Working Paper 24-021

Bringing Science to Market: Knowledge Foundations and Performance

Justine Boudou
Maria Roche

Harvard Business School

Bringing Science to Market: Knowledge Foundations and Performance

Justine Boudou
Harvard Business School

Maria Roche
Harvard Business School

Working Paper 24-021

Funding for this research was provided in part by Harvard Business School.

Bringing Science to Market: Knowledge Foundations and Performance*

Justine Boudou
Harvard Business School

Maria Roche
Harvard Business School

October 27, 2023

Abstract

Possessing unique knowledge is widely considered a critical source of competitive advantage. In this paper, we assess whether new ventures that exploit their founders' technologically unique knowledge perform better on the exit market than those that do not (or only little). Using a panel dataset of 510 academic startups in biomedicine created between 2005 and 2015, we find that, contrary to expectations, startups that rely heavily on their founders' academic work are less likely to be acquired, while we do not observe marked differences with regards to their ability to generate valuable technologies. Besides their outcomes on the commercial market, we further examine the potential ramifications on the academic market. Applying a difference-in-difference design, we find that founders who build extensively on their academic work as the foundation of their startup experience a decrease in both the number of publications and top publications. Our crude back-of-the-envelope calculations suggest that this decrease represents a potential loss of value generated from publications of 40,000 - 333,000 dollars per year per academic entrepreneur that relies strongly on their own research as the basis of the firm. Additional analyses reveal a notable shift in the research approach of these founders after entering entrepreneurship targeted towards increasing economic payoffs. Of note are our observations of a reduction in collaboration with other scientists, both within and outside their institution, a narrowing of their research focus, though not at the expense of exploring new concepts, and an increase in patenting activity.

Keywords: *Firm Performance*, *Exits*, *Knowledge Foundations*, *Academic Startups*,

*The authors thank Annamaria Conti for data access. Many thanks to participants at the Barcelona School of Economics Entrepreneurship Working Group, and DRUID. We are indebted to Ashish Arora, George Chondrakis, Shane Greenstein, Jacqueline Lane, Brian Silverman, and Keyvan Vakili for feedback on early drafts. We gratefully acknowledge funding from the Harvard Business School Division of Research and Faculty Development. All errors and omissions are our own. Contact: jboudou@hbs.edu; mroche@hbs.edu

1 Introduction

Possessing unique knowledge has long been heralded a critical component in achieving competitive advantage (Barney, 1991). Departing from the vision of perfect competition, where information symmetries exist and every firm can copy what the other is doing, knowledge asymmetries as embodied in unique – often tacit – knowledge can help a firm differentiate. The exploitation of such special information may even provide the basis for a new company to form in the first place (Alvarez and Busenitz, 2001) and have a persistent impact on the performance of a firm (Conti and Roche, 2021; Geroski et al., 2010).

In this paper, we examine the extent to which startups that exploit their founders' technologically unique knowledge perform better than those that do not (or only little) on the exit market. We study this question in the particularly suitable context of academic entrepreneurship. We do so provided the bulk of specialized knowledge emanates from academia (National Science Board, 2018), science's increasing role in the market for technologies (Arora, Belenzon and Suh, 2022), and the importance of specialized knowledge in the formation of a startup in this setting. Long recognized to be among the major drivers of economic growth (Dasgupta and David, 1994), universities have been increasingly expanding their traditional role as producers of knowledge into the commercialization of scientific discoveries (Bhaskarabhatla and Hegde, 2014; Crespi et al., 2011; Hsu et al., 2007; Mowery et al., 2004). Academic spinouts in particular – defined as startups founded by faculty and students around technological knowledge developed within academic laboratories (Shah and Pahnke, 2014) – have been increasing at a stunning rate (Grimaldi et al., 2011; Roche et al., 2020).[1]

However, despite large public investment, and broad enthusiasm for such activity (Roach, 2017), academic spinouts do not perform particularly well on the market on average (Roche et al., 2020). Although some studies have highlighted that critical heterogeneity among academic spinouts may exist (mainly highlighting the role of outliers, so called "stars") (Zucker et al., 2002, 1998), we have yet to better understand the sources of performance differentials. The one source of potential advantage we seek to examine in this paper is related to academic founders' own unique technological knowledge. Leveraging such expertise acquired over years

[1]Some of the more famous examples of academic spinouts include Tableau (founded by Pat Hanrahan, professor at Stanford), Genentech (launched by Herbert Boyer, professor at the University of California), and Duolingo (initiated by Luis von Ahn, professor at Carnegie Mellon University).

spent in their occupational training (Colombo and Piva, 2012) as the foundation of their new venture could be an important antecedent for success. Building on established strategy theories, we may expect such access to unique resources to present an important source of competitive advantage on the market.

To test this, we create a panel data set of 510 academic startups in biomedicine created between 2005 and 2015 using sources such as CrunchBase. This sector is particularly well suited for our study because it is tightly linked to academic research and has a relatively high propensity to patent, which enables us to capture the initial knowledge base of a startup. In addition, we have detailed information about each startup regarding its founding date and place, founders, patents and exit events. We further link each academic founder to their research and patenting output using Dimensions AI (*https://www.dimensions.ai/*).

Our main variable of interest is the extent to which academic founders rely on their own academic work when creating their startup, which we also conceptualize as the distance between founders' academic work and the initial knowledge base of their startup. We view this measure as a proxy for the degree to which technology vs market demand, is a driving force in the creation of a startup (Bikard et al., 2019). Proximity between the two knowledge-bases should indicate the degree to which a technology-centric approach to founding is pursued, while a larger distance should capture a more market driven approach for finding a solution an academic founder may have developed as part of their research agenda for a known problem. We capture founders' academic work by considering their published papers before startup creation. We capture the initial knowledge base of a startup by looking at its first granted patents, at or close to the time of creation. To operationalize the distance variable, we then calculate the percentage of citations that these patents make to their founders' academic papers using the Reliance on Science dataset (Marx and Fuegi, 2020, 2022).[2] We then analyze the impact of this measure on exit events – differentiating between acquisition and IPO. In estimating this relationship between knowledge-base difference and performance outcomes we take a step-wise approach. We first control for as many confounding factors as possible including a wider set of fixed effects, and then perform several subsequent tests to show that our results remain robust to different empirical and sample specifications. In addition, we provide evidence that selection is unlikely the main driver of our results, and implement an

[2]We cross-validate our measure with others, which we describe in more detail in the data section. There is strong correlation with these alternative proxies.

instrumental variable approach that reveals consistent results with our more naive model.

Our analyses reveal noteworthy patterns. On the exit market, contrary to what we may expect, we find that startups that are closest to their founder's academic knowledge base are the least successful. This appears to be driven by a lower likelihood of being acquired, with a 10p.p decrease in distance being associated with a 2.4% lower probability of acquisition. We do not find any differential effect on the likelihood to IPO. This result does not appear to be driven by a lower invention quality nor a more nascent technology: we show that a 10p.p increase in similarity between founders' academic work and their startup increases the total number of patents by 18%, representing about one additional patent on average, and has no significant impact on the number of citations these patents receive nor on the amount of funds raised by these startups. Consistent with prior work, this implies that startups closer to their founders' academic knowledge have higher inventive output on average (Roche et al., 2020). Our results also provide suggestive evidence that this lower likelihood of acquisition is not related to a more nascent or less developed technology: indeed, startups that are closest to their founder's academic knowledge base seem to transfer their technology to the market faster. Taken together, it appears that founders who first identify a problem in the market and then try to find solutions by using previous knowledge (i.e. those relying relatively less on their own work) are more successful on the market than founders who adopt a more technology-centric approach to market, driven by the technology but potentially still looking for a problem to solve. In similar vein, it may be that startups created around the specialized knowledge of their founders lack complementary assets in the market rendering them less valuable to potential acquirers (Arora, Fosfuri and Roende, 2022).

In addition, provided that academic entrepreneurship is a departure from the more traditional academic job description (Cohen et al., 2020), we further examine potential ramifications for academics' research output. Ideally, commercial and academic activities should be reinforcing, such as work focusing on understanding the impact of academic patenting (Azoulay et al., 2009; Crespi et al., 2011) or university licensing (Thursby and Thursby, 2011) on university research suggests. Though similar in the sense that these are all commercial activities, entrepreneurship is likely a stronger departure from academia, so it is not guaranteed the relationship is similar. In fact, the much scarcer literature on academic spinouts and its relationship with research output is rather ambiguous. On the one hand, Toole and Czarnitzki

(2010) provide evidence of a significant decrease in research performance within their sample of academic entrepreneurs after they begin working in for-profit firms. On the other hand, results from Fini et al. (2022) suggest that entrepreneurship may lead to more impactful research because founders will engage in greater exploration. As such, it remains an empirical question to understand how the extent to which an academic founder exploits their own unique technological knowledge in creating their startup relates to her academic output.

Implementing a staggered difference-in-difference design with individual-level fixed effects, our results indicate that, on average, there is no significant difference in terms of academic output following startup creation between founders that build and those who do not build on their academic work in their startups. However, this masks important heterogeneity. In particular, we find a decrease in the number of publications and top publications after startup creation for founders whose firm's knowledge base is closest to their academic work compared to founders who are more distant, highlighting the potential trade-off generated by the dual involvement in academia and industry.

Several mechanisms could be at play. A possible one could be that these results are capturing an escalation of commitments (Arkes and Blumer, 1985; Staw, 1976) leading to negative performance outcomes in both entrepreneurship and academia for founders relying the most on their own academic work. Alternatively, they could have higher payoff concerns, leading them to keep greater control over what they publicly disclose. Our findings that founders who rely more on their previous academic work experience a) a decrease in the number of co-authors after entering entrepreneurship, b) a narrowing of their research breadth but not at the expense of exploration, and c) an increase in their propensity to patent, provides some suggestive evidence for the latter. Such refocusing (more narrow, less collaborative, less open) of the research agenda could entail broader consequences for the overall production of science (Aghion et al., 2008).[3]

[3] Interviews with academic entrepreneurs reveal that our results are not because academic entrepreneurs want to remain in control of their companies and do not want to go public or be acquired.

2 Conceptual framework

2.1 A knowledge-based advantage

The central role of resources in shaping not only established firms', but also new ventures' strategic outcomes has been well-documented (Barney, 1991). Entrepreneurial opportunities are presumed to exist primarily because of differences in beliefs of the relative value of resources, where entrepreneurs can leverage their specialist knowledge to create rents (Schumpeter, 1912; Shane and Venkataraman, 2000). In particular, new firms are created by allocating resources to novel ends (Alvarez and Busenitz, 2001), where initial resource endowment, such as social capital, can serve as a critical foundation for the long-term performance of new ventures (Shane and Stuart, 2002). Although a notable body of work (Agarwal et al., 2004; Åstebro et al., 2011; Sørensen, 2007; Stenard and Sauermann, 2016) has made fundamental strides in understanding the impact of capability differentials in shaping both the decision to become an entrepreneur and the entry mode into entrepreneurship, what remains understudied is how initial *technical knowledge endowments* may shape the subsequent performance of firms. Highly unique technical knowledge endowments may present an important factor in establishing a competitive advantage early on in the life of a firm.

Academic entrepreneurship provides a suitable setting to examine the extent to which startups that exploit their founders' technologically unique knowledge perform better. The number of academic startups has been increasing rapidly in the US over the past decades (Rothaermel et al., 2007; Audretsch, 2014; Roche et al., 2020) coming to represent a fundamental engine for innovation, especially in knowledge-intensive areas (Acs and Audretsch, 1990). They are most prevalent in fields characterized by a close alignment between scientific research and commercialization prospects (Stokes, 2011) and where appropriability regimes are strong, such as the biological and life sciences. Given the specialized, often tacit knowledge, that academic entrepreneurs possess (Bercovitz and Feldman, 2006) granting them the ability to "recognize the value of new, external information, assimilate it, and apply it to commercial ends [...]" (Cohen and Levinthal 1990, p.128), specialized knowledge is particularly important in the formation of academic startups. As such, they play a critical role in bridging the gap between the academic and the private sector: startups emanating from academia are typically based on scientific advances made within a laboratory, and not surprisingly, target inventions that the private-sector would have not otherwise pursued because of a lack of technical knowledge. From this, we may expect academic startups that are founded based primarily on the scientific work of their founder(s) to fair best provided their advantage based on knowledge about a specific technology. Such knowledge could help place these firms at the frontier of knowledge, may provide a critical source for differentiation, and may thus create more value for customers. Indeed, academic scientists often have extensive knowledge in a narrow field of science no one else possesses (Jones, 2009). From this, it appears, we should expect that *the more a startup exploits the knowledge-base of its founders, the more likely these startups are to be successful.*

2.2 Dual commitment in entrepreneurship and academia

Given their dual involvement in academia and entrepreneurship, academic founders share their time between two core tasks: i) knowledge production (and training) at universities, and ii) creating their venture (Roche, 2023). The extent to which a founder builds a startup on their own unique technological knowledge may thereby also have important consequences for their academic performance. The line of work examining dual involvement of academics in academia and another activity have primarily focused on patenting or licensing (e.g., Azoulay et al. 2009; Crespi et al. 2011; Thursby and Thursby 2011), with the aim of getting a better understanding of the reasons behind professors' involvement in commercial activity (Perkmann et al., 2013), the characteristics of those professors who do (Agrawal and Henderson, 2002), and what the implications of commercialization are for professors' time, knowledge, norms, and resources (Shibayama et al., 2012). The empirical evidence suggests that the most productive academic life scientists are those involved in commercialization (Agrawal and Henderson, 2002) where specifically in the case of biotechnology, an influential stream of research points to the fundamental role "star" scientists play in transferring new academic knowledge to industry (Higgins et al., 2011; Toole and Czarnitzki, 2009; Zucker et al., 2002, 1998). From this we may expect that both activities could be reinforcing.

Other work has investigated the impact of the adoption of commercial attitudes and behaviors by academic researchers on a number of outcomes (Dasgupta and David, 1994; Etzkowitz, 2003; Powell and Owen-Smith, 1998; Powell and Snellman, 2004; Stephan, 2012; Stuart and Ding, 2006), such as sharing behaviors (Shibayama et al., 2012), and shifts in the amount, direction, and quality of scientific research. Some studies highlight potential risks for academic research, such as changes in the content of scientific research toward more applied topics (Blumenthal et al., 1986), a slowing-down of open knowledge diffusion (Murray and Stern, 2007; Nelson, 2004), or even an exodus of academic scientists to industry (Azoulay et al., 2009). Other work finds that commercialization may enhance traditional scholarship (Goldfarb et al., 2009) and does not seem to distract from academic knowledge production (Abramo et al., 2012; Thursby and Thursby, 2007).

The literature on academic spinouts and its relationship with research output remains scarce. On the one hand, work by Czarnitzki and Toole (2010) and Toole and Czarnitzki (2010) find a significant decrease in research performance among their sample of academic entrepreneurs after they begin working in for-profit firms. On the other hand, Fini et al. (2022) finds that entrepreneurship leads to more impactful research because founders will engage in greater exploration, and Ambos et al. (2008) provide evidence that these successful "ambidextrous" individuals are more highly cited.

Provided the state of extant literature, it is *unclear ex-ante if the extent to which an academic entrepreneur exploits their own unique technological knowledge impacts their scientific productivity and/or research agenda.* Whether the activities are mutually reinforcing or lead to changes in the rate and direction of research remains an empirical question, which we examine in the following sections.

3 Data

3.1 Academic startups dataset

Our analysis relies on a panel dataset of 510 academic startups that was created using Crunchbase as well as a variety of other sources such as companies' website, LinkedIn and Bloomberg. Following this approach, we identify over 1,700 US startups in the biomedicine (i.e., biotechnology and medical devices) sector, as well as their patents. We then retain only those with at least one professor in the founding team, which we define as academic startups. Our final sample is composed of 510 academic startups corresponding to 1,083 unique founders, among whom 676 are identified as professors. We complement this dataset with PatentsView to retrieve patent-level information such as the number of forward citations. We also match each founder to their publication and patenting output using Dimensions AI based on last name, middle name, first name and institution (see Appendix A for more details about the matching algorithm). Among the 640 professors who founded one startup only (see section 4.3), we uniquely matched 561 of them.

3.2 Calculating the distance between the knowledge-base of a startup and that of its founders

Our main independent variable corresponds to the distance between the knowledge-base of a startup and that of its founders. We conceptualize this measure as a way to capture the extent to which technology, vs market demand, is a driving force in the creation of a startup. A lower distance (higher proximity) between the two knowledge-bases suggests a more technology-centric approach, while a larger distance (lower proximity) suggests a more market driven approach.

To operationalize our independent variable, we first need to define the knowledge-base of a startup at the time of a creation and then compare it to its founders' academic work. We capture the knowledge-base of a startup at the time of creation by considering its first granted patents. We capture founders' academic work by considering the pool of papers they have published before entering entrepreneurship. We then calculate the distance between founders' academic work and the knowledge base of their startup by leveraging the patenting process: when a company applies for a patent, it has to list all the knowledge on which it builds on, including scientific papers. This allows us to differentiate between citations that patents make to the founders' academic work vs citations that patents make to other researchers' work. We then calculate the distance between the knowledge base of a startup and its founders' academic work by computing the percentage of scientific citations that their patents make to their founders' academic papers.

In practice, we take the first granted patent(s) a startup applied to[4] and we match them to the Reliance on Science (RoS) dataset (Marx and Fuegi, 2020, 2022), which provides a publicly-available set of citations from U.S. patents to scientific articles. For each academic article cited by a patent, we create a self-citation dummy equal to 1 when at least one author of the academic article is matched to an inventor with a confidence score above 50[5]. This

[4]For each startup, we consider the granted patent with the earliest application year. In case there are several granted patents with the same earliest application year, we consider all of them.

[5]The match is performed based on last, middle and first name (Marx and Fuegi, 2020, 2022). Results are robust to the use of more stringent thresholds, such as 75.

identifies instances where we can reasonably be confident that at least one inventor of the patent is citing their own academic work. Results are robust to using a more demanding definition where the average confidence score of all authors in matching with an inventor is used. We then calculate the percentage of self-cites at the patent level by dividing the total number of self-cites by the total number of scientific citations.

$$\text{Percentage self-cites} = \frac{\text{Number of self-cites in RoS}}{\text{Total number of cites in RoS}} \tag{1}$$

Finally, we average this measure at the firm level for startups with multiple first patents. The higher the percentage of self-cites, the closer the startup's and founder's knowledge bases are[6].

3.3 Summary statistics

Our main independent variable of interest is skewed, with an average value of 6% (Figure A1 presents the histogram of *Percentage Self-Cites*). A majority of startups have patents that do not rely on their founders' academic work: the treatment variable takes a value of 0, implying a relative high distance between the knowledge base of the startup and that of its founders' academic work[7].

<Insert Table 1 here>

Table 1 displays summary statistics for our sample of 510 firms. Startups have on average 4.2 patents, and on average 3.8 cite scientific literature. There are 1.4 professors per startup, for an average team size of 2.3 people. 16% of firms have at least one female founder and 66%

[6]In order to provide additional support for our knowledge-base proximity variable, we also construct a similarity measure between founders' publications and their venture's patents. To that end, we use the *pmra* probabilistic topic-based model for content similarity developed by Lin and Wilbur (2007) to calculate the similarity between each founder publication abstract published before startup creation and each patent of his venture. We then aggregate this measure at the venture level, using several measures such as the mean, the median or the max. We find a positive correlation between this similarity measure and our treatment variable that relies on the number of self-cites, providing support that our measure captures an overlap between academic and commercial output in a systematic way, rather than, e.g., narcissism. We reiterate this analysis by comparing venture websites and founders' publication abstracts, using the dataset of Guzman and Li (2023) and find similar results. However, we remain cautious of using one of these similarity measures as treatment because they compare different types of goods (publication vs patent or publication vs website) which use a different vocabulary that is not directly comparable between each other and does not entirely reflect the link to knowledge-bases.

[7]Note that there are two cases where the distance value could be null. First, if a startup cites scientific papers but none of them come from the founders' previous academic work. Second, if a startup does not cite any scientific papers, which implies that it also does not cite any previous work from its founders. In robustness test, we show that our results are robust to excluding startups from this last case

have at least one founder who graduated from a top-tier university[8]. 6% of the startups in our sample are acquired and 6% of them go public through an IPO. The amount of funds raised within 5 years of inception is skewed, with an average of USD 12.6 millions and a median of USD 0.5 million.

4 Estimation strategy and results

Estimating the causal effect of founders' reliance on their academic work when founding a venture is subject to the classic problem of selection: researchers choose whether to enter entrepreneurship and conditional on entering it, choose whether they predominantly rely on their previous work or not. In an ideal experiment, we would have access to a sample of professors that we would randomly assign to creating a startup that relies more or less on their academic work. We would then observe outcomes on the exit market – such as whether the startups get acquired or go public through an IPO – as well as the subsequent academic output of the founders. This is not feasible in practice for numerous cost related and ethical reasons, and we must rely on observational data instead. Our main empirical strategy consists in controlling for as many confounding factors as possible (exit market results) and difference-in-difference analysis (academic market). We also provide evidence that "selection into treatment" does not seem to be correlated with variables that would raise obvious endogeneity concerns. While our results cannot be interpreted as definitely causal, we test their robustness in a series of subsequent analyses, including an instrumental variable strategy, that convey a similar story as our baseline results and echo our discussions with academic founders. In what follows, we begin by presenting the "naive" OLS results, that we corroborate with robustness tests that include difference-in-differences models, and individual-level fixed effects specifications.

4.1 Determinants of the reliance on one's academic work

In Table 2, we start by presenting results related to the probability of relying more or less on previous academic work when founding the venture. Overall, conditional on entering entrepreneurship, there does not seem to be any significant differences between professors who rely more or less on their academic work.

