrics*, 1–20.

A Probit first stage

	Team affiliation present
T_{size}	-0.203***
	(0.003)
N_{refs}	0.029***
	(0.004)
Pub_{fields}	0.012***
	(0.003)
$N_{journals}$	0.119***
	(0.004)
Constant	0.476***
	(0.003)
Observations	254,107

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A1: Probit First Stage estimates.

B Subsample Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.283***	-0.273***	-0.217***	-1.040***	-1.035***	-0.936***	-0.052***	-0.051***	-0.050***
	(0.018)	(0.017)	(0.016)	(0.049)	(0.048)	(0.047)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.075***	-0.077***	-0.081***	-0.352***	-0.351***	-0.356***	-0.453***	-0.453***	-0.453***
	(0.021)	(0.021)	(0.021)	(0.010)	(0.010)	(0.009)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.061***	0.062***	0.096***	0.080***	0.080***	0.094***	0.012***	0.012***	0.013***
	(0.022)	(0.022)	(0.021)	(0.009)	(0.008)	(0.007)	(0.003)	(0.003)	(0.003)
$International$	0.037***	0.038***	-0.033***	0.048***	0.050***	0.009	-0.012**	-0.010**	-0.011**
	(0.013)	(0.013)	(0.012)	(0.009)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.688***	0.686***	0.729***	0.665***	0.665***	0.659***	0.275***	0.274***	0.274***
	(0.031)	(0.031)	(0.030)	(0.032)	(0.033)	(0.032)	(0.011)	(0.011)	(0.011)
N_{orgs}	0.014	0.042***	0.042***	0.009**	0.021***	0.020***	0.000	0.005	0.006*
	(0.010)	(0.009)	(0.011)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.538***	-0.537***	-0.564***	-0.213***	-0.212***	-0.220***	-0.074***	-0.073***	-0.076***
	(0.078)	(0.077)	(0.079)	(0.026)	(0.026)	(0.021)	(0.011)	(0.011)	(0.011)
$Industry$		-0.242***	-0.115***		-0.125***	-0.055***		-0.052***	-0.050***
		(0.023)	(0.012)		(0.010)	(0.007)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.270***			-0.148***			
			(0.048)			(0.015)			
$PIndAuth$									-0.132***
									(0.008)
IMR				8.267***	8.273***	7.729***			
				(0.445)	(0.445)	(0.434)			
Constant				-4.196***	-4.176***	-3.883***	0.224***	0.233***	0.204***
				(0.234)	(0.234)	(0.229)	(0.011)	(0.011)	(0.011)
Cut1:	1.143***	1.090***	1.099***						
Not Novel|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.887***	2.836***	2.854***						
Low Novelty|High Novelty	(0.065)	(0.065)	(0.062)						
Ln(sd(Constant))							-1.217***	-1.223***	-1.350***
							(0.025)	(0.025)	(0.028)
Ln(sd(Residual))							-0.137***	-0.137***	-0.137***
							(0.004)	(0.004)	(0.004)
Observations	927,019	927,019	927,019	172,215	172,215	172,215	274,363	274,363	274,363
Groups	1,526	1,526	1,526	1,440	1,440	1,440	1,518	1,518	1,518
R^2				0.186	0.188	0.206			
AIC	1014083	1012927	1005759	453207	452749	448908	706204	706075	705868
BIC	1014189	1013045	1005888	453297	452850	449019	706309	706191	705995

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B1: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for STEM publications excluding Biotech publications.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.292***	-0.283***	-0.236***	-0.141***	-0.136***	-0.098***	-0.055***	-0.054***	-0.053***
	(0.018)	(0.017)	(0.014)	(0.008)	(0.007)	(0.006)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.079***	-0.081***	-0.092***	-0.456***	-0.456***	-0.453***	-0.444***	-0.443***	-0.443***
	(0.022)	(0.022)	(0.022)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.084***	0.084***	0.112***	0.037***	0.037***	0.050***	0.012***	0.012***	0.013***
	(0.023)	(0.023)	(0.021)	(0.008)	(0.008)	(0.007)	(0.004)	(0.004)	(0.004)
$International$	0.044***	0.046***	-0.020*	0.030***	0.032***	-0.007	-0.012**	-0.010*	-0.011*
	(0.013)	(0.013)	(0.012)	(0.009)	(0.009)	(0.007)	(0.006)	(0.006)	(0.006)
$N_{journals}$	0.670***	0.667***	0.706***	0.174***	0.172***	0.197***	0.260***	0.259***	0.259***
	(0.031)	(0.030)	(0.030)	(0.014)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)
N_{orgs}	0.022**	0.049***	0.052***	0.017***	0.028***	0.027***	0.003	0.008**	0.008***
	(0.010)	(0.009)	(0.010)	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.499***	-0.502***	-0.575***	-0.197***	-0.197***	-0.234***	-0.087***	-0.086***	-0.091***
	(0.081)	(0.080)	(0.080)	(0.027)	(0.027)	(0.023)	(0.012)	(0.012)	(0.012)
$Industry$		-0.235***	-0.125***		-0.115***	-0.052***		-0.056***	-0.055***
		(0.023)	(0.012)		(0.009)	(0.007)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.289***			-0.163***			
			(0.060)			(0.020)			
$PIndAuth$									-0.133***
									(0.011)
Constant				0.137***	0.160***	0.204***	0.190***	0.200***	0.202***
				(0.020)	(0.019)	(0.014)	(0.012)	(0.012)	(0.012)
Cut1:	1.187***	1.133***	1.078***						
Not Novelty|Low Novelty	(0.059)	(0.059)	(0.056)						
Cut2:	2.966***	2.913***	2.868***						
Low Novelty|High Novelty	(0.068)	(0.068)	(0.065)						
Ln(sd(Constant))							-1.253***	-1.257***	-1.351***
							(0.027)	(0.027)	(0.030)
Ln(sd(Residual))							-0.135***	-0.135***	-0.135***
							(0.004)	(0.004)	(0.004)
Observations	830,837	830,837	830,837	154,611	154,611	154,611	240,238	240,238	240,238
Groups	1,191	1,191	1,191	1,152	1,152	1,152	1,186	1,186	1,186
R^2				0.170	0.172	0.190			
AIC	908885	907884	901625	409583	409231	405660	618973	618837	618720
BIC	908990	908001	901753	409663	409320	405760	619077	618951	618844

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for field's industry involvement between 10 and 45.

C Measures

C.1 Alternative novelty measures

Following J. Wang et al. (2017), we calculate two alternative measures of novelty. In the first variation, the *sum measure*, unlike the measure for novelty defined in Section 4, the difficulty of recombining new journal pairs is not scaled by the number of total journal pairs. This measure captures the degree of novelty of a publication assuming that more number of journal pairs need not be associated with higher novelty.

$$Novelty_{Sum/measure} = \sum_{ij_{new}} (1 - CosineSimilarity_{ij})$$

In the second variation, the *max measure*, instead of adding the difficulty of recombining a new journal pair across all new journal pairs, we captures the farthest new journal combination.

