Beyond the Lab: The Effect of PhD Programs on Innovation

The latest version of this paper is available at this link

Manfredi Aliberti*

March 11, 2026

Abstract

This paper estimates the causal impact of PhD programs, designed to train individuals to advance the frontier of knowledge, on innovation. I exploit the centrally planned and staggered rollout of doctoral programs across Italian universities and construct a new dataset linking program openings to local patenting activity. The introduction of PhD programs increased patenting by 21% between 1986 and 2001. Using admission exam scores in a regression discontinuity design, I show that about 22% of this effect is direct and driven by the increased patenting of program graduates, while most of the remainder reflects spillovers to local firms. These findings indicate that PhD programs stimulate technological progress both by increasing graduates' inventive output and by strengthening the surrounding innovation ecosystem. A cost–benefit analysis based on patent valuations suggests that the social return to these programs exceeds total costs by at least 46%. Finally, I estimate that PhD programs raised Italy's GDP by 0.6% to 4.7% over the same period.

*Rome Economics Doctorate (RED). Email: maliberti@redphd.it.

I am deeply grateful to Fabiano Schivardi, Barbara Biasi, and Michela Giorcelli for their invaluable guidance. I extend my gratitude to Gaia Dossi for the extremely helpful advice and comments. Special thanks go to the EIEF faculty, including Stefano Gagliarducci, Luigi Guiso, Claudio Michelacci, Joao Monteiro, Matteo Paradisi, Luigi Pascali, Mounu Prem, Facundo Piguillem, Andrea Pozzi, Bernardo Ribeiro, Massimo Sbracia, Giancarlo Spagnolo, and Emanuele Tarantino, and to Fabio Bertolotti, Gabriele Cristelli, Arnaud Dyèvre, Ina Ganguli, Daniel Gross, Dietmar Harhoff, Kyle Myers, Jacob Moscona, and Michael Rose for their valuable feedback. I also thank all participants to various seminars and conferences for feedback and helpful discussions. Any errors are my own.

1 Introduction

Innovation drives long-run economic growth, but what drives innovation itself? Human capital is one of the main engines of innovation (Romer, 1990, 1994), and education is one of the key mechanisms through which human capital is formed. Among different forms of education, PhD programs stand out because they are designed to train researchers to expand the frontier of knowledge. Yet, while PhD holders are 20 to 40 times more likely to be inventors than the average individual (Aghion et al., 2017; Akcigit et al., 2024; Jung and Ejermo, 2014), little is known about the causal effect of PhD programs on innovation.

Isolating the impact of PhD programs on innovation is challenging at both the aggregate and individual levels, and three key questions remain open. First, do PhD programs increase local innovation? Addressing this question is difficult because program establishment may be endogenous to local conditions and historical data on their rollout are scarce, particularly in advanced economies where doctoral education dates back to the nineteenth century. Second, do PhD programs improve graduates' inventive capacity, or do they only select the most talented applicants? Descriptive statistics conflate selection and training effects, and the lack of data on admission rankings has limited the ability to disentangle these channels. Finally, even if PhD programs foster innovation, are they a cost-effective policy instrument? Estimating cost-effectiveness requires translating innovation outcomes into monetary terms, which remains empirically challenging.

This paper estimates the causal effects of PhD programs on innovation by exploiting the staggered introduction of doctoral programs across Italian universities starting in 1983. Using newly digitized data on the panel of Italian PhD programs, I find that they significantly increased local patenting. To assess how much of this increase can be attributed to the training received by PhD graduates, I assemble a novel dataset linking PhD admission exam scores to applicants' patenting activity. The results show that PhD training enhances graduates' innovation, but that most of the effects stem from spillovers to non-PhD inventors employed in R&D-intensive firms and socially connected to PhD graduates through networks formed among former university peers. Finally, I estimate the monetary value of the patents induced by the reform and find that it exceeds program costs, even under conservative assumptions.

PhD programs were introduced in Italy in 1983. Before their introduction, Italian universities offered only four- and five-year degrees with no further opportunities for research training. Graduates willing to pursue an academic career would work as professors' assistants, employed mostly in teaching duties. The new doctoral degree was introduced to create a dedicated track for young researchers to conduct original research, as PhD students were not allowed to teach and the programs did not entail courses or exams. Thus, this reform represents a discrete policy shock that I exploit to identify the causal effects of doctoral programs on local innovation.

I identify the effect of PhD programs by exploiting the centralized allocation mechanism in place between 1983 and 1998, when doctoral programs could operate only if included in the annual national plan drafted by the Ministry of University. After this reform, PhD positions and universities offering doctoral training increased sharply, suggesting that the centrally determined supply had constrained universities' demand. The data on pre-reform local characteristics indicate that the ministry targeted more populous and service-oriented provinces with a more educated workforce. However, trends in patenting reveal that provinces opening programs earlier were not on a differential trajectory of innovation relative to those opening after 1998. I control for initial conditions throughout the analysis and define as treated the provinces whose first PhD program opened before 1999, and as controls those where it opened later.

I compile a new dataset linking the opening of PhD programs to local patent production. I first construct a comprehensive series of all PhD programs established in Italy between 1983 and 2012 by digitizing public calls for applications up to 1998 and merging them with the official database available from 1999 onward. Then, I merge this series with local patent production data from Bianchi and Giorcelli (2019), using universities' and inventors' provinces and the patent filing year. The patent data include inventors' and assignees' names, inventors' locations, technology (International Patent Classification, IPC), and assignees' tax codes from 1990 on.

PhD programs significantly raise local patenting, with an increase of about 21% between 1986 and 2001 when restricting the treatment to STEM PhD programs. I estimate the effects in a Difference-in-Differences (DiD) setting using the Wooldridge (2021) estimator, robust to cohort-specific dynamics and heterogeneous treatment effects. Dynamic estimates show that the impact builds up gradually and stabilizes after about ten years, consistent with a permanent increase in

local innovative capacity through the expansion of the stock of PhD graduates.

The results are robust to alternative estimation methods and aggregations of patent data, and are not driven by large cities or migration from control provinces. First, using a classical two-way fixed effects (TWFE) specification yields estimates similar to the baseline, as does aggregating innovation outcomes at the province level. Second, scaling patent counts by local population or by the number of establishments does not alter the results, and excluding major cities from the analysis produces similar effects. Finally, omitting treated provinces that border control areas does not affect the estimates, consistent with individual-level evidence showing that only about 2% of applicants were born in control provinces.

Treatment effects concentrate in technologies more exposed to local PhD programs. I construct an exposure index that weights the stock of PhD graduates by the proximity between each PhD subject and technology. I measure knowledge proximity using text data from about 1.3 million US theses in the *ProQuest* database and 2.1 million patents from the *USPTO*, computing the average Term Frequency–Inverse Document Frequency (TF–IDF) cosine similarity for each subject–technology pair. The resulting province–year–technology measure proxies the number of local PhD researchers in each technology. The results show that PhD programs generate innovation in more similar technologies, indicating that their subject composition shapes where innovation occurs.

PhD programs can influence innovation through both direct and indirect channels. Directly, PhD graduates may contribute to technological progress by transforming the research skills and scientific knowledge acquired during their training into patentable discoveries. Indirectly, the introduction of PhD programs might strengthen the local innovation environment by diffusing advanced research methods and frontier knowledge beyond academia.

I estimate the direct effect of completing a PhD on patenting using a fuzzy regression discontinuity design (RDD) that instruments completion with crossing exam-specific admission cutoffs. Admission-exam data allow me to compare applicants who all exhibit a revealed preference for research, while the RD design isolates the causal effect of PhD training from ability-based selection.

I implement this design compiling a new dataset that links applicants' performance in PhD admission exams to PhD completion and patenting outcomes. Specifically, I digitize admission-

exam rankings and scores from the archives of five Italian universities, covering about 25% of all PhD positions opened in Italy between 1983 and 1998. Completion status is derived by linking applicants to thesis records at the National Central Library of Florence.

The RDD estimates show that PhD training from applied STEM programs explains about 22% of the total local effect on patenting. To obtain this result, I first classify programs as applied or non-applied based on the 1982 subject-specific share of public R&D expenditure devoted to applied research. The RDD results indicate that completing an applied STEM PhD significantly increases individual patenting, while no effect emerges for non-applied programs. I then perform a back-of-the-envelope calculation that multiplies the estimated LATE for applied STEM programs by the number of graduates, assuming it approximates the average training effect. This product quantifies the individual-level contribution of PhD training to the overall local patenting impact.

Corporate patents by local non-PhD inventors account for about 54% of the overall effect of PhD programs, pointing to substantial spillovers. To capture these effects, I restrict the outcome to patents with no PhD graduates among their inventors and limit the sample to patents owned by limited liability firms. These restrictions ensure that no PhD graduate participated in the patent and that results are not driven by research organizations where PhD-trained researchers may be informally involved, such as universities or public research centers. Although RD estimates show that PhD graduates themselves contribute to the rise in patenting, more than half of the overall impact stems from spillovers to local firms, demonstrating that doctoral programs strengthen local innovation environments.

The mechanism behind these effects remains unclear. Recent studies (Beaman et al., 2021; Prato, 2024) emphasize the role of social ties in the diffusion of knowledge. However, because PhD students were not allowed to teach and the observed spillovers stem from corporate patents with no PhD inventors, the transmission channel cannot rely on knowledge transfer through university teaching or workplace interactions. Thus, I test if the mechanism operates through social connections between PhD graduates and local workers that predate PhD enrollment. To capture these links, I exploit subject-specific local flows of university graduates into the labor market and PhD programs to estimate the social links of different industries to local PhD programs.

Local spillovers concentrate in R&D-intensive industries with stronger social ties to local PhD

graduates. To explore this mechanism, I classify industries according to their pre-reform R&D intensity and their expected social links to PhD programs. R&D intensity is measured as the pre-reform share of workers employed in R&D positions. The expected social links are proxied by the product of the expected number of local recent graduates employed in an industry and the probability that one of their former university peers enrolls in a local PhD program, using ISTAT data on university graduates and my compiled dataset on PhD positions. The results indicate that spillovers materialize only where both links to PhD programs and R&D investments are present, highlighting the complementarity between human capital formation and firms' research capacity.

The final part of the paper shows that PhD programs are a cost-effective instrument to foster innovation. I begin by collecting information on the accounting costs of PhD programs including scholarships, research equipment, and administrative expenses. Dividing the overall effects of PhD programs by their total costs, I find that doctoral programs generated 4.27 additional patents for every $1 million spent in 2020 dollars, of which 0.92 can be directly attributed to PhD graduates. These estimates are higher than those reported in the literature but remain within a comparable order of magnitude to other innovation policies, such as the US Department of Energy SBIR program (Myers and Lanahan, 2022) and the NIH grants (Azoulay et al., 2018).

I estimate the monetary value of patents using the method of Kogan et al. (2017). This method infers patent values from changes in the assignee's stock price in the days following the patent announcement. The key assumption is that financial markets incorporate into firms' valuations the discounted flow of profits expected from newly granted patents. This approach allows me to compare the monetary benefits of PhD programs with their costs, overcoming the limitation of purely quantitative measures of innovation.

Even under conservative assumptions, the patent-based benefits of PhD programs exceed their costs by 46%. I estimate patent values by matching Italian listed firms to their patents and compare these benefits to program costs, including students' opportunity costs proxied by field-specific earnings of recent graduates. I compute a lower bound assuming that non-listed firms' patents have no value and an upper bound assigning them the mean value of listed-firm patents in the same technology field. The resulting benefit–cost ratio captures the social returns on public investment and implies that PhD spending raised Italy's GDP by between 0.6 and 4.7% between 1986 and 2001.

Related Literature This paper contributes to the literature on how public policies foster innovation (see (Bloom et al., 2019) for a review). Recent research has examined several types of interventions: increases in public funding for universities and research centers (Azoulay et al., 2018; Babina et al., 2023; Bergeaud et al., 2022), R&D grants to firms (Howell, 2017; Myers and Lanahan, 2022), mission-oriented programs and large-scale public R&D initiatives (Fieldhouse and Mertens, 2024; Gross and Sampat, 2023; Kantor and Whalley, 2025), and fiscal policies affecting innovation incentives (Akcigit et al., 2021; Dechezleprêtre et al., 2023; Moretti and Wilson, 2017; Mukherjee et al., 2017). These policies operate primarily through financial or incentive channels. In contrast, the policy I study isolates the human capital channel: it expanded the local supply of research-intensive workers without altering infrastructure, funding, or R&D incentives.

This paper contributes to the literature on the returns to higher education. Previous research has documented its effects on wages (Dale and Krueger, 2002; Duryea et al., 2023; Kirkebøen et al., 2016; Mountjoy, 2024; Zimmerman, 2014), social mobility (Mountjoy, 2022), marriage (Ge et al., 2022; Kirkebøen et al., 2021), and health (Buckles et al., 2016). I instead provide causal evidence on the effects of doctoral education on individuals' propensity to innovate and their patenting outcomes.

This research is most closely related to the literature examining how universities affect innovation. Prior work has studied the establishment of new institutions (Andrews, 2023; Toivanen and Väänänen, 2016), the local presence of universities (Valero and Van Reenen, 2019), and changes in intellectual property regimes (Hausman, 2022). Unlike these studies, my analysis isolates the effect of research-oriented training within existing universities, capturing the impact of doctoral education itself rather than changes in institutional structure or incentive frameworks. The closest empirical contribution is Bianchi and Giorcelli (2019), who evaluate the innovation effects of expanding access to STEM undergraduate education. While they find limited or mixed impacts on patenting, I show that STEM PhD programs significantly raise individual propensity to patent and overall patenting. This contrast may stem from the nature of the policies: while their reform broadened access to scientific education, mine specifically targets research training. On the theoretical side, Akcigit et al. (2024) develop a model linking PhD education to innovation through selection into the research

sector. My findings complement this theory by providing causal evidence on the training effect and documenting spillover effects to local firms and non-PhD inventors.