[8] Top-tier universities are determined following the 2016 Academic Ranking of World Universities ("Shanghai Ranking", accessible at shanghairanking.com)

<Insert Table 2 here>

4.2 Startup level outcomes on the exit market

In Table 3, we focus on a set of venture outcomes and their relationship with the reliance of a startup's knowledge base on the academic work of its founders. Our regressions are of the form:

$$Y_i = \beta \text{Percentage self-cites}_i + \gamma X_i + \delta_{\text{State}} + \delta_{\text{Founding year}} + \delta_{\text{Sector}} + \epsilon_i$$

where i indexes startups. X_i includes the following list of controls: the log number of patents and the log number of patents relying on scientific literature which both proxy for startup inventive quality, the log of team size calculated with the number of founders at inception which captures the impact of venture size at founding; an indicator equal to 1 if there is at least one female in the founding team and an indicator equal to 1 if at least one founder graduated from a top-tier university. We also control for sector, state, founding year and state $\times$ founding year fixed effects. We cluster standard errors at the startup and state level[9].

Our dependent variables of interest are: *Success*, an indicator equal to 1 if the startup was either acquired *or* went public via an Initial Public Offering (IPO), *Acquired*, an indicator equal to 1 if the startup was acquired, and *IPO*, an indicator equal to 1 if the startup went public via an IPO.

<Insert Table 3 here>

While column (1) shows no significant effect of distance on the overall success rate, column (2) shows that a lower distance between a startup's knowledge-base and its founders' academic work (i.e., an increased percentage of self-cites) decreases the likelihood of being acquired: a 10p.p decrease in distance (i.e., a 10p.p increase in self-cites) is associated with a 2.4% decrease in acquisition likelihood. We do not find any significant effect for the probability of IPO.

In the following subsections, we provide suggestive evidence that the lower acquisition likelihood experienced by startups whose knowledge-base is closer to their founder's academic work does not seem to be explained by worse or more nascent technologies. Rather, this result is consistent with the idea that the knowledge they embody results from a more technology-centric (vs a more market driven) approach, making their technology potentially less easily

[9] Results are robust to clustering at the state level

integrated by the private sector (Bikard et al., 2019). Although using science in invention may be an advantage on the market for technologies, on average (Arora, Belenzon and Suh, 2022), the more the startup is based on the specialized knowledge of their founders, the more likely it may be that potential acquirers do not possess suitable complementary assets rendering the startup less valuable on the exit market (Arora, Fosfuri and Roende, 2022).

For further robustness, we also implement an instrumental variable approach, where we use the academic network size of founders as an instrument for the *Percentage of self-cites* variable. A bigger academic network could make a founder more likely to rely on others' work vs their own, as they have been exposed to a more diverse and likely broader set of ideas. To proxy for network size, we use the number of unique collaborators in research articles published before startup creation and average this measure at the firm level for startups with multiple academic founders. For network size to be a suitable instrument, it should not impact the likelihood of acquisition other than through its influence on the extent to which founders rely on their own work when entering entrepreneurship. We see little reason why a higher number of academic co-authors would influence the likelihood of acquisition, as this network is made of academic researchers who should not have a differential influence on acquirers' decision. In addition, to qualify as a suitable instrument, academic network size should not be correlated with a variable that would happen to be also correlated with the likelihood of acquisition. It is possible that a higher network could be correlated with location, which could also influence the number of prospective acquirers, which is why we include state fixed effects in our regression. It could also be that gender or type of university could affect acquisition likelihood, which is why we include controls for such variables. While we cannot prove our instrument is perfectly valid, we believe it provides a compelling opportunity to test the robustness of our OLS results.

Table 4 presents the 2SLS result and shows suggestive evidence that the previous negative correlation between *Percentage self-cites* and *Acquired* is robust to using this network instrument.[10] Column (1) shows the first-stage regression which suggests, as hypothesized, that a bigger network size decreases founder's propensity to rely on their own work. Column (2) shows the 2SLS result, which shows a negative impact of *Percentage self-cites* on *Acquired.* Note that the coefficient on the reduced form is about 10 times larger than that of the coefficient from our more naive approach described above. Two credible reasons could explain this

[10]Table A9 shows the 2SLS results for *Success* and *IPO*. Similar as *Acquired*, we find negative magnitudes.

finding: 1) the instrument could be weak, or 2) the instrument is shifting the behavior of a subpopulation more than another. Though the F-stat of our instrument is above the rule of thumb, it is not the strongest. However, we believe it is probably more likely that the local average treatment effect that we are estimating is larger than the average treatment effect of our OLS regression because of heterogeneity in the sample we are analyzing. Thus, the compliers that we are shifting - those founders who relied less on their own work because they happened to have a bigger academic network - may have higher returns than the average founder of our sample. While we do not aim to use this instrumental variable approach as a way to pinpoint a specific magnitude, we believe it is reassuring to find that a lower distance between a founder's academic work and their startup leads to a lower likelihood of acquisition, similar to our previous OLS result.[11]

<Insert Table 4 here>

4.3 Startup level invention outcomes

In Table 5, we explore the relationship between our main independent variable of interest and invention and funding outcomes.

<Insert Table 5 here>

Column (1) reports that startups whose knowledge-base is closer to their founders' academic work have a higher number of patents on average: a 10p.p decrease in distance is associated with a 18% increase in the number of patents. Column (2) uses the same dependent variable but with a Poisson model in order to account for the count nature of patents, and leads to a similar conclusion. We also report in Table A1 of the Appendix that there is no significant difference in the number of forward citations between patents which are more or less distant to their founders' previous academic work. Columns (3) and (4) indicate that there is no significant relationship between distance and the amount of funding raised following a startup

[11]In results left unreported we apply further instruments for the percentage of self-cites (our treatment variable), such as using the number of examiner-added scientific citations. Similar to the judge leniency instrument (Arnold et al., 2018; Dobbie et al., 2018), some examiners have a higher propensity than others to add scientific references. Conditioning on art-unit and application year, the random assignment of examiners to patents can in theory allow us to create an instrument capturing the propensity of examiners to add scientific citations. This would instrument for the number of scientific citations associated with each patent (the denominator of our treatment variable) and hence instrument for our treatment variable (*Percentage of self-cites*). However, despite results that show similar magnitudes and signs than our baseline estimates, our first-stage is too weak, probably because examiners add very few citations on average. Ideally, we would have liked to implement the concept of idea twins (Bikard, 2020) where founders with different academic knowledge bases would create startups embodying almost identical ideas, but is not a viable option given our already small sample size.

creation. Overall, these results suggest that the lower likelihood of acquisition experienced by startups whose knowledge base is closer to their founders' academic work is not related to a lower inventive quality or valuation by investors.

4.4 Can our results be explained by different stages of technological development?

While we do not find any significant difference in inventive quality, the lower success rate experienced by startups whose knowledge base is closer to their founders' academic work could be explained by a more nascent and less developed technology. In order to test this explanation, we construct a proxy for the development time of each startup's technology and examine its relationship with the percentage of self-cites variable. For each startup, we take the application year of its first patent, which can be conceptualized as the time when the technology was ready to be commercialized. We then consider the publication year of the first self-cite of this patent, which can be thought of as the closest approximation to the scientific start date of research around the technology. This should capture when the idea around the technology was first studied in an academic lab but was not yet ready to be commercialized. We then subtract the self-cite publication year from the patent application year and use this difference as a proxy for the time needed to develop the technology. The higher this difference, the more time should have been needed to bring a scientific invention to a commercializable state.[12] Columns (1) and (2) of Table A2 display that there is a negative association between the percentage of self-cites variable and the time of technology development: a 10p.p decrease in distance (i.e. a 10p.p increase in self-cites) is associated with about 1.6 to 1.4 years less of technological development. We reiterate this analysis by considering the time between the publication year of the first self-cite and the year of startup creation and find the same negative association in columns (3) and (4). These results suggest that the lower acquisition likelihood experienced by startups that are closer to their founders' academic work does not seem to be due to a more nascent technology. If anything, their technology seems to be transferred faster to the market. These findings are consistent with our conceptual arguments: founders who first identify a problem in the market and then try to find solutions by using previous knowledge combining a broad range of work (i.e., those founders with a lower percentage of

[12]We can only calculate this measure for the 300 startups for which have at least one self-cite

self-cites) may take longer to develop their technology, but seem to be more successful on the market than those who are more driven by a technology resulting from their own lab's work but potentially still looking for a problem to solve.

4.5 Founder level academic outcomes

Besides their entrepreneurial endeavors, academic entrepreneurs have another job to fulfill: produce knowledge. In order to assess the relationship between the distance in knowledge bases and academic output, we implement a two-way fixed effects specification:

$$Y_{it} = Treated_i \times Post_t + \alpha_i + \delta_t + X_{it} + \epsilon_{it}$$

where i indexes individuals and t indexes year. In our specification, $Treated_i$ equal 1 if individual i is part of our treatment group (that we define below) and 0 otherwise, while $Post_t$ equal 1 for years following startup creation. X_{it} comprises bins of experience dummies where each dummy has a size of 4 years.[13] We account for differences between the treatment and control groups by including a full set of individual fixed effects. Hence, our identification strategy does not require that professors in the treatment and control groups were similar when they created their startup. Rather, we assume that academic activity in these two groups would have evolved in parallel. We also present evidence in support of this assumption with a staggered event study specification where we interact the treatment group indicator with a full set of time leads and lags:

$$Y_{it} = \sum_{\tau \neq -1} \beta_\tau \mathbf{1}[\tau = t - t_S^*] + \alpha_i + \delta_t + X_{it} + \epsilon_{it}$$

where t_S^* indexes the year of startup creation. We will use $\{\beta_\tau, \tau < -1\}$ to identify potential pre-trends.

Our main dependent variables of interest are the number of publications, citations, top publications (defined as papers which receive more citations than the average number of citations of other papers in the same area of research and year)[14] and citations per publication.

[13] One dummy for years of experience 1-5, one dummy for years of experience 6-10, one dummy for years of experience 11-15, one dummy for years of experience 16-20, one dummy for years of experience 21-25 and one dummy for years of experience above 25. Year 1 of experience corresponds to the first year for which we observe an academic publication.

[14] This measure is provided by Dimensions AI and is available for PubMed publications. The area of research

Conditional on publishing, we also examine the percentage of papers in which founders appear as first author – which often indicates that the individual performed the major share of the work, and the percentage of papers in which founders appear as last author – which often indicates that the individual was the Principal Investigator of the project.[15]

Among the 676 professors, we keep the 640 founders who have founded only one startup in order to have clear and easily identifiable pre- and post-treatment periods. Among these professors, we were able to uniquely match 561 of them (88%). In our baseline specification, our treatment group is composed of individuals whose startup has patents that cite at least one of their academic papers (i.e., *Percentage self-cites* > 0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites* $= 0$). We will explore the sensitivity of our results to other definitions of the treatment and control groups in subsequent analysis.

<Insert Table 6 here>

Table 6 provides summary statistics for the treatment and control groups. Panel A indicates that startups whose founder is in the treatment group have more patents and more scientific patents on average than the startups whose founder is in the control group, which echoes the previous findings about invention outcomes. Panel B explores academic output at the time of startup creation. Both groups have, on average, slightly more than 20 years of experience when they began their entrepreneurial activity. There is no significant difference between the treatment and the control groups with regards to the number of publications, top publications, citations and authorship, prior to startup creation. Though our empirical strategy does not require both groups to be comparable in terms of level, it is reassuring to see that there is no systematic difference in academic output between the treatment and control groups.

Figure A2 in the Appendix displays raw graphs of the main dependent variables of interest as a function of time from (negative values) and since (positive values) startup creation. The number of publications increases for both groups until startup creation. After that, it stabilizes and then seems to slightly decrease for the treatment group, while it first increases and then plateaus at about 11 publications per year for the control group. The number of top publications appears to be decreasing for both groups after startup creation, but to a greater

is defined by publications co-cited with the article of interest and is therefore dynamically created.

[15] A robust social norm in the Life Sciences and Engineering assigns last authorship to the most senior faculty running the lab and leading projects, while first authorship is given to younger individuals involved in writing the paper (Azoulay et al. (2009)), and conducting the bulk of the research work

extent for the treatment group. As expected, the number of cites is on a decreasing trend as papers published more recently had less time to gather citations. The percentage of papers written as last author varies between 0.35 and 0.55. It is noteworthy that this value decreases for both groups after startup creation, potentially suggesting some change in research strategy. Overall, from the graphs, it seems that the number of publications and top publications of the treatment group is negatively impacted compared to the control group after entering entrepreneurship.

We now go beyond the raw data and explore results in more detail in a regression framework. Table 7 shows the interaction term of the difference-in-difference regression. There is no significant difference across the six main outcomes between the treatment and control groups.

<Insert Figure 1 and Table 7 here>

We also run the event study regression for the six outcomes of interest and plot the coefficients in Figure 1. There is no apparent pre-trend, and the coefficient on the interaction term is not significantly different from 0 in the post-period, except for the number of publications and citations which seem to be significantly negative in the longer term.

Next, we examine the heterogeneity of this result as a function of the degree of reliance of founders on their academic work. First, we keep the same definition for the control group (*Percentage self-cites*= 0) but we vary the treatment group by looking separately at professors' academic output whose distance with their startup falls in different percentiles of the distribution of the *Percentage self-cites* variable. A plot of the coefficient on the interaction term is presented in Figure 2.

<Insert Figure 2 here>

Results indicate that there is no significant difference between the treatment and control groups for individuals whose academic work is relatively far from the knowledge base of their startup (up to the 75^{th} percentile). However, we find a decrease on the number of publications for individuals who are above the 75^{th} percentile of closeness post-startup creation compared to the control group, and a decrease on the number of publications and top publications for individuals who are above the 90^{th} percentile of closeness. This represents a decrease of about 2 publications and 1 top publication from the mean per year for founders that are closest to the knowledge base of their startup. Taking rough estimates from prior work on the grant dollar to paper relationship, and assuming that this amount should proxy for the lower bound of what a paper is worth (investment should be equal or below expected returns) our back-of-the-envelope calculations suggest that this could represent a loss of value generated from publications of 40,000 - 333,000 dollars per year per academic entrepreneur.[16] Naturally, these numbers are crude and do not take into account the gains from pursuing other activities. These gains could, quite possibly also drive down the costs that go into producing a paper.

We repeat this exercise, this time by varying the threshold below/above which we categorize the control and treatment groups. Figure 3 plots the coefficient of the difference-in-difference interaction term.

<Insert Figure 3 here>

Each point corresponds to a specific threshold for defining the treatment and control groups: for instance the point corresponding to p(1) on the x-axis is the interaction term of the difference-in-difference regression when we define the control (treatment) group as being below (above) the 1st percentile of the treatment variable. Results convey a similar story as before: in particular, professors who tend to be very close to their startups in the knowledge space experience a decrease in the number of publications and top publications.

4.6 Exploring mechanisms behind the change in academic output

So far, our findings indicate that professors who rely extensively on their academic work to create their startup experience (i) poorer exit prospects on the entrepreneurial market and (ii) a decrease in their overall number of publications and top publications after startup creation, compared to individuals who do not rely (or who rely less extensively) on their previous academic work. To get a better understanding of what may be driving poorer performance on the exit market and the changes observed in academia, we examine further individual level outcomes. These are: co-authorship, research direction and patenting activity.

First, we explore founders' co-authorship networks and how they evolve with entrepreneurship. We calculate the number of unique co-authors that each founder has in a given year - conditional on having at least one paper that year - differentiating between co-authors who

[16] Based on Boyack and Börner (2003), Druss and Marcus (2005) and Leydesdorff and Wagner (2009), the range lies between 0.6 and 5 published papers per \$100k in funding. Taking (100,000/5)*2 = 40,000 and (100,000/0.6)*2 = 333,000, we obtain our estimated range.

share the same affiliation and co-authors working at different institutions. Results are presented in Table 8 and show the difference-in-difference coefficient for the yearly number of unique co-authors (column 1), the yearly number of unique co-authors from the same institution (column 2) and the yearly number of unique co-authors from a different institution (column 3).

<Insert Table 8 here>

Overall, we find no significant difference between the treatment and control groups. However, once again, these results mask significant heterogeneity. As before, we rerun the difference-in-difference regressions by changing the definition of the treatment and control groups. Results are presented in Figure 4 Panel (a), where we keep the definition of the control group as individuals with a *Percentage of self-cites* variable equal to 0 but vary the definition of the treatment group and Figure 4 Panel (b), where we vary the threshold defining the treatment and control groups.

<Insert Figure 4 here>

In both cases, we find that founders who rely extensively on their previous academic output (above the 75^{th} percentile of the *Percentage of self-cites* variable) experience a decrease in the number of co-authors after startup creation compared to the control group. Interestingly, this result seems driven both by a decrease in the number of co-authors coming from the same institution as well as a decrease in the number of co-authors working at a different university. Overall, this suggests that founders who tend to use a technology-centric strategy narrow their network of co-authorship after entering entrepreneurship.

Next, we explore changes in research focus after startup creation for professors relying more or less on their previous academic work. The Dimensions AI database uses machine learning techniques to derive concepts in papers' abstracts and rank them based on their relevance on a scale from 0 (not relevant) to 1 (very relevant). We first calculate for each paper the number of concepts with a score above 0.5, which informs us about the diversity of main ideas present in a paper. We then average this measure at the professor-year level. Similarly, among concepts with a score above 0.5, we calculate the number of new concepts, defined as concepts that do not appear in any previous papers published by the founder in the 5 preceding years. We then calculate the share of new concepts by dividing the number of new concepts by the overall number of unique concepts and average this measure at the professor-year level.[17] Figure 5 shows the interaction term of the difference-in-difference regression for both the number of unique concepts and the share of new concepts, looking at heterogeneity across founders.

<Insert Figure 5 here>

As before, Figure 5 Panel (a) keeps the control group as being individuals with a *Percentage of self-cites* variable equal to 0 and varies the treatment group definition, while Figure 5 Panel (b) varies the threshold defining the treatment and control groups. In both cases, we find that founders who are very close to their startup's knowledge base tend to decrease the number of concepts they use in their papers after startup creation compared to the treatment group. However, there does not seem to be any difference in the share of new concepts they use in their papers, suggesting that this narrowing does not seem to come at the expense of topic exploration.

We then reiterate the analysis using the number of unique Medical Subject Headings (MeSH)[18] terms and the share of new MeSH terms[19] used by founders yearly. Results are presented in Figure 6 and show similar patterns: founders who are very close to their startup's knowledge base tend to decrease the number of terms they use in their papers after startup creation compared to the treatment group, but there is no significant impact on the share of new terms used.

<Insert Figure 6 here>

Overall, this suggests that founders who tend to rely more on their own academic work, with a more technology-centric approach to entrepreneurship, experience a change in their research agenda after founding, focusing on a narrower set of ideas. This interpretation is in line with canonical theoretical models suggesting that private sector research is more focused than academic research, where the fundamental trade-off lies in giving up creative control for higher, more certain payoffs (Aghion et al., 2008). This may occur through specialization around a smaller, defined, sliver of the research "pie" (Jones, 2009). It is interesting to find that this narrowing does not seem to come at the expense of exploration.

[17] Using the share rather than the number of new concepts allows to control for differences in research breadth. E.g., 5 new concepts could be high if the research area of a paper is narrow with an average of e.g., 10 unique concepts (50% of novelty) but is low if the research area of a paper is broad with e.g., 50 unique concepts (10% of novelty).

[18] MeSH terms are a set of keywords maintained by the National Library of Medicine that indexes the intellectual content of PubMed articles. See for example Azoulay et al. (2019).

[19] New MeSH terms are those that haven't been used in the previous 5 years.

Following the same difference-in-difference strategy, we also explore changes in patenting activity following startup creation. We use Dimensions AI to match each founder to their patenting output and reiterate the difference-in-difference analysis. Results are presented in Figure 7. Varying the definition of treatment and control groups, we find that founders who rely more extensively on their previous academic output increase their patenting activity after startup creation compared to founders who rely less on it.

<Insert Figure 7 here>

While suggestive only, we believe that, in totality, the previous results offer tentative evidence for an economic payoff concern explanation. Because founders' research is directly linked to the knowledge base of their startup that they seek to commercialize, it becomes more critical to control what is publicly disclosed. At the same time such control may be at odds with the demands of potential acquirers hoping to integrate the new technology with, e.g., their existing technology stack or other complementary assets (Arora, Fosfuri and Roende, 2022). In our interviews, an acquirer referred to this as "grafting on new technology to the [firm's established] tech[nology] stem."

Naturally, this is not necessarily the exclusive driver behind our presented results. It could also be, for example, that founders with a higher reliance on their previous academic work suffer from an escalation of commitments: they experience more difficulties on the entrepreneurship side, which leads them to progressively invest less of their time in academia, shrink their lab size and hence write less papers. This mechanism would also be consistent with a decrease in network size and research breadth.

However, our interviews suggest that escalating commitment is not the main driver of these results. During our conversations, a founder relatively close to his startup's knowledge base explained that he decreased his collaborations with colleagues, from within and outside his department, because the type of research he was conducting had shifted and was more catered towards his new entrepreneurship project. He mentioned this was also the reason why he narrowed his research agenda, though it did not come at the expense of exploration because of the interdisciplinary nature of the project. The fact that patenting activity increases and the share of new knowledge concepts used in papers does not significantly decrease tend to give greater support to an economic payoff concerns story.

5 Robustness

Goodman-Bacon (2021) provide evidence that when there is variation in treatment timing, the two-way fixed effects estimator is a weighted average of all possible 2×2 difference-in-difference estimators, which can lead to bias when there is dynamic treatment effect. Hence, we re-estimate the more traditional difference-in-difference results by creating several datasets (one per year) where we drop already-treated units from the control group, and stack these datasets together. Results are presented in Table 9 and appear robust to this specification (i.e., on average, there is no significant difference in academic output between founders who rely on their academic work and those who do not rely on it).

<Insert Table 9 here>

Next, we provide evidence that our results are robust to excluding startups whose patents do not rely on science. Indeed, these startups were assigned a distance value of 0 as treatment since they (mechanically) were not citing their founders' academic work, but they could be different from startups relying on science but not citing their founders' academic work that have a distance value of 0 too. Excluding startups that do not rely on science leaves us with a sample of 318 startups, with a mean *Percentage of self-cites* variable of 9.0% (see Figure A3 in the Appendix for an histogram of the treatment variable).