$$Novelty_{Max/measure} = Max_{ij_{new}} (1 - CosineSimilarity_{ij})$$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.309***	-0.296***	-0.209***	-1.080***	-1.074***	-0.957***	-0.056***	-0.054***	-0.053***
	(0.018)	(0.017)	(0.016)	(0.039)	(0.039)	(0.037)	(0.004)	(0.003)	(0.004)
N_{refs}	-0.097***	-0.097***	-0.089***	-0.415***	-0.414***	-0.409***	-0.447***	-0.447***	-0.446***
	(0.019)	(0.019)	(0.019)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.065***	0.065***	0.096***	0.096***	0.096***	0.103***	0.011***	0.011***	0.011***
	(0.019)	(0.019)	(0.017)	(0.007)	(0.007)	(0.006)	(0.003)	(0.003)	(0.003)
$International$	0.020	0.021	-0.040***	0.045***	0.046***	0.018***	-0.008*	-0.007	-0.007
	(0.013)	(0.013)	(0.012)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.726***	0.722***	0.751***	0.794***	0.793***	0.764***	0.296***	0.295***	0.295***
	(0.027)	(0.027)	(0.026)	(0.028)	(0.029)	(0.028)	(0.010)	(0.010)	(0.010)
N_{orgs}	0.006	0.034***	0.023**	0.005	0.016***	0.010**	-0.005*	-0.001	-0.001
	(0.010)	(0.009)	(0.011)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.759***	-0.750***	-0.630***	-0.253***	-0.248***	-0.184***	-0.099***	-0.099***	-0.091***
	(0.070)	(0.069)	(0.068)	(0.022)	(0.021)	(0.019)	(0.010)	(0.010)	(0.009)
$Industry$		-0.262***	-0.104***		-0.125***	-0.055***		-0.053***	-0.049***
		(0.021)	(0.011)		(0.009)	(0.006)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.315***			-0.143***			
			(0.041)			(0.012)			
imr				8.454***	8.462***	7.825***			
				(0.350)	(0.350)	(0.341)			
$PIndAuth$									-0.164***
									(0.008)
Constant				-4.187***	-4.172***	-3.892***	0.226***	0.233***	0.148***
				(0.181)	(0.181)	(0.177)	(0.009)	(0.009)	(0.009)
Cut1:	0.891***	0.845***	0.981***						
Not Novel|Low Novelty	(0.046)	(0.045)	(0.046)						
Cut2:	2.551***	2.507***	2.657***						
Low Novelty|High Novelty	(0.055)	(0.054)	(0.046)						
Ln(sd(Constant))							-1.257***	-1.263***	-1.427***
							(0.023)	(0.023)	(0.024)
Ln(sd(Residual))							-0.140***	-0.140***	-0.140***
							(0.004)	(0.004)	(0.004)
Observations	1,058,812	1,058,812	1,058,812	217,504	217,504	217,504	343,454	343,454	343,454
Groups	2158	2158	2158	2061	2061	2061	2151	2151	2151
R^2				0.203	0.205	0.220			
AIC	1233028	1231525	1220280	567901	567381	563290	882163	882015	881608
BIC	1233134	1231644	1220411	567994	567484	563403	882270	882133	881737

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the combined sample of STEM and non-STEM publications.

The measure for novelty as described in the Section 4

The estimates for Ordered Logistic regression, OLS with Heckman correction and Hierarchical Linear regression for the two measures are as follows.

C.2 Classification of author affiliations

We define a university as an institution whose primary objective is education and/or advancing science (Florida and Cohen, 1999).

GRID classifies research organizations into one of the following types: 'Government', 'Company', 'Bank', 'Healthcare', 'Archive', 'Education', 'Facility', 'Nonprofit', 'Other'. The classifications of the categories 'Government', 'Company', 'Bank', 'Healthcare' and 'Education' are self-explanatory. The category 'Archives' consists of organizations which hold archives like libraries and museums. In accordance with our definition of a university, we classify the category of 'Education' as university. We classify organizations in the categories of 'Government', 'Company', 'Bank', 'Healthcare', 'Healthcare' and 'Archive' as industry. We classify Government organizations, for example 'Health Canada' and

	(1) $Novelty_{ordinal}$	(2) $Novelty_{ordinal}$	(3) $Novelty_{ordinal}$	(4) $Ln(Novelty)$	(5) $Ln(Novelty)$	(6) $Ln(Novelty)$	(7) $Ln(Novelty)$	(8) $Ln(Novelty)$	(9) $Ln(Novelty)$
main									
T_{size}	-0.292***	-0.282***	-0.226***	-0.382***	-0.376***	-0.269***	-0.052***	-0.050***	-0.049***
	(0.018)	(0.018)	(0.016)	(0.043)	(0.043)	(0.043)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.042**	-0.043**	-0.046**	0.008	0.009	0.005	-0.018***	-0.017***	-0.017***
	(0.021)	(0.021)	(0.022)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)
Pub_{fields}	0.062***	0.063***	0.098***	0.043***	0.043***	0.059***	0.014***	0.013***	0.014***
	(0.023)	(0.022)	(0.021)	(0.009)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.038***	0.039***	-0.032***	0.046***	0.048***	0.004	-0.003	-0.001	-0.002
	(0.013)	(0.013)	(0.012)	(0.009)	(0.009)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.700***	0.698***	0.742***	0.381***	0.380***	0.374***	0.349***	0.348***	0.348***
	(0.031)	(0.031)	(0.031)	(0.026)	(0.026)	(0.026)	(0.013)	(0.014)	(0.013)
N_{orgs}	0.016	0.044***	0.044***	0.018***	0.032***	0.031***	0.004	0.011***	0.011***
	(0.010)	(0.009)	(0.011)	(0.005)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
$Applied_f$	-0.549***	-0.549***	-0.575***	-0.236***	-0.235***	-0.244***	-0.080***	-0.080***	-0.083***
	(0.078)	(0.078)	(0.079)	(0.029)	(0.028)	(0.022)	(0.012)	(0.012)	(0.011)
$Industry$		-0.244***	-0.116***		-0.140***	-0.065***		-0.065***	-0.063***
		(0.023)	(0.012)		(0.011)	(0.007)		(0.006)	(0.006)
$PIndAuth_{max}$			-0.273***			-0.159***			
			(0.049)			(0.017)			
IMR				2.144***	2.151***	1.569***			
				(0.380)	(0.379)	(0.382)			
$PIndAuth$									-0.130***
									(0.009)
Constant				-0.972***	-0.949***	-0.635***	0.233***	0.243***	0.215***
				(0.203)	(0.202)	(0.204)	(0.012)	(0.012)	(0.011)
/									
Cut1:	1.133***	1.081***	1.090***						
Not Novel|Low Novelty	(0.058)	(0.058)	(0.058)						
Cut2:	2.840***	2.789***	2.807***						
Low Novelty|High Novelty	(0.062)	(0.062)	(0.059)						
Ln(sd(Constant))							-1.192***	-1.198***	-1.315***
							(0.026)	(0.026)	(0.029)
Ln(sd(Residual))							-0.079***	-0.079***	-0.079***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,445	1,445	1,445	1,518	1,518	1,518
R^2				0.069	0.072	0.093			
AIC	1016235	1015065	1007828	477637	477138	473239	738023	737844	737658
BIC	1016341	1015183	1007957	477728	477239	473350	738128	737960	737784

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C1: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the *sum measure* of novelty.

'Indian Navy', as industry because such organizations may not have commercial objectives, but have similar norms. Government organizations are interested in applied work and often keep their knowledge secret for national security reasons.

The category of 'Nonprofit' consists of organizations like 'Baker IDI Heart and Diabetes Institute' and 'IIT Research Institute' and the category of 'Facility' includes organizations like 'Roswell Park Cancer Institute' and 'Special Astrophysical Observatory'. The category of 'Other' includes organizations like 'Groupe SEB' and 'Royal Meteorological Society'. The three categories include some organizations which can be classified as university, while others as industry. However, the percentage of organizations in these categories is relatively small (20%). We developed three classification systems to sort authors as university or industry authors, which are as follows:

Classification A University: 'Education', 'Facility', 'Nonprofit'

Industry: 'Government', 'Company', 'Bank', 'Healthcare', 'Archive', 'Other'