The remainder of the paper is structured as follows. Section 2 describes the historical background of doctoral programs in Italy. Section 3 presents the data and descriptive statistics. Section 4 documents the local effects of PhD programs on innovation. Section 5 shows that the increase in patenting is concentrated in technologies that cover similar topics to local PhD programs. Section 6 examines mechanisms through an individual-level RDD and a spillover analysis. Section 7 provides estimates for the cost-effectiveness of the introduction of PhD programs in terms of cost-benefit analysis. Section 8 concludes.

2 The Institution of PhD Programs in Italy

2.1 Background

After the end of World War II, Italy experienced a period of economic growth locally known as the Economic Miracle (*il Miracolo Economico*). Between 1951 and 1972, real GDP per capita grew at an average annual rate of 4.8 percent. This sustained growth enabled Italy to catch up with the other major European economies: by 1973, output per hour worked exceeded that of the United Kingdom, and GDP per capita reached 90 percent of the British level (Bastasin and Toniolo, 2023). The resulting increase in demand for STEM graduates from the growing industrial sector prompted a series of reforms of the university system between 1961 and 1969 (see Bianchi and Giorcelli, 2019).

In 1961, the Italian university system was still governed by a 1933 law[1], which imposed a rigid hierarchy among students based on their secondary-school track. Only classics and sciences high school graduates could enroll in STEM university courses (see Appendix Table C.1). The system also lacked any structured form of postgraduate education: the only pathway to an academic career was to be appointed as a professor's assistant. Although the position of assistant had been formally defined by a 1948 decree[2], these roles were often unpaid and the duties assigned were left entirely to the discretion of the supervising professor.

[1] R.D. 31 Agosto 1933, n. 1592
[2] D.L. 7 Maggio 1948, n. 1172

Between 1961 and 1969, a series of reforms liberalized university admissions by allowing graduates from any type of secondary school to enroll in all university programs (Bianchi, 2020). These changes led to a more than threefold increase in university enrollment between 1959 and 1979 (Fazio, 1980) and the ministry had to increasingly divert more resources from research to accommodate the increasing need for teaching personnel.

This reallocation of resources coincided with a marked decline in academic research and patenting activity in Italy. Appendix Figure B.1 shows these opposing trends: as the student–faculty ratio in STEM disciplines rose sharply, the number of patents filed by Italian residents fell. The expansion of university teaching thus appears to have constrained research capacity within academia, contributing to a slowdown in innovation.

In response, the government introduced PhD programs in 1980, though their actual implementation began in 1983 due to administrative delays. These programs were meant to create a space for young researchers to undertake original research without the burden of teaching duties, as teaching was forbidden to PhD students by law.

2.2 The Introduction of PhD Programs

The 1980 reform[3] formally introduced PhD programs into the Italian higher education system.[4] The decree defined the PhD degree "as an academic degree to be valued only within the scope of scientific research", explicitly stating that it was not a prerequisite for academic positions.[5] PhDs could be awarded only "following the performance of research activities [...] that have resulted in original contributions to knowledge." This highlights that the main purpose of PhD programs was to foster scientific research by creating a new academic role devoted exclusively to advancing knowledge.

PhD programs were designed as pure research programs, without coursework or exams. Training

[3]D.P.R. 11 Luglio 1980, n. 382.

[4]Before the 1980 reform, two PhD-like programs already existed in Pisa and Trieste. In Pisa, the Scuola Normale Superiore offered postgraduate *Corsi di Perfezionamento* in Humanities and Sciences, often requiring an original research thesis. In Trieste, the International School for Advanced Studies had established a PhD in Theoretical Physics in 1978 with the support of the International Centre for Theoretical Physics. Both programs were officially recognized as PhD courses in 1987.

[5]The PhD became a formal requirement for entry-level faculty positions (*ricercatori a tempo determinato*) only in 2010 under Law L. 30 Dicembre 2010, n. 240.

consisted exclusively of seminars and research under the supervision of a faculty member, typically within the supervisor's laboratory. This design allowed PhD graduates to focus on research and to work in close collaboration with established researchers in their fields, facilitating knowledge transfer from mentors to students.

This law prescribed how the admission procedure was to be structured. Each department administering a PhD program was required to open a public call for applications to which any individual with a recognized university degree could partake. The following admission process consisted of an exam presided by a committee of university professors, with at least one member coming from a different university.

The decree established a centralized allocation mechanism for PhD programs, under which PhD positions could only be opened within an annual ministerial plan (Appendix Figure B.2 outlines the process). Each year, the Ministry of University issued a call for applications inviting universities to propose new or renewed PhD programs, specifying the subject and the number of positions requested. Based on these applications and the available budget, the Ministry determined the overall number of positions for the following academic year in each field. The National University Council (CUN)[6] was then responsible for allocating these positions across universities and specific subjects. This highly centralized procedure governed the organization of PhD programs from 1983 to 1998. Due to administrative delays within the Ministry, however, no programs were offered in 1984 or 1987.

The ministerial budget constraint and the centralized allocation mechanism led to a slow diffusion of PhD programs. When the first programs were launched in 1983, they were hosted by 36 universities across 32 provinces; by 1998, 58 universities across 48 provinces offered at least one PhD program. A 1998 reform[7] dismantled the centralized allocation system, liberalizing both the funding and organization of doctoral education. From that point onward, universities could independently establish PhD programs provided that departments met minimum standards of faculty size and research infrastructure set by the Ministry.

Evidence of under-supply during the centralized period emerges clearly from Appendix Figure

[6] An advisory body within the Ministry composed of elected professors and researchers with equal representation across disciplines.

[7] L. 3 Luglio 1998, n. 210

B.3. Panel B.3a shows that the number of universities offering PhD programs increased linearly until 1998, followed by a sharp jump from 58 to 66 in 1999. Panel B.3b shows a similar discontinuity in the number of available PhD positions, which nearly doubled after liberalization. These patterns indicate that the Ministry's allocation process constrained both the number and expansion of PhD programs. This under-supply will be at the core of my identification strategy, as outlined in section 4.1.

3 Data and Descriptive Statistics

I assemble two novel datasets to analyze the effects of PhD programs on innovation at both the aggregate and individual levels. The first is a panel dataset at the province–year–technology level, which links the staggered rollout of doctoral programs to local patenting activity. The second is an individual-level dataset that matches applicants to PhD programs with the patents they subsequently produced, enabling me to estimate the causal effect of completing a PhD on patent output. Finally, I report descriptive statistics on the rollout of PhD programs and on the role of PhD graduates in innovation in Italy.

PhD Programs Data

I collect the data on PhD programs from two different sources based on the starting year of the programs. I use the calls for application published on the *Official Gazette of the Italian Republic* for programs starting before 1999. I employ the official database of PhD programs for programs starting in 1999 or later. Appendix Table C.2 shows some descriptive statistics.

The data cover the universe of PhD programs in Italy, except for the years 1985 and 1986 when only a subset of programs published public calls for applications. For programs that appear in 1988 but not in 1983, I backdate their local opening to 1985. This is a conservative choice: it may anticipate the treatment to years without actual exposure, therefore biasing estimates toward zero by including periods with no treatment effect. Additionally, excluding the 1985 cohort from the analysis does not alter the results.

The final dataset include the subject, the university offering the program, the ministry-defined

scientific sectors, and the number of positions made available[8].

To assign PhD programs to Italian provinces, I link each program to a university department using the university name and the scientific sectors. When multiple scientific sectors are reported, I link the program to all departments that are related to the sectors. When the linked departments have multiple locations in different provinces, I link the program to all relevant provinces. Appendix A.1 describes in more detail how this dataset was constructed.

Patents and Inventors Data

I measure local innovation through the data on patents collected by Bianchi and Giorcelli (2019). This dataset includes the patents issued by the IPO between 1946 and 2010, retrieved from the IPO archives and the European Patent Office's (EPO) PATSTAT database.

I complement this dataset with the inventors' names retrieved by Google Patents and the online database of the IPO and the International Patent Classification (IPC) classes reported by the EPO. Additionally, I retrieve from PATSTAT the total number of citations received by Italian patents from foreign-filed ones.[9] From the resulting dataset, I use the subset of patents filed between 1976 and 2001. Descriptive statistics are shown in Appendix Table C.3.

The data include the date of application, names and locations of the inventors, and International Patent Classification (IPC) class. Inventors' and firms' social security number are consistently provided for patents published after 1989.

I use the location of inventors to assign patents to provinces. If a patent has multiple inventors located in different provinces, the patent is linked to all relevant provinces. Patents are assigned to years based on the application date.

Finally, I exclude from all samples the patents whose IPC class was entered incorrectly or introduced into the classification after 1976, thus retaining 132,109 patents (99.94% of the original data) from the IPO data. Each IPC class constitutes a technology in my dataset.

[8] A sample entry: PhD in Historical Economics; University of Rome; Applied Economics, Economic History; 4 positions

[9] Information on citations made by Italian patents is unavailable before 2008, as the IPO only assessed the bureaucratic correctness of patents and not the prior art until July 1st 2008

PhD Graduates

I obtain the data on all individual PhD graduates in Italy from 1986 to 2019 from the website of the National Central Library of Florence (BCNF). All Italian universities are required by law to deposit a copy of each PhD dissertation at the BCNF and BCNF has created a public online catalogue of all PhD dissertations based on the deposits.

I retrieve 174,218 unique dissertations, corresponding to 65.71% of the PhD positions opened between 1983 and 2016. The dataset includes name of PhD graduates, dissertation title and year of deposit for each dissertation. University, subject of the PhD program, starting year of the PhD program, year of graduation, and the names of the supervisors are present for a subset of dissertations. Appendix A.2 describes how this dataset was built and shows the frequency of each piece of information. Appendix A.3 explains in detail how this dataset was matched to the inventors and PhD entry exam data.

PhD Entry Exam Data

To estimate the individual returns from completing a PhD, I collect data on PhD entry exams held between 1983 and 1998 in five major Italian universities (Rome La Sapienza, Milan La Statale, Padua, Genoa, and Catania) which together account for approximately 25% of all PhD positions offered in Italy during this period. I digitize the final rankings for the near-universe of PhD programs offered by these universities, recording for each applicant their name, surname, subject, year, gender, and score. The resulting dataset contains 19,567 applications from 17,023 distinct individuals.

Using the information on name, surname, subject, year of application, and university, I merge this dataset with the data on PhD graduates and inventors as explained in detail in Appendix A.3. The resulting dataset includes for each applicant the number of patents published and if they have completed any PhD program between 1986 and 2001. Appendix Table C.4 present some descriptive statistics.

Additional Data

I complement the dataset on patents and PhD programs with province-level information on local economic and demographic conditions, and academic quality. I use the 1981 General Census of Industry, Trade, Services and Handicrafts to collect information on the industrial specialization and size of local firms before the reform.

I obtain information on local human capital and demographics from the 1981 General Population Census. This census contains information on local population, demographic structure and share of graduates among young people at the provincial level.

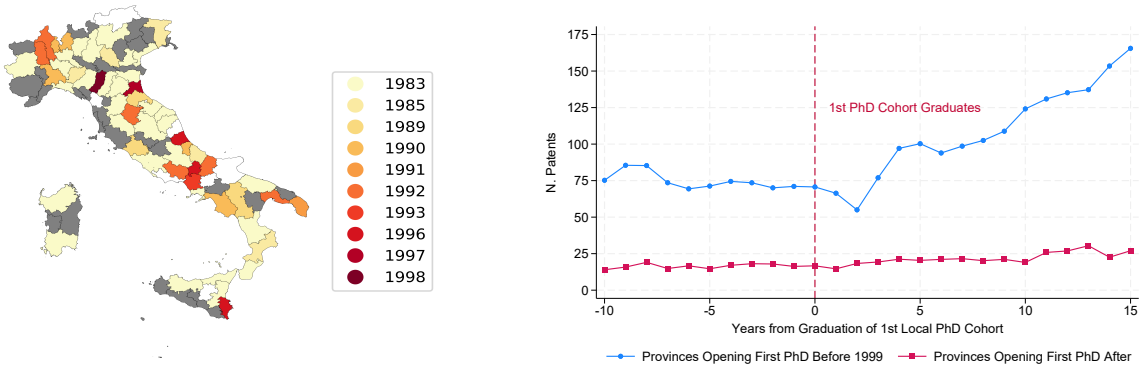
In addition to the main datasets used throughout the analysis, I will also use complementary data sources described later in the relevant sections. Appendix Table C.5 lists all the different data sources employed in the paper.

Descriptive Statistics

This section presents descriptive statistics on patenting trends in Italy and examines the contribution of PhD graduates to innovation between 1976 and 2001. It also describes the introduction of PhD programs, documenting their geographical diffusion and field composition.

Figure 1 shows the distribution of PhD programs across Italy and the trends in patenting for provinces that opened their first PhD program under the centralized mechanism compared to those that did so after its dismantling. Panel 1a shows that PhD programs were opened across the entire country and that their diffusion across universities was highly staggered. Panel 1b evidences how provinces that opened under the centralized mechanism experienced a sharp rise in local patenting between 1986 and 2001, while those opening later saw only a modest increase. This pattern suggests that the opening of PhD programs lead to an increase in local patenting, with small positive spillovers on the provinces with no openings similar to other findings in the literature (Myers and Lanahan, 2022; Valero and Van Reenen, 2019). Appendix Figure B.5 shows that also aggregate patent production increased more in Italy than in other European countries after the introduction of PhD programs.