Table A3 and A4 replicate the previous regressions that examine exit and invention outcomes. The results on the exit market are similar to our baseline: the coefficient on the acquisition likelihood remains negative and is roughly the same magnitude (-0.213 vs -0.240), while there is no significant effect on *IPO* and *Success* overall. As before, this cannot be explained by a lower inventive quality: Table A4 shows that there is no significant difference regarding the number of patents (columns (1) and (2)) or funding (columns (3) and (4)).

We now explore the robustness of our results regarding academic outcomes. As before, we replicate our heterogeneity analysis, defining the treatment and control groups of founders based on whether their percentage of self-cites lies above or below a specific percentile of the treatment variable. Results are presented in Figure A4 and are very similar to the main ones: the number of publications and top publications for founders who rely extensively on their previous academic work decreases post-startup creation compared to the control group.

A further concern is that startups with several professors in the founding team might exhibit systematically different results compared to firms with a single professor-founder. In Table A5, we replicate the exit analysis by differentiating between startups with one professor only in the founding team (columns 1 to 3) and startups with more than one professor in the founding team (columns 4 to 6). In both cases, we find that startups that are closer to their founders' academic work have a lower likelihood of being acquired. Table A6 replicates the results on the invention and funding outcomes, differentiating again by founding team composition. For startups with one professor in the founding team, we find - similar as before - that a lower distance is positively related to patenting activity but there is no significant effect on funding. Ventures with multiple professors whose knowledge base is closer to their founders' academic work appear less likely to receive funds, but the coefficients on patents are not statistically significant. Overall, there is no consistent evidence that inventive quality systematically differs for firms relying more or less on their founders' academic work in both samples.

We then look at differences in academic output, replicating the previous heterogeneity analysis for ventures with one professor only in the founding team (Figures A5 and A6) and ventures with several professors in the founding team (Figures A7 and A8). In both cases, we see a decrease in the number of publications for founders that are closest to their startup in knowledge space. The decrease in the number of top publications experienced by founders closest to their startup seems to be driven by those individuals part of a single-professor team, while founders who are part of a multiple-professor team who closely rely on their academic work experience a decrease in citations. Figures A9 (firms with one professor in the founding team) and A10 (firms with multiple professors in the founding team) indicate that the number of co-authors decreases for founders who are closer to their startup's knowledge base. Interestingly, this decrease is mainly driven by co-authors from different affiliations for ventures with one professor in the founding team and by co-authors from the same affiliation for ventures with several professors. Though the results are noisier given the smaller sample size, Figures A11 (firms with one professor in the founding team) and A12 (firms with multiple professors in the founding team) suggest that the number of unique concepts used in papers published after startup creation does not differ or decreases for founders who rely more on their previous academic work (compared to the control group), while the coefficient for the share of new concepts is not statistically significant. Finally, Figures A13 (firms with one professor in the founding team) and A14 (firms with several professors in the founding team) show an increase in patenting activity for founders close to their startup knowledge base after startup creation, compared to the control group, though the effect is bigger for founders in a single-professor team. Overall, it does not seem that startups with several professors in the founding team exhibit systematically different results compared to firms with a single professor-founder.

We also replicate our results by using all patents pertaining to a startup to calculate the distance measure. Said differently, we calculate the percentage of self-cites at the startup level by considering all its patents, and not just its oldest ones. This measure captures a more dynamic idea of the distance between founders' academic work and the knowledge base of their startup. Results are unchanged. Table A7 reports that startups closer to their founders' academic work experience a lower likelihood of getting acquired, with a magnitude similar to our main result (coefficient of -0.171 vs -0.240). Table A8 shows that if anything, these startups have a higher inventive output, highlighting again that this lower success on the exit market is not driven by a lower quality of their technology. Figures A15 and A16 replicate the results on the academic market: once again, founders who are closer to their startup's knowledge base experience a decrease in publications and top publications after entering entrepreneurship, compared to the control group. We also find the same results regarding the decrease in the number of co-authors (Figure A17), both within and outside the focal institution and the decrease in the number of concepts used, although as before, this does not come at the expense of greater exploration (Figures A18 and A19). As before, our results suggest that patenting activity increases for those founders that are closest to the knowledge base of their startup (Figure A20).

6 Discussion and conclusion

Herein, we examine how the extent to which an academic founder exploits their own technologically unique knowledge in the formation of their startup relates to entrepreneurial success and academic output after creation. To do so, we analyze a sample of 510 academic startups corresponding to 676 founder-professors in biomedicine. We find evidence that startups which rely more on their founders' academic work are less likely to be acquired. However, this does not seem to result from lower invention quality or an earlier stage of technological development. Rather, these results provide suggestive evidence that, while more or less distant startups do not differ with regards to their ability to generate valuable technologies, startups that are closer to their founders' academic work may embody a different type of knowledge that is less easily transferable to the private sector.

In a next step, we examine the difference in academic output between founders who created startups that rely on their academic work and those who founded startups that do not. Assessing the consequences of entrepreneurial strategy on founders' core output is critical to better comprehend potential trade-offs associated with commercial engagement. While we do not find any significant difference in academic output after startup creation between these two groups on average, our results exhibit significant heterogeneity. In particular, founders who rely more extensively on their previous academic work show a decrease in the number of publications and top publications after entering entrepreneurship. In subsequent analysis, we find suggestive evidence that those founders with shorter distances to their startup's knowledge base tend to decrease their collaborations with other scientists – both within and outside their own institution – after venture creation compared to founders who rely less on their previous academic work. As a consequence, less "offspring" lines of inquiry are pursued; an interpretation further supported by the reduction in citations. They also tend to narrow their research focus, though this does not appear to come at the expense of greater exploration. We further detect an increase in their patenting activity.

While we cannot draw any definitive conclusion, we believe these findings provide tentative evidence that IP control and related economic payoff concerns may be a mechanism at play. In other words, founders who rely heavily on their own research strive to reap the benefits of their work; work over the direction of which they decide. Aiming for creative control and focus appears at odds with either the academic or private sector development model leaving these founders, what appears to be, stuck in the middle (Aghion et al., 2008). Our extensive interviews with academic founders confirm our conjectures – at least anecdotally.

As a consequence, our findings draw attention to important economic implications for founders, potential acquirers and universities, with suggestive evidence that some types of startups (and founders) may have better commercialization prospects. In particular, founders whose startups rely on knowledge beyond their own seem to achieve better performance on the exit market, while at the same time remaining comparatively more productive in academia compared to those founders who rely deferentially more on their own work. While the central role of resources – knowledge being one of the most critical – in shaping established firm's strategic outcomes generally has been well-documented (Barney, 1991), less is known about the role of initial knowledge endowments. Building on the work by Agarwal et al. (2004), Sørensen (2007), Åstebro et al. (2011), and Stenard and Sauermann (2016) who examine the role of capability differentials in shaping both the decision to become an entrepreneur and the entry mode into entrepreneurship, our findings suggest that the knowledge foundation of a nascent firm and its composition may thereby be particularly critical to ensure successful exits. Contrary to our expectations, however, heavy reliance on one's own highly specialized, unique knowledge may not lead to an overall advantage on the exit market.

Importantly, we also highlight the trade-off academic founders who rely heavily on their own work appear to be making. With entry into entrepreneurship, we perceive a notable shift in the research approach of the professors involved. Namely, their work becomes more narrow, more focused, and more targeted at achieving economic payoff (such as through patenting activity). Although, this may come at a lower "disutility cost" for the professors themselves, the ramifications for academia as a whole may be long-lasting, especially should high powered incentives be put in place to encourage academics to start companies building heavily on their own research (Roach, 2017). Following the theoretical model proposed by Aghion et al. (2008), as a consequence, society as a whole may miss out on the discovery of entirely new lines of inquiry that are not pursued given early privatization of research. This is a conjecture we leave for future work to examine.

As such, our study may aid in developing more nuanced and targeted policies for academic innovation, acknowledging that some founders may benefit from deeper involvement and support in order to be successful. Moreover, our work directly replies to the call of understanding what the role of intermediaries (e.g., Technology Transfer Offices) may be (https://twitter.com/heidilwilliams_/status/1646889128242593792). From our results, it appears that initiatives targeted at helping "non-academic" founders better find solutions in academic science to solve existing problems, may be a viable path forward. Perhaps, more counter-intuitively, is that to be successful on the exit market, founders may be better advised to build their nascent firms around a knowledge-base beyond their own, which to some academic scientist represents a departure from their occupational mental model (El-Awad et al., 2022).

Our study is not without limitations. Although we take great care in controlling for potential confounders and show that our results are robust to different specifications, we cannot formally claim causality. Future research could try to find clever ways to implement quasi-experimental strategies that could help deal with identification challenges. Future work could also try to study features other than knowledge distance and their impact on entrepreneurial and academic success. Given the complexity of firm creation, and persistent consequences of early decisions around the resources used as the foundation of the firm (Geroski et al., 2010), opportunities for follow-on work abound.

References

Abramo, G., D'Angelo, C. A., Ferretti, M. and Parmentola, A. (2012), 'An individual-level assessment of the relationship between spin-off activities and research performance in universities', *R&D Management* **42**(3), 225–242.

Agarwal, R., Echambadi, R., Franco, A. M. and Sarkar, M. B. (2004), 'Knowledge transfer through inheritance: Spin-out generation, development, and survival', *Academy of Management journal* **47**(4), 501–522.

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

Agrawal, A. and Henderson, R. (2002), 'Putting patents in context: Exploring knowledge transfer from MIT', *Management Science* **48**(1), 44–60.

Alvarez, S. A. and Busenitz, L. W. (2001), 'The entrepreneurship of resource-based theory', *Journal of Management* **27**(6), 755–775.

Ambos, T. C., Mäkelä, K., Birkinshaw, J. and d'Este, P. (2008), 'When does university research get commercialized? creating ambidexterity in research institutions', *Journal of Management Studies* **45**(8), 1424–1447.

Arkes, H. R. and Blumer, C. (1985), 'The psychology of sunk cost', *Organizational Behavior and Human Decision Processes* **35**(1), 124–140.

Arnold, D., Dobbie, W. and Yang, C. S. (2018), 'Racial bias in bail decisions', *The Quarterly Journal of Economics* **133**(4), 1885–1932.

Arora, A., Belenzon, S. and Suh, J. (2022), 'Science and the market for technology', *Management Science* **68**(10), 7176–7201.

Arora, A., Fosfuri, A. and Roende, T. (2022), Caught in the middle: The bias against startup innovation with technical and commercial challenges, Working Paper 29654, National Bureau of Economic Research. **URL:** *http://www.nber.org/papers/w29654*

Åstebro, T., Chen, J. and Thompson, P. (2011), 'Stars and misfits: Self-employment and labor market frictions', *Management Science* **57**(11), 1999–2017.

Audretsch, D. B. (2014), 'From the entrepreneurial university to the university for the entrepreneurial society', *The Journal of Technology Transfer* **39**, 313–321.

Azoulay, P., Ding, W. and Stuart, T. (2009), 'The impact of academic patenting on the rate, quality and direction of (public) research output', *The Journal of Industrial Economics* **57**(4), 637–676.

Azoulay, P., Fons-Rosen, C. and Zivin, J. S. G. (2019), 'Does science advance one funeral at a time?', *American Economic Review* **109**(8), 2889–2920.

Barney, J. (1991), 'Firm resources and sustained competitive advantage', *Journal of Management* **17**(1), 99–120.

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

Bhaskarabhatla, A. and Hegde, D. (2014), 'An organizational perspective on patenting and open innovation', *Organization Science* **25**(6), 1744–1763.

Bikard, M. (2020), 'Idea twins: Simultaneous discoveries as a research tool', *Strategic Management Journal* **41**(8), 1528–1543.

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

Blumenthal, D., Epstein, S. and Maxwell, J. (1986), 'Commercializing university research', *New England Journal of Medicine* **314**(25), 1621–1626.

Boyack, K. W. and Börner, K. (2003), 'Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers', *Journal of the American Society for Information Science and Technology* **54**(5), 447–461.

Cohen, W. M. and Levinthal, D. A. (1990), 'Absorptive capacity: A new perspective on learning and innovation', *Administrative Science Quarterly* **35**(1), 128–152.

Cohen, W. M., Sauermann, H. and Stephan, P. (2020), 'Not in the job description: The commercial activities of academic scientists and engineers', *Management Science* **66**(9), 4108–4117.

Colombo, M. G. and Piva, E. (2012), 'Firms' genetic characteristics and competence-enlarging strategies: A comparison between academic and non-academic high-tech start-ups', *Research Policy* **41**(1), 79–92.

Conti, A. and Roche, M. (2021), 'Lowering the bar? external conditions, opportunity costs, and high-tech start-up outcomes', *Organization Science* **32**(4), 965–986.

Crespi, G., D'Este, P., Fontana, R. and Geuna, A. (2011), 'The impact of academic patenting on university research and its transfer', *Research Policy* **40**(1), 55–68.

Czarnitzki, D. and Toole, A. A. (2010), 'Is there a trade-off between academic research and faculty entrepreneurship? evidence from us nih supported biomedical researchers', *Economics of Innovation and New Technology* **19**(5), 505–520.

Dasgupta, P. and David, P. A. (1994), 'Toward a New Economics of Science', *Research Policy* **23**(5), 487–521.

Dobbie, W., Goldin, J. and Yang, C. S. (2018), 'The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges', *American Economic Review* **108**(2), 201–240.

Druss, B. G. and Marcus, S. C. (2005), 'Tracking publication outcomes of national institutes of health grants', *The American journal of medicine* **118**(6), 658–663.

El-Awad, Z., Brattström, A. and Breugst, N. (2022), 'Bridging cognitive scripts in multidisciplinary academic spinoff teams: A process perspective on how academics learn to work with non-academic managers', *Research Policy* **51**(10), 104592.

Etzkowitz, H. (2003), 'Research groups as "quasi-firms": The invention of the entrepreneurial university', *Research Policy* **32**(1), 109–121.

Fini, R., Perkmann, M. and Ross, J.-M. (2022), 'Attention to exploration: The effect of academic entrepreneurship on the production of scientific knowledge', *Organization Science* **33**(2), 688–715.

Geroski, P. A., Mata, J. and Portugal, P. (2010), 'Founding conditions and the survival of new firms', *Strategic Management Journal* **31**(5), 510–529.

Goldfarb, B., Marschke, G. and Smith, A. (2009), 'Scholarship and inventive activity in the university: Complements or substitutes?', *Economics of Innovation and New Technology* **18**(8), 743–756.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Grimaldi, R., Kenney, M., Siegel, D. S. and Wright, M. (2011), '30 years after bayh–dole: Reassessing academic entrepreneurship', *Research Policy* **40**(8), 1045–1057.

Guzman, J. and Li, A. (2023), 'Measuring founding strategy', *Management Science* **69**(1), 101–118.

Higgins, M., Stephan, P. and Thursby, J. (2011), 'Conveying Quality and Value in Emerging Industries: Star Scientists and the Role of Signals in Biotechnology', *Research Policy* **40**(4), 605–617.

Hsu, D. H., Roberts, E. B. and Eesley, C. E. (2007), 'Entrepreneurs from technology-based universities: Evidence from MIT', *Research Policy* **36**(5), 768–788.

Jones, B. F. (2009), 'The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?', *The Review of Economic Studies* **76**(1), 283–317.

Leydesdorff, L. and Wagner, C. (2009), 'Research funding and research output: A bibliometric contribution to the us federal research roadmap', *J. of Informetrics* .

Lin, J. and Wilbur, W. J. (2007), 'Pubmed related articles: A probabilistic topic-based model for content similarity', *BMC bioinformatics* **8**(1), 1–14.

Marx, M. and Fuegi, A. (2020), 'Reliance on science: Worldwide front-page patent citations to scientific articles', *Strategic Management Journal* **41**(9), 1572–1594.

Marx, M. and Fuegi, A. (2022), 'Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations', *Journal of Economics & Management Strategy* **31**(2), 369–392.

Mowery, D. C., Nelson, R. R., Sampat, B. and Ziedonis, A. (2004), *Ivory Tower and Industrial Innovation*, Palo Alto: Stanford Business Books.

Murray, F. and 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.

National Science Board (2018), *Science and Engineering Indicators 2016*, Arlington, VA: National Science Foundation.

Nelson, R. R. (2004), 'The market economy, and the scientific commons', *Research Policy* **33**(3), 455–471.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A. and Sobrero, M. (2013), 'Academic engagement and commercialisation: A review of the literature on university–industry relations', *Research Policy* **42**(2), 423–442.

Powell, W. W. and Owen-Smith, J. (1998), 'Universities and the Market for Intellectual Property in the Life Sciences', *Journal of Policy Analysis and Management* **17**(2), 253–277.

Powell, W. W. and Snellman, K. (2004), 'The knowledge economy', *Annual Review of Sociology* **30**(1), 199–220.

Roach, M. (2017), 'Encouraging entrepreneurship in university labs: Research activities, research outputs, and early doctorate careers', *PLOS ONE* **12**(2).

Roche, M., Conti, A. and Rothaermel, F. T. (2020), 'Different founders, different venture outcomes: A comparative analysis of academic and non-academic startups', *Research Policy* **49**(10), 104062.

Roche, M. P. (2023), 'Academic entrepreneurship: Entrepreneurial advisors and their advisees' outcomes', *Organization Science* **34**(2), 959–986.

Rothaermel, F. T., Agung, S. D. and Jiang, L. (2007), 'University entrepreneurship: a taxonomy of the literature', *Industrial and corporate change* **16**(4), 691–791.

Schumpeter, J. A. (1912), 'Theorie der Wirtschaftlichen Entwicklung: Eine Untersuchung ueber Unternehmergewinn', *Kaptial, Kredit, Zins und den Konjunkturzyklus, Duncker&Humblot, Berlin* .

Shah, S. K. and Pahnke, E. C. (2014), 'Parting the ivory curtain: Understanding how universities support a diverse set of startups', *The Journal of Technology Transfer* **39**, 780–792.

Shane, S. and Stuart, T. (2002), 'Organizational endowments and the performance of university start-ups', *Management Science* **48**(1), 154–170.

Shane, S. and Venkataraman, S. (2000), 'The promise of entrepreneurship as a field of research', *Academy of Management Review* **25**(1), 217–226.

Shibayama, S., Walsh, J. P. and Baba, Y. (2012), 'Academic Entrepreneurship and Exchange of Scientific Resources: Material Transfer in Life and Materials Sciences in Japanese Universities', *American Sociological Review* **77**(5), 804–830.

Singh, V. K., Singh, P., Karmakar, M., Leta, J. and Mayr, P. (2021), 'The journal coverage of web of science, scopus and dimensions: A comparative analysis', *Scientometrics* **126**, 5113–5142.

Sørensen, J. B. (2007), 'Bureaucracy and entrepreneurship: Workplace effects on entrepreneurial entry', *Administrative Science Quarterly* **52**(3), 387–412.

Staw, B. M. (1976), 'Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action', *Organizational Behavior and Human Performance* **16**(1), 27–44.

Stenard, B. S. and Sauermann, H. (2016), 'Educational mismatch, work outcomes, and entry into entrepreneurship', *Organization Science* **27**(4), 801–824.

Stephan, P. E. (2012), *How Economics Shapes Science*, Vol. 1, Cambridge: Harvard University Press.

Stokes, D. E. (2011), *Pasteur's quadrant: Basic science and technological innovation*, Brookings Institution Press.

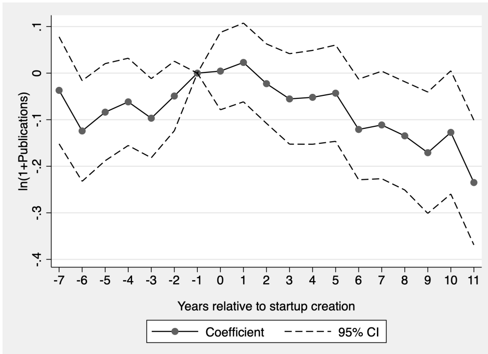
Stuart, T. and Ding, W. (2006), 'When Do Scientists Become Entrepreneurs? The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences', *American Journal of Sociology* **112**(1), 97–144.

Thursby, J. G. and Thursby, M. C. (2007), ‘University Licensing’, *Oxford Review of Economic Policy* **23**(4), 620–639.
Thursby, J. G. and Thursby, M. C. (2011), ‘Faculty participation in licensing: implications for research’, *Research Policy* **40**(1), 20–29.
Toole, A. A. and Czarnitzki, D. (2009), ‘Exploring the relationship between scientist human capital and firm performance: The Case of Biomedical Academic Entrepreneurs in the SBIR Program’, *Management Science* **55**(1), 101–114.
Toole, A. A. and Czarnitzki, D. (2010), ‘Commercializing science: Is there a university “brain drain” from academic entrepreneurship?’, *Management Science* **56**(9), 1599–1614.
Zucker, L. G., Darby, M. R. and Armstrong, J. S. (2002), ‘Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology’, *Management Science* **48**(1), 138–153.
Zucker, L. G., Darby, M. R. and Brewer, M. B. (1998), ‘Intellectual human capital and the birth of U.S. biotechnology enterprises’, *The American Economic Review* **88**(1), 290–306.