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.285***	-0.275***	-0.217***	-0.353***	-0.346***	-0.203***	-0.033***	-0.031***	-0.030***
	(0.018)	(0.018)	(0.016)	(0.037)	(0.037)	(0.031)	(0.005)	(0.005)	(0.005)
N_{refs}	-0.040**	-0.041**	-0.044**	0.017***	0.018***	0.013*	-0.013***	-0.012***	-0.012***
	(0.020)	(0.021)	(0.021)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
Pub_{fields}	0.062***	0.062***	0.099***	0.034***	0.033***	0.055***	0.016***	0.016***	0.016***
	(0.023)	(0.022)	(0.021)	(0.011)	(0.010)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.042***	0.043***	-0.032***	0.068***	0.070***	0.011	-0.008	-0.006	-0.007
	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.008)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.659***	0.657***	0.704***	0.072***	0.071***	0.063***	0.070***	0.069***	0.069***
	(0.030)	(0.029)	(0.029)	(0.013)	(0.013)	(0.012)	(0.005)	(0.005)	(0.005)
N_{orgs}	0.016*	0.045***	0.045***	0.045***	0.060***	0.058***	0.032***	0.038***	0.038***
	(0.010)	(0.009)	(0.011)	(0.006)	(0.006)	(0.005)	(0.003)	(0.004)	(0.004)
$Applied_f$	-0.523***	-0.522***	-0.550***	-0.174***	-0.173***	-0.185***	-0.068***	-0.067***	-0.070***
	(0.078)	(0.077)	(0.078)	(0.033)	(0.032)	(0.023)	(0.011)	(0.011)	(0.011)
$Industry$		-0.246***	-0.113***		-0.151***	-0.050***		-0.054***	-0.051***
		(0.023)	(0.012)		(0.014)	(0.009)		(0.007)	(0.007)
$PIndAuth_{max}$			-0.285***			-0.214***			
			(0.048)			(0.021)			
IMR				1.881***	1.888***	1.107***			
				(0.271)	(0.269)	(0.238)			
$PIndAuth$									-0.134***
									(0.009)
Constant				-0.878***	-0.853***	-0.431***	0.244***	0.252***	0.221***
				(0.149)	(0.147)	(0.128)	(0.011)	(0.010)	(0.011)
/									
Cut1:	1.156***	1.103***	1.112***						
Not Novel|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.785***	2.733***	2.751***						
Low Novelty|High Novelty	(0.070)	(0.071)	(0.065)						
Ln(sd(Constant))							-1.239***	-1.244***	-1.374***
							(0.036)	(0.036)	(0.039)
Ln(sd(Residual))							-0.068***	-0.068***	-0.068***
							(0.018)	(0.018)	(0.018)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,445	1,445	1,445	1,518	1,518	1,518
R^2				0.025	0.029	0.066			
AIC	1027720	1026520	1018541	485593	485034	478258	744036	743917	743708
BIC	1027826	1026637	1018670	485683	485134	478369	744141	744033	743835

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the *max measure* of novelty.

Classification B

University: ‘Education’, ‘Facility’

Industry: ‘Government’, ‘Company’, ‘Bank’, ‘Healthcare’, ‘Archive’, ‘Nonprofit’, ‘Other’

Classification C

University: ‘Education’, ‘Facility’, ‘Nonprofit’, ‘Other’

Industry: ‘Government’, ‘Company’, ‘Bank’, ‘Healthcare’, ‘Archive’

We use the classification system A to calculate whether a publication is authored by industry ($Industry$) and the degree of industry involvement in publishing in a field ($PIndAuth_f$ and $PIndAuth_{max}$) in the paper. We also estimate the main regression models the author affiliation classifications B and C below.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.283***	-0.271***	-0.210***	-1.040***	-1.034***	-0.931***	-0.052***	-0.050***	-0.050***
	(0.018)	(0.017)	(0.016)	(0.049)	(0.049)	(0.047)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.075***	-0.076***	-0.081***	-0.352***	-0.352***	-0.356***	-0.453***	-0.453***	-0.453***
	(0.021)	(0.021)	(0.021)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.061***	0.062***	0.100***	0.077***	0.076***	0.092***	0.012***	0.012***	0.013***
	(0.022)	(0.022)	(0.020)	(0.009)	(0.008)	(0.007)	(0.003)	(0.003)	(0.003)
$International$	0.037***	0.043***	-0.036***	0.048***	0.052***	0.010	-0.012**	-0.009*	-0.010*
	(0.013)	(0.013)	(0.012)	(0.009)	(0.008)	(0.006)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.688***	0.686***	0.740***	0.664***	0.663***	0.659***	0.275***	0.274***	0.274***
	(0.031)	(0.031)	(0.029)	(0.033)	(0.033)	(0.032)	(0.011)	(0.011)	(0.011)
N_{orgs}	0.014	0.047***	0.047***	0.009**	0.023***	0.022***	0.000	0.006*	0.006*
	(0.010)	(0.009)	(0.011)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.538***	-0.539***	-0.585***	-0.213***	-0.213***	-0.230***	-0.074***	-0.073***	-0.077***
	(0.077)	(0.076)	(0.078)	(0.026)	(0.025)	(0.020)	(0.011)	(0.011)	(0.011)
$Industry$		-0.266***	-0.130***		-0.126***	-0.057***		-0.056***	-0.053***
		(0.022)	(0.011)		(0.009)	(0.007)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.296***			-0.152***			
			(0.047)			(0.014)			
$PIndAuth$									-0.129***
									(0.008)
IMR				8.234***	8.239***	7.666***			
				(0.446)	(0.446)	(0.439)			
Constant				-4.183***	-4.160***	-3.845***	0.224***	0.234***	0.211***
				(0.234)	(0.235)	(0.232)	(0.011)	(0.011)	(0.011)
Cut1:	1.142***	1.079***	1.078***						
Not Novel|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.887***	2.825***	2.835***						
Low Novelty|High Novelty	(0.064)	(0.065)	(0.061)						
Ln(sd(Constant))							-1.217***	-1.223***	-1.350***
							(0.025)	(0.025)	(0.028)
Ln(sd(Residual))							-0.137***	-0.137***	-0.137***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,532	1,532	1,532	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.186	0.188	0.207			
AIC	1016481	1015012	1006327	454513	454025	449981	706204	706050	705841
BIC	1016586	1015129	1006456	454603	454125	450091	706309	706166	705967

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B1: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification B.

C.3 Estimates for Minimum of degree of industry involvement in publishing among fields of publication

C.4 Alternative $Novelty_{ordinal}$

We calculated $Novelty_{ordinal}$ from different percentiles of novelty scores and re-estimated the Ordered Logistic regression model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.292***	-0.281***	-0.219***	-0.382***	-0.375***	-0.264***	-0.052***	-0.050***	-0.049***
	(0.018)	(0.017)	(0.016)	(0.045)	(0.045)	(0.045)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.042**	-0.043**	-0.046**	0.008	0.009	0.004	-0.018***	-0.017***	-0.017***
	(0.021)	(0.021)	(0.021)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)
Pub_{fields}	0.062***	0.063***	0.101***	0.043***	0.043***	0.059***	0.014***	0.013***	0.014***
	(0.023)	(0.022)	(0.020)	(0.009)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.038***	0.044***	-0.035***	0.046***	0.050***	0.005	-0.003	0.000	-0.000
	(0.013)	(0.013)	(0.012)	(0.009)	(0.009)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.700***	0.698***	0.754***	0.381***	0.380***	0.376***	0.349***	0.348***	0.348***
	(0.032)	(0.032)	(0.030)	(0.028)	(0.028)	(0.027)	(0.013)	(0.014)	(0.013)
N_{orgs}	0.016*	0.050***	0.050***	0.018***	0.033***	0.032***	0.004	0.012***	0.012***
	(0.010)	(0.009)	(0.011)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.549***	-0.550***	-0.597***	-0.236***	-0.236***	-0.254***	-0.080***	-0.080***	-0.084***
	(0.078)	(0.077)	(0.079)	(0.028)	(0.027)	(0.022)	(0.012)	(0.012)	(0.011)
$Industry$		-0.268***	-0.132***		-0.141***	-0.066***		-0.069***	-0.066***
		(0.023)	(0.011)		(0.010)	(0.007)		(0.006)	(0.006)
$PIndAuth_{max}$			-0.299***			-0.164***			
			(0.048)			(0.016)			
$PIndAuth$									-0.127***
									(0.008)
IMR				2.144***	2.150***	1.534***			
				(0.400)	(0.399)	(0.408)			
Constant				-0.972***	-0.946***	-0.607***	0.233***	0.245***	0.222***
				(0.213)	(0.212)	(0.217)	(0.012)	(0.012)	(0.011)
Cut1:	1.133***	1.069***	1.069***						
Not Novel|Low Novelty	(0.058)	(0.057)	(0.058)						
Cut2:	2.840***	2.778***	2.789***						
Low Novelty|High Novelty	(0.062)	(0.062)	(0.059)						
Ln(sd(Constant))							-1.192***	-1.199***	-1.313***
							(0.026)	(0.026)	(0.028)
Ln(sd(Residual))							-0.079***	-0.079***	-0.079***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,532	1,532	1,532	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.069	0.072	0.094			
AIC	1016235	1014749	1005964	477637	477102	473010	738023	737813	737627
BIC	1016341	1014867	1006093	477728	477202	473121	738128	737929	737754

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification B with alternative 'Sum' measure of novelty.