Figure 1: PhD Programs Rollout and Trends in Patenting



(a) Year of Opening of First Local PhD

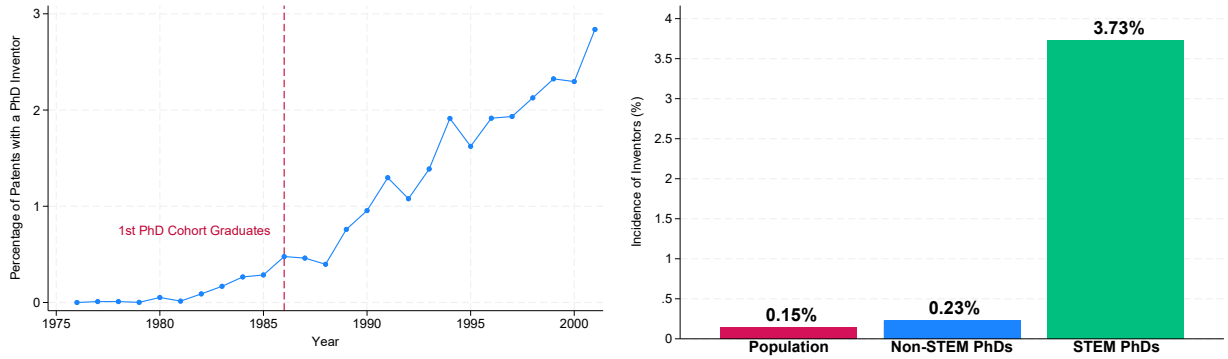
(b) Patenting by Year of Opening

Notes: The left panel shows the year of the first local PhD opening in each Italian province. Provinces in white had their first local PhD activated between 1999 and 2012, while provinces in gray had no local PhD graduates before 2012. The right panel plots the average number of granted patents across provinces between 1976 and 2001, aligned by distance from the year of the first local PhD graduation. The blue line with circles corresponds to provinces where the first local graduation occurred between 1986 and 2001. The red line with squares corresponds to provinces where it occurred between 2002 and 2014. For comparability, the red line is computed as a weighted average over years, weighting by the number of treated provinces in each graduation year.

Turning to the direct contribution of PhD graduates to innovation, Figure 2 reports the share of patents with at least one PhD inventor between 1986 and 2001 and the incidence of inventors by educational attainment. The share of patents with at least one PhD inventor grows steadily from the mid-1980s onward, reaching nearly 3% by 2001 even though PhD graduates made up less than 0.1% of the working-age population at the time. STEM PhD graduates are particularly likely to innovate: the incidence of inventors among them is about 28 times higher than that of the rest of the population. These patterns suggest that STEM PhD graduates are the most innovative individuals of the population but do not allow us to isolate the causal effect of PhD training due to a potential selection bias. As PhD graduates are selected based on their abilities at the time of enrollment and self-select into applying to PhD because of their taste for research, any simple comparison would be biased.

Finally, Appendix Figure B.4 documents how doctoral programs were introduced between 1983 and 2011. During the centralized allocation period, positions were distributed evenly across Italy, and the composition of graduates by field also remained broadly constant over time. This suggests that the type of treatment was fairly homogeneous across territories and time, with the main

Figure 2: The Role of PhD Graduates in Patenting



(a) Share of Patents with a PhD Inventor (b) Probability of Being an Inventor by Education

Notes: The figures highlight the role of PhD graduates in producing patent in Italy. Panel (a) reports the share of patents produced each year with at least one PhD inventor, that is an inventor who obtained a PhD between 1986 and 2001 (even after the patent submission). Panel (b) shows the probability of being an inventor for PhD graduates and the rest of the working age population, calculated as the probability of having at least one granted patent submitted between 1976 and 2001. Both figures have been made using the method described in A.4

variation arising from the local size of the stock of PhD graduates.

4 The Effects of PhD Programs on Innovation

This section describes the empirical strategy to identify the effects of PhD programs on local innovation and presents the main results. section 4.1 explains the identification strategy and the methodology. In section 4.2, I show and discuss the main results.

4.1 Empirical Strategy

The first aim of the paper is to estimate the effect of PhD programs on innovation. A natural starting point would be to compare areas hosting PhD programs with those that do not. However, such comparisons would be biased due to selection issues. Both the decision to open a PhD program and the presence of a local university, which is a prerequisite for activating doctoral programs, are endogenous to local economic conditions. In particular, universities and PhD programs are more likely to be established in areas with higher demand for skilled labor and research activities.

I address this selection bias by exploiting the institutional features of the centralized allocation

Table 1: Pre-Reform Descriptive Statistics

Variable	STEM PhD	All PhD	Control	Variable	STEM PhD	All PhD	Control
'82 IPO Patents	0.68	0.65	0.15	'81 % Workers in Wholesale and Retail	18.9	18.76	19.99
	(3.49)	(3.35)	(0.54)		(2.93)	(2.97)	(1.79)
'81 % Workers in Large Firms	10.18	10.01	6.29	'81 % Workers in Hospitality	3.78	3.79	6.14
	(6.30)	(6.06)	(2.77)		(1.29)	(1.37)	(5.63)
81 % of Young Graduates	7.92	7.82	6.77	'81 % Workers in Logistics	6.65	6.51	5.74
	(1.48)	(1.48)	(1.11)		(2.29)	(2.28)	(1.38)
'81 % Workers in STEM R&D	0.07	0.07	0.04	'81 % Workers in Financial Sector	2.22	2.21	1.74
	(0.10)	(0.10)	(0.03)		(0.81)	(0.78)	(0.31)
'81 % Workers in non-STEM R&D	0.06	0.05	0.02	'81 % Workers in Business Services	3.21	3.18	2.82
	(0.08)	(0.07)	(0.01)		(0.86)	(0.83)	(0.31)
'81 % Workers in IT	0.21	0.20	0.14	'81 % Workers in Public Administration	4.89	4.71	5.66
	(0.16)	(0.15)	(0.06)		(2.33)	(2.30)	(2.47)
'81 % Workers in Agriculture	0.64	0.64	0.67	'81 % Workers in Education	9.36	9.05	11.12
	(0.77)	(0.74)	(0.30)		(3.63)	(3.61)	(4.91)
'81 % Workers in Fishing	0.18	0.18	0.26	'81 % Workers in Healthcare	5.28	5.17	4.99
	(0.21)	(0.21)	(0.31)		(1.13)	(1.15)	(1.43)
'81 % Workers in Mining	0.40	0.41	1.22	'81 % Workers in Public and Social Services	3.79	3.80	4.03
	(0.45)	(0.44)	(1.49)		(0.70)	(0.79)	(0.76)
'81 % Workers in Manufacturing	31.24	32.12	26.02	'81 % Workers in Other Industries	0.05	0.04	0.00
	(11.39)	(11.57)	(11.52)		(0.20)	(0.19)	(0.00)
'81 % Workers in Public Utilities	1.06	1.04	0.96	'81 Total Population ('000)	811.39	784.72	394.97
	(0.37)	(0.36)	(0.19)		(790.20)	(761.61)	(141.34)
'81 % Workers in Construction	8.22	8.26	8.56				
	(2.82)	(2.71)	(1.19)				

Notes: The table reports descriptive statistics for the treatment and the control groups in the pre-treatment period. The STEM PhD columns present the statistics of the treatment group when only STEM PhDs are used for the definition of treatment, All PhD refers to the broader treatment definition. The statistical unit is $province \times technology \times year$. The variables included are local patent activity registered at the IPO in 1982, share of university graduates among people aged between 25 and 34 in 1981, share of workers employed in R&D institutions in 1981 by discipline, share of workers employed in IT firms in 1981, share of workers employed in firms with 500 or more employees in 1981, industrial composition in 1981, and total population in 1981. The data reported are the mean and the standard error (in parenthesis).

mechanism described in section 2.2. Because the Ministry under-supplied PhD positions relative to local demand, I interpret provinces that established their first doctoral program after its dismantling as areas where demand for PhD programs was previously unmet by ministerial planning. Accordingly, I define as treated all provinces that opened their first PhD program by 1998 within a ministerial plan,[10] and as control those that did so later. The treatment start is defined as the graduation year of the first local cohort, three years after the opening of the first local PhD, reflecting the minimum legal duration of PhD programs which covers about 71% of graduates (see Figure B.6).

Since the allocation of PhD positions was not random under the centralized mechanism, it is essential to understand which local characteristics guided the ministerial committee's choices. Identifying these determinants helps isolate the causal impact of PhD programs from confounding local factors. Table 1 shows that treated and control provinces differed before the reform, particularly in industrial composition and population size. Regression analyses reported in Appendix Figures B.7

[10] Thus, Pisa and Trieste are excluded. See section 2.2.

and B.8 confirm that the committee tended to select more populous, service-based provinces.

For my identification strategy to be valid, treated and control provinces must not exhibit differential trends in patenting before the treatment. Figure 1b shows that patenting evolved similarly across the two groups prior to the reform, indicating that the ministerial committee did not target provinces because they were already on a faster innovation trajectory. Rather, their choices appear to have been driven by pre-reform local characteristics. In the next section, I show that controlling for these pre-reform characteristics does not alter the results.

To estimate the effects of PhD programs on local patenting, I adopt the estimator proposed by Wooldridge (2021). This method estimates treatment effects separately by treatment cohort and event time, which the researcher can then aggregate into any desired parameter by explicitly defining the weighting scheme. This framework avoids the negative weighting issues that affect the standard TWFE estimator (Athey and Imbens, 2022; Borusyak et al., 2023; Goodman-Bacon, 2021). Specifically, I estimate:

$$y_{p,c,t} = \alpha_{p,c} + \alpha_{c,t} + \sum_{j=1986}^{2001} \sum_{i \neq -1} \beta_{i,j} \cdot \mathbb{1}\left\{t = g\left(p\right) + i,\ g\left(p\right) = j\right\} + \gamma \cdot \alpha_t \times \mathbf{X}_{p,1981} + \varepsilon_{p,c,t} \qquad (1)$$

Where $y_{p,c,t}$ denotes the number of patents produced in IPC class c and year t by inventors located in province p. The coefficient $\beta_{i,j}$ captures the effect of PhD programs on patenting activity i years after the graduation of the first local PhD cohort, $g\left(p\right)$, for provinces whose first cohort graduated in year j.The term $\alpha_{p,c}$ represents province–technology fixed effects, controlling for time-invariant differences in local technological specialization, while $\alpha_{c,t}$ denotes technology–year fixed effects, absorbing aggregate trends in patenting that are specific to each technological field. Finally, $\alpha_t \times \mathbf{X}_{p,1981}$ captures the interaction between year dummies and pre-reform local socio-economic conditions, including industrial composition[11], the share of employees in large firms, the share of graduates among individuals aged 25–34, total population, and the shares of workers employed in IT, and in STEM and non-STEM R&D separately. These interactions allow for differential time trends in patenting across provinces, as long as they depend on these pre-reform local characteristics.

[11]Defined using the 17 one-digit industries in the 1991 ATECO classification

Since the treatment is defined at the province level while the dependent variable varies at the province–technology level, the coefficient $\beta_{i,j}$ captures the average effect of opening a PhD program across technologies and provinces within the same treatment cohort. This choice regarding the definition of treatment is dictated by the lack of a clear mapping between technologies and PhD subjects. In section 5, I address this limitation by constructing a province–technology-specific index of exposure to local PhD programs, based on a text-based measure of knowledge proximity and the subject-specific stock of graduates.

I use the coefficients $\beta_{i,j}$ to compute the average dynamic effects, denoted by β_i, which correspond to the coefficients of a dynamic DiD specification. For each event time i, I calculate a weighted average of the cohort-specific effects $\beta_{i,j}$, using weights that reflect the relative size of each treatment cohort. I then compute the overall average effect of PhD programs as the mean across all post-treatment event times:

$$\beta_i = \sum_{j=1986}^{2001} w_{i,j}\beta_{i,j} \tag{2}$$

$$\beta_{POST} = \frac{1}{16}\sum_{i=0}^{15} \beta_i \tag{3}$$

Where $w_{i,j}$ is equal to the share of treated provinces belonging to treatment cohort j among those having $Grad\,Year_p + i \leq 2001$, and zero if $j + i > 2001$.

Since the outcome in the main analysis is the annual count of new local inventions, the coefficient β_{POST} identifies the average change in the local innovation flow induced by the introduction of a PhD program. Interpreted through the lens of standard growth models, this represents a shift in the local Poisson arrival rate of innovations attributable to the introduction of doctoral programs.

4.2 PhD Programs Increase Local Patenting

PhD programs significantly increase local patenting. As shown in Table 2, the opening of any PhD program raises annual patenting in the province by 18 to 20 patents.[12] Since Figure 2b shows that STEM PhD holders are much more likely to participate in innovation than non-STEM ones, I

[12]These effects are obtained by multiplying the estimated $\hat{\beta}_{POST}$ coefficients in Table 2 by 118, the number of technological fields.