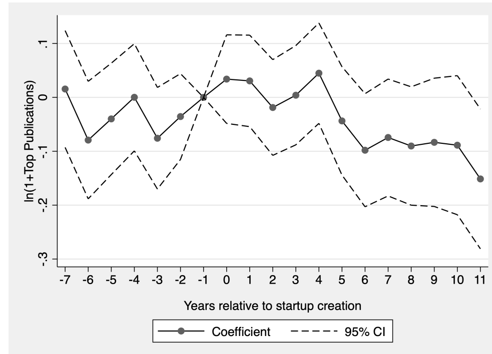
Figures

Figure 1: Event study graphs

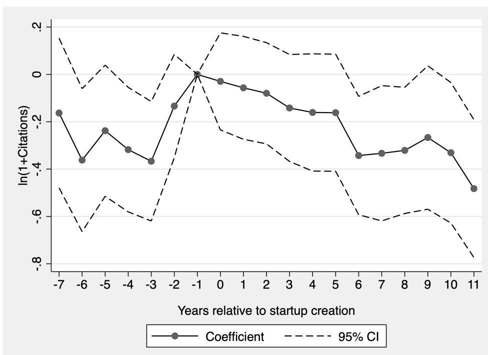
(a) Log(1+Publications)



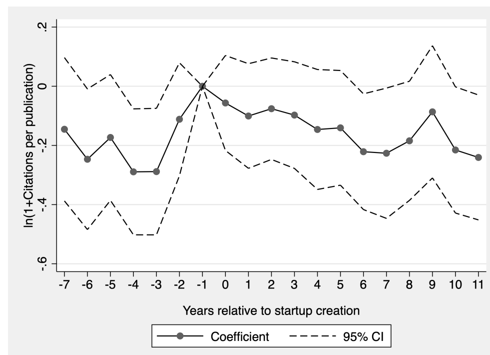
(b) Log(1+Top Publications)



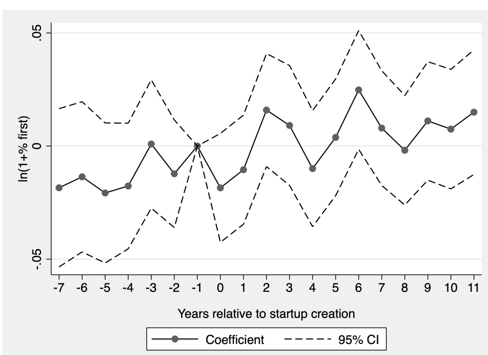
(c) Log(1+Citations)



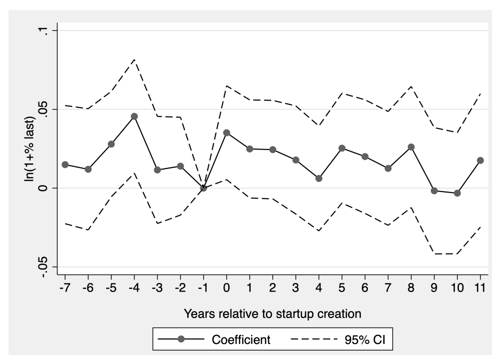
(d) Log(1+Citations per publication)



(e) Log(1+% papers as first author)



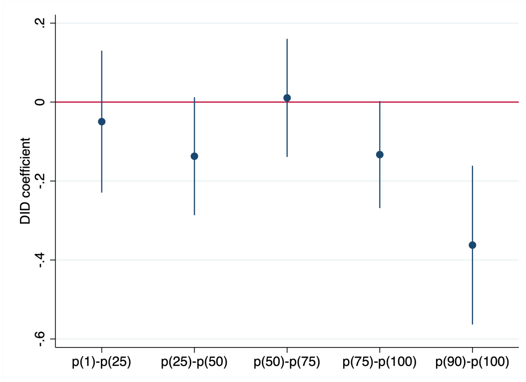
(f) Log(1+% papers as last author)



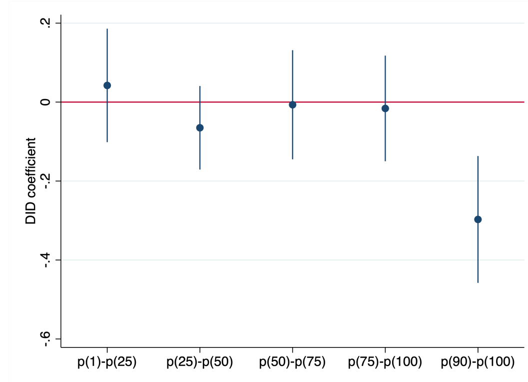
Notes: This figure shows the difference-in-difference coefficients over time, before startup creation (negative values of the x-axis) and after startup creation (positive values of the x-axis). We normalize all coefficients with respect to the year preceding startup creation. The treatment group is composed of individuals whose startup has a patent that cites at least one of their academic papers (i.e., *Percentage self-cites* > 0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites* $= 0$). Outcomes are the log yearly number of publications (panel (a)), the log yearly number of top publications (panel (b)), the log yearly number of citations (panel (c)), the log yearly number of citations per publication (panel (d)), the log yearly number of publications where the founder appears as first author (panel (e)) and the log yearly number of publications where the founder appears as last author (panel (f)). Dotted lines represent the 95% confidence interval.

Figure 2: Heterogeneity Analysis of Academic Output, varying the definition of the treatment group

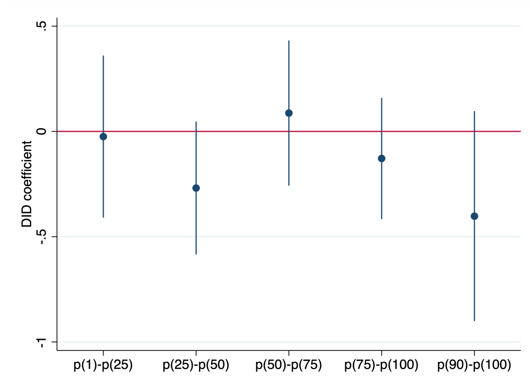
(a) Log(1+Publications)



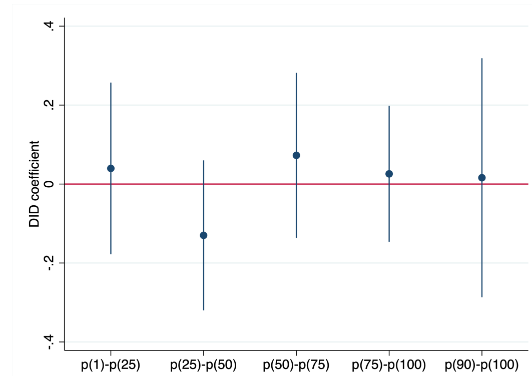
(b) Log(1+Top Publications)



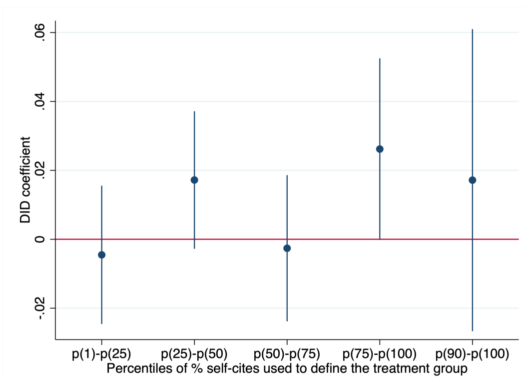
(c) Log(1+Citations)



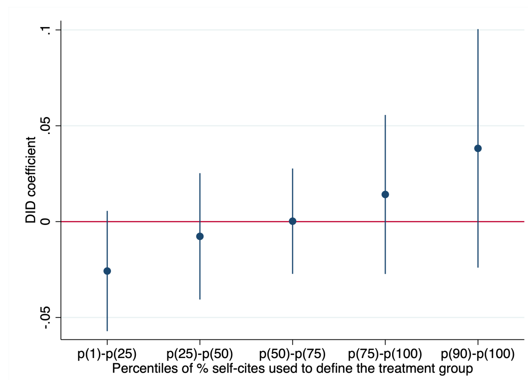
(d) Log(1+Citations per publication)



(e) Log(1+% papers as first author)



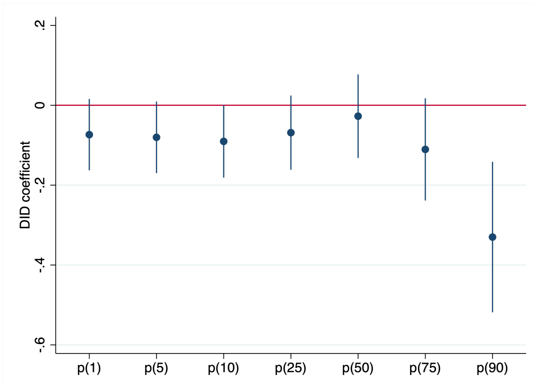
(f) Log(1+% papers as last author)



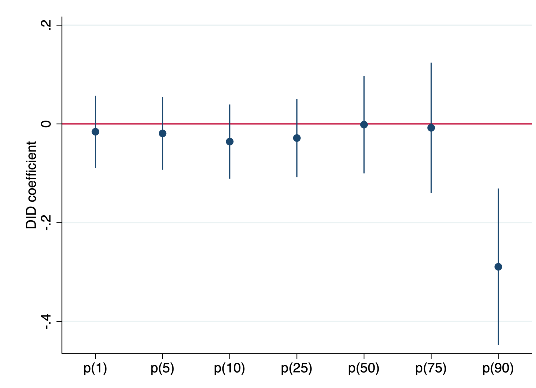
Notes: This figure shows the difference-in-difference coefficients varying the definition of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. For instance, p(1)-p(25) means that all individuals whose *Percentage self-cites* variable falls between the 1^{st} and the 25^{th} percentiles are included in the treatment group. The control group does not vary and includes all individuals with Percentage self-cites= 0. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log yearly number of citations. In Panel (d), the outcome is the log yearly number of citations per publication. In Panel (e), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (f), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure 3: Heterogeneity Analysis of Academic Output, varying the threshold to define treatment and control groups

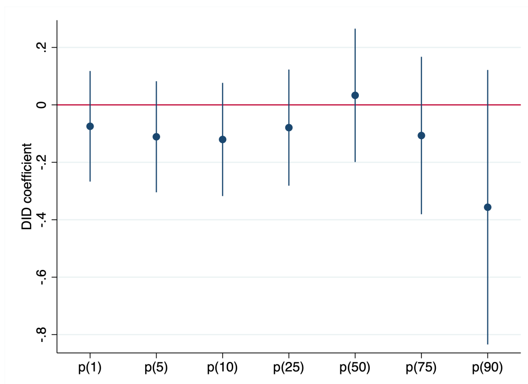
(a) Log(1+Publications)



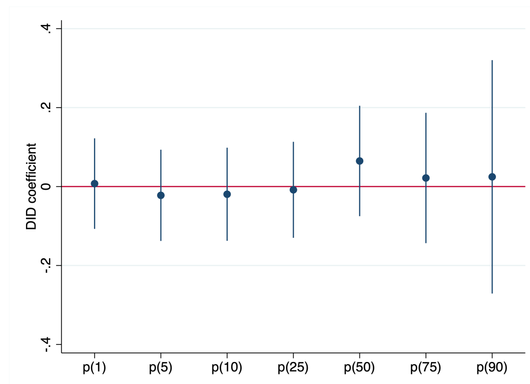
(b) Log(1+Top Publications)



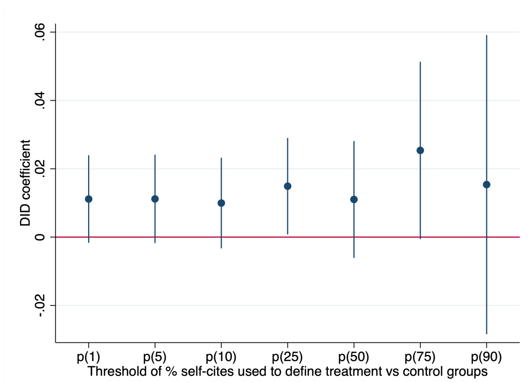
(c) Log(1+Citations)



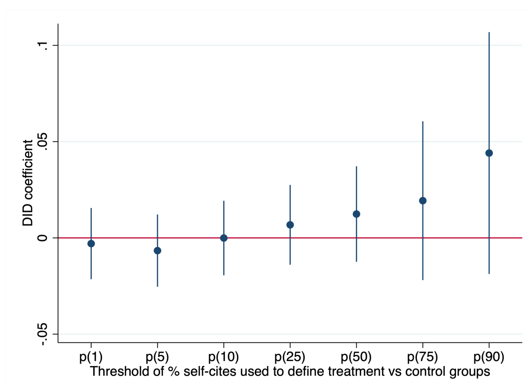
(d) Log(1+Citations per publication)



(e) Log(1+% papers as first author)



(f) Log(1+% papers as last author)

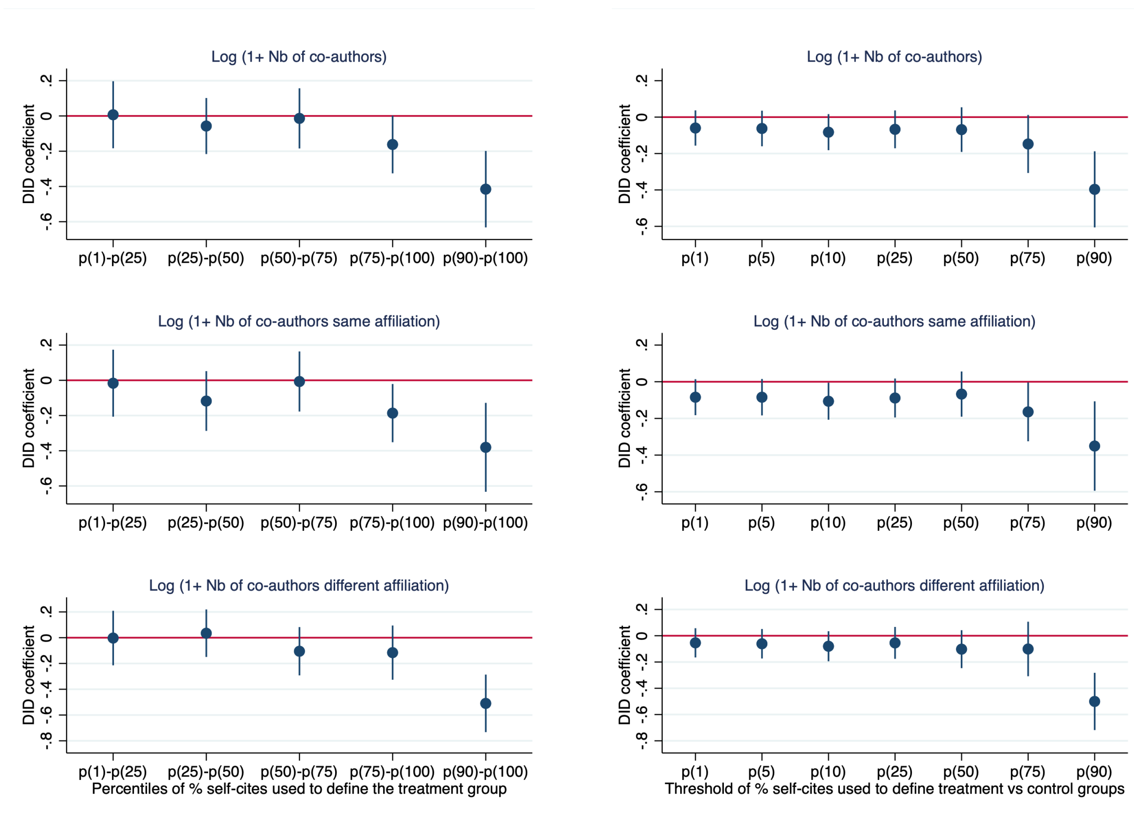


Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log yearly number of citations. In Panel (d), the outcome is the log yearly number of citations per publication. In Panel (e), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (f), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure 4: Heterogeneity Analysis of the Number of Co-Authors

(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups

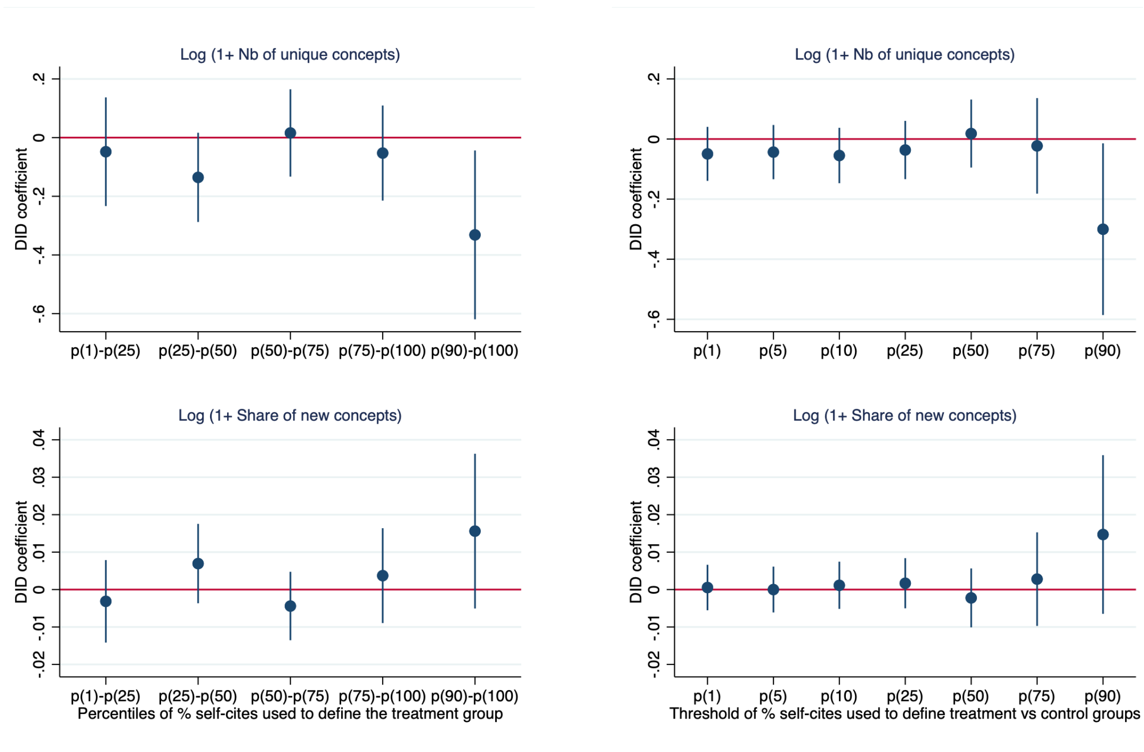


Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique co-authors founders published with in a given year. In the second row, the outcome is the log number of unique co-authors from the same institution founders published with in a given year. In the third row, the outcome is the log number of unique co-authors from a different institution founders published with in a given year. Each bar denotes the 90% confidence interval.

Figure 5: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts

(a) Using the same control group and varying the treatment group

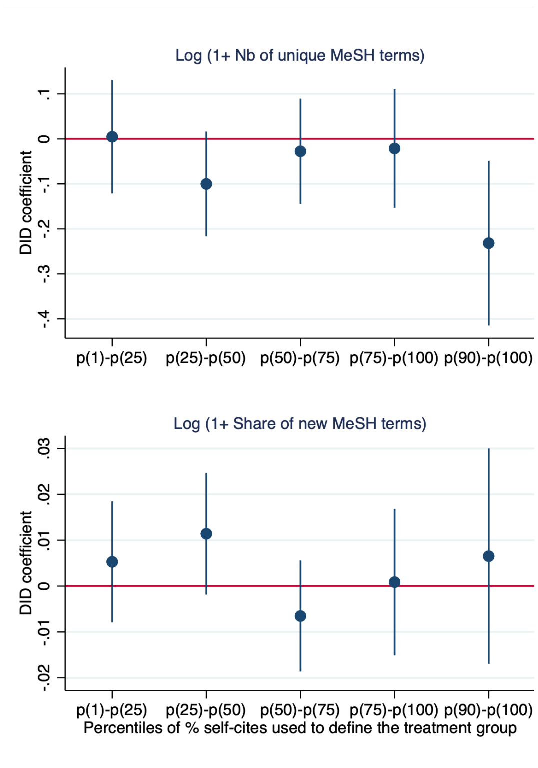
(b) Varying the threshold to define treatment and control groups



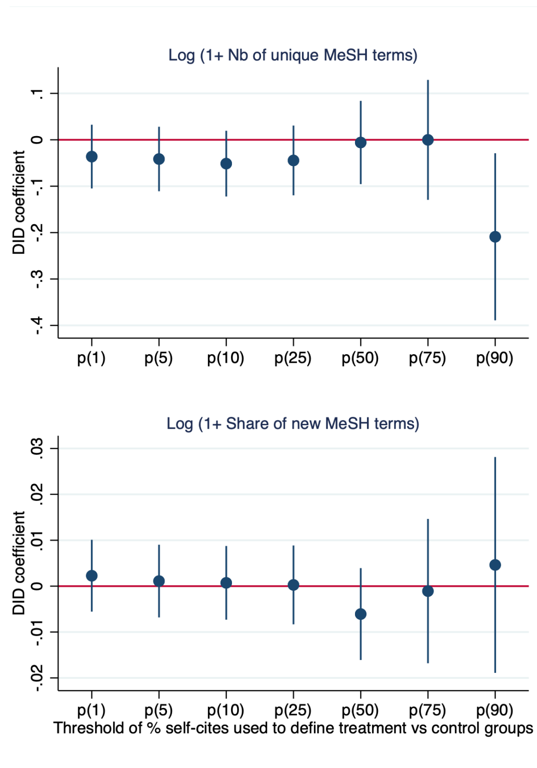
Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique concepts founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new concepts founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure 6: Heterogeneity Analysis of Research Focus, using MeSH terms

(a) Using the same control group and varying the treatment group

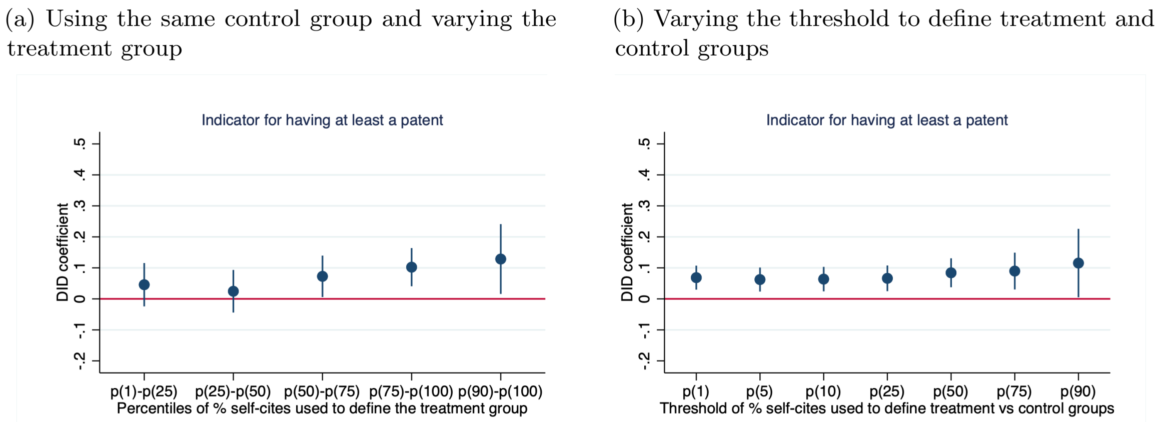


(b) Varying the threshold to define treatment and control groups



Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique MeSH terms founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new MeSH terms founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure 7: Heterogeneity Analysis, Patenting Activity



Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. The outcome is an indicator variable equal to 1 if the founder filed a patent in a given year and 0 otherwise. Each bar denotes the 90% confidence interval.