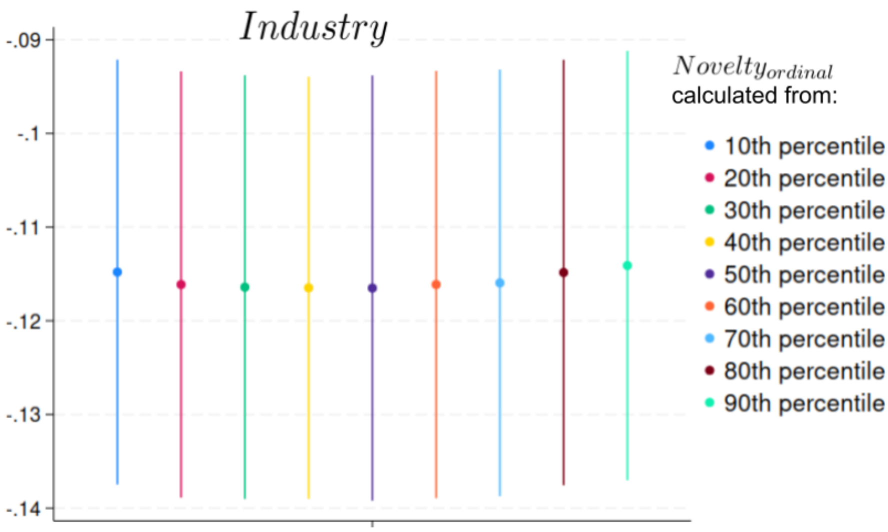


Figure 6: Ordered Logistic regression estimates for $Industry$.

D Moderators

	(1) $Novelty_{ordinal}$	(2) $Novelty_{ordinal}$	(3) $Novelty_{ordinal}$	(4) $Ln(Novelty)$	(5) $Ln(Novelty)$	(6) $Ln(Novelty)$	(7) $Ln(Novelty)$	(8) $Ln(Novelty)$	(9) $Ln(Novelty)$
T_{size}	-0.285***	-0.273***	-0.210***	-0.353***	-0.346***	-0.199***	-0.033***	-0.031***	-0.031***
	(0.018)	(0.017)	(0.016)	(0.037)	(0.037)	(0.030)	(0.005)	(0.005)	(0.005)
N_{refs}	-0.040**	-0.041**	-0.044**	0.017***	0.018***	0.012*	-0.013***	-0.012***	-0.012***
	(0.020)	(0.020)	(0.021)	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)
Pub_{fields}	0.062***	0.062***	0.102***	0.034***	0.033***	0.055***	0.016***	0.016***	0.016***
	(0.023)	(0.022)	(0.020)	(0.011)	(0.010)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.042***	0.048***	-0.034***	0.068***	0.072***	0.012	-0.008	-0.005	-0.006
	(0.013)	(0.013)	(0.012)	(0.011)	(0.011)	(0.008)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.659***	0.657***	0.715***	0.072***	0.071***	0.066***	0.070***	0.069***	0.069***
	(0.030)	(0.030)	(0.028)	(0.013)	(0.013)	(0.012)	(0.005)	(0.005)	(0.005)
N_{orgs}	0.016*	0.050***	0.051***	0.045***	0.061***	0.059***	0.032***	0.038***	0.038***
	(0.010)	(0.009)	(0.010)	(0.006)	(0.006)	(0.006)	(0.003)	(0.004)	(0.004)
$Applied_f$	-0.523***	-0.524***	-0.572***	-0.174***	-0.174***	-0.198***	-0.068***	-0.068***	-0.071***
	(0.077)	(0.076)	(0.077)	(0.032)	(0.031)	(0.023)	(0.011)	(0.011)	(0.011)
$Industry$		-0.268***	-0.127***		-0.146***	-0.047***		-0.050***	-0.048***
		(0.022)	(0.011)		(0.013)	(0.008)		(0.007)	(0.007)
$PIndAuth_{max}$			-0.312***			-0.217***			
			(0.047)			(0.020)			
$PIndAuth$									-0.130***
									(0.009)
IMR				1.881***	1.887***	1.071***			
				(0.265)	(0.263)	(0.229)			
Constant				-0.878***	-0.851***	-0.402***	0.244***	0.252***	0.228***
				(0.146)	(0.144)	(0.123)	(0.011)	(0.010)	(0.011)
Cut1:	1.156***	1.092***	1.090***						
Not Novel|Low Novelty	(0.057)	(0.056)	(0.056)						
Cut2:	2.785***	2.723***	2.733***						
Low Novelty|High Novelty	(0.070)	(0.070)	(0.065)						
Ln(sd(Constant))							-1.239***	-1.244***	-1.370***
							(0.036)	(0.036)	(0.038)
Ln(sd(Residual))							-0.068***	-0.068***	-0.068***
							(0.018)	(0.018)	(0.018)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,532	1,532	1,532	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.025	0.029	0.067			
AIC	1027720	1026218	1016642	485593	485045	478136	744036	743928	743722
BIC	1027826	1026335	1016771	485683	485146	478247	744141	744043	743848

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification B with alternative 'Max' measure of novelty.

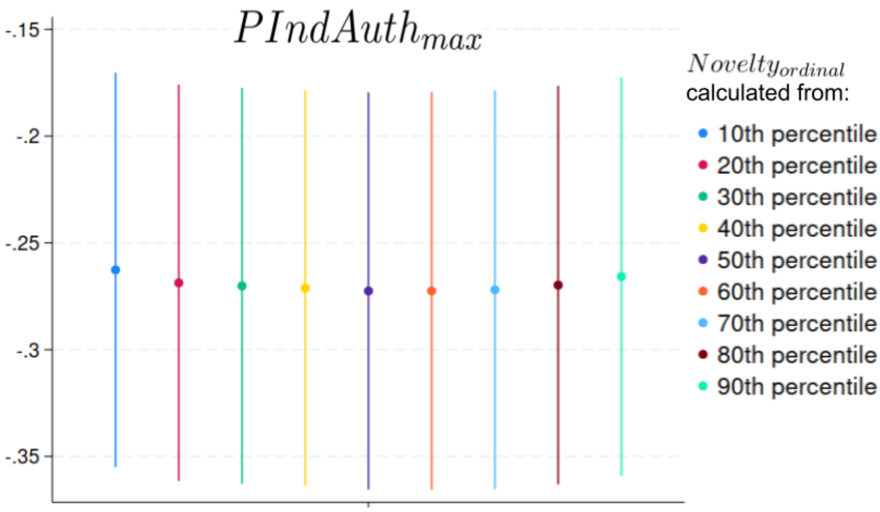


Figure 7: Ordered Logistic regression estimates for $PIndAuth_{max}$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
main									
T_{size}	-0.283***	-0.272***	-0.215***	-1.040***	-1.034***	-0.934***	-0.052***	-0.050***	-0.050***
	(0.018)	(0.017)	(0.016)	(0.048)	(0.048)	(0.046)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.075***	-0.076***	-0.081***	-0.352***	-0.352***	-0.356***	-0.453***	-0.453***	-0.453***
	(0.021)	(0.021)	(0.021)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Pub_{fields}	0.061***	0.062***	0.097***	0.077***	0.076***	0.091***	0.012***	0.012***	0.013***
	(0.023)	(0.022)	(0.021)	(0.009)	(0.009)	(0.007)	(0.003)	(0.003)	(0.003)
$International$	0.037***	0.038***	-0.034***	0.048***	0.049***	0.008	-0.012**	-0.010**	-0.011**
	(0.013)	(0.013)	(0.012)	(0.009)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.688***	0.686***	0.730***	0.664***	0.663***	0.657***	0.275***	0.274***	0.274***
	(0.031)	(0.031)	(0.030)	(0.032)	(0.032)	(0.031)	(0.011)	(0.011)	(0.011)
N_{orgs}	0.014	0.041***	0.041***	0.009**	0.021***	0.020***	0.000	0.005*	0.006*
	(0.010)	(0.009)	(0.011)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$Applied_f$	-0.538***	-0.537***	-0.558***	-0.213***	-0.212***	-0.217***	-0.074***	-0.073***	-0.076***
	(0.077)	(0.077)	(0.078)	(0.026)	(0.026)	(0.020)	(0.011)	(0.011)	(0.011)
$Industry$		-0.246***	-0.117***		-0.127***	-0.057***		-0.054***	-0.052***
		(0.023)	(0.012)		(0.010)	(0.007)		(0.005)	(0.005)
$PIndAuth_{max}$			-0.273***			-0.149***			
			(0.048)			(0.015)			
$PIndAuth$									-0.134***
									(0.008)
IMR				8.234***	8.239***	7.691***			
				(0.440)	(0.439)	(0.427)			
Constant				-4.183***	-4.163***	-3.870***	0.224***	0.232***	0.201***
				(0.232)	(0.232)	(0.226)	(0.011)	(0.011)	(0.011)
Cut1:	1.142***	1.091***	1.104***						
Not Novelty|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.887***	2.837***	2.859***						
Low Novelty|High Novelty	(0.064)	(0.065)	(0.061)						
Ln(sd(Constant))							-1.217***	-1.223***	-1.354***
							(0.025)	(0.025)	(0.028)
Ln(sd(Residual))							-0.137***	-0.137***	-0.137***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.186	0.188	0.206			
AIC	1016481	1015312	1008012	454513	454051	450163	706204	706070	705856
BIC	1016586	1015430	1008141	454603	454151	450273	706309	706185	705982