Table 2: Effects of PhD Programs on Local Innovation

	(1)	(2)	(3)	(4)
Treatment	All PhDs	All PhDs	STEM PhDs	STEM PhDs
Effects of PhD Programs:				
$\hat{\beta}_{POST}$	0.1732***	0.1547**	0.1660***	0.2069***
	(0.0530)	(0.0680)	(0.0578)	(0.0665)
$\hat{\beta}_{+15}$	0.1929*	0.2679**	0.2099**	0.3169***
	(0.0990)	(0.1159)	(0.1066)	(0.1161)
Mean of DV, Treat=1	0.6925	0.6925	0.7232	0.7232
Control Variables:				
Province-Tech FE	X	X	X	X
Tech-Year FE	X	X	X	X
Full Controls		X		X
N. Clusters	63	63	58	58
Adj. R^2	0.8438	0.8476	0.8513	0.8554
N. [Prov × Tech × Year]	193,284	193,284	177,944	177,944

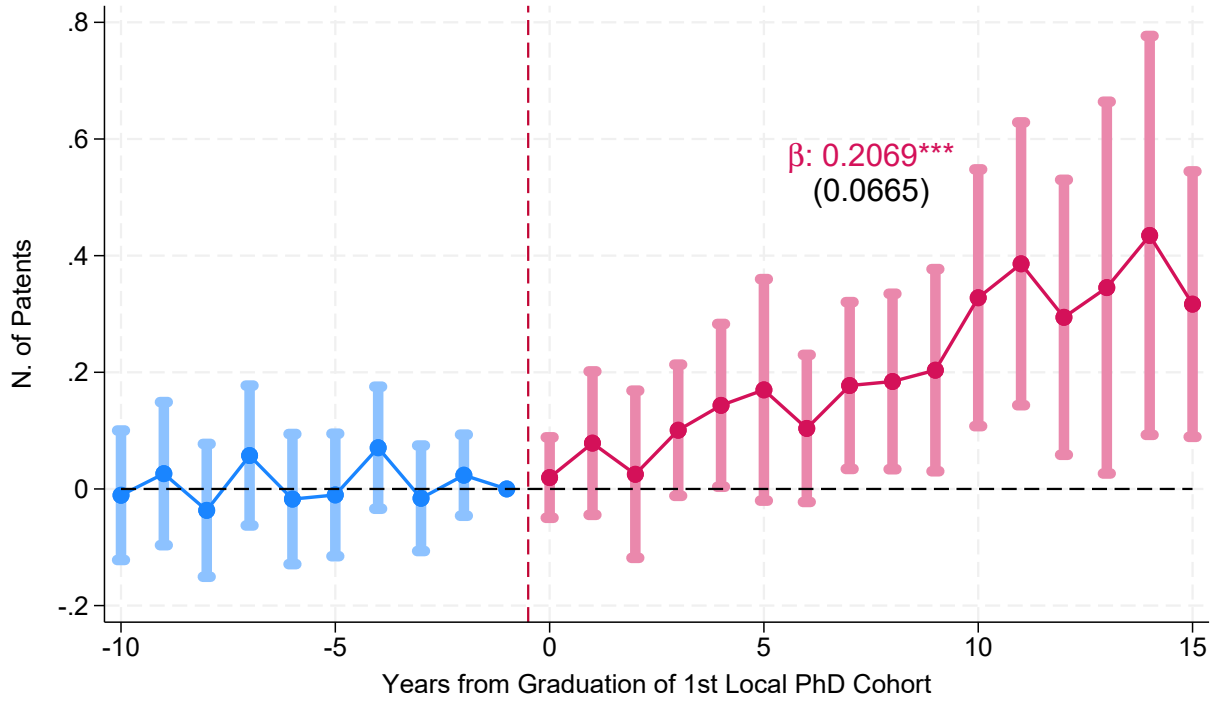
Notes: OLS estimates of Equation 1, one observation is a province-technology-year cell. The dependent variable is the number of patents produced in a province and technology in a year. Columns (1) and (2) use as treatment the opening of any PhD program. Columns (3) and (4) use as treatment the opening of a STEM PhD program and exclude from the analysis the provinces that opened only non-STEM doctoral programs before 1999. Columns (2) and (4) control for local characteristics in 1981 interacted with year fixed effects: industrial composition using the 17 one-digit 1991 ATECO industries, local population, share of workers employed in firms with 500 or more employees, share of workers in IT, share of workers in STEM R&D, share of workers in non-STEM R&D, and share of university graduates among people aged between 25 and 34. Standard errors are clustered at the province level and reported in parentheses. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

restrict the treatment definition to STEM PhD programs only. The results confirm that the effect is driven by STEM programs, which increase local annual patenting by 20 to 24 patents.

Using the specification with STEM PhD programs as the treatment and pre-reform local characteristics as controls, I estimate that doctoral programs generated 13,667 additional patents between 1986 and 2001. This aggregate effect is obtained by multiplying each event-time coefficient $\beta_{i,j}$ by the number of treated provinces in treatment group j and summing across all post-treatment periods. This effect corresponds to a 21% increase relative to a counterfactual equal to total patent production in treated provinces after the first local PhD graduation, net of the estimated PhD-induced patents. Details of the computation are reported in Appendix A.5.

Figure 3 shows no evidence of differential pre-trends in patenting activity between treated and

Figure 3: Effects of STEM PhD Programs on Local Patent Activity



Notes: The figure shows the effects of PhD programs on local patent activity. The plotted coefficients come from the estimation of the model in Column (4) of Table 2. Standard errors are clustered at the province level.

control provinces, suggesting that the allocation of PhD programs did not target areas already experiencing rising innovation before the reform. Consistent with this, Figure B.9 plots the treatment effect at the start of treatment for each cohort and shows no relationship between treatment timing and early effects. If the ministry had strategically located programs in more promising areas, we would expect treatment effects to decline monotonically across later cohorts. The graphs with the full set of dynamic effects for the other specifications of table 2 are shown in Figure B.10 and exhibit similar patterns.

As the descriptive statistics in section 3 show that PhD graduates gradually entered the pool of inventors as more people graduated and suggest that the patenting in treated provinces increased over time, it is natural to ask if the effects of PhD programs built up over the years. Table 2 shows that effects 15 years after the first local PhD graduation exceed the average effect, indicating that the impact of doctoral programs grows as the graduate stock builds. Figure 3 confirms this as coefficients increase consistently over the years.

One potential concern is that the baseline results depend on the choice of the estimator. As shown in Table C.6, a standard TWFE specification yields estimates similar to the baseline. Aggregating patent counts at the province level produces comparable results, indicating that the findings are not driven by cells with zero patents in the province–technology panel. Excluding the imputed 1988 treatment cohort leaves the estimates virtually unchanged, suggesting that the effects are not driven by this specific group of provinces. Finally, the results are robust to using the synthetic DiD estimator of Arkhangelsky et al. (2021), which addresses potential bias from heterogeneous pre-treatment trends.

I further test the robustness of my results to potential violations of the parallel trends assumption in Figure B.11. I apply the method of Rambachan and Roth (2023) to the estimates in column (4) of Table 2. This approach uses deviations from parallel trends in the pre-treatment period to bound possible violations in the post-treatment period and to assess how large they would need to be to overturn the findings. The results indicate that the estimated effects are robust to linear deviations from parallel pre-trends.

As the dependent variable is the count of patents and the DiD specification relies on a linear functional form, one might worry that results are driven by large cities where patenting activity is mechanically higher. To address this, I perform three robustness checks. First, I rescale the dependent variable by local population in 1981 to reduce the weight of the largest cities. Second, I normalize by the number of local establishments in 1981. Third, I exclude the four largest cities in Italy: Milan, Naples, Rome, and Turin. The estimates in Table C.7 confirm that the results are not driven by these large urban areas.

A remaining concern is that the results might be affected by violations of the stable unit treatment value assumption (SUTVA) if PhD applicants or inventors migrated from control provinces to treated ones. Such movements would bias the estimates upward, as they would capture both the positive effects in treated provinces and the negative effects in control areas. Descriptive evidence in section 3 shows that control provinces experienced a small increase in patenting after the reform, suggesting that any SUTVA violations are more likely due to positive geographical spillovers, in line with Myers and Lanahan (2022) and Valero and Van Reenen (2019). Moreover, Table C.8 shows that only 2% of applicants to PhD programs in Rome and Milan were born in control provinces,

while locally born applicants constitute the majority. To assess the role of migration formally, I re-estimate the baseline excluding all treated provinces bordering control ones, under the assumption that migration costs are lowest for neighboring areas. Table C.7 shows that the results are nearly identical to the baseline, indicating that migration is unlikely to bias the estimates.

One limitation of the analysis is the absence of citation data from the Italian Patent Office (IPO), which prevents a direct assessment of whether PhD programs affect patent quality. To address this constraint, I construct a quality-adjusted measure of patenting using citations from foreign patent offices to the patent families of Italian patents. Specifically, I apply the fixed-effects methodology developed by Hall et al. (2001) to purge citation counts from systematic differences across technology classes and filing years. Figure B.13 shows that the increase in patenting following the introduction of PhD programs remains statistically and economically significant when patents are weighted by quality-adjusted citations. This suggests that the observed rise in patent counts is not driven solely by marginal or low-impact innovations.

A related concern is whether PhD programs shift the technological composition of innovation toward more or less novel domains. Appendix A.6 uses the average importance of IPC classes from US patents constructed by Kelly et al. (2021) as a proxy for technological novelty. Kelly et al. (2021) employ text analysis to assess how similar each patent is to prior and subsequent patents, defining importance as the ratio of forward to backward similarity. The idea is that breakthrough patents are more similar to subsequent innovations than to earlier ones, as they break from older technologies and lay the foundation for future advances. I use the average importance by technology to test whether the increase in patenting occurs disproportionately in more or less novel technologies. The rise in patenting is evenly distributed across the novelty spectrum, suggesting that PhD programs increase overall inventive activity without altering its quality composition.

The main results show that PhD programs significantly increase local patenting, particularly when the new programs are in STEM fields. However, since treatment is defined at the province level by the local opening of a PhD program, the estimated effects could reflect local dynamics correlated with the timing of these openings rather than the programs themselves. If this were the case, effects would not vary across fields and would be unrelated to the technological proximity between local PhD programs and the patenting technologies.

In the next section, I show that treated provinces experience an increase in academic publications precisely in the fields where PhD programs are established. Consistently, the rise in patenting is concentrated in technologies most exposed to these programs, those with a larger local stock of PhD graduates in closely related fields.

5 Effects Concentrate in Fields of Local PhD Programs

In this section, I show that the effects of PhD programs concentrate in the technological and academic areas that are closer to the subjects of local doctoral programs. In section 5.1, I show that PhD programs increased local academic output in terms of publications and citations in the precise subjects in which they were opened. In section 5.2, I show that the effects on patenting concentrate in the technologies that are more exposed to local PhD programs through a larger stock of PhD graduates in close fields.

5.1 Effects on Academic Output

As a starting point to assess the heterogeneity of PhD program effects across disciplines, I turn to academic output. Examining the impact on publications and citations serves two purposes. First, publications can be directly linked to the academic field of each PhD program, allowing me to redefine treatment at the province–field level. This enables a within-province comparison across fields and a test of whether the effects of PhD programs are concentrated in the subjects where they are introduced. Second, analyzing academic output provides a robustness check for the main mechanism: since PhD programs operate within universities, one would expect an immediate impact on academic research activity, with subsequent effects on innovation and patenting.

I construct a panel of academic output using data from the Scopus database, which covers journal articles published by researchers based in Italy between 1976 and 2001. Each publication is assigned to a province according to the authors' affiliations and to an academic field based on the journal classification. Aggregating these data at the province–field–year level allows me to examine how the introduction of PhD programs affected local research production across disciplines.

I use this dataset to redefine treatment as the opening of a PhD program in a given academic field and province, where treatment starts in the year of the first local graduation in that field. The outcome variable, $y_{p,f,t}$, is measured either as the count of publications or as the citation-weighted number of publications produced by researchers based in province p, published in field f, and year t. Formally, I estimate:

$$y_{p,f,t} = \alpha_{p,f} + \alpha_{f,t} + \alpha_{p,t} + \sum_{j=1986}^{2001} \sum_{i \neq -1} \beta_{i,j} \cdot \mathbb{1}\left\{t = g\left(p,f\right) + i,\ g\left(p,f\right) = j\right\} + \varepsilon_{p,s,t} \tag{4}$$

Where $g(p, f)$ denotes the year of the first PhD graduation in field f within province p. The specification includes province–field fixed effects $\alpha_{p,f}$ to control for time-invariant local academic specialization, field-year fixed effects $\alpha_{f,t}$ to absorb any field-specific shock over time, and province–year fixed effects $\alpha_{p,t}$ to capture any province-specific trends.

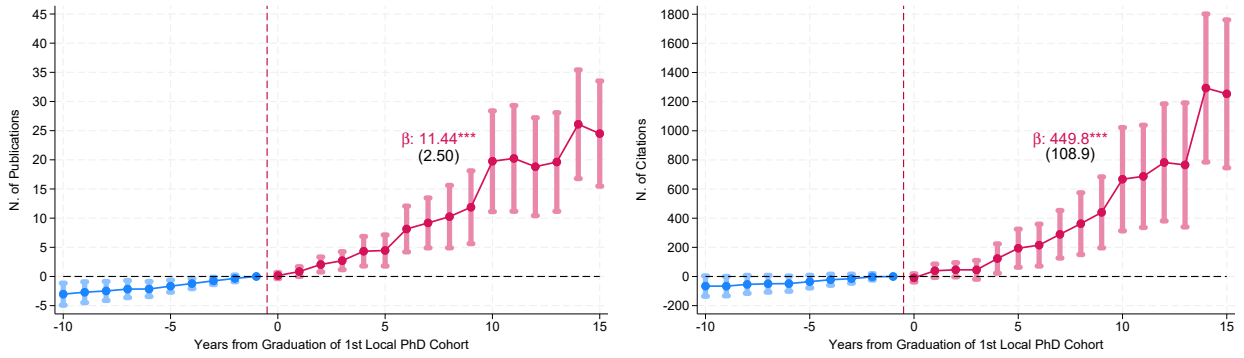
The refinement of the treatment and the inclusion of the province-year fixed effects allow me to compare fields within the same province over time, so that I can test if the effects of PhD programs are concentrated in the subjects in which they are introduced.

Figure 4 shows that the opening of PhD programs increases both the number of publications and their citation impact within the treated fields. To quantify the effects, the opening of PhD programs led to 76,682 more publications between 1986 and 2001 in treated departments[13], corresponding to about 1.5 publications per PhD position. These results indicate that PhD programs foster knowledge creation in their specific disciplines, suggesting that the content of doctoral programs may impact patenting activity as well.

The event-study coefficients suggest, however, the presence of significant pretrends, specifically for the number of publications. To test if these patterns undermine the causal interpretation of the results, I implement the method developed by Rambachan and Roth (2023). As shown in Appendix Figure B.12, the estimated effects remain robust across a wide range of deviations, confirming that the observed post-treatment increases are not driven by pre-existing trends.

[13]The overall number is obtained by multiplying the estimated average post-treatment effect on annual publications, $\hat{\beta}_{POST} = 11.44$, by the total number of treated province–field–year observations (6,703).

Figure 4: Effects of PhD Programs on Local Academic Output



(a) Publications (b) Citations

Notes: The figure shows the effects of PhD programs on local academic output. The coefficients shown come from estimations of Equation 4. Standard errors are clustered at the province-field level.