Tables

Table 1: Summary statistics, Startup level

	Min	p(50)	Mean	Max
Percentage of self-cites	0	0	.06	1
Number patents	0	1	4.2	73
Number patents in RoS	0	1	3.8	67
Number professors	1	1	1.4	6
Team size	1	2	2.3	7
At least one female founder	0	0	0.16	1
At least one top-tier university	0	1	0.66	1
Biotechnology sector	0	1	0.82	1
Founding year	2005	2008	2008	2012
Acquisition	0	0	.06	1
IPO	0	0	.06	1
Amount of funds raised within 5y ($million)	0	.49	12.6	181
Observations	510			

Notes: This table shows summary statistics for our sample of 510 academic startups.

Table 2: Determinants of founders' reliance on previous academic work

	(1)	(2)	(3)
	Percentage self-cites		
Experience=[6,10]	-0.00996	0.00949	0.00431
	(0.0290)	(0.0294)	(0.0301)
Experience=[11,15]	-0.0225	-0.000490	-0.00678
	(0.0285)	(0.0280)	(0.0289)
Experience=[16,20]	-0.0101	0.0140	0.00938
	(0.0298)	(0.0294)	(0.0296)
Experience$\geq$20	-0.00281	0.0226	0.0144
	(0.0299)	(0.0298)	(0.0305)
Top Institution	0.00474	0.00343	0.00288
	(0.00902)	(0.00914)	(0.00917)
Associate prof.	-0.00294	-0.00723	-0.0105
	(0.0193)	(0.0199)	(0.0198)
Full prof.	-0.00939	-0.0151	-0.0224
	(0.0193)	(0.0193)	(0.0202)
Female	-0.00687	-0.0113	-0.0105
	(0.0116)	(0.0116)	(0.0109)
Cum. grants before venture		3.01e-11	3.31e-11
		(5.03e-11)	(4.96e-11)
Publications Stock$_{t-1}$		-5.65e-5	4.92e-7
		(0.000569)	(0.000545)
Publications Stock$_{t-2}$		2.31e-5	-7.13e-6
		(0.000591)	(0.000569)
Sector FE	Yes	Yes	Yes
Founding Year FE	No	No	Yes
Observations	541	535	535
R-sq.	0.00546	0.00888	0.0507

Notes: This table shows the determinants of founders' reliance on their previous academic work when creating their startup. The dependent variable is the percentage of scientific citations in a startup's (first granted) patents that comes from papers written by founders themselves. *Publications Stock*$_{t-k}$ is the aggregate number of publications up to k years preceding startup creation. We add sector fixed effects in each model and startup creation year fixed effects in model (3). Standard errors (in parentheses) are clustered at the founder level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: OLS Performance Outcomes, Firm Level

	(1) Success	(2) Acquired	(3) IPO
Percentage self-cites	-0.0547	-0.240**	0.142
	(0.224)	(0.110)	(0.196)
Number patents (log)	-0.0349	-0.0577	0.0164
	(0.0949)	(0.0704)	(0.0968)
Number scient. patents (log)	0.0815	0.0654	0.0212
	(0.0995)	(0.0755)	(0.101)
Team size (log)	0.0295	0.0597	-0.0174
	(0.0776)	(0.0584)	(0.0611)
At least one female founder	-0.0419	-0.0474	-0.00270
	(0.0511)	(0.0359)	(0.0393)
At least one top-tier university	0.0721	0.0229	0.0504
	(0.0495)	(0.0370)	(0.0390)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
State $\times$ Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	510	510	510
R-sq.	0.303	0.247	0.222

Notes: Each observation corresponds to a startup. In column (1), the outcome *Success* is an indicator variable equal to 1 if the startup is acquired or went public via an IPO. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we add sector, state, startup creation year and state $\times$ startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: 2SLS Performance Outcomes, Firm Level

Dep. Var.	(1) Percentage self-cites	(2) Acquired
Model	First Stage	IV
Log(1+Network size)	-0.0578***	
	(0.0173)	
Percentage self-cites		-2.927***
		(0.834)
Founder controls	Yes	Yes
State FE	Yes	Yes
Founding Year FE	Yes	Yes
State × Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	326	326
F-Stat		11.2

Notes: This table shows the result of our 2SLS estimation using the number of unique co-authors before entering entrepreneurship as an instrument. The outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. Each model includes as controls the log number of patents, the log number of patents relying on scientific literature, the log of team size calculated with the number of founders at inception, an indicator equal to 1 if there is at least one female in the founding team and an indicator equal to 1 if at least one founder graduated from a top-tier university. We also include state, founding year, state × founding year and sector fixed effects. We cluster standard errors at the state and startup level. Results are robust to clustering only at the startup level. Column (1) shows the first-stage regression. Column (2) shows the 2SLS result. We report the F-statistic of the first-stage in the last row of column (2).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: OLS Invention and Funding Outcomes, Firm Level

	(1) Nb patents (log)	(2) Nb patents (Poisson)	(3) 5y funds (log)	(4) 3y funds (log)
Percentage self-cites	1.808***	1.260**	-5.446	1.683
	(0.526)	(0.530)	(3.637)	(3.324)
Team size (log)	0.206	0.203	1.721	1.669
	(0.208)	(0.249)	(1.618)	(1.628)
At least one female founder	-0.164	-0.373*	0.321	0.555
	(0.160)	(0.196)	(1.227)	(1.256)
At least one top-tier university	0.0138	0.0994	0.565	-0.684
	(0.139)	(0.195)	(1.036)	(0.953)
Number patents (log)			-0.195	2.327
			(1.449)	(1.604)
Number scient. patents (log)			0.839	-1.952
			(1.538)	(1.713)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	510	510	510	510
R-sq.	0.416		0.398	0.424

Notes: Each observation corresponds to a startup. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In column (3), the outcome variable is the amount of funds a startup raised in the first 5 years after inception (expressed in natural logarithm). In column (4), the outcome variable is the amount of funds a startup raised in the first 3 years after inception (expressed in natural logarithm). In each model, we add sector, state, startup creation year and state × startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Summary statistics for the treatment and control groups of the difference-in-difference analysis

	Control group % self-cites= 0	Treatment group % self-cites> 0	p-value of difference
Panel A: Time invariant characteristics			
Number patents	2.6	6.2	<0.001
Number patents in RoS	2.6	6.0	<0.001
Female	0.08	0.09	0.57
Top-tiers university	0.39	0.42	0.45
Biotechnology sector	0.82	0.88	0.05
Panel B: At time of startup creation			
Academic experience	21.1	21.6	0.50
Publications (cumulative)	134	133	0.94
Top publications (cumulative)	58	58	0.93
Citations ('000s) (cumulative)	17.1	16.1	0.64
Citations per publication (average)	104	104	0.95
% papers last author (average)	38.5	40.0	0.36
% papers first author (average)	26.6	24.7	0.13
Number of unique individuals	331	230	

Cumulative: values of the variable are summed from career start year to startup creation year.

Average: values of the variable from career start year to startup creation year are averaged.

Table 7: Difference-in-difference analysis of academic output

	(1) Publications (log)	(2) Top (log)	(3) Cites (log)	(4) Cites per pub (log)	(5) % first (log)	(6) % last (log)
$\text{Treated}_i \times \text{Post}_t$	-0.0644	-0.00474	-0.0605	0.0105	0.00981	-0.00372
	(0.0551)	(0.0452)	(0.118)	(0.0704)	(0.00791)	(0.0114)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,532	15,063	19,532	19,532	17,574	17,574
R-sq.	0.553	0.530	0.516	0.437	0.345	0.320

Notes: In column (1), the outcome is the number of publications published by founders in a specific year. In column (2), the outcome is the number of top publications, defined as publications which received more citations than other articles in the same area of research. In column (3), the outcome is the cumulative number of cites received by articles published in a specific year. In column (4), the outcome is the average number of citations per paper received in a year. In column (5), the outcome is the share of papers where the founder appears as first author. In column (6), the outcome is the share of papers where the founder appears as the last author. All outcomes are logged. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Difference-in-difference analysis of number of co-authors by affiliation

	(1) All co-authors (log)	(2) Same affiliation (log)	(3) Different affiliation (log)
$\text{Treated}_i \times \text{Post}_t$	-0.0456	-0.0689	-0.0402
	(0.0598)	(0.0607)	(0.0686)
Year FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes
Observations	17,369	17,369	17,369
R-sq.	0.641	0.631	0.507

Notes: In column (1), the outcome is the number of unique co-authors each founder has in a given year. In column (2), the outcome is the number of unique co-authors from the same institution that each founder has in a given year. In column (3), the outcome is the number of unique co-authors from a different institution that each founder has in a given year. Columns (4), (5) and (6) use the log outcomes of columns (1), (2) and (3) respectively. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

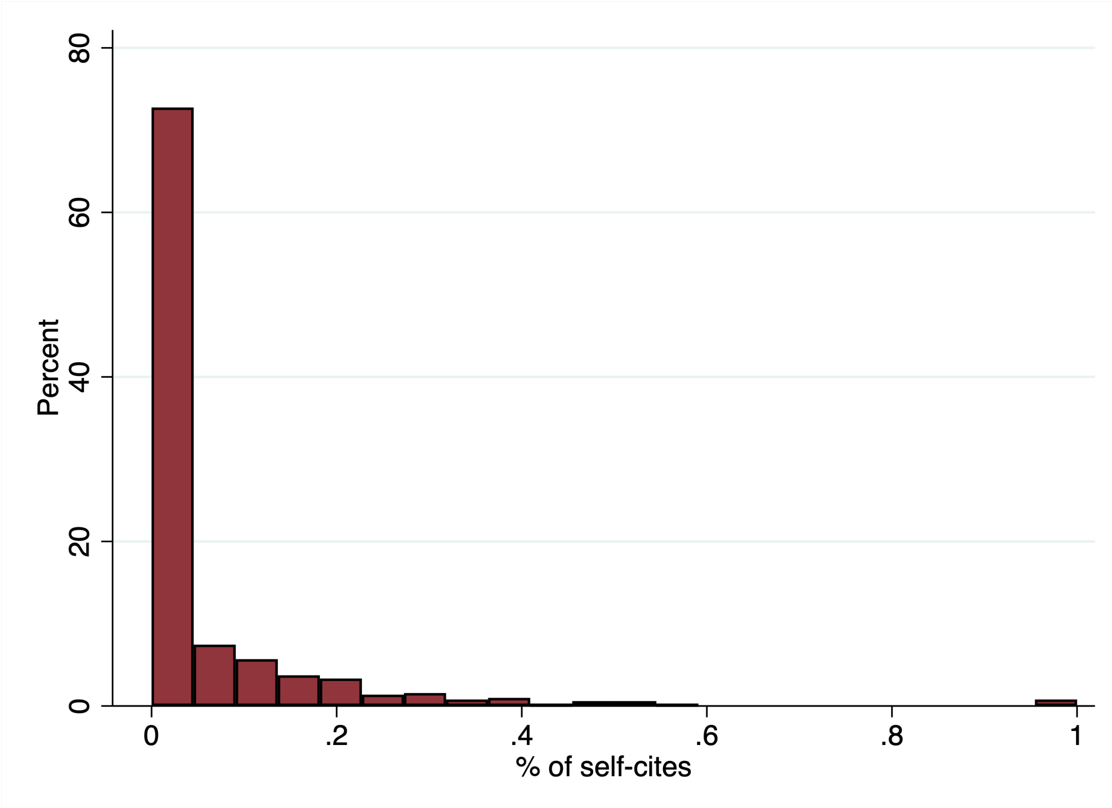
Table 9: Difference-in-difference analysis of academic output - Robustness check

	(1) Publications (log)	(2) Top (log)	(3) Cites (log)	(4) Cites per pub (log)	(5) % first (log)	(6) % last (log)
$\text{Treated}_i \times \text{Post}_t$	-0.0519	0.00464	-0.0482	0.00780	0.00760	-0.0000631
	(0.0563)	(0.0460)	(0.119)	(0.0700)	(0.00809)	(0.0116)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,206	87,870	114,206	114,206	102,315	102,315
R-sq.	0.560	0.539	0.513	0.432	0.352	0.326

Notes: In this specification, we follow Goodman-Bacon (2021) and exclude the post-period of earlier treated groups when using them as control groups. In column (1), the outcome is the number of publications published by founders in a specific year. In column (2), the outcome is the number of top publications, defined as publications which received more citations than other articles in the same area of research. In column (3), the outcome is the cumulative number of cites received by articles published in a specific year. In column (4), the outcome is the average number of citations per paper received in a year. In column (5), the outcome is the share of papers where the founder appears as first author. In column (6), the outcome is the share of papers where the founder appears as the last author. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

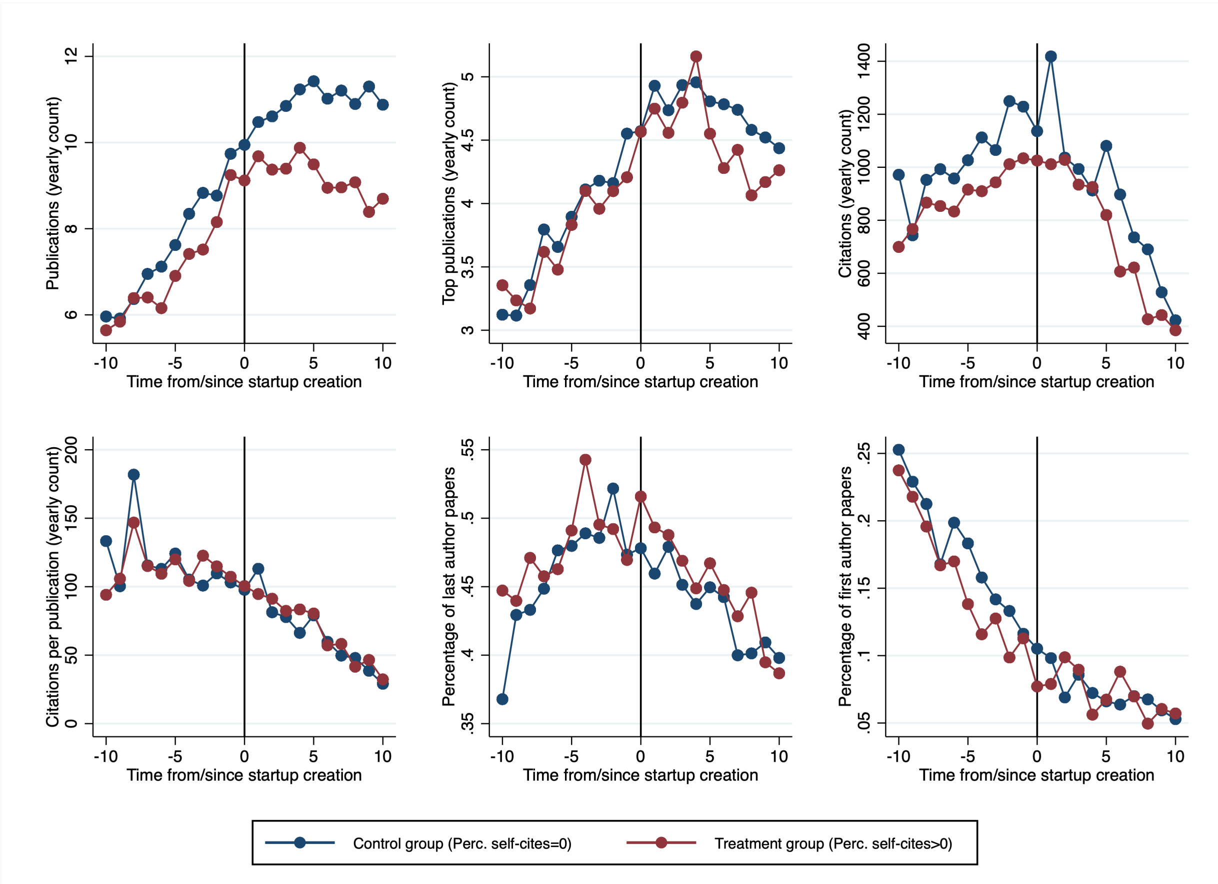
Online Appendix for: Bringing Science to Market: Knowledge Foundations and Performance

Figure A1: Histogram of the treatment variable



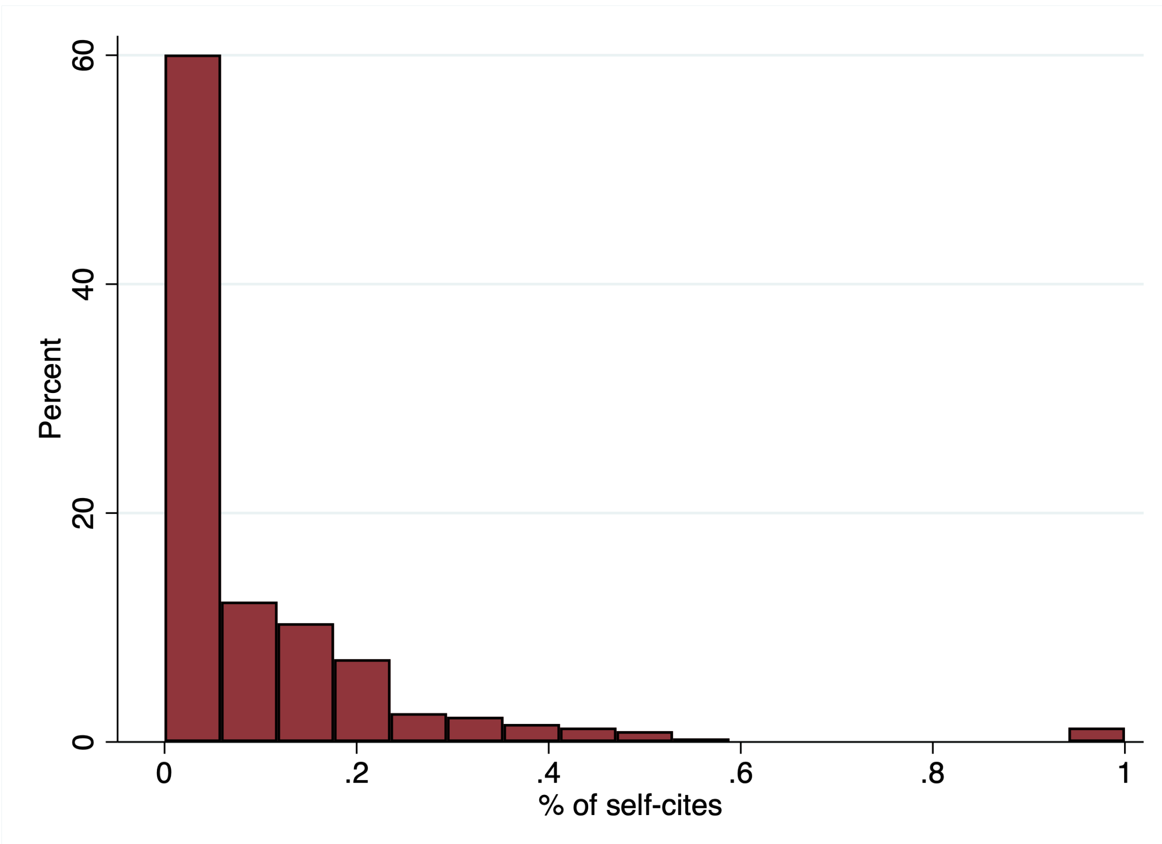
Notes: This figure shows the distribution of the *Percentage self-cites* variable. To calculate this variable, we calculate the percentage of citations that the first granted patents of a startup makes to its founders' academic papers. We then average this measure at the startup level.