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B4: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification C.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
main									
T_{size}	-0.292***	-0.282***	-0.225***	-0.382***	-0.376***	-0.268***	-0.052***	-0.050***	-0.049***
	(0.018)	(0.018)	(0.016)	(0.045)	(0.045)	(0.045)	(0.004)	(0.004)	(0.004)
N_{refs}	-0.042**	-0.043**	-0.046**	0.008	0.009	0.005	-0.018***	-0.017***	-0.017***
	(0.021)	(0.021)	(0.022)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)
Pub_{fields}	0.062***	0.063***	0.098***	0.043***	0.043***	0.059***	0.014***	0.013***	0.014***
	(0.023)	(0.023)	(0.021)	(0.009)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.038***	0.039***	-0.033***	0.046***	0.048***	0.004	-0.003	-0.001	-0.002
	(0.013)	(0.013)	(0.012)	(0.009)	(0.009)	(0.007)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.700***	0.698***	0.743***	0.381***	0.380***	0.374***	0.349***	0.348***	0.348***
	(0.031)	(0.031)	(0.031)	(0.028)	(0.027)	(0.027)	(0.013)	(0.014)	(0.014)
N_{orgs}	0.016	0.044***	0.044***	0.018***	0.032***	0.030***	0.004	0.011***	0.011***
	(0.010)	(0.009)	(0.011)	(0.005)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
$Applied_f$	-0.549***	-0.548***	-0.570***	-0.236***	-0.235***	-0.240***	-0.080***	-0.079***	-0.082***
	(0.078)	(0.077)	(0.079)	(0.028)	(0.028)	(0.022)	(0.012)	(0.012)	(0.011)
$Industry$		-0.247***	-0.118***		-0.142***	-0.067***		-0.067***	-0.064***
		(0.023)	(0.012)		(0.010)	(0.007)		(0.006)	(0.006)
$PIndAuth_{max}$			-0.275***			-0.160***			
			(0.049)			(0.017)			
$PIndAuth$									-0.132***
									(0.009)
IMR				2.144***	2.151***	1.561***			
				(0.399)	(0.397)	(0.403)			
Constant				-0.972***	-0.949***	-0.634***	0.233***	0.243***	0.213***
				(0.213)	(0.212)	(0.215)	(0.012)	(0.011)	(0.011)
Cut1:	1.133***	1.082***	1.095***						
Not Novel|Low Novelty	(0.057)	(0.057)	(0.057)						
Cut2:	2.840***	2.790***	2.813***						
Low Novelty|High Novelty	(0.062)	(0.062)	(0.059)						
Ln(sd(Constant))							-1.192***	-1.198***	-1.318***
							(0.026)	(0.026)	(0.029)
Ln(sd(Residual))							-0.079***	-0.079***	-0.079***
							(0.004)	(0.004)	(0.004)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.069	0.072	0.093			
AIC	1016235	1015056	1007692	477637	477131	473191	738023	737838	737647
BIC	1016341	1015173	1007821	477728	477232	473302	738128	737954	737773

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification C with alternative 'Sum' measure of novelty.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.285***	-0.274***	-0.215***	-0.353***	-0.346***	-0.201***	-0.033***	-0.031***	-0.030***
	(0.018)	(0.018)	(0.016)	(0.038)	(0.038)	(0.031)	(0.005)	(0.005)	(0.005)
N_{refs}	-0.040*	-0.041**	-0.044**	0.017***	0.018***	0.012*	-0.013***	-0.012***	-0.012***
	(0.021)	(0.021)	(0.021)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
Pub_{fields}	0.062***	0.062***	0.099***	0.034***	0.033***	0.055***	0.016***	0.016***	0.016***
	(0.023)	(0.023)	(0.021)	(0.011)	(0.011)	(0.008)	(0.004)	(0.004)	(0.004)
$International$	0.042***	0.043***	-0.032***	0.068***	0.070***	0.011	-0.008	-0.006	-0.007
	(0.013)	(0.013)	(0.011)	(0.012)	(0.011)	(0.008)	(0.005)	(0.005)	(0.005)
$N_{journals}$	0.659***	0.657***	0.704***	0.072***	0.071***	0.063***	0.070***	0.069***	0.069***
	(0.030)	(0.029)	(0.029)	(0.013)	(0.013)	(0.011)	(0.005)	(0.005)	(0.005)
N_{orgs}	0.016*	0.044***	0.045***	0.045***	0.060***	0.058***	0.032***	0.038***	0.038***
	(0.010)	(0.009)	(0.011)	(0.006)	(0.006)	(0.005)	(0.003)	(0.004)	(0.004)
$Applied_f$	-0.523***	-0.522***	-0.544***	-0.174***	-0.172***	-0.180***	-0.068***	-0.067***	-0.069***
	(0.077)	(0.077)	(0.077)	(0.033)	(0.032)	(0.023)	(0.011)	(0.011)	(0.011)
$Industry$		-0.249***	-0.115***		-0.155***	-0.054***		-0.058***	-0.055***
		(0.023)	(0.012)		(0.014)	(0.009)		(0.007)	(0.007)
$PIndAuth_{max}$			-0.288***			-0.215***			
			(0.048)			(0.021)			
$PIndAuth$									-0.136***
									(0.010)
IMR				1.881***	1.888***	1.095***			
				(0.273)	(0.270)	(0.236)			
Constant				-0.878***	-0.853***	-0.429***	0.244***	0.252***	0.219***
				(0.150)	(0.148)	(0.127)	(0.011)	(0.010)	(0.011)
Cut1:	1.156***	1.104***	1.117***						
Not Novel|Low Novelty	(0.056)	(0.056)	(0.056)						
Cut2:	2.785***	2.734***	2.757***						
Low Novelty|High Novelty	(0.070)	(0.070)	(0.064)						
Ln(sd(Constant))							-1.239***	-1.244***	-1.378***
							(0.036)	(0.036)	(0.039)
Ln(sd(Residual))							-0.068***	-0.068***	-0.068***
							(0.018)	(0.018)	(0.018)
Observations	928,787	928,787	928,787	172,675	172,675	172,675	274,363	274,363	274,363
Groups	1,531	1,531	1,531	1,448	1,448	1,448	1,518	1,518	1,518
R^2				0.025	0.029	0.067			
AIC	1027720	1026512	1018401	485593	485018	478158	744036	743903	743688
BIC	1027826	1026630	1018531	485683	485118	478269	744141	744019	743814

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B6: Ordered Logistic regression (Models 1-3), OLS regression (Models 4-6) and Hierarchical Linear regression (Models 7-9) models for the Classification C with alternative 'Max' measure of novelty.