5.2 Effects Concentrate in Technologies with Similar Topics to Local PhDs

The previous section shows that academic output increases in the fields where a PhD program is opened, suggesting that the program's disciplinary focus may also shape its effects on patenting. Establishing this link, however, is far from straightforward. Patent classes are organized by technological function and often combine insights from several scientific areas. Conversely, PhD programs are defined by topics and methods, and advances in a single field may translate into applications across multiple technological domains.

I address this mapping issue using text analysis to empirically measure the proximity between PhD subjects and technological fields. Following Myers and Lanahan (2022) and Biasi and Ma (2022), who estimate pairwise text similarities between patents and grant descriptions and between university courses' syllabi and academic publications, respectively, I compute TF–IDF–based cosine similarities between the text of PhD theses and patent abstracts. I then average these scores within each technology–subject pair to obtain a measure of their knowledge proximity.

I compute the measure of knowledge proximity using short text descriptions from 2.1 million USPTO patents and abstracts from 1.3 million US PhD theses available in the *ProQuest Dissertations & Theses Global* database. Using US sources allows me to address potential endogeneity concerns that would arise from constructing the measure with Italian texts. If PhD graduates in

Italy produce more patents in specific technological areas as a result of their PhD training, this would mechanically increase the observed proximity between their PhD subject and those technologies. Consequently, the measure of knowledge proximity would become a function of the very effect of PhD programs on patenting, generating endogeneity. The identifying assumption behind the use of US texts is that the introduction of PhD programs in Italy did not affect the structural relationship between PhD subjects and technologies in the United States.

I use these texts to compute TF–IDF–weighted pairwise cosine similarities between all patents and PhD theses. Then, I calculate the knowledge proximity between any technology-subject couple as the average cosine similarity. PhD subjects are defined according to the 28 scientific areas established by the Italian Ministry of University, while technologies are identified by their IPC classes. Additional details on the construction of the measure are provided in Appendix A.7.

This measure captures the extent to which a PhD subject and a technology refer to similar topics, allowing me to rank all subject–technology pairs by their degree of proximity. For instance, *biochemistry–biology* and *musical instruments–art* are among the most similar pairs, while *hydraulics–philosophy* and *combustion engines–pedagogy* are among the least similar. Additional examples of high- and low-proximity pairs are reported in Appendix Table A.4.

I calculate a province–technology–year exposure index as a proximity-weighted sum of the stock of PhD graduates in different PhD subjects:

$$E_{p,c,t} = \sum_{f=1}^{28} N_{p,f,t} \times \sigma_{c,f} \tag{5}$$

Where $N_{p,f,t}$ is the stock of PhD graduates in academic field f from universities in province p in year t, and $\sigma_{c,f}$ is the knowledge proximity between IPC class c and PhD field f. Intuitively, $E_{p,c,t}$ can be thought of as the stock of potential inventors trained in technology c by year t due to the local opening of PhD programs in province p.

As a first test of the relevance of the exposure measure, I re-estimate the baseline specification by interacting the treatment variable with the full set of technology dummies. This approach allows me to estimate the average treatment effect of PhD programs separately for each technology. Appendix Figure B.14 plots these technology-specific effects against the corresponding average exposure index,

showing a positive correlation between the two.

To formally assess the relevance of exposure, I interact the treatment variable with indicators for the quartile of initial exposure:

$$\begin{aligned} patents_{p,c,t} &= \alpha_{p,c} + \alpha_{c,t} + \gamma \cdot \alpha_t \times \mathbf{X}_{p,1981} + \\ &+ \sum_{q=1}^{4} \sum_{j=1986}^{2001} \sum_{i \neq -1} \beta_{i,j,q} \cdot 1\left\{Q\left(E_{p,c,g(p)}\right) = q\right\} \times 1\left\{t = g\left(p\right) + i,\ g\left(p\right) = j\right\} + \varepsilon_{p,c,t} \end{aligned} \tag{6}$$

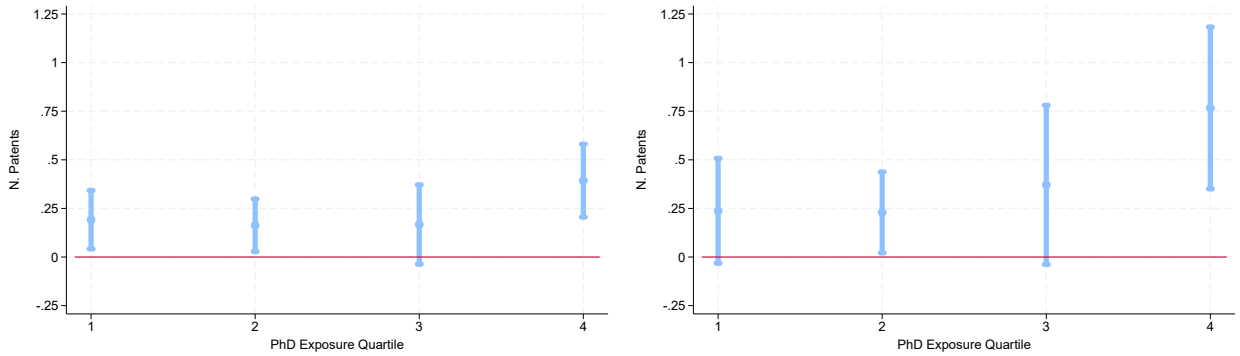
Here $Q\left(E_{p,c,g(p)}\right)$ denotes the exposure quartile that IPC class c in province p belongs to, based on the first local PhD cohort. Using only the first cohort avoids biases from potential adjustments in the allocation of PhD positions over time, specifically if we expect the ministry to assign more spots in the future to universities whose first cohort was successful in fostering innovation.

This regression can tell us whether the baseline effects are concentrated in the technology-province combinations more exposed to PhD programs, as suggested by Appendix Figure B.14 and the results on academic output.

Figure 5 presents the results, both on average and fifteen years after the first local PhD graduation. The effects are concentrated in the top exposure quartile, indicating that PhD programs stimulate innovation disproportionately in technologies most closely related to the subjects of local PhD programs. The difference between the average effects and those observed fifteen years after the first PhD graduations is consistent with the expected dynamics: as the stock of PhD graduates builds up over time and exposure diverges across province–technology cells, the impact of PhD programs on patenting becomes increasingly differentiated across exposure quartiles. This pattern suggests that both the intensive margin (number of graduates) and knowledge complementarity (field–technology proximity) are critical in linking doctoral education to local inventive activity.

A potential caveat in interpreting the exposure index $E_{p,c,t}$ is that it combines two components: a size factor given by the total number of local PhD graduates, and a composition factor, reflecting how closely these graduates' fields align with a given technology. To isolate the role of composition,

Figure 5: Effects of PhD Programs by Technology-Province Exposure



(a) Average Effects

(b) Effects 15 Yrs After

Notes: The figure shows the effects of PhD programs on patent activity, clustering technology-province cells into quartiles of initial exposure to local PhD programs. The plotted coefficients come from the estimation of Equation 6. Standard errors are clustered at the province level.

I estimate:

$$patents_{p,c,t} = \alpha_{p,c} + \alpha_{c,t} + \alpha_{p,t} + \beta \cdot \underbrace{\sum_{f=1}^{28} N_{p,f,t} \times \sigma_{c,f}}_{=E_{p,c,t}} + \varepsilon_{p,c,t} \tag{7}$$

Where province–year fixed effects $\alpha_{p,t}$ ensure comparisons are made across technologies within the same province and year. Thus, the coefficient β captures whether differences in the composition of the graduate stock translate into differential patenting outcomes, holding constant the overall local size of the stock.

I also use the number of publications and of citations generated by the PhD programs as predicted by Equation 4 instead of the stock of graduates to measure exposure. This tests whether the local increases in patenting are concentrated in technologies closer to academic academic scientific output. Although this tests a different transmission channel, both results point to a direct connection between the fields of the local doctoral programs and the technologies that experience increases in patenting.

Table 3 shows that the effects of PhD programs on patenting are concentrated in the technologies most exposed to local PhD fields. Column (1) reports OLS estimates using the stock of PhD graduates to measure exposure. The positive and significant coefficient indicates that technologies closer to the fields of local PhD programs experience an increase in patenting. To illustrate magni-

Table 3: Effects of Exposure to Local PhD Programs on Innovation

	(1)	(2)	(3)
Stock Used for Exposure:	Graduates	$\widehat{\text{Publications}}$	$\widehat{\text{Citations}}$
Effect of Exposure $\hat{\beta}$:	0.0184**	0.0172**	0.0157**
	(0.0081)	(0.0069)	(0.0064)
Control Variables:			
Province-Tech FE	X	X	X
Tech-Year FE	X	X	X
Province-Year FE	X	X	X
N. Clusters	58	58	58
Adj. R^2	0.8513	0.8511	0.8511
N. [Prov × Tech × Year]	160,160	160,160	160,160

Notes: OLS and TSLS estimates of Equation 7, one observation is a province-technology-year cell. The dependent variable is the number of patents produced in a province and technology in a year. Column (1) presents the estimates from an OLS regression. The coefficient has been rescaled so that it represents the effect of the average PhD position opened between 1983 and 1998. Columns (2) and (3) show the results from a TSLS estimation where the exposure have been calculated using the academic output predicted by the DiD model represented by Equation 4. In these columns, the normalization implies that the coefficients represent the average effect of opening a PhD program. Standard errors are clustered at the province level and reported in parentheses. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

tudes, replacing a Mathematics PhD position with one in Industrial Engineering in the first PhD cohort graduating in 1986 would generate roughly four additional local patents over the following sixteen years.[14]

Columns (2) and (3) instrument exposure with predicted publications and citations (from Equation 4), capturing a slightly different channel: whether patenting increases in technologies more exposed to fields that experience higher academic output.The results again show positive and significant effects, reinforcing the view that both the composition of PhD graduates and academic spillovers play a role in connecting doctoral programs to patent activity.

Overall, these results confirm that the subject composition of doctoral programs matters for innovation outcomes. PhD programs do not simply increase the volume of patenting, they shape where new inventions emerge, favoring technologies thematically closer to the academic fields of

[14]This value is obtained from the difference between the average normalized proximity score of Industrial Engineering (0.32) and Mathematics (0.19) across technologies, multiplied by the estimated coefficient $\hat{\beta}$ (0.0184), the sixteen-year period from 1986 to 2001, and the number of technologies considered (110): $(0.32 - 0.19) \times 0.0184 \times 16 \times 110 \approx 4$.

local doctoral training.

6 Mechanism

In the previous sections, I showed that PhD programs foster local innovation by increasing patenting in technological fields aligned with the subjects of local doctoral programs. However, the mechanisms driving this effect remain to be understood. In this section, I investigate *how* PhD programs influence innovation. Section 6.1 estimates the individual-level causal effect of PhD training on patent output using a novel dataset linking PhD admission exam performance to patenting output in a RDD. Section 6.2 then explores the indirect effects of PhD programs focusing on local firms' characteristics.

6.1 PhD Programs Increase Graduates' Patenting

Descriptive statistics show that PhD graduates are the most innovative individuals in the population. However, disentangling the causal effect of PhD training from selection bias is not straightforward. Admission to doctoral programs involves both a selective entry process on research abilities and self-selection of applicants with preferences for research-oriented activities. Since these traits are likely correlated with patenting outcomes, simple comparisons between PhD and non-PhD graduates would return biased estimates of the effect of doctoral training.

I isolate the causal effect of PhD programs through a RDD using the admission exam data. Since these data only include applicants to PhD programs, their use restricts the analysis to individuals with comparable preferences for research-oriented activities. Additionally, I can exploit the scores to the admission exams to select a subset of applicants around the thresholds to be admitted. This way, I only compare individuals with similar taste and abilities ex-ante, whose differences in patent output are only imputable to the PhD training.

To structure the RDD, I exploit the fact that the number of positions of each program was set by the ministry before the publication of the calls for application. As a first step, I normalize the realized exam scores at the program level to enhance comparability across programs. Then, I define

my running variable as the difference from the last useful position to enter each program. To give an example, for a program with 3 positions, I would calculate the running variable as the difference from the third positioned applicant. By design, this process of normalization creates a cutoff at zero for each PhD program. Figure B.15 shows that there is a smooth distribution of scores on both sides of the cutoff.

To estimate the effects on patent activity, I use two outcome variables: a binary indicator equal to one if an individual has ever produced a patent, and the total number of patents produced by each individual. The first captures the effect of completing a PhD program on the probability of becoming an inventor, while the second measures the impact on patent productivity. To ensure that outcomes reflect post-graduation innovation, I restrict the definition to patents filed at least three years after the start of the program. Since information on the exact graduation year is not consistently available, I use the legal minimum duration of PhD programs of three years as a proxy for the completion date.

However, admission exam scores cannot be used to estimate the effect of obtaining a PhD directly because of noncompliance: some applicants above the admission cutoff do not enroll or enroll elsewhere, and some enrollees do not graduate. In a similar way, it could be the case that individuals who enrolled do not graduate from the PhD program. Appendix Figure B.16 shows the scope of noncompliance by plotting the share of PhD graduates by distance from the cutoff. To account for noncompliance, I employ a fuzzy RDD in which the outcome of the first stage will be whether the individual has completed a PhD program in any Italian university between 1986 and 2001. I implement it using the robust RDD estimator developed by Calonico et al. (2014), which uses local polynomials with kernel weighting and bias correction to consistently estimate the effect at the cutoff.