Figure A2: Raw plot of the outcomes for the treatment and control groups



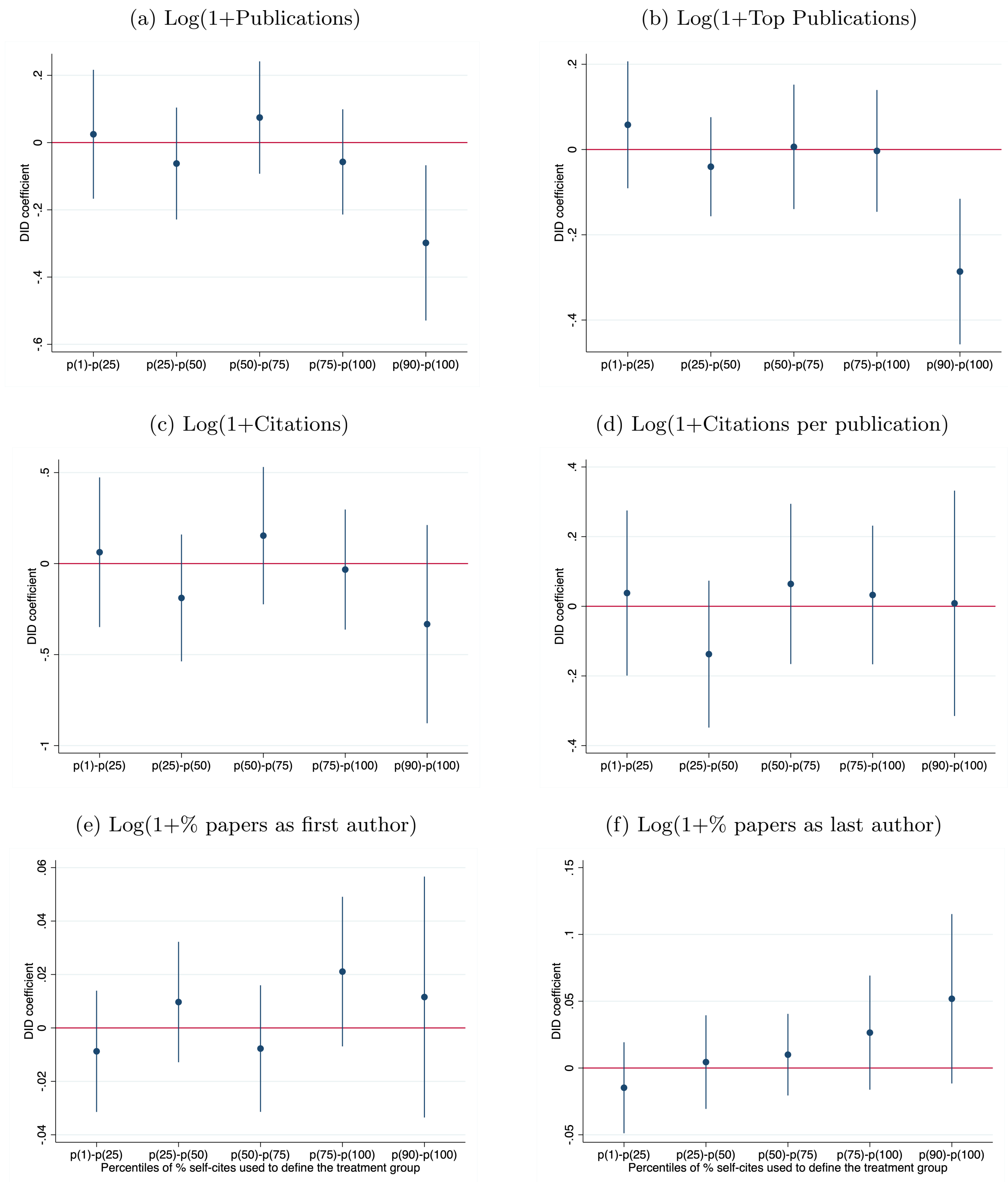
Notes: This figure shows the mean values before startup creation (negative values of the x-axis) and after startup creation (positive values of the x-axis) of several outcome variables for the treatment (red line) and control (blue line) groups. The treatment group is composed of individuals whose startup has a patent that cites at least one of their academic papers (i.e., *Percentage self-cites* > 0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites* $= 0$). Outcomes are (from left to right, top to bottom): the log yearly number of publications, the log yearly number of top publications, the log yearly number of citations, the log yearly number of citations per publication, the log yearly number of publications where the founder appears as first author and the log yearly number of publications where the founder appears as last author.

Figure A3: Histogram of the treatment variable for startups that rely on Science



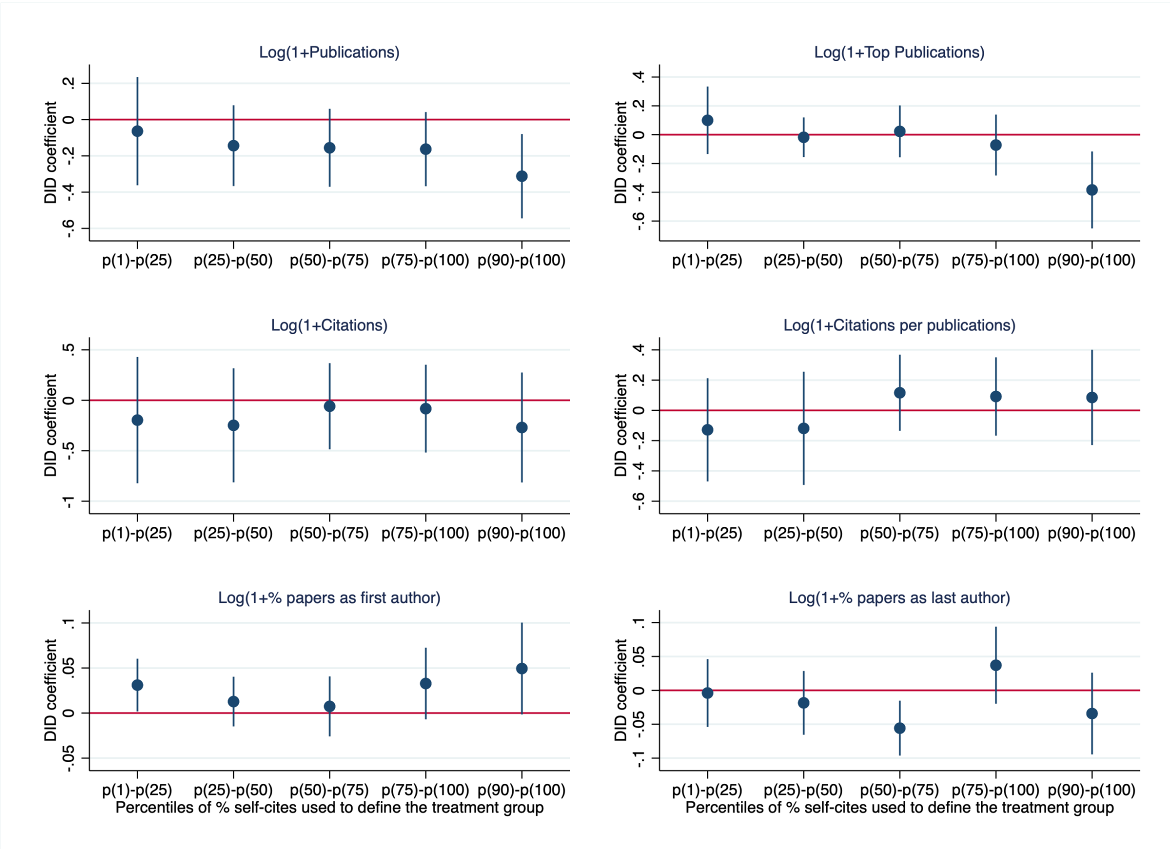
Notes: This figure shows the distribution of the *Percentage self-cites* variable when restricted to startups that rely on Science. To calculate this variable, we calculate the percentage of citations that the first granted patents of a startup makes to its founders' academic papers. We then average this measure at the startup level.

Figure A4: Heterogeneity Analysis, Academic Output, startups relying on Science



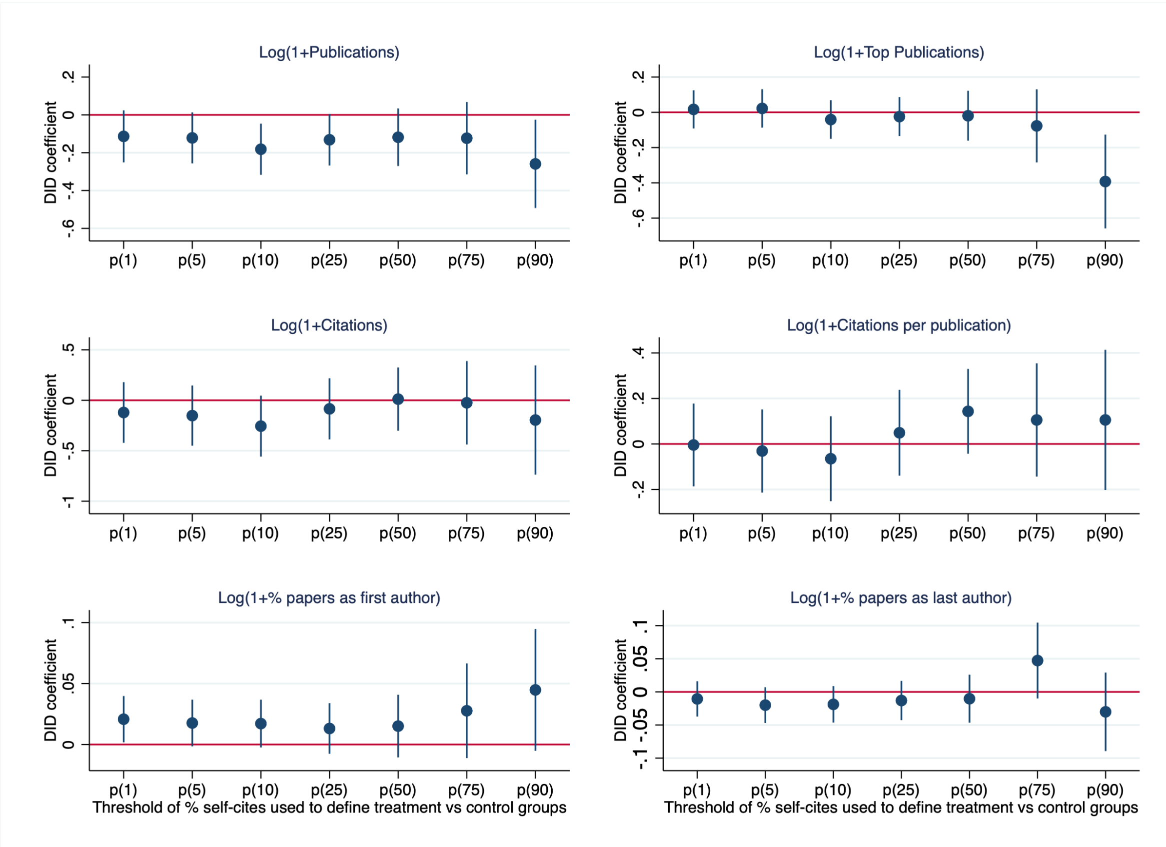
Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups, when restricting the sample to individuals whose startup relies on science. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log yearly number of citations. In Panel (d), the outcome is the log yearly number of citations per publication. In Panel (e), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (f), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure A5: Heterogeneity Analysis of Academic Output, using the same control group and varying the treatment group - Single professor in the founding team



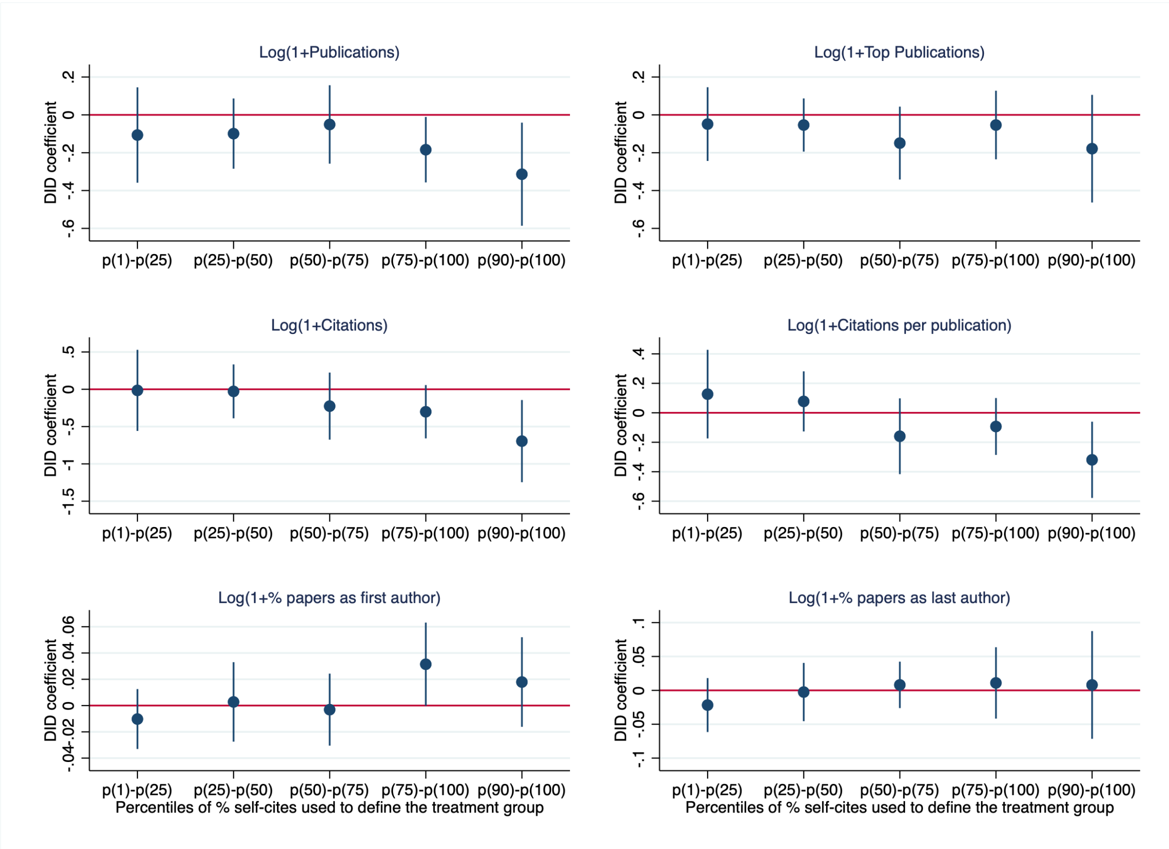
Notes: This figure shows the difference-in-difference coefficients varying the definition of the treatment group for startups with only one professor in the founding team. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. For instance, p(1)-p(25) means that all individuals whose *Percentage self-cites* variable falls between the 1^{st} and the 25^{th} percentiles are included in the treatment group. The control group does not vary and includes all individuals with *Percentage self-cites*$= 0$. In the first row, outcomes are the log yearly number of publications (left) and the log yearly number of top publications (right). In the second row, outcomes are the log yearly number of citations (left) and the log yearly number of citations per publication (right). In the third row, outcomes are the log yearly number of publications where the founder appears as first author (left) and the log yearly number of publications where the founder appears as last author (right). Each bar denotes the 90% confidence interval.

Figure A6: Heterogeneity Analysis of Academic Output, varying the threshold to define treatment and control groups - Single professor in the founding team



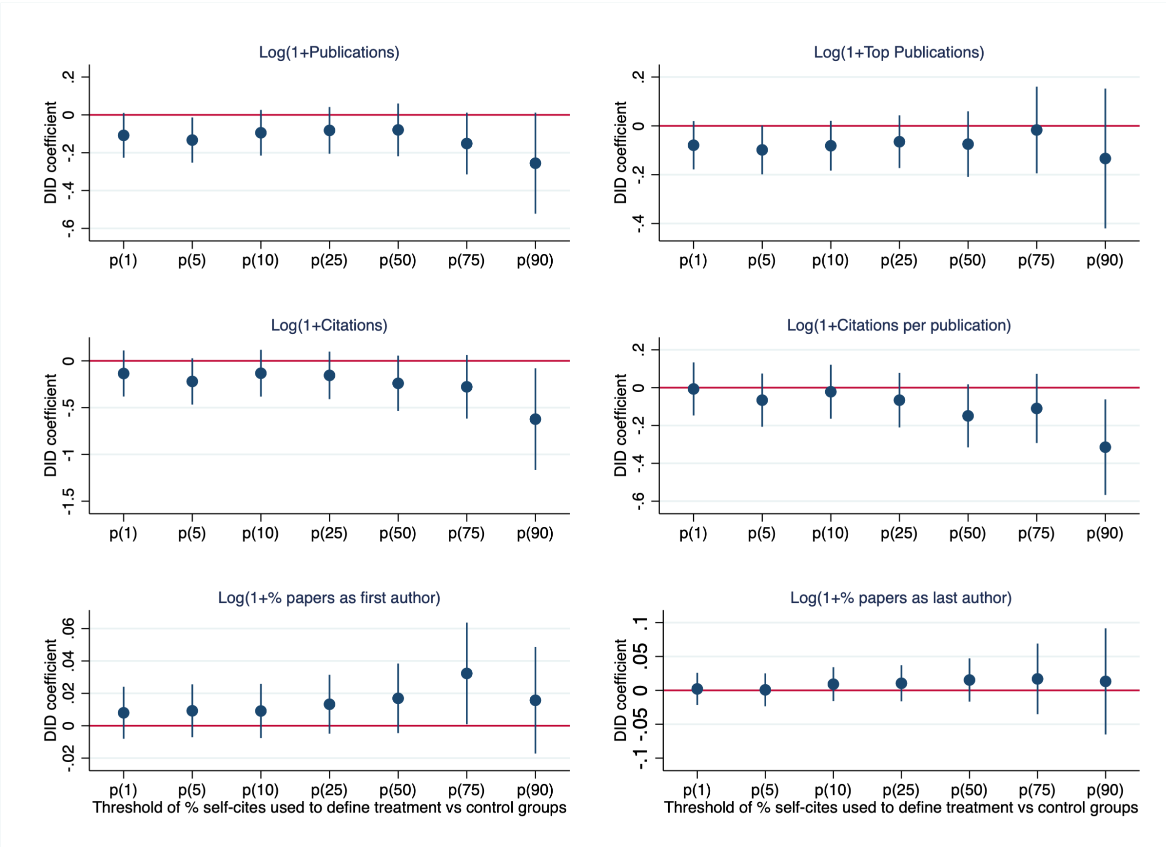
Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups for startups with only one professor in the founding team. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In the first row, outcomes are the log yearly number of publications (left) and the log yearly number of top publications (right). In the second row, outcomes are the log yearly number of citations (left) and the log yearly number of citations per publication (right). In the third row, outcomes are the log yearly number of publications where the founder appears as first author (left) and the log yearly number of publications where the founder appears as last author (right). Each bar denotes the 90% confidence interval.

Figure A7: Heterogeneity Analysis of Academic Output, using the same control group and varying the treatment group - Several professors in the founding team



Notes: This figure shows the difference-in-difference coefficients varying the definition of the treatment group for startups with several professors in the founding team. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. For instance, p(1)-p(25) means that all individuals whose *Percentage self-cites* variable falls between the 1^{st} and the 25^{th} percentiles are included in the treatment group. The control group does not vary and includes all individuals with *Percentage self-cites*$= 0$. In the first row, outcomes are the log yearly number of publications (left) and the log yearly number of top publications (right). In the second row, outcomes are the log yearly number of citations (left) and the log yearly number of citations per publication (right). In the third row, outcomes are the log yearly number of publications where the founder appears as first author (left) and the log yearly number of publications where the founder appears as last author (right). Each bar denotes the 90% confidence interval.

Figure A8: Heterogeneity Analysis of Academic Output, varying the threshold to define treatment and control groups - Several professors in the founding team

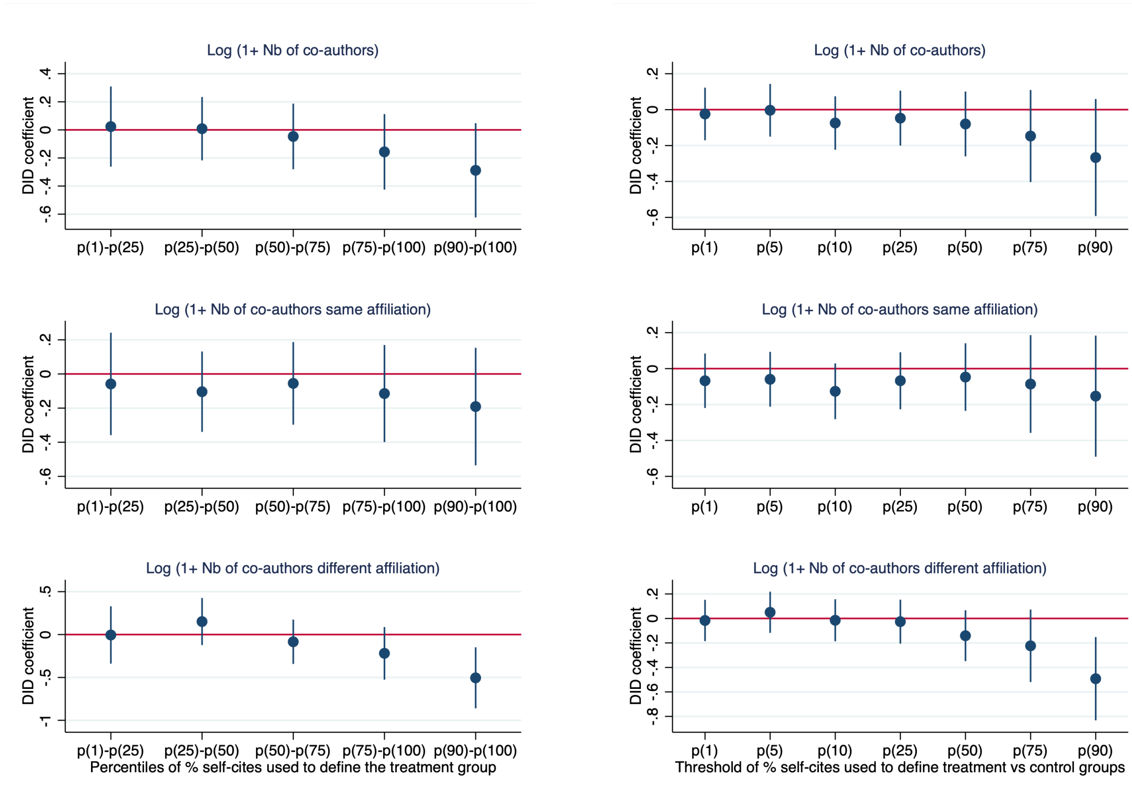


Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups for startups with several professors in the founding team. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In the first row, outcomes are the log yearly number of publications (left) and the log yearly number of top publications (right). In the second row, outcomes are the log yearly number of citations (left) and the log yearly number of citations per publication (right). In the third row, outcomes are the log yearly number of publications where the founder appears as first author (left) and the log yearly number of publications where the founder appears as last author (right). Each bar denotes the 90% confidence interval.