	(1)	(2)	(3)	(4)	(5)	(6)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.283***	-0.273***	-0.205***	-1.040***	-1.034***	-0.925***
	(0.017)	(0.017)	(0.015)	(0.049)	(0.049)	(0.046)
N_{refs}	-0.075***	-0.076***	-0.081***	-0.352***	-0.352***	-0.357***
	(0.021)	(0.021)	(0.022)	(0.009)	(0.009)	(0.008)
Pub_{fields}	0.061***	0.062***	0.038*	0.077***	0.076***	0.054***
	(0.022)	(0.022)	(0.020)	(0.009)	(0.008)	(0.007)
$International$	0.037***	0.038***	-0.043***	0.048***	0.050***	0.006
	(0.012)	(0.013)	(0.011)	(0.009)	(0.008)	(0.007)
$N_{journals}$	0.688***	0.686***	0.737***	0.664***	0.663***	0.658***
	(0.032)	(0.031)	(0.032)	(0.034)	(0.034)	(0.033)
N_{orgs}	0.014	0.042***	0.040***	0.009**	0.022***	0.019***
	(0.010)	(0.009)	(0.010)	(0.004)	(0.004)	(0.005)
$Applied_f$	-0.538***	-0.537***	-0.552***	-0.213***	-0.212***	-0.210***
	(0.076)	(0.077)	(0.076)	(0.026)	(0.025)	(0.020)
$Industry$		-0.243***	-0.093***		-0.125***	-0.047***
		(0.023)	(0.012)		(0.009)	(0.006)
$PIndAuth_{min}$			-0.030***			-0.162***
			(0.004)			(0.015)
IMR				8.234***	8.240***	7.646***
				(0.452)	(0.451)	(0.427)
Constant				-4.183***	-4.163***	-3.851***
				(0.237)	(0.237)	(0.226)
Cut1:	1.142***	1.090***	0.455***			
Not Novel|Low Novelty	(0.057)	(0.057)	(0.113)			
Cut2:	2.887***	2.836***	2.213***			
Low Novelty|High Novelty	(0.066)	(0.065)	(0.126)			
Observations	928,787	928,787	928,787	172,675	172,675	172,675
R^2				0.186	0.188	0.209
AIC	1016481	1015322	1005788	454513	454057	449470
BIC	1016586	1015439	1005917	454603	454158	449581

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C7: Ordered Logistic regression (Models 1-3), and OLS regression (Models 4-6) models for the Minimum of degree of industry involvement in publishing among fields of a publication.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$
T_{size}	-0.217***	-0.217***	-0.131***	-0.143***	-0.142***	-0.217***	-0.222***	-0.222***	-0.150***
	(0.016)	(0.016)	(0.021)	(0.020)	(0.020)	(0.016)	(0.016)	(0.016)	(0.019)
N_{refs}	-0.081***	-0.081***	-0.043*	-0.043*	-0.043*	-0.081***	-0.081***	-0.081***	-0.043*
	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.022)
Pub_{fields}	0.096***	0.096***	0.102***	0.102***	0.103***	0.096***	0.097***	0.097***	0.103***
	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.022)
$International$	-0.034***	-0.036***	-0.050***	-0.049***	-0.050***	-0.045***	-0.035***	-0.035***	-0.074***
	(0.012)	(0.012)	(0.015)	(0.015)	(0.015)	(0.013)	(0.012)	(0.012)	(0.018)
$N_{journals}$	0.729***	0.729***	0.752***	0.748***	0.748***	0.729***	0.727***	0.727***	0.746***
	(0.030)	(0.030)	(0.032)	(0.032)	(0.032)	(0.030)	(0.030)	(0.030)	(0.032)
N_{orgs}	0.042***	0.037***	0.020	0.021	0.022	0.042***	0.042***	0.042***	0.019
	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)	(0.014)
$Applied_f$	-0.564***	-0.564***	-0.541***	-0.537***	-0.537***	-0.564***	-0.563***	-0.563***	-0.536***
	(0.079)	(0.079)	(0.086)	(0.086)	(0.086)	(0.079)	(0.078)	(0.078)	(0.086)
$Industry$	-0.115***	-0.058***	-0.101***	-0.101***	-0.063**	-0.132***	-0.116***	-0.119***	-0.052*
	(0.012)	(0.012)	(0.017)	(0.017)	(0.029)	(0.013)	(0.012)	(0.012)	(0.028)
$PIndAuth_{max}$	-0.270***	-0.269***	-0.332***	-0.335***	-0.336***	-0.270***	-0.272***	-0.272***	-0.336***
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
$First\ Author\ Industry$		-0.105***							-0.074***
		(0.015)							(0.021)
$Female$				0.060***	0.069***				0.064***
				(0.022)	(0.024)				(0.023)
$Repeated\ Collaboration$							-0.021***	-0.023***	-0.022***
							(0.006)	(0.006)	(0.006)
$Female \times Industry$					-0.060*				-0.061*
					(0.035)				(0.032)
$International \times Industry$						0.045***			0.079***
						(0.016)			(0.026)
$Repeated\ Collaboration \times Industry$								0.014	0.007
								(0.009)	(0.009)
Cut1:	1.099***	1.099***	1.089***	1.129***	1.134***	1.096***	1.090***	1.089***	1.116***
Not Novel|Low Novelty	(0.057)	(0.057)	(0.063)	(0.065)	(0.065)	(0.057)	(0.057)	(0.057)	(0.064)
Cut2:	2.854***	2.854***	2.790***	2.830***	2.835***	2.851***	2.846***	2.845***	2.818***
Low Novelty|High Novelty	(0.061)	(0.061)	(0.066)	(0.066)	(0.066)	(0.062)	(0.061)	(0.061)	(0.065)
Observations	928,787	928,787	423,956	423,956	423,956	928,787	928,787	928,787	423,956
AIC	1008144	1008077	466985	466943	466937	1008136	1008002	1007998	466816
BIC	1008273	1008218	467105	467074	467080	1008277	1008143	1008151	467002

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D1: Ordered Logistic regression estimates for moderators for $Industry$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.935***	-0.936***	-0.999***	-1.003***	-1.004***	-0.935***	-0.941***	-0.942***	-1.009***
	(0.046)	(0.046)	(0.079)	(0.079)	(0.080)	(0.046)	(0.047)	(0.047)	(0.080)
N_{refs}	-0.356***	-0.356***	-0.361***	-0.360***	-0.360***	-0.356***	-0.355***	-0.355***	-0.360***
	(0.009)	(0.009)	(0.014)	(0.014)	(0.014)	(0.009)	(0.009)	(0.009)	(0.014)
Pub_{fields}	0.091***	0.091***	0.099***	0.099***	0.099***	0.091***	0.091***	0.091***	0.099***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
$International$	0.009	0.008	0.002	0.003	0.002	0.005	0.008	0.008	-0.001
	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.012)
$N_{journals}$	0.658***	0.658***	0.720***	0.721***	0.722***	0.657***	0.660***	0.660***	0.723***
	(0.031)	(0.031)	(0.047)	(0.047)	(0.047)	(0.031)	(0.032)	(0.032)	(0.048)
N_{orgs}	0.020***	0.019***	0.017***	0.017***	0.018***	0.020***	0.020***	0.021***	0.017***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
$Applied_f$	-0.220***	-0.220***	-0.214***	-0.213***	-0.213***	-0.220***	-0.220***	-0.220***	-0.213***
	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)	(0.020)	(0.020)	(0.022)
$Industry$	-0.055***	-0.035***	-0.061***	-0.061***	-0.039***	-0.060***	-0.055***	-0.058***	-0.033*
	(0.007)	(0.008)	(0.009)	(0.009)	(0.015)	(0.009)	(0.007)	(0.007)	(0.019)
$PIndAuth_{max}$	-0.148***	-0.148***	-0.155***	-0.156***	-0.156***	-0.148***	-0.148***	-0.148***	-0.156***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$First\ Author\ Industry$		-0.037***							-0.018
		(0.011)							(0.016)
$Female$				0.009	0.014				0.013
				(0.009)	(0.010)				(0.010)
$Repeated\ Collaboration$							-0.007***	-0.009***	-0.006**
							(0.002)	(0.002)	(0.003)
$Female \times Industry$					-0.034*				-0.034*
					(0.017)				(0.018)
$International \times Industry$						0.013			0.008
						(0.012)			(0.020)
$Repeated\ Collaboration \times Industry$								0.011***	0.001
								(0.003)	(0.005)
IMR	7.699***	7.703***	8.637***	8.659***	8.666***	7.694***	7.739***	7.747***	8.696***
	(0.430)	(0.430)	(0.747)	(0.748)	(0.748)	(0.430)	(0.434)	(0.435)	(0.753)
Constant	-3.871***	-3.874***	-4.352***	-4.370***	-4.376***	-3.868***	-3.891***	-3.894***	-4.389***
	(0.227)	(0.228)	(0.395)	(0.396)	(0.396)	(0.227)	(0.229)	(0.230)	(0.398)
Observations	172,675	172,675	81,212	81,212	81,212	172,675	172,675	172,675	81,212
R^2	0.206	0.206	0.204	0.204	0.204	0.206	0.206	0.206	0.204
AIC	450207	450195	212444	212444	212443	450208	450189	450181	212440
BIC	450318	450316	212546	212556	212564	450329	450310	450312	212598