I estimate the individual returns to completing a PhD program in terms of patent activity using a fuzzy RDD. Formally, the specification is:

$$\left.\begin{aligned} \text{First Stage:} \quad & g_i = \alpha_1 + \delta \cdot 1\{SCORE_i \geq 0\} + X_i'\theta_g + \varepsilon_{1,i}, \\ \text{Reduced Form:} \quad & y_i = \alpha_2 + \psi \cdot 1\{SCORE_i \geq 0\} + X_i'\theta_y + \varepsilon_{2,i}, \end{aligned}\right\} \tau = \frac{\psi}{\delta} \tag{8}$$

Where g_i is a binary indicator equal to 1 if individual i obtained a PhD in Italy between 1986 and 2001, and y_i is an innovation-related outcome such as being listed as an inventor or the number of patents filed. The parameter δ captures the effect of crossing the cutoff on PhD completion, while ψ captures the effect of crossing the cutoff on innovation outcomes. Then, the Local Average Treatment Effect (LATE) of completing a PhD program on the innovativeness of marginal applicants can be derived as the Wald ratio $\tau = \frac{\psi}{\delta}$. X_i is a set of control variables including gender, university, and field fixed effects.

Since the raw scores underlying the running variable are discrete and patenting is an extremely rare outcome, the standard continuity-based RDD framework is not well suited to this setting. For this reason, I use a local randomization framework which states that, within a sufficiently narrow window around the cutoff, crossing the threshold is assumed to generate as good as random variation in PhD completion. Identification therefore does not rely on continuity of potential outcomes in the running variable, but on exchangeability of applicants just above and below the cutoff.

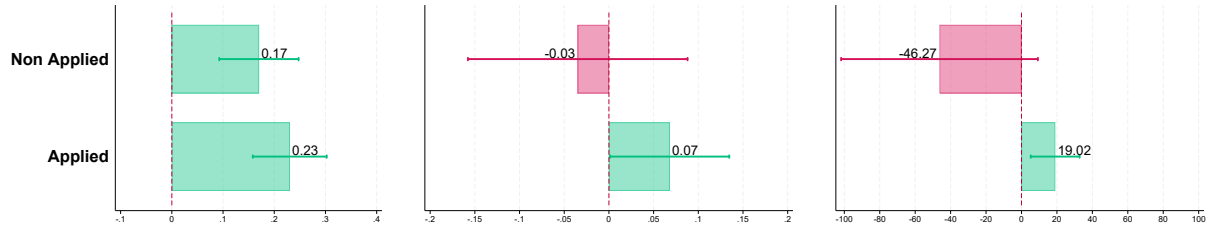
Not all PhD programs are equally close to the technology frontier. Some fields are intrinsically more applied and likely to generate innovations with patenting potential, while others are more theoretical. To capture this heterogeneity, I use the 1982 ISTAT Survey on Scientific Research, which reports total public R&D expenditure by topic and broken down into three categories following the OECD Frascati Manual: theoretical, applied, or development research[15]. I classify a STEM subject as applied when theoretical research does not receive the relative majority of research funding. Because these data include both universities and public research institutes, they closely reflect the research environment in which PhD students were trained. Pooling all fields together would instead yield an average effect across programs, obscuring potentially large differences in the extent to which research training translates into patentable innovation.

Finally, since some candidates applied to multiple PhD programs, I weight each observation by the inverse of the number of applications made by an individual.

Figure 6 reports the RDD estimates, showing both the first-stage effects and the corresponding Wald ratios. The first stage indicates that the probability of completing a PhD increases by about

[15]Figures are presented in separate tables for the three categories and research topics are more specific than mere fields, e.g. *research on cancer* rather than medicine or *nuclear fission* rather than physics and industrial engineering

Figure 6: Individual-Level Effects of Completing a PhD Program



Notes: This figure presents the results of the RDD analysis, separately by category of the program: applied versus non-applied. The optimal bandwidths were calculated using the *rdbwselect* package and the estimation was made using *rdrobust* (Calonico et al., 2017). The plotted coefficients are bias-corrected and the reported confidence intervals are calculated using robust bias-corrected standard errors clustered at the individual level.

20 percentage points across fields, an approximately 40% rise relative to baseline. In contrast, the effects on patenting outcomes are markedly heterogeneous. Graduates from applied STEM PhD programs exhibit a significant increase in both the propensity to patent and the total number of patents produced, whereas no significant effects emerge among non-applied STEM programs. This heterogeneity mirrors the underlying nature of research across fields: applied programs emphasize problem-oriented research that directly translates into patentable knowledge, while non-applied programs focus on basic research whose contributions to technological progress are more indirect.

This differential effect of PhD programs based on their content also informs us on the channel through which doctoral programs affect graduates' inventiveness. Indeed, we could think of PhD programs as a bundled treatment combining increased time for research with the accumulation of research-specific human capital, through mentorship from the supervisor and exposure to specialized research seminars. The differential effects observed across STEM programs indicate that time alone cannot account for the results, since all programs extend research time. Consistent with a human capital accumulation channel and a lasting impact of PhD training on graduates' research abilities, Figure B.17 shows that the estimated effects remain positive a decade after the end of the program.

To quantify the causal effect of PhD programs on graduates, I assume that the LATE estimated in the RDD can be extrapolated to the entire population of PhD graduates. Economically, this corresponds to assuming homogeneous treatment effects across individuals, regardless of their ex-ante ability. Because data on applicants' abilities beyond their admission test scores are unavailable,

it is not possible to estimate heterogeneous effects or extrapolate the treatment effect away from the cutoff.

By multiplying the effect of applied STEM programs with the corresponding number of PhD positions, I estimate that PhD training causally contributed to approximately 2,953 additional patents, about 22% of the total increase in patenting found in the baseline specification. Detailed estimates are reported in Appendix Table C.9.

As a robustness check, I test for potential manipulation of exam scores using the test proposed by McCrary (2008). The results, displayed in Figure B.18, show no evidence of manipulation in any of the categories considered. Table C.10 reports covariate balance tests around the cutoff. While certain characteristics show imbalances, these are accounted for in the RDD specification through the inclusion of fixed effects, which absorb average differences in innovativeness across groups. Furthermore, since the pattern of imbalance is similar across the two PhD categories whereas the treatment effects differ markedly, this reinforces the interpretation that the estimated effects are not driven by compositional differences among applicants near the cutoff.

6.2 Spillovers to Local R&D Firms through Social Links

The RDD results establish that PhD programs directly foster innovation by increasing graduates' innovativeness. Yet, these direct effects do not fully explain the overall increase in patenting induced by the introduction of PhD programs. In this section, I focus on patents by non-PhD inventors to assess the spillovers of PhD programs on the local innovation environment. Specifically, I exploit information on patents' corporate assignees and firms' locations to estimate the extent to which local firms benefited from PhD programs even without employing doctoral graduates among their inventors.

To estimate these spillovers, I construct a measure of local patenting by non-PhD corporate inventors. The procedure involves three steps. First, I estimate for each patent the probability that it does not include a PhD inventor. I train a probabilistic logit model on the manually linked dataset of applicants used in the RDD analysis. The model predicts the likelihood that a name match between an inventor and a PhD graduate is correct, using information on the patent's

technology, the PhD field, the time distance between program start and patent filing, and the commonness of the name in Italy. I then apply the estimated probabilities to all inventor–graduate matches and compute for each patent the probability of having no PhD inventor as the product of the probabilities of *incorrect* matches. Appendix A.4 provides further details on the probabilistic model and final computation.

Second, I extend the coverage of firms' tax codes to the pre-1990 period using a matching algorithm. I exploit inconsistencies in the reporting of assignees' names to train a fuzzy-matching model, which uses the unique tax code available from 1990 onward to link all variants of a firm's name. I then apply the trained model to connect firms observed before 1990 to those appearing in later years. Appendix A.8 provides additional details on the algorithm.

Third, I use the tax codes to match firms with their headquarters' location and industry[16] from the CERVED dataset, a detailed dataset including financial information on the universe of Italian limited liability firms.

After these three steps, I exclude all patents not matched to firms in CERVED. This restriction ensures that the analysis focuses on patents owned by limited liability firms, rather than universities, research centers, or public institutions, which may have benefited more directly from PhD programs. I then aggregate the data at the province–year–industry and province–year–technology levels. Within each cell, I sum the patent-specific probabilities of having no PhD inventors, which provides the expected number of corporate patents attributable to non-PhD inventors.

To quantify spillovers, I re-estimate Equation 1 with the expected number of corporate patents not attributable to PhD inventors as the dependent variable. In this specification, the coefficients measure the impact of PhD programs on local firms' innovative output, net of doctoral graduates' direct contributions. Since the outcome is a count variable aggregated at the same level as in the baseline, the results are directly comparable.

Table 4 shows that the establishment of doctoral programs led to a positive effect on patenting by local non-PhD inventors. Column (1) documents a significant increase in patenting from non-PhD inventors overall, while columns (2) and (3) restrict the sample to patents assigned to limited

[16]Industry is defined at the 2-digit NACE level

Table 4: Effects of PhD Programs on Local Non-PhD Corporate Innovation

	(1)	(2)	(3)
Dependent Var.	Non-PhD Patents	Non-PhD Corporate Patents	
Effects of PhD Programs:			
$\hat{\beta}_{POST}$	0.2060***	0.1376**	0.1292**
	(0.0662)	(0.0621)	(0.0650)
Mean of DV, Treat=1	0.7193	0.1104	0.1162
Control Variables:			
Province-Tech FE	X	X	X
Tech-Year FE	X	X	X
Full Controls	X	X	X
N. Clusters	58	58	58
Aggregation	Technology	Technology	Industry
Obs.	177,944	177,944	128,180

Notes: OLS estimates of Equation 1. One observation is a province-technology-year cell in columns (1) and (2), a province-industry-year cell in column (3). The dependent variable is the number of patents produced in a province and technology in a year by non-PhD inventors in column (1). In column (2), the set of patents is restricted to those owned by Cerved firms. In column (3), the depndent variable is the number of patents produced by non-PhD inventors in a province in a year owned by firms in an industry. All columns control for local characteristics in 1981 interacted with year fixed effects: industrial composition using the 17 one-digit 1991 ATECO industries, local population, share of workers employed in firms with 500 or more employees, share of workers in IT, share of workers in STEM R&D, share of workers in non-STEM R&D, and share of university graduates among people aged between 25 and 34. Standard errors are clustered at the province level and reported in parentheses. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

liability firms. The results indicate that PhD programs generated substantial indirect effects on innovation through local inventors and firms. Using column (2) for comparability with the baseline specification, I estimate that the increase in corporate patents between 1986 and 2001 amounts to approximately 7,395, corresponding to about 54% of the total effect.

It is important to note that spillovers are defined with respect to local firms. This restriction stems from the challenge of identifying geographic spillovers beyond the areas where PhD programs are established. Nevertheless, as knowledge can diffuse across provinces, positive spillovers likely extend beyond these boundaries[17]. In this case, my estimates would represent a lower bound of the true aggregate spillovers generated by PhD programs.

Although the estimated effects are generally positive and significant, they may mask important

[17]Consistent with Myers and Lanahan (2022), who find positive effects also outside the counties of grant recipients.

heterogeneity across industry–province cells. Section 5 showed that the subject of local PhD programs affect their impact on innovation. This suggests that similar heterogeneity may arise across industries, as different PhD fields are more closely related with certain sectors. Moreover, industries vary substantially in their research input intensity, which I refer to as research intensity. I therefore explore heterogeneity in effects along two dimensions: industries' research intensity and their differential exposure to local PhD programs.

Identifying the channel through which PhD programs affect researchers in private firms is challenging. Recent literature (Beaman et al., 2021; Prato, 2024) has highlighted the role of social connections in knowledge diffusion. However, since PhD students were not allowed to teach, my results cannot stem from vertical knowledge transmission from PhD graduates to university students. Given that most of the increase in patenting derives from patents without PhD inventors, the mechanism is unlikely to involve direct coworker interactions. Consistent with this, official ISTAT data[18] indicate that only 1,306 PhD graduates were employed in R&D within private firms in 2005, out of about 89,000 PhD graduates[19] in Italy. Thus, if the mechanism operates through social connections between PhD graduates and local workers that predate PhD enrollment.

I construct a measure of industry-level exposure to local PhD programs as the *expected social links* between workers and PhD graduates. As a first step, I use the ISTAT Surveys on Graduates' Employment (1989–2004), which report the first job placement of university graduates from 1989 to 2001. Using these data, I reconstruct the national flows of graduates from each university subject to the different industries.

Next, I collect information on the number of final-year university students by subject and university in the year preceding the first local opening of a PhD program, which I use as a proxy for the number of graduates. These data come from the ISTAT Statistical Yearbook on Education, which provide annual information on the Italian education system.

Finally, I measure the *expected social links* between PhD programs and local industries as the number of potential peer connections between university graduates who entered each industry and their former classmates who pursued a PhD locally. For each industry–province pair, this measure

[18] ISTAT Historical Series on R&D are available at https://seriestoriche.istat.it.

[19] 89,056 PhD positions were opened between 1983 and 2002, implying graduation by 2005.

is defined as the sum over university fields of the expected number of local graduates entering the industry multiplied by the probability that a peer in the same field enrolled in a local PhD program. Formally:

$$Links_{p,s} = \sum_{f=1}^{28} \frac{P_{p,f,o(p)}}{G_{p,f,o(p)}} \times \left(G_{p,f,o(p)-P_{p,f,o(p)}} \right) \times Share_{s,f} \tag{9}$$

Where $\frac{P_{p,f,o(p)}}{G_{p,f,o(p)}}$ denotes the ratio between the number of PhD positions $P_{p,f,o(p)}$ opened in province p and field f in the first local cohort $o\,(p)$ and the number of local university graduates $G_{p,f,o(p)}$ in the same field and year. This ratio approximates the share of local graduates in field f who start a PhD. The term $G_{p,f,o(p)} - P_{p,f,o(p)}$ represents the expected number of graduates entering the labor market, which, when multiplied b $Share_{s,f}$, the national share of graduates in field f whose first job after graduation is in industry s, yields the expected number of graduates in field f working in industry s. Multiplying these two components gives a proxy for the number of social links that workers in industry s and province p have with local PhD students in field f through their university studies. Summing across fields returns the total number of *expected social links* between industry s and local PhD programs in province p.