Figure A9: Heterogeneity Analysis of the Number of Co-Authors - Single professor in the founding team

(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups

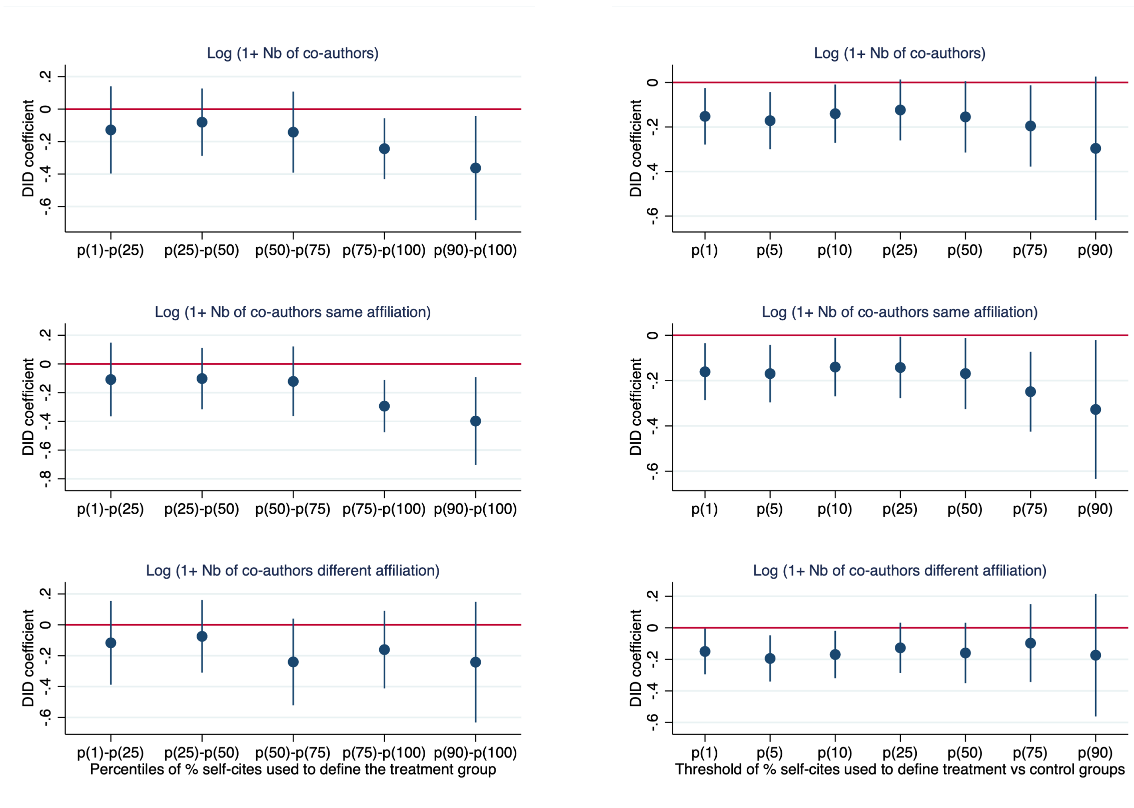


Notes: This figure shows the difference-in-difference coefficients for startups with only one professor in the founding team. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique co-authors founders published with in a given year. In the second row, the outcome is the log number of unique co-authors from the same institution founders published with in a given year. In the third row, the outcome is the log number of unique co-authors from a different institution founders published with in a given year. Each bar denotes the 90% confidence interval.

Figure A10: Heterogeneity Analysis of the Number of Co-Authors - Several professors in the founding team

(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups

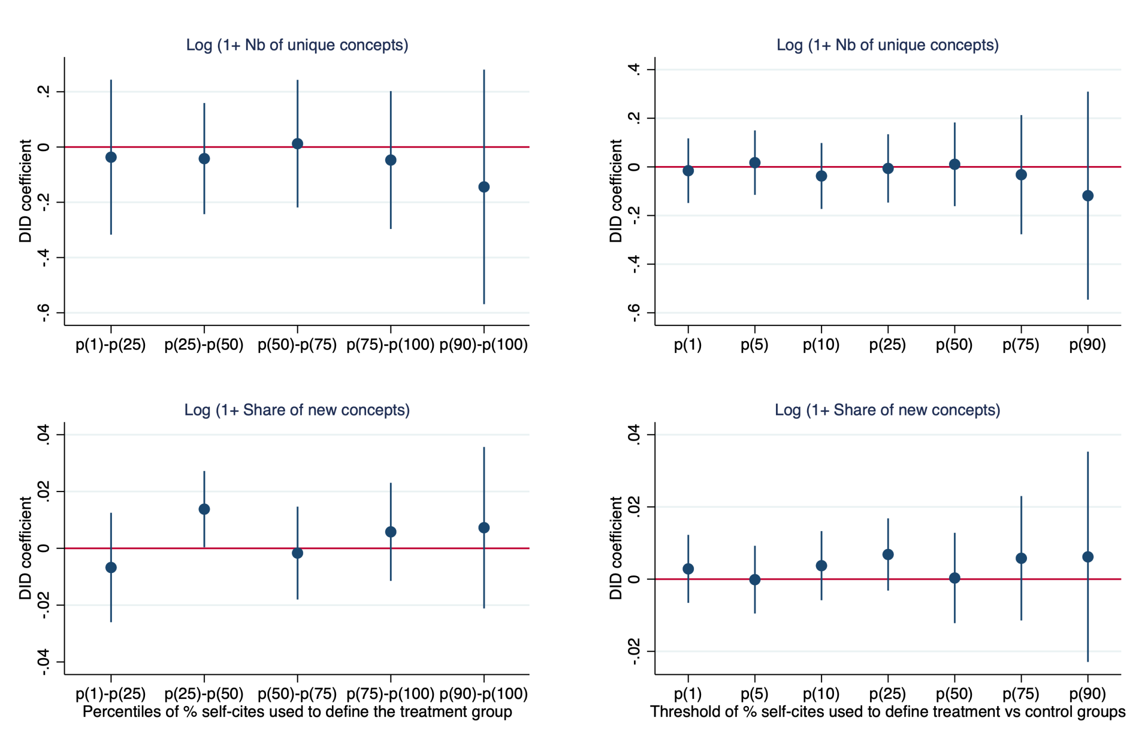


Notes: This figure shows the difference-in-difference coefficients for startups with several professors in the founding team. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique co-authors founders published with in a given year. In the second row, the outcome is the log number of unique co-authors from the same institution founders published with in a given year. In the third row, the outcome is the log number of unique co-authors from a different institution founders published with in a given year. Each bar denotes the 90% confidence interval.

Figure A11: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts - Single professor in the founding team

(a) Using the same control group and varying the treatment group

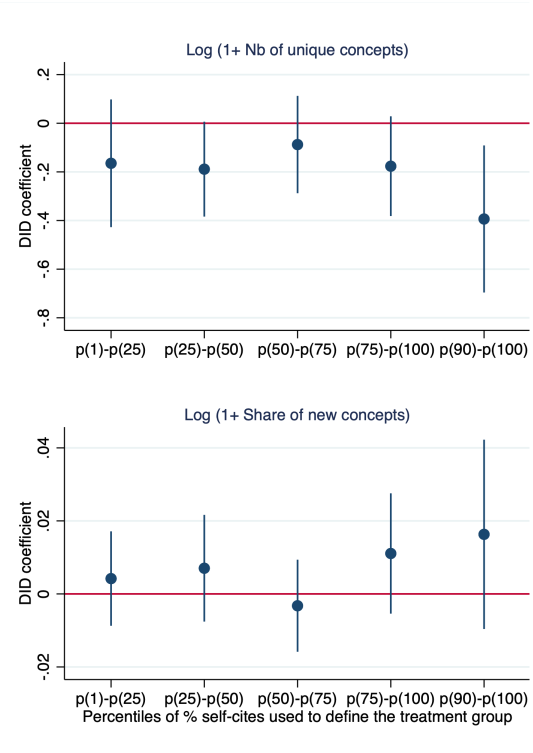
(b) Varying the threshold to define treatment and control groups



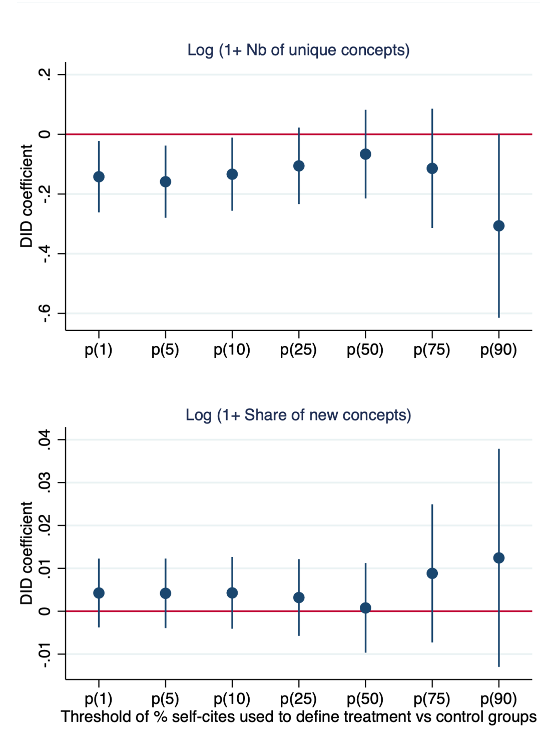
Notes: This figure shows the difference-in-difference coefficients for startups with a single professor in the founding team. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique concepts founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new concepts founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure A12: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts - Several professors in the founding team

(a) Using the same control group and varying the treatment group



(b) Varying the threshold to define treatment and control groups

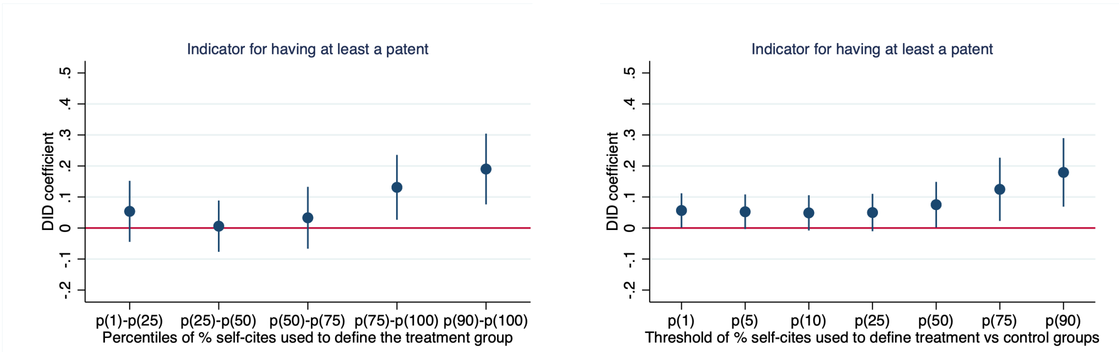


Notes: This figure shows the difference-in-difference coefficients for startups with several professors in the founding team. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique concepts founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new concepts founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure A13: Heterogeneity Analysis, Patenting Activity - Single professor in the founding team

(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups

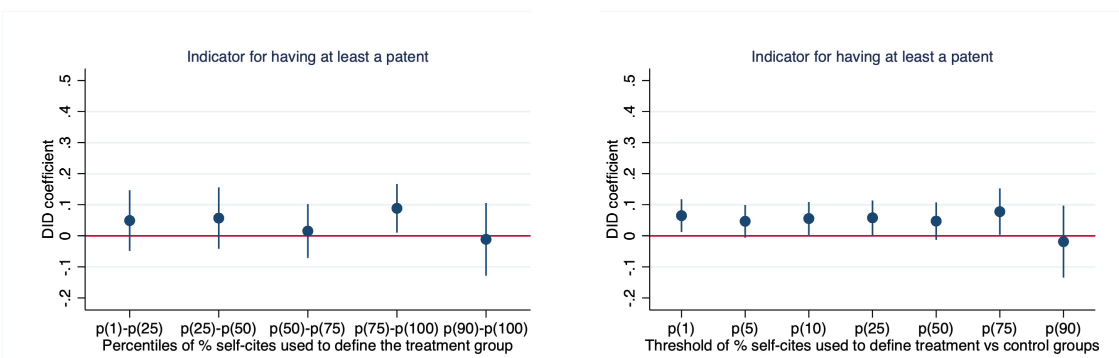


Notes: This figure shows the difference-in-difference coefficients using different definitions of the treatment group (left figure) and different thresholds for defining the treatment and control groups (right figure). In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In both panels, the outcome is an indicator variable equal to 1 if the founder filed a patent that year and 0 otherwise. Each bar denotes the 90% confidence interval.

Figure A14: Heterogeneity Analysis, Patenting Activity - Several professors in the founding team

(a) Using the same control group and varying the treatment group

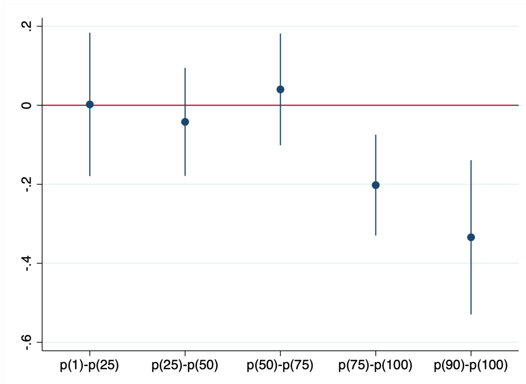
(b) Varying the threshold to define treatment and control groups



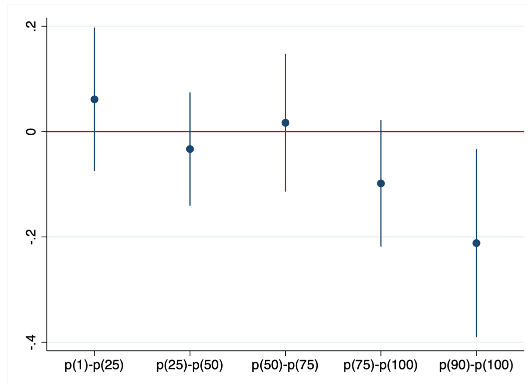
Notes: This figure shows the difference-in-difference coefficients using different definitions of the treatment group (left figure) and different thresholds for defining the treatment and control groups (right figure). In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In both panels, the outcome is an indicator variable equal to 1 if the founder filed a patent that year and 0 otherwise. Each bar denotes the 90% confidence interval.

Figure A15: Heterogeneity Analysis of Academic Output, varying the definition of the treatment group - All patents

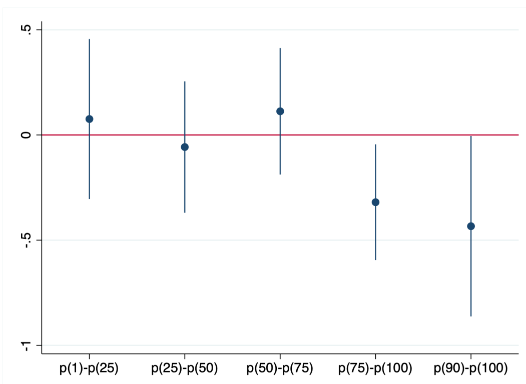
(a) Log(1+Publications)



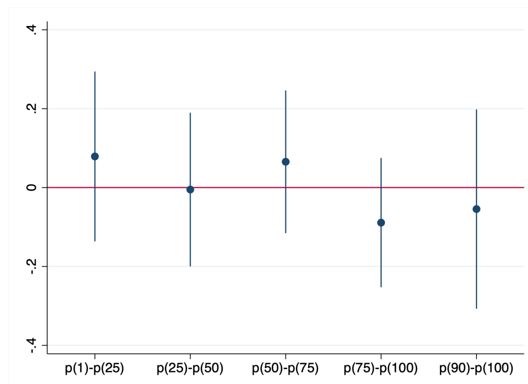
(b) Log(1+Top Publications)



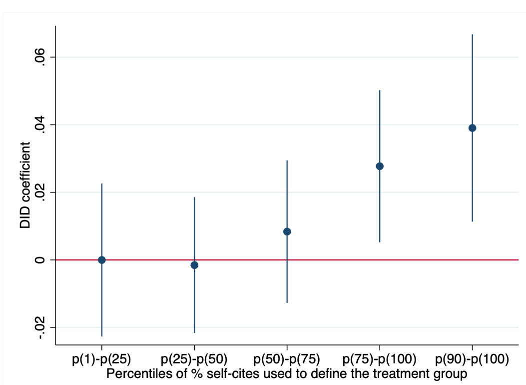
(c) Log(1+Citations)



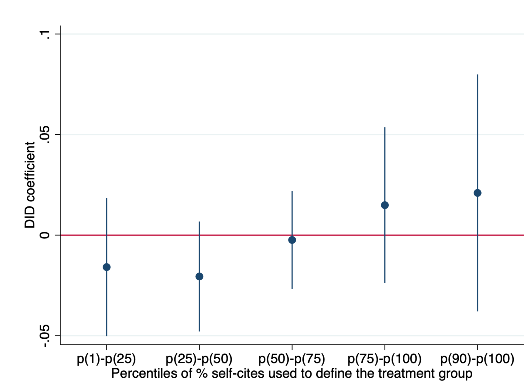
(d) Log(1+Citations per publication)



(e) Log(1+% papers as first author)



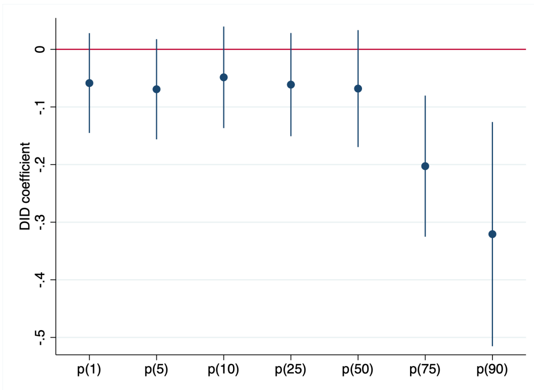
(f) Log(1+% papers as last author)



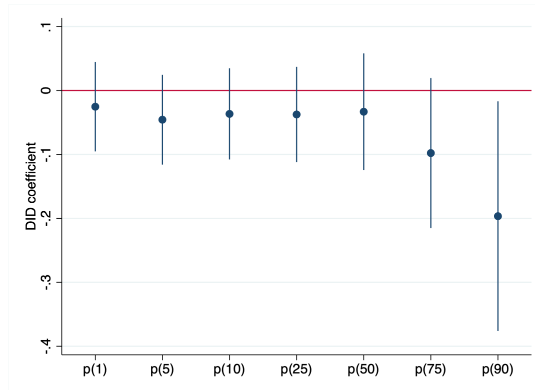
Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients varying the definition of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. For instance, p(1)-p(25) means that all individuals whose *Percentage self-cites* variable falls between the 1^{st} and the 25^{th} percentiles are included in the treatment group. The control group does not vary and includes all individuals with Percentage self-cites= 0. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log yearly number of citations. In Panel (d), the outcome is the log yearly number of citations per publication. In Panel (e), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (f), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure A16: Heterogeneity Analysis of Academic Output, varying the threshold to define treatment and control groups - All patents

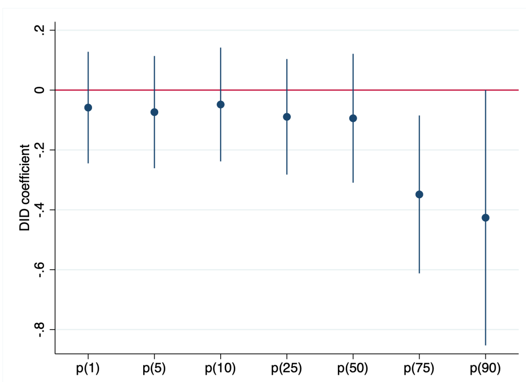
(a) Log(1+Publications)



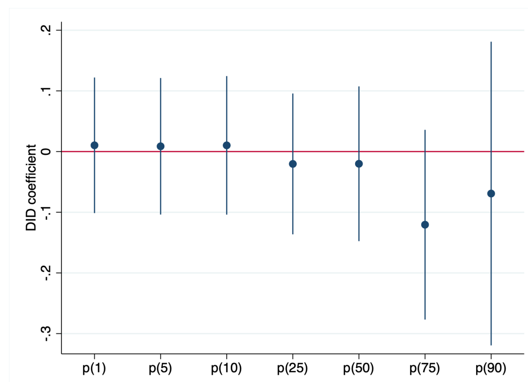
(b) Log(1+Top Publications)



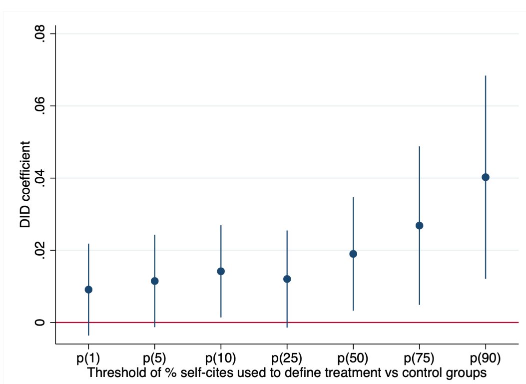
(c) Log(1+Citations)



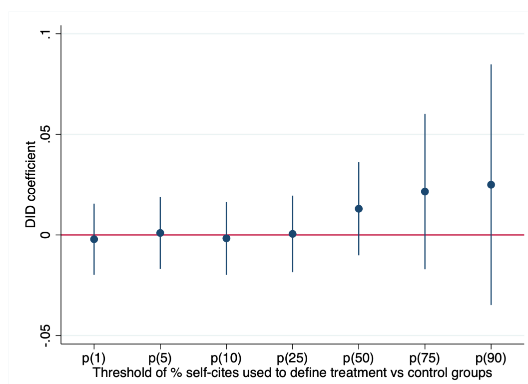
(d) Log(1+Citations per publication)



(e) Log(1+% papers as first author)



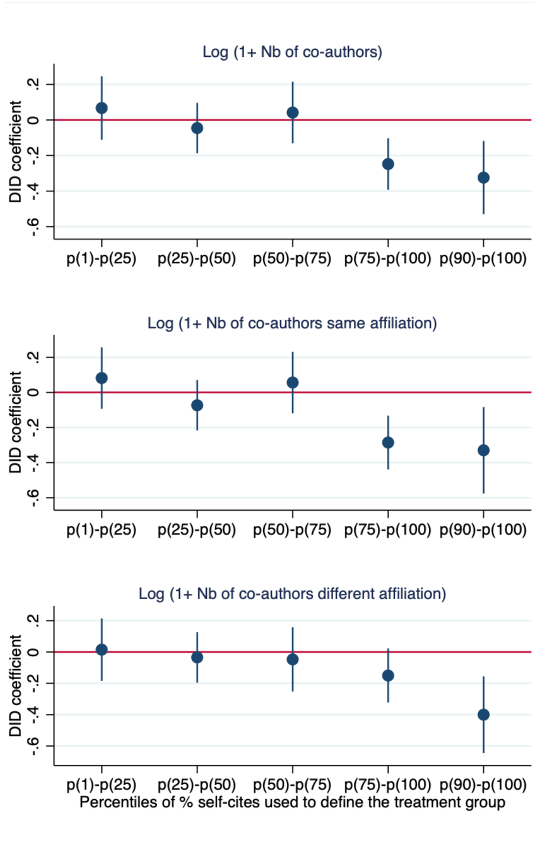
(f) Log(1+% papers as last author)