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2: OLS estimates for moderators for $Industry$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.050***	-0.050***	-0.051***	-0.052***	-0.052***	-0.050***	-0.050***	-0.050***	-0.052***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
N_{refs}	-0.453***	-0.453***	-0.467***	-0.467***	-0.467***	-0.453***	-0.453***	-0.453***	-0.467***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
Pub_{fields}	0.013***	0.013***	0.019***	0.019***	0.019***	0.013***	0.013***	0.013***	0.019***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
$International$	-0.011**	-0.011**	-0.012	-0.012	-0.013	-0.017***	-0.011**	-0.011**	-0.019**
	(0.005)	(0.005)	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.008)
$International$						0.000			
						(.)			
$N_{journals}$	0.274***	0.274***	0.291***	0.291***	0.291***	0.274***	0.274***	0.274***	0.291***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)
N_{orgs}	0.006*	0.005	0.004	0.004	0.005	0.006*	0.006*	0.006*	0.005
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
$Applied_f$	-0.076***	-0.076***	-0.102***	-0.102***	-0.101***	-0.076***	-0.076***	-0.076***	-0.101***
	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)	(0.014)
$Industry$	-0.050***	-0.041***	-0.056***	-0.056***	-0.034***	-0.059***	-0.050***	-0.052***	-0.041***
	(0.005)	(0.007)	(0.008)	(0.008)	(0.012)	(0.007)	(0.005)	(0.005)	(0.015)
$PIndAuth$	-0.132***	-0.132***	-0.146***	-0.147***	-0.147***	-0.132***	-0.132***	-0.132***	-0.147***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
$First\ Author\ Industry$		-0.017**							-0.003
		(0.009)							(0.012)
$Female$				0.005	0.011				0.011
				(0.006)	(0.007)				(0.007)
$Repeated\ Collaboration$							-0.001	-0.003*	-0.001
							(0.002)	(0.002)	(0.002)
$Female \times Industry$					-0.034***				-0.035***
					(0.013)				(0.013)
$International \times Industry$						0.023**			0.023*
						(0.009)			(0.013)
$Repeated\ Collaboration \times Industry$								0.009***	0.001
								(0.002)	(0.004)
Constant	0.204***	0.204***	0.210***	0.206***	0.204***	0.205***	0.204***	0.205***	0.205***
	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.013)
Ln(sd(Constant))	-1.350***	-1.351***	-1.381***	-1.381***	-1.380***	-1.350***	-1.350***	-1.350***	-1.381***
	(0.028)	(0.028)	(0.034)	(0.034)	(0.034)	(0.028)	(0.028)	(0.028)	(0.034)
Ln(sd(Residual))	-0.137***	-0.137***	-0.132***	-0.132***	-0.132***	-0.137***	-0.137***	-0.137***	-0.132***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	274,363	274,363	129,483	129,483	129,483	274,363	274,363	274,363	129,483
Groups	1,518	1,518	1,479	1,479	1,479	1,518	1,518	1,518	1,479
AIC	705868	705865	334879	334880	334876	705864	705870	705860	334880
BIC	705995	706002	334997	335007	335013	706001	706006	706007	335056

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D3: Hierarchical Linear regression estimates for moderators for $Industry$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$	$Novelty_{ordinal}$
T_{size}	-0.217***	-0.216***	-0.216***	-0.143***	-0.143***	-0.217***	-0.222***	-0.223***	-0.151***
	(0.016)	(0.016)	(0.016)	(0.020)	(0.021)	(0.016)	(0.016)	(0.016)	(0.019)
N_{refs}	-0.081***	-0.081***	-0.081***	-0.043*	-0.043*	-0.081***	-0.081***	-0.082***	-0.043*
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.022)
Pub_{fields}	0.096***	0.096***	0.096***	0.102***	0.103***	0.096***	0.097***	0.097***	0.103***
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.022)
$International$	-0.034***	-0.030**	-0.030**	-0.049***	-0.049***	-0.031**	-0.035***	-0.035***	-0.043***
	(0.012)	(0.012)	(0.012)	(0.015)	(0.015)	(0.012)	(0.012)	(0.012)	(0.016)
$N_{journals}$	0.729***	0.731***	0.731***	0.748***	0.748***	0.729***	0.727***	0.727***	0.747***
	(0.030)	(0.030)	(0.030)	(0.032)	(0.032)	(0.030)	(0.030)	(0.030)	(0.031)
N_{orgs}	0.042***	0.032***	0.032***	0.021	0.021	0.041***	0.042***	0.042***	0.013
	(0.011)	(0.010)	(0.010)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)	(0.014)
$Applied_f$	-0.564***	-0.563***	-0.563***	-0.537***	-0.537***	-0.563***	-0.563***	-0.564***	-0.536***
	(0.079)	(0.079)	(0.079)	(0.086)	(0.086)	(0.079)	(0.078)	(0.078)	(0.085)
$Industry$	-0.115***	-0.070***	-0.072***	-0.101***	-0.101***	-0.115***	-0.116***	-0.115***	-0.063***
	(0.012)	(0.016)	(0.017)	(0.017)	(0.017)	(0.012)	(0.012)	(0.012)	(0.022)
$PIndAuth_{max}$	-0.270***	-0.288***	-0.287***	-0.335***	-0.333***	-0.274***	-0.272***	-0.276***	-0.361***
	(0.048)	(0.049)	(0.049)	(0.048)	(0.050)	(0.048)	(0.048)	(0.048)	(0.049)
$First\ Author\ Industry$		-0.112***	-0.107***						-0.093***
		(0.015)	(0.016)						(0.022)
$Female$				0.060***	0.059***				0.053***
				(0.022)	(0.020)				(0.020)
$Repeated\ Collaboration$							-0.021***	-0.021***	-0.023***
							(0.006)	(0.005)	(0.005)
$Industry \times PIndAuth_{max}$		0.082***	0.088***						0.058*
		(0.027)	(0.027)						(0.035)
$First\ Author\ Industry \times PIndAuth_{max}$			-0.013						0.001
			(0.017)						(0.022)
$Female \times PIndAuth_{max}$					-0.004				-0.001
					(0.041)				(0.038)
$International \times PIndAuth_{max}$						0.016			0.030
						(0.015)			(0.021)
$Repeated\ Collaboration \times PIndAuth_{max}$								0.008	0.011***
								(0.005)	(0.004)
Cut1:	1.099***	1.105***	1.106***	1.129***	1.128***	1.100***	1.090***	1.089***	1.119***
Not Novel|Low Novelty	(0.057)	(0.058)	(0.058)	(0.065)	(0.063)	(0.057)	(0.057)	(0.056)	(0.062)
Cut2:	2.854***	2.861***	2.861***	2.830***	2.829***	2.855***	2.846***	2.844***	2.820***
Low Novelty|High Novelty	(0.061)	(0.061)	(0.061)	(0.066)	(0.067)	(0.061)	(0.061)	(0.061)	(0.065)
Observations	928,787	928,787	928,787	423,956	423,956	928,787	928,787	928,787	423,956
AIC	1008144	1007960	1007960	466943	466944	1008140	1008002	1007975	466755
BIC	1008273	1008112	1008125	467074	467087	1008281	1008143	1008128	466963