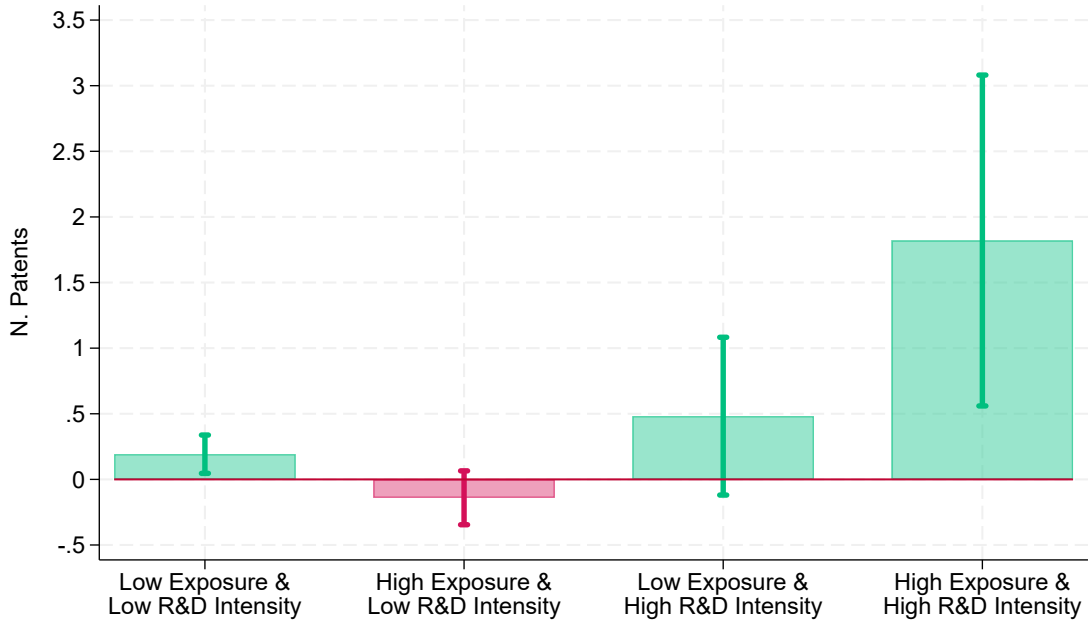
The number of graduates in the year preceding the opening of the first local PhD program and the number of PhD positions offered in the first cohort ensure that the measure does not incorporate adjustments driven by the performance of local doctoral programs. Since data from pre-reform years are unavailable, the flows from university to employment are computed as averages between 1989 and 2001.

I define R&D intensity as the share of the workforce employed in R&D within each industry, based on the 1982 ISTAT survey on scientific research and the 1981 ISTAT industry census.

Using R&D intensity and the measure of *expected social links*, I classify province–industry cells into four groups according to whether their R&D workforce share is in the top decile and whether their connections to PhD programs are above or below the median. Given the highly skewed distribution of R&D effort, discretizing based on the top decile rather than the median allows meaningful heterogeneity across industries to be captured.

This classification directly relates to the mechanism of social links between corporate inventors and PhD students. If spillovers operate through social interaction between PhD students and their

Figure 7: Spillovers by Industries' R&D Intensity and Social Links to PhD Programs



Notes: The figure shows the heterogeneous spillover effects of PhD programs on local industries by their R&D intensity and *expected social links* to local PhD programs. Industries are divided into four groups based on whether they are or not in the top decile of R&D intensity, defined as the share of workforce employed in R&D, and above or below the median of *expected social links* to local PhD programs. Standard errors are clustered at the province level.

former peers, both the extent of social links and R&D intensity should matter. When an industry has few links to PhD students, knowledge transmission is limited. Conversely, when many links exist but firms' R&D intensity is low, workers have no incentives to acquire the research-specific human capital developed by PhD students. Only when both conditions are met, high connectivity and high R&D intensity, do workers have both the opportunity and the incentive to learn from PhD students.

I test this mechanism by estimating heterogeneous effects of PhD programs on corporate non-PhD patenting, interacting the treatment variable from column (4) of Table 4 with indicators for the four province–industry groups described above. Figure 7 shows that spillover effects are highly concentrated among firms that are both R&D-intensive and more exposed to local PhD graduates. This pattern supports the idea that workers require both connections and incentives to acquire human capital from PhD graduates.

Because the measure of social links is constructed based on the local openings of PhD programs,

a potential concern is that the estimated effects may simply reflect local dynamics. To address this, Appendix Figure B.19 redefines the treatment as an indicator for belonging to the top decile of R&D intensity and above the median of expected social links, controlling for province–year fixed effects. The results confirm that the effects are not driven by province-specific dynamics.

7 Cost-Benefit Analysis

Having estimated the causal effects of PhD programs on patenting, an important question remains: how do these programs compare to other innovation policies in terms of cost-effectiveness? To answer this, I quantify their costs and benefits and construct several measures of the benefit–cost ratio. Section 7.1 estimates cost-effectiveness as the number of patents generated per dollar spent, while Section 7.2 translates the induced increase in patenting into monetary terms using microdata and presents a range of benefit–cost ratios. Section 7.3 concludes by calculating the Marginal Value of Public Funds (Hendren, 2016) of PhD programs and comparing them to a broader set of public policies.

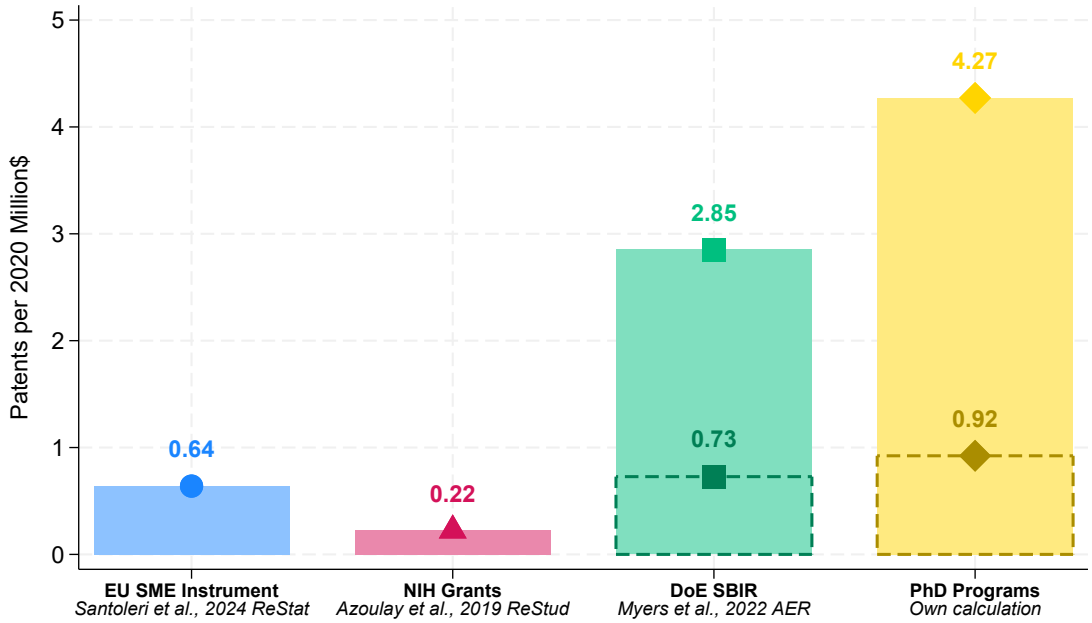
7.1 Cost-Effectiveness of PhD Programs

As a first step, I calculate the total costs of PhD programs. I estimate the accounting cost of PhD programs by adding the scholarship amount to the per-capita university expenditure on research capital and the per-capita administrative cost of PhD programs.

I obtain the scholarship amount from a 1998 law which increased the scholarship amount for the following three years. As I could not find any specific amount for previous years but only the establishment of PhD scholarships in the law creating PhD programs in 1980, I extend the nominal amount for the year 1998 to all previous years. This is most likely an overestimate of the true amount of scholarships in earlier years, as Italy exhibited average annual inflation of 6% between 1983 and 1998 and the amount might have increased over the years to compensate this.

Per-capita university expenditure on research capital comes from the ISTAT time series web archive, which registers yearly university research expenditure by type of expense and size of the

Figure 8: Cost-Effectiveness of Different Innovation Policies



Notes: The figure shows the ratio between the policy-induced increase in patent production and the cost of a set of policies studied in the literature. The first policy is the European Union Small and Medium Enterprise (EU SME) Instrument (Santoleri et al., 2024), the second are the US National Institute of Health (NIH) grants (Azoulay et al., 2018), and the third is the US Department of Energy program called Small Business Innovation and Research (SBIR) (Myers and Lanahan, 2022). The fourth policy is the establishment of PhD programs in Italy analyzed in this paper. For the first two policies only the effect on grant recipients is provided. For the remaining two the dashed area represents the effects on recipients or graduates, the full bars are based on the total effects. All values are reported as patents per 2020 million dollar spent on the policy.

research personnel. Finally, I calculate per-capita administrative expenditure on PhD programs from the balance sheets of La Sapienza University of Rome, which I then extend to all other universities. As La Sapienza university offered about 12% of all PhD programs between 1983 and 1998, I take it as a fair representative of the actual average administrative cost incurred by Italian universities for the PhD programs.

To provide a consistent definition of the real value of these costs, I convert all values into 2020 dollars using the official exchange rates provided by the Bank of Italy and the CPI from the Saint Louis Federal Reserve.

The resulting measure of patents per million dollar spent is in Figure 8, where PhD programs are compared to other policies studied in the innovation literature. For the Italian PhD programs and the US Department of Energy Small Business Innovation and Research (DoE SBIR) program,

I show both the direct effects with a dashed line, and the total effects with a solid bar. As the other studies only estimate the effects for grant recipients, the inclusion of the direct effects allows me to compare cost-effectiveness across policies.

The figure shows that PhD programs led to an increase in patenting of 0.92 for each 2020 million dollars spent, when I only consider patents made by PhD graduates. This result puts PhD programs on the same order of magnitude as other innovation policies but on the upper end of the cost-effectiveness scale. The cost-effectiveness considering total effects is 4.27, also higher but in the same scale as the one estimated for the DoE SBIR program.

One issue of these measures is that they do not take into account the value of the patents produced thanks to these policies. In the next section, I use microdata to estimate the monetary value of patents and use it to estimate a benefit-cost ratio of PhD programs as innovation policies.

7.2 Benefits from PhD Programs Exceed Costs

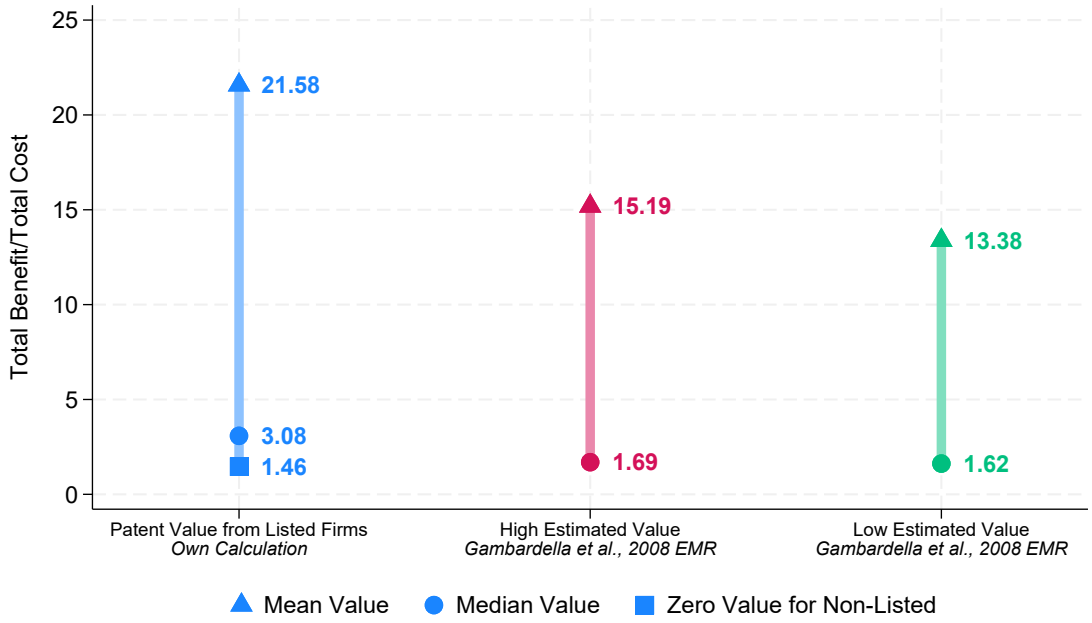
I estimate the monetary value of PhD-induced patent activity using the methodology of Kogan et al. (2017). This method estimates the value of patents as the excess return of the assignee's stocks in the days following the announcement of a patent.

To apply this methodology to Italy, I match listed firms from the CERVED database to the historical stock market data of LSEG Datastream and to their patents. Then, I use the COMIT stock market index for the Milan stock exchange to calculate the stocks' excess returns and estimate patents' value. This way, I calculate the value of 3,718 patents published in treated provinces after the first local STEM PhD graduation, corresponding to 6% of the respective total. Appendix A.9 explains in more detail how I calculate the patent value and shows some descriptive statistics.

I use these patent values to calculate the approximate average value of patents for each IPC class. As a first step, I estimate the mean patent value within each IPC across patents from listed firm and apply it to all patents of the same technology. Although this measure is based on patent-level estimates, it is likely to be an over-estimation of the true average value of patents, as patents from listed firms are likely to be of higher value than the average patent.