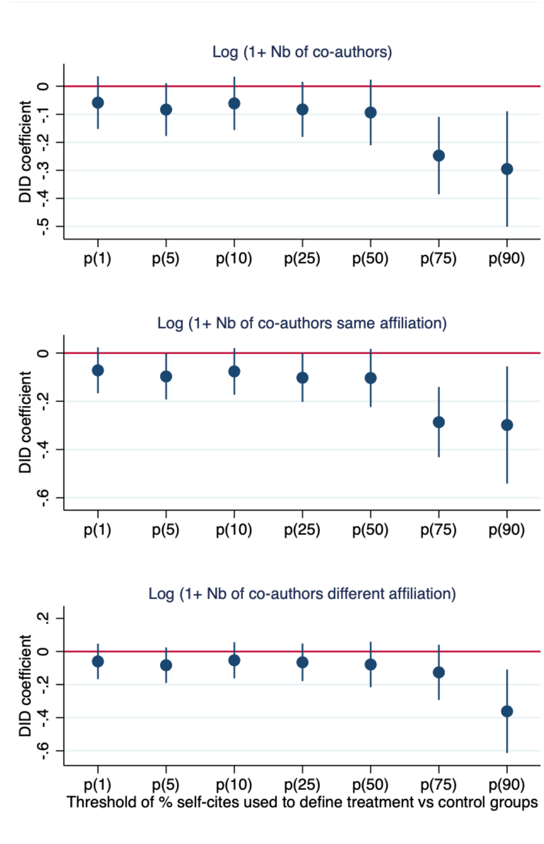
Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log yearly number of citations. In Panel (d), the outcome is the log yearly number of citations per publication. In Panel (e), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (f), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure A17: Heterogeneity Analysis of the Number of Co-Authors - All patents

(a) Using the same control group and varying the treatment group



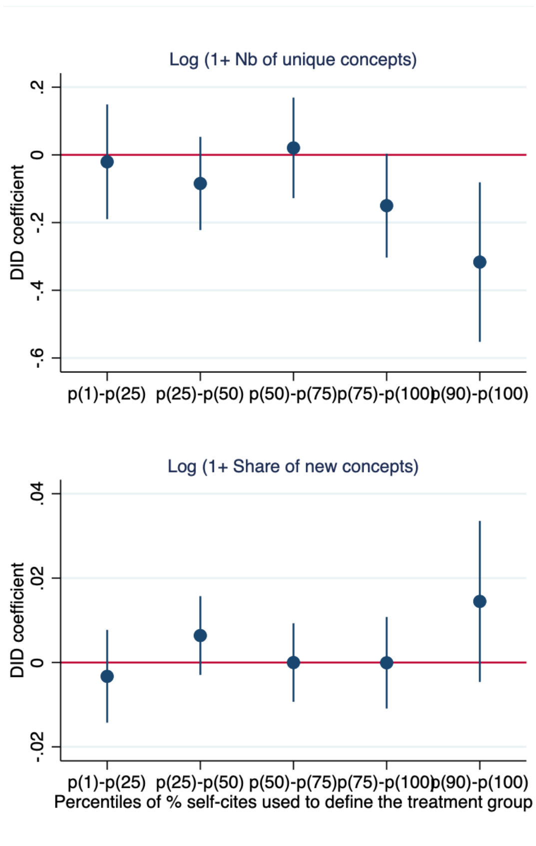
(b) Varying the threshold to define treatment and control groups



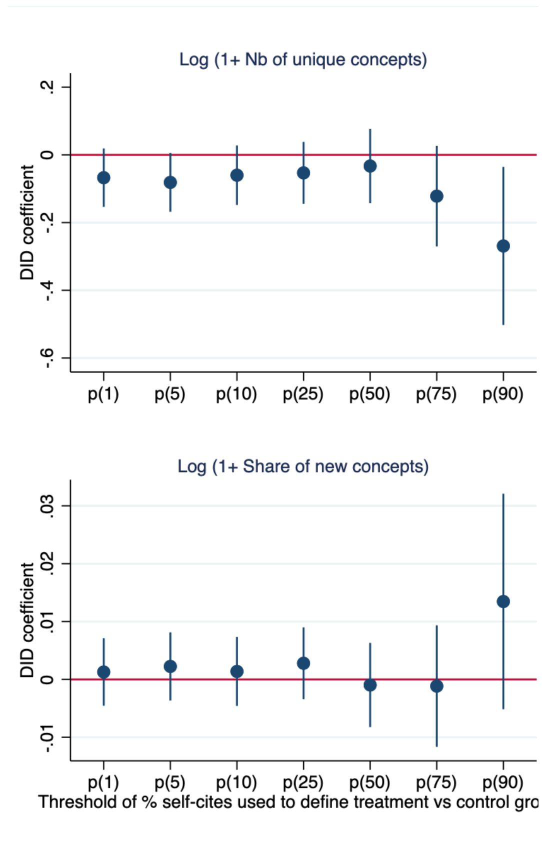
Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique co-authors founders published with in a given year. In the second row, the outcome is the log number of unique co-authors from the same institution founders published with in a given year. In the third row, the outcome is the log number of unique co-authors from a different institution founders published with in a given year. Each bar denotes the 90% confidence interval.

Figure A18: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts - All patents

(a) Using the same control group and varying the treatment group



(b) Varying the threshold to define treatment and control groups

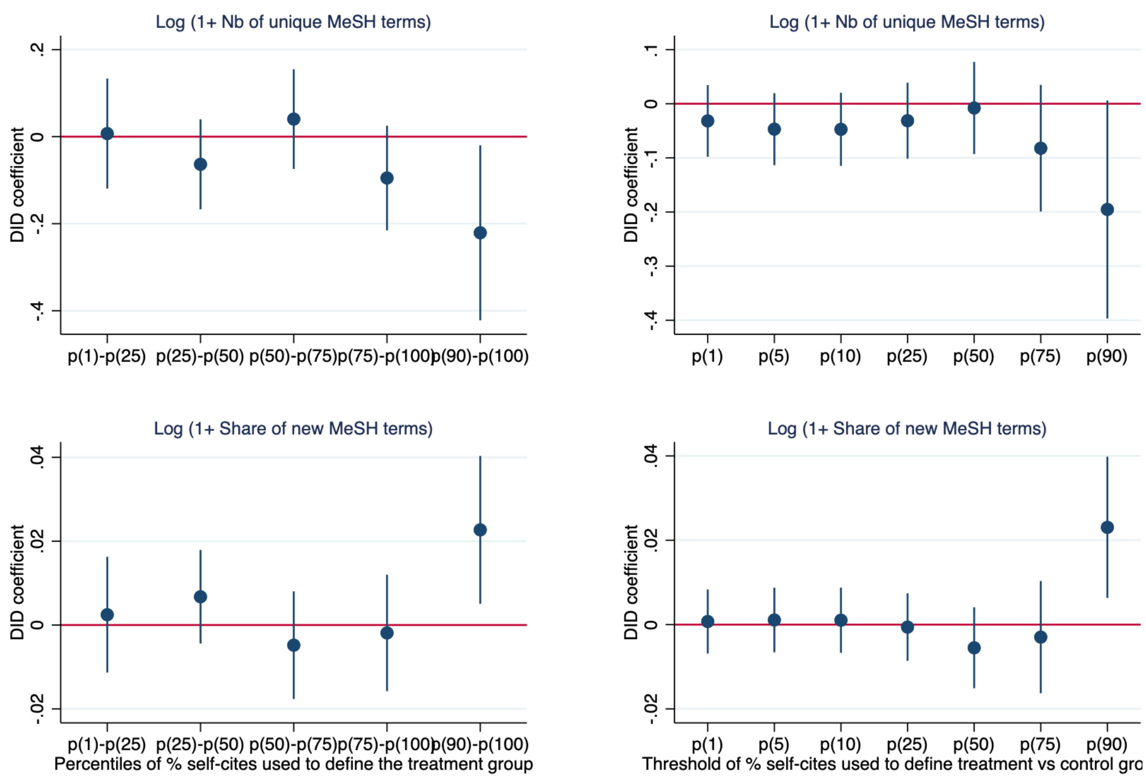


Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique concepts founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new concepts founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure A19: Heterogeneity Analysis of Research Focus, using MeSH terms - All patents

(a) Using the same control group and varying the treatment group

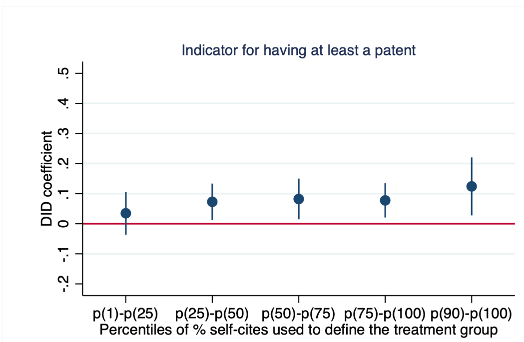
(b) Varying the threshold to define treatment and control groups



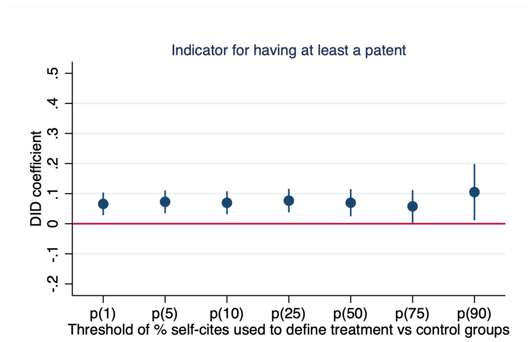
Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique MeSH terms founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new MeSH terms founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure A20: Heterogeneity Analysis, Patenting Activity - All patents

(a) Using the same control group and varying the treatment group



(b) Varying the threshold to define treatment and control groups



Notes: We use all patents pertaining to a startup to calculate the *Percentage self-cites* variable. This figure shows the difference-in-difference coefficients using different definitions of the treatment group (left figure) and different thresholds for defining the treatment and control groups (right figure). In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In both panels, the outcome is an indicator variable equal to 1 if the founder filed a patent that year and 0 otherwise. Each bar denotes the 90% confidence interval.

Table A1: Forward Citations Patent Level

	(1)	(2)	(3)
	Forward citations (log)		
Percentage self-cites	-0.478	-0.495	-0.400
	(0.428)	(0.432)	(0.486)
At least one scient. cite		0.0601	0.311
		(0.255)	(0.326)
Application Year Controls	Yes	Yes	Yes
Sector Control	Yes	Yes	Yes
Patent Class Control	No	No	Yes
Observations	454	454	446
R-sq.	0.214	0.214	0.473

Notes: The outcome variable is the log number of forward citations received by patents. In each model, we add patent application year and sector fixed effects. The last column adds IPC class fixed effects. We cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: OLS Stage of Development Outcomes, Firm Level

	Time from paper to patent		Time from paper to startup	
	(1)	(2)	(3)	(4)
Percentage self-cites	-16.52***	-13.77***	-13.80***	-9.925*
	(4.626)	(4.844)	(5.060)	(5.347)
Team size (log)		1.884		1.462
		(2.655)		(2.736)
At least one female founder		-3.828		-2.449
		(2.392)		(2.443)
At least one top-tier university		-1.061		-0.861
		(1.967)		(2.012)
State FE	No	Yes	No	Yes
Founding Year FE	No	Yes	No	Yes
Sector FE	No	Yes	No	Yes
Observations	300	300	300	300
R-sq.	0.0323	0.208	0.0225	0.199

Notes: This table shows the correlation between *Percentage of self-cites* and proxies of the time needed to bring a startup technology from Academia to the private sector. Columns (1) and (2) use the time between a startup's first patent application year and its oldest scientific self-cite. Columns (3) and (4) use the time between a startup's creation year and the oldest scientific self-cite of its first patent. In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: OLS Performance Outcomes, start-ups relying on Science

	(1) Success	(2) Acquired	(3) IPO
Percentage self-cites	-0.0963	-0.213***	0.108
	(0.159)	(0.0747)	(0.141)
Number patents (log)	-0.00469	-0.0479	0.0383
	(0.101)	(0.0531)	(0.0952)
Number scient. patents (log)	0.0708	0.0521	0.0248
	(0.0998)	(0.0575)	(0.0922)
Team size (log)	0.0463	0.0725	-0.00981
	(0.0708)	(0.0524)	(0.0551)
At least one female founder	-0.0763	-0.0849***	-0.000681
	(0.0505)	(0.0304)	(0.0433)
At least one top-tier university	0.0423	0.00534	0.0362
	(0.0458)	(0.0321)	(0.0388)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	318	318	318
R-sq.	0.221	0.156	0.146

Notes: Each observation corresponds to a startup. We restrict the sample to startups that rely on science. In column (1), the outcome *Success* is an indicator variable equal to 1 if the startup is acquired or went public via an IPO. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Innovation and Funding Outcomes, start-ups relying on Science

	(1) Nb patents (log)	(2) Nb patents (Poisson)	(3) 5y funds (log)	(4) 3y funds (log)
Percentage self-cites	-0.276	-0.141	-2.779	-0.199
	(0.280)	(0.434)	(2.758)	(2.632)
Team size (log)	0.0445	0.0806	2.008	3.309**
	(0.153)	(0.232)	(1.615)	(1.497)
At least one female founder	-0.206*	-0.314*	-0.529	0.0587
	(0.124)	(0.165)	(1.231)	(1.254)
At least one top-tier university	0.0416	0.0719	0.228	-1.056
	(0.102)	(0.156)	(1.062)	(0.979)
Number patents (log)			0.404	0.714
			(1.979)	(1.871)
Number scient. patents (log)			0.758	-0.892
			(1.972)	(1.904)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	318	318	318	318
R-sq.	0.172		0.197	0.202

Notes: Each observation corresponds to a startup. We restrict the sample to startups that rely on science. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In column (3), the outcome variable is the amount of funds a startup raised in the first 5 years after inception (expressed in natural logarithm). In column (4), the outcome variable is the amount of funds a startup raised in the first 3 years after inception (expressed in natural logarithm). In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: OLS Performance Outcomes Firm Level, for startups with a single vs several professors in the founding team

	One professor			Multiple professors		
	(1) Success	(2) Acquired	(3) IPO	(4) Success	(5) Acquired	(6) IPO
Percentage self-cites	-0.0206	-0.211**	0.172	-0.360	-0.305**	-0.0806
	(0.205)	(0.0941)	(0.185)	(0.222)	(0.153)	(0.126)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	353	353	353	157	157	157
R-sq.	0.187	0.148	0.126	0.255	0.177	0.237

Notes: Each observation corresponds to a startup. Columns (1), (2) and (3) are restricted to startups with a single professor in the founding team. Columns (4), (5) and (6) are restricted to startups with multiple professors in the founding team. In columns (1) and (4), the outcome *Success* is an indicator variable equal to 1 if the startup is acquired or went public via an IPO. In columns (2) and (5), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In columns (3) and (6), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: OLS Innovation and Funding Outcomes Firm Level, for startups with a single vs several professors in the founding team

	(1) Nb patents (log)	(2) Nb patents (Poisson)	(3) 5y funds (log)	(4) 3y funds (log)
Panel A: One professor				
Percentage self-cites	2.106***	1.002***	-2.802	3.526
	(0.448)	(0.383)	(3.536)	(3.316)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
Observations	353	353	353	353
R-sq.	0.237		0.200	0.213
Panel B: Multiple professors				
Percentage self-cites	0.521	1.186	-9.124**	-7.603**
	(0.793)	(0.969)	(3.948)	(3.680)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
Observations	157	157	157	157
R-sq.	0.311		0.347	0.253

Notes: Each observation corresponds to a startup. Panel A restricts the sample to startups with a single professor in the founding team. Panel B restricts the sample to startups with multiple professors in the founding team. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In column (3), the outcome variable is the amount of funds a startup raised in the first 5 years after inception (expressed in natural logarithm). In column (4), the outcome variable is the amount of funds a startup raised in the first 3 years after inception (expressed in natural logarithm). In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: OLS Performance Outcomes, Firm Level - All patents

	(1) Success	(2) Acquired	(3) IPO
Percentage self-cites	-0.202**	-0.171**	-0.0567
	(0.128)	(0.0786)	(0.114)
Number patents (log)	-0.0188	-0.0409	0.0177
	(0.0895)	(0.0516)	(0.0870)
Number scient. patents (log)	0.0708	0.0527	0.0224
	(0.0932)	(0.0559)	(0.0896)
Team size (log)	0.00875	0.0592**	-0.0362
	(0.0492)	(0.0358)	(0.0392)
At least one female founder	-0.0422	-0.0460	-0.00290
	(0.0365)	(0.0254)	(0.0283)
At least one top-tier university	0.0603***	0.00340	0.0490
	(0.0324)	(0.0239)	(0.0260)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	510	510	510
R-sq.	0.172	0.116	0.102

Notes: Each observation corresponds to a startup. In column (1), the outcome *Success* is an indicator variable equal to 1 if the startup is acquired or went public via an IPO. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: OLS Invention and Funding Outcomes, Firm Level - All patents

	(1) Nb patents (log)	(2) Nb patents (Poisson)	(3) 5y funds (log)	(4) 3y funds (log)
Percentage self-cites	2.228***	1.347**	-2.754	-0.854
	(0.585)	(0.480)	(3.031)	(2.843)
Team size (log)	0.0734	0.0942	1.675	1.650
	(0.150)	(0.245)	(1.227)	(1.174)
At least one female founder	-0.0845	-0.259*	0.0474	1.196
	(0.116)	(0.181)	(0.910)	(0.951)
At least one top-tier university	0.0726	0.143	0.782	-0.474
	(0.104)	(0.203)	(0.795)	(0.745)
Number patents (log)			-0.343	1.256
			(1.739)	(1.685)
Number scient. patents (log)			0.902	-0.954
			(1.806)	(1.763)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	510	510	510	510
R-sq.	0.195		0.174	0.163

Notes: Each observation corresponds to a startup. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In column (3), the outcome variable is the amount of funds a startup raised in the first 5 years after inception (expressed in natural logarithm). In column (4), the outcome variable is the amount of funds a startup raised in the first 3 years after inception (expressed in natural logarithm). In each model, we add state, sector and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: 2SLS Performance Outcomes for Success and IPO, Firm Level

	(1)	(2)	(3)
Dep. Var.	Percentage self-cites	Success	IPO
Model	First Stage	IV	IV
Log(1+Network size)	-0.0578***		
	(0.0173)		
Percentage self-cites		-5.502***	-2.612*
		(1.970)	(1.492)
Founder controls	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
State × Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	326	326	326
F-Stat		11.2	11.2

Notes: This table shows the result of our 2SLS estimation using the number of unique co-authors before entering entrepreneurship as an instrument. Each model includes as controls the log number of patents, the log number of patents relying on scientific literature, the log of team size calculated with the number of founders at inception, an indicator equal to 1 if there is at least one female in the founding team and an indicator equal to 1 if at least one founder graduated from a top-tier university. We also include state, founding year, state × founding year and sector fixed effects. We cluster standard errors at the state and startup level. Results are robust to clustering only at the startup level. Column (1) shows the first-stage regression. Column (2) shows the 2SLS results for *Success*, an indicator variable equal to 1 if the startup is acquired or went public via an IPO. Column (3) shows the 2SLS results for *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. We report the F-statistic of the first-stage in the last row of columns (2) and (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Academic and patenting output using Dimensions AI

In order to identify publications and patents associated with each professor in our sample, we use the Dimensions AI database, which is very similar to Scopus and Web of Science (Singh et al., 2021). Interestingly for us, Dimensions provides a disambiguation of researchers where they take into account a variety of variables such as existing person IDs, name variants, affiliation data, research topics, journals, co-authors, and active years in order to aggregate publications, patents and grants into author profile[20]. Our algorithm relies on identifying the Dimensions researcher IDs associated with each of the professors in our sample, in order to then retrieve information about publications and patents associated with these IDs. Note that several IDs can be associated with one individual. Our algorithm matches each professor with one (or more) researcher ID(s) using the following procedure:

- We first clean institution names in our sample and match each institution to its equivalent in Dimensions. Each institution is therefore associated with a "grid.id" identifier
- We clean first names in the obvious cases where the nickname was used as first name (e.g., some individuals in our sample have first name "Bill" that we transform into "William")
- At this point, each row in our dataset corresponds to an individual for whom we know last name, first name, and grid.id. The goal of the procedure is now to find the researcher ID(s) associated with each row
- For this, for each individual (or equivalently row), we perform an exact match based on last name and grid.id. This gives us all the potential researcher ID(s) that could be a match because they worked at the same institution as our focal professor at one point during their career and have an identical last name. We do not perform an exact match on the first name as this point as Dimensions may combine first and middle names into first name, which would make us miss potential matches when our main dataset does not include middle name
- For each individual, we search among his potential matches and keep only those whose associated first name in Dimensions includes the first name of our focal professor. For example, if our individual has first name "Carl" and we have 3 potential matches for this individual whose first names are respectively "Carl K", "Tom" and "Carter", only the first match will be kept
- Among the individuals for which the previous procedure led to no match, we perform a fuzzy match on first name. This is useful for individuals for which the first-name in our dataset has an hyphen (for instance "kwok-kin" should be matched to the Dimension researcher "kwok kin" but this was missed in the previous step. Similarly, "Robert" is a match for the first name "Rob" but would have been missed otherwise). We test several thresholds for the fuzzy match and end up selecting matches whose score is above 80. We manually check every match to ensure accuracy
- We exclude 2 individuals for which there was a high number of researcher IDs matches (11 and 12 matches respectively)
- For researchers where there is no middle name in our dataset, we manually search online to find it and keep researcher IDs matches whose middle name is the same as the one identified

This procedure leads us with 561 professors for whom we have identified one or more researcher IDs. Given that a professor can be associated with several researcher IDs, we then collapse publications, citations and patents at the professor level.

[20]https://dimensions.freshdesk.com/support/solutions/articles/23000018779-how-are-researchers-unified-disambiguated-in-dimensions-