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D4: Ordered Logistic regression estimates for moderators for $PIndAuth_{max}$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.936***	-0.935***	-0.935***	-1.003***	-1.001***	-0.935***	-0.941***	-0.942***	-1.007***
	(0.046)	(0.046)	(0.046)	(0.079)	(0.080)	(0.046)	(0.047)	(0.047)	(0.080)
N_{refs}	-0.356***	-0.356***	-0.356***	-0.360***	-0.360***	-0.356***	-0.355***	-0.355***	-0.360***
	(0.009)	(0.009)	(0.009)	(0.014)	(0.014)	(0.009)	(0.009)	(0.009)	(0.014)
Pub_{fields}	0.091***	0.091***	0.091***	0.099***	0.099***	0.091***	0.091***	0.091***	0.099***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
$International$	0.008	0.008	0.008	0.003	0.002	0.010	0.008	0.008	0.001
	(0.007)	(0.007)	(0.007)	(0.010)	(0.011)	(0.006)	(0.007)	(0.007)	(0.010)
$N_{journals}$	0.658***	0.657***	0.658***	0.721***	0.721***	0.657***	0.660***	0.660***	0.722***
	(0.031)	(0.031)	(0.031)	(0.047)	(0.047)	(0.031)	(0.032)	(0.032)	(0.048)
N_{orgs}	0.019***	0.019***	0.019***	0.017***	0.017***	0.020***	0.020***	0.020***	0.016***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
$Applied_f$	-0.220***	-0.220***	-0.220***	-0.213***	-0.213***	-0.220***	-0.220***	-0.220***	-0.212***
	(0.021)	(0.021)	(0.020)	(0.022)	(0.022)	(0.021)	(0.020)	(0.020)	(0.022)
$Industry$	-0.035***	-0.037***	-0.035***	-0.061***	-0.061***	-0.055***	-0.055***	-0.055***	-0.047***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.010)	(0.007)	(0.007)	(0.007)	(0.014)
$PIndAuth_{max}$	-0.148***	-0.149***	-0.149***	-0.156***	-0.144***	-0.152***	-0.148***	-0.151***	-0.156***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)	(0.015)	(0.013)
$First\ Author\ Industry$	-0.037***	-0.038***	-0.040***						-0.024
	(0.011)	(0.011)	(0.011)						(0.016)
$Female$				0.009	0.006				0.005
				(0.009)	(0.009)				(0.009)
$Repeated\ Collaboration$							-0.007***	-0.007***	-0.007***
							(0.002)	(0.002)	(0.002)
$Industry \times PIndAuth_{max}$		0.004	0.001						-0.006
		(0.008)	(0.009)						(0.013)
$First\ Author\ Industry \times PIndAuth_{max}$			0.006						0.011
			(0.011)						(0.014)
$Female \times PIndAuth_{max}$					-0.019				-0.016
					(0.012)				(0.011)
$International \times PIndAuth_{max}$						0.015**			0.025***
						(0.006)			(0.009)
$Repeated\ Collaboration \times PIndAuth_{max}$								0.008***	0.008***
								(0.002)	(0.002)
IMR	7.703***	7.699***	7.699***	8.659***	8.652***	7.690***	7.739***	7.746***	8.676***
	(0.430)	(0.430)	(0.430)	(0.748)	(0.749)	(0.430)	(0.434)	(0.435)	(0.754)
Constant	-3.874***	-3.871***	-3.871***	-4.370***	-4.363***	-3.867***	-3.891***	-3.893***	-4.373***
	(0.228)	(0.228)	(0.228)	(0.396)	(0.396)	(0.228)	(0.229)	(0.230)	(0.399)
Observations	172,675	172,675	172,675	81,212	81,212	172,675	172,675	172,675	81,212
R^2	0.206	0.206	0.206	0.204	0.204	0.206	0.206	0.206	0.204
AIC	450195	450197	450198	212444	212437	450200	450189	450152	212402
BIC	450316	450327	450339	212556	212558	450321	450310	450282	212579

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D5: OLS estimates for moderators for $PIndAuth_{max}$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$	$Ln(Novelty)$
T_{size}	-0.050***	-0.050***	-0.050***	-0.052***	-0.051***	-0.050***	-0.050***	-0.050***	-0.076***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
N_{refs}	-0.453***	-0.453***	-0.453***	-0.467***	-0.467***	-0.453***	-0.453***	-0.453***	-0.468***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.009)
Pub_{fields}	0.013***	0.013***	0.013***	0.019***	0.019***	0.013***	0.013***	0.013***	0.037***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)
$International$	-0.011**	-0.013**	-0.013**	-0.012	-0.012	-0.010**	-0.011**	-0.012**	-0.016**
	(0.005)	(0.005)	(0.005)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.008)
$N_{journals}$	0.274***	0.274***	0.274***	0.291***	0.291***	0.274***	0.274***	0.274***	0.237***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.014)
N_{orgs}	0.006*	0.006*	0.006*	0.004	0.004	0.005	0.006*	0.006*	0.016***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)
$Applied_f$	-0.076***	-0.076***	-0.076***	-0.102***	-0.102***	-0.076***	-0.076***	-0.076***	-0.217***
	(0.011)	(0.011)	(0.011)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)	(0.021)
$Industry$	-0.050***	-0.035***	-0.037***	-0.056***	-0.056***	-0.051***	-0.050***	-0.050***	-0.040***
	(0.005)	(0.006)	(0.007)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.010)
$PIndAuth$	-0.132***	-0.128***	-0.128***	-0.147***	-0.137***	-0.136***	-0.132***	-0.135***	-0.170***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.014)
$First\ Author\ Industry$		-0.016*	-0.012						-0.027**
		(0.009)	(0.009)						(0.013)
$Female$				0.005	0.003				-0.011
				(0.006)	(0.006)				(0.008)
$Repeated\ Collaboration$							-0.001	-0.002	-0.003*
							(0.002)	(0.001)	(0.002)
$Female \times PIndAuth$					-0.017***				-0.019*
					(0.006)				(0.010)
$International \times PIndAuth$						0.015***			0.022***
						(0.004)			(0.007)
$Repeated\ Collaboration \times PIndAuth$								0.007***	0.008***
								(0.001)	(0.002)
$Industry \times PIndAuth$		-0.020***	-0.016***						-0.000
		(0.004)	(0.005)						(0.010)
$First\ Author\ Industry \times PIndAuth$			-0.007						0.013
			(0.007)						(0.011)
Constant	0.204***	0.205***	0.205***	0.206***	0.209***	0.203***	0.204***	0.205***	0.225***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.014)
Ln(sd(Constant))	-1.350***	-1.349***	-1.349***	-1.381***	-1.381***	-1.351***	-1.350***	-1.351***	
	(0.028)	(0.028)	(0.028)	(0.034)	(0.034)	(0.028)	(0.028)	(0.028)	
Ln(sd(Residual))	-0.137***	-0.137***	-0.137***	-0.132***	-0.132***	-0.137***	-0.137***	-0.137***	-0.105***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	274,363	274,363	274,363	129,483	129,483	274,363	274,363	274,363	129,483
Groups	1,518	1,518	1,518	1,479	1,479	1,518	1,518	1,518	1,479
AIC	705868	705847	705848	334880	334872	705855	705870	705821	340324
BIC	705995	705994	706005	335007	335008	705992	706006	705969	340510

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D6: Hierarchical Linear regression estimates for moderators for $PIndAuth$.

	(1)	(2)	(3)	(4)
	$Novelty : NotNovel\|Novel$	$Novelty : NotNovel\|Novel$	$Novelty : LowNovelty\|HighNovelty$	$Novelty : LowNovelty\|HighNovelty$
T_{size}	-0.214***	-0.213***	-0.199***	-0.199***
	(0.016)	(0.016)	(0.016)	(0.016)
N_{refs}	-0.045**	-0.044**	-0.862***	-0.862***
	(0.022)	(0.022)	(0.016)	(0.016)
Pub_{fields}	0.091***	0.091***	0.101***	0.101***
	(0.020)	(0.020)	(0.012)	(0.012)
$International$	-0.033***	-0.027**	-0.021	-0.023
	(0.012)	(0.012)	(0.019)	(0.019)
$N_{journals}$	0.717***	0.718***	0.485***	0.485***
	(0.029)	(0.029)	(0.022)	(0.022)
N_{orgs}	0.040***	0.037***	0.040***	0.041***
	(0.011)	(0.010)	(0.012)	(0.012)
$Applied_f$	-0.539***	-0.539***	-0.445***	-0.445***
	(0.077)	(0.078)	(0.042)	(0.042)
$Industry$	-0.112***	-0.127***	-0.109***	-0.109***
	(0.012)	(0.016)	(0.019)	(0.018)
$PIndAuth_{max}$	-0.254***	-0.271***	-0.321***	-0.317***
	(0.047)	(0.048)	(0.029)	(0.030)
$Industry \times PIndAuth_{max}$		0.074***		-0.023
		(0.026)		(0.020)
Constant	-1.121***	-1.127***	-1.087***	-1.085***
	(0.057)	(0.057)	(0.026)	(0.026)
Observations	928,787	928,787	172,675	172,675
R^2				
AIC	826925	826829	167746	167746
BIC	827042	826958	167847	167857

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D7: Logistic regression estimates including interaction between $Industry$ and $PIndAuth_{max}$.