To account for this potential over-estimation, I compute two alternative measures. For the first

Figure 9: Benefit-Cost Ratio of PhD Programs



Notes: The figure shows the ratio between the monetary value of the PhD-induced increase in patenting between 1986 and 2001 and the total cost of PhD programs. The results shown on the left blue line are based on value of patents from listed firms, calculated using the method of Kogan et al. (2017). The results on the middle red line and the right green one are calculated starting from estimates of patent values provided by Gambardella et al. (2008). The estimates represented by triangles are calculated starting from the mean patent values, those represented by circles from medians. The square in the left blue line represents is a limit case estimate in which the value of patents from non-listed firms is assumed to be zero.

one, I calculate the median patent value across listed firms' patents. This measure has the benefit of not being directly affected by the high skewness of the distribution of patent values. The second measure uses the average value of patents from listed firms' and multiplies it by the share of listed firms' patents in the specific technology. This provides me with an estimate of the average value of patents in a limit case in which the value of patents from non-listed firms is assumed to be zero.

The last measure likely underestimates the actual monetary value of the innovation induced by PhD programs, providing a conservative estimate. Indeed, if we multiply the total number of patents produced in Italy between 1986 and 2001 by the average value estimated in this way, and compare it to the total R&D expenditure of Italian firms from 1983 to 1998[20], we obtain a negative return on investment of -84%.

Finally, I calculate the total monetary value of PhD-induced innovation by multiplying each

[20]Provided by ISTAT's Historical Series on R&D

of these measures with the class-specific average effect of PhD programs on patent production, summing across technologies. Assuming that the estimated variation in patent flow induced by PhD programs in treated provinces represents the total change in patents resulting from the introduction of PhD programs, this calculation corresponds to their aggregate effect on the value of the stock of Italian patents.

To make the cost estimate more comprehensive, I add graduates' cost-opportunity of enrolling into a PhD to the accounting cost of the programs. I calculate the field-specific opportunity cost as the average salary earned by university graduates in their first occupation after graduation and I multiply by the average length of PhD programs and the number of PhD positions opened in the field.[21]

Finally, I define the benefit-cost ratio dividing the total value of PhD-induced patenting by the total cost of PhD programs. Figure 9 shows the value of the ratio for the three different estimates of the average value of patents. To provide a benchmark for my estimation, I add to the graph two sets of values calculated starting from the data of Gambardella et al. (2008). Their research provides estimates of the median and mean values of European patents published between 1993 and 1997.

The graph shows that the benefits from the introduction of PhD programs exceed their cost, with a range of the benefit-cost ratio between 3.08 and 21.58 under standard assumptions. It is noteworthy that, even under the extreme assumption that only listed firms produce valuable innovation, benefits still exceed costs with a ratio equal to 1.46.

Finally, the overlap between the estimates from the microdata and the ones calculated starting from Gambardella et al. (2008) suggest that the range of values for the cost-benefit ratio is line with previous estimations of the value of European innovation.

Using these ratios as back-of-the-envelope multipliers of the public expenditure in PhD programs suggest that the establishment of doctoral programs led to an increase in GDP in the range of 0.6% and 4.7% between 1986 and 2001. These correspond to 1.6% and 11.5% of Italian GDP growth

[21] Additionally, it is important to note that the measure of patent value used in this analysis is based on the reaction of financial markets to the patents, meaning the benefit measure already accounts for the average expected return from the R&D expenditure of innovating firms.

during this period, respectively.

7.3 PhD Programs are High-MVPF Policies

The previous sections show that PhD programs sit at the top end in terms of impact on innovation among the innovation policies analyzed in the literature, and that they provide more innovation-bound benefits than they cost. In this section, I calculate the MVPF of PhD programs and use it to compare them to a broader set of public policies.

To measure the benefits, I subtract the foregone wages of PhD graduates (due to their enrollment in the programs) from the monetary value of PhD-induced innovation, while adding the scholarships received by graduates. This measure acts as a proxy for the willingness to pay of both the graduates and firms benefiting from the innovation.

On the cost side, I sum the accounting cost of PhD programs with the foregone labor taxes from graduates' delayed entry into the labor market, and subtract the increase in corporate taxes resulting from the new innovations driven by the PhD programs. I calculate the foregone labor taxes by multiplying graduates' foregone labor income times the tax schedules, assuming those jobs would constitute the sole employment of PhD graduates. I calculate the increase in corporate profits' taxes by multiplying the PhD-induced variation in the value of the stock of patents times the corporate income tax. All sources for the tax schedules are listed in Table C.5.

These definitions of benefits and costs are used to calculate the MVPF of PhD programs under two key assumptions: i) PhD programs do not lead to any increase in future income or productivity for graduates, they earn exactly as much as they would have in the counterfactual scenario without the PhD program; and ii) the costs and benefits scale linearly with the number of positions, allowing us to use the average benefit in the MVPF calculation.

The MVPF is calculated with the following formula comparing net benefits and costs:

$$\text{MVPF} = \frac{\Delta_{PhD}\ Patent\ Stock\ Value - (Foregone\ Labor\ Income\ -\ Scholarships)}{PhD\ Accounting\ Cost\ +\ Foregone\ Labor\ Tax\ - Innovation\ Profit\ Tax} \tag{10}$$

The resulting MVPF values range from 8.4, under the conservative assumption that only listed

firms generate valuable innovation, to infinity in the more generous cases. These estimates are consistent with those reported in the literature for adult college policies (the Policy Impacts website reports a median MVPF of 8.35 for adult college education Hendren and Sprung-Keyser 2020)[22] and for education policies more broadly, as shown in Appendix Figure B.20. These results reinforce the conclusion that PhD programs generate benefits exceeding their costs, and suggest that under broader definitions of benefits, such programs may even pay for themselves.

8 Conclusions

Even though PhD graduates are over-represented among inventors, the existing literature does not provide an estimate of the causal effects of PhD programs on innovation, largely due to data limitations and the absence of a clean setting allowing for causal inference. This paper addresses this gap by exploiting the 1983 introduction of PhD programs in Italy and compiling two novel datasets linking the establishment of programs to local patenting and PhD applicants to their subsequent patent output.

The results show that PhD programs raise local patenting through multiple channels. On the individual side, PhD programs increase graduates' propensity to patent, challenging the notion that doctoral education merely selects more innovative individuals. Yet, most of the aggregate impact on innovation arises from spillovers to local firms. These findings indicate that PhD programs are an effective instrument to stimulate local innovation and that their effects extend beyond the training of doctoral graduates themselves, as they diffuse knowledge, research methods, and best practices to the private sector. Finally, my estimates on the patent-related benefits of PhD programs show that, even under conservative assumptions, the monetary value of innovation generated by PhD programs exceeds their total cost.

These results have important policy implications. While some governments invest billions of dollars each year in doctoral education (Italian Government, 2021; European Commission, 2022), others have reduced university funding, leading to cuts in PhD positions (Nature, 2025). This paper provides new causal evidence on the potential impact of these policies on innovation, shedding light

[22] https://policyimpacts.org

on the channels through which doctoral education shapes technological progress. Indeed, while the funding itself directly targets PhD programs, its impact extends beyond them, encompassing the broader local innovation ecosystem.

While this paper focuses on innovation outcomes, future work could investigate the broader economic consequences of PhD programs. Promising directions include examining their effects on occupational sorting, wage trajectories, and entrepreneurship. Another avenue concerns the optimal geographical allocation of doctoral programs. Since the analysis suggests that spillover effects depend on local social linkages and R&D investments, an open question is how pre-existing industrial specialization shapes the local effects of doctoral programs.

This paper shows that PhD programs are an engine of growth by increasing innovation both directly and indirectly, highlighting their role as a central component of modern innovation policy. More broadly, it provides a foundation for a research agenda on how investments in advanced education can promote technological progress and ensure that human capital formation translates into sustained innovation and growth.

References

Aghion, P., Akcigit, U., Hyytinen, A., and Toivanen, O. (2017). The social origins of inventors. Working Paper 24110, National Bureau of Economic Research.

Akcigit, U., Grigsby, J., Nicholas, T., and Stantcheva, S. (2021). Taxation and Innovation in the Twentieth Century. *The Quarterly Journal of Economics*, 137(1):329–385.

Akcigit, U., Pearce, J., and Prato, M. (2024). Tapping into Talent: Coupling Education and Innovation Policies for Economic Growth. *The Review of Economic Studies*, page rdae047.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? evidence from the establishment of us colleges. *American Economic Journal: Economic Policy*, 15(2):1–41.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Athey, S. and Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79. Annals Issue in Honor of Gary Chamberlain.

Azoulay, P., Graff Zivin, J. S., Li, D., and Sampat, B. N. (2018). Public R amp;D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies*, 86(1):117–152.

Babina, T., He, A. X., Howell, S. T., Perlman, E. R., and Staudt, J. (2023). Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications. *The Quarterly Journal of Economics*, 138(2):895–954.

Bastasin, C. and Toniolo, G. (2023). *The Rise and Fall of the Italian Economy*. Cambridge University Press.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, 111(6):1918–43.

Bergeaud, A., Guillouzouic, A., and Malgouyres, C. (2022). From Public Labs to Private Firms: Magnitude and Channels of R&D Spillovers. CEPR Discussion Papers 17487, C.E.P.R. Discussion Papers.

Bianchi, N. (2020). The indirect effects of educational expansions: Evidence from a large enrollment increase in university majors. *Journal of Labor Economics*, 38(3):767–804.

Bianchi, N. and Giorcelli, M. (2019). Scientific Education and Innovation: From Technical Diplomas to University Stem Degrees. *Journal of the European Economic Association*, 18(5):2608–2646.

Biasi, B. and Ma, S. (2022). The education-innovation gap. Working Paper 29853, National Bureau of Economic Research.

Bloom, N., Van Reenen, J., and Williams, H. (2019). A toolkit of policies to promote innovation. *Journal of Economic Perspectives*, 33(3):163–84.

Borusyak, K., Jaravel, X., and Spiess, J. (2023). Revisiting event study designs: Robust and efficient estimation.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M., and Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics*, 50:99–114.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables*. *The Quarterly Journal of Economics*, 117(4):1491–1527.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T., and Van Reenen, J. (2023). Do tax incentives increase firm innovation? an rd design for rd, patents, and spillovers. *American Economic Journal: Economic Policy*, 15(4):486–521.

Duryea, S., Ribas, R. P., Sampaio, B., Sampaio, G. R., and Trevisan, G. (2023). Who benefits from tuition-free, top-quality universities? evidence from brazil. *Economics of Education Review*, 95:102423.

European Commission (2022). Horizon europe work programme 2023-2024, 2. marie skłodowska-curie actions.

Fazio, D. (1980). L'Università Italiana: Situazione e Problemi. *Universitas: Studi e Documentazione di Vita Universitaria*, 1:10–46.

Fieldhouse, A. J. and Mertens, K. (2024). The returns to government r&d: Evidence from u.s. appropriations shocks. Working Paper 2305, Federal Reserve Bank of Dallas, Research Department.

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

Ge, S., Isaac, E., and Miller, A. (2022). Elite schools and opting in: Effects of college selectivity on career and family outcomes. *Journal of Labor Economics*, 40(S1):S383–S427.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277. Themed Issue: Treatment Effect 1.

Gross, D. P. and Sampat, B. N. (2023). America, jump-started: World war ii rd and the takeoff of the us innovation system. *American Economic Review*, 113(12):3323–56.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools. Working Paper 8498, National Bureau of Economic Research.

Hausman, N. (2022). University Innovation and Local Economic Growth. *The Review of Economics and Statistics*, 104(4):718–735.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, 30(1):51–89.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies*. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Howell, S. T. (2017). Financing innovation: Evidence from rd grants. *American Economic Review*, 107(4):1136–64.

Italian Government (2021). Piano nazionale di ripresa e resilienza.

Jung, T. and Ejermo, O. (2014). Demographic patterns and trends in patenting: Gender, age, and education of inventors. *Technological Forecasting and Social Change*, 86:110–124.

Kantor, S. and Whalley, A. (2025). Moonshot: Public rd and growth. *American Economic Review*, 115(9):2891–2925.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–20.

Kirkebøen, L., Leuven, E., Mogstad, M., and Mountjoy, J. (2021). College as a marriage market. Working Paper 28688, National Bureau of Economic Research.

Kirkebøen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection*. *The Quarterly Journal of Economics*, 131(3):1057–1111.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth*. *The Quarterly Journal of Economics*, 132(2):665–712.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714. The regression discontinuity design: Theory and applications.

Moretti, E. and Wilson, D. J. (2017). The effect of state taxes on the geographical location of top earners: Evidence from star scientists. *American Economic Review*, 107(7):1858–1903.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review*, 112(8):2580–2630.

Mountjoy, J. (2024). Marginal returns to public universities. Working Paper 32296, National Bureau of Economic Research.

Mukherjee, A., Singh, M., and Žaldokas, A. (2017). Do corporate taxes hinder innovation? *Journal of Financial Economics*, 124(1):195–221.

Myers, K. R. and Lanahan, L. (2022). Estimating spillovers from publicly funded rd: Evidence from the us department of energy. *American Economic Review*, 112(7):2393–2423.

Nature (2025). Us universities cut phd admissions amid trump research funding reductions. By Dan Garisto. Accessed August 2025.

Prato, M. (2024). The global race for talent: Brain drain, knowledge transfer, and growth. *The Quarterly Journal of Economics*, 140(1):165–238.

Raffo, J. (2015). Matchit: Stata module to match two datasets based on similar text patterns. Statistical Software Components, Boston College Department of Economics.

Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, 90(5):2555–2591.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2):S71–S102.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives*, 8(1):3–22.

Santoleri, P., Mina, A., Di Minin, A., and Martelli, I. (2024). The causal effects of r&d grants: Evidence from a regression discontinuity. *The Review of Economics and Statistics*, 106(6):1495–1510.

Toivanen, O. and Väänänen, L. (2016). Education and Invention. *The Review of Economics and Statistics*, 98(2):382–396.

Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68:53–67.

Wasi, N. and Flaaen, A. (2015). Record linkage using stata: Preprocessing, linking, and reviewing utilities. *The Stata Journal*, 15(3):672–697.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

A Appendix

A.1 PhD Programs Dataset

The data on PhD programs come from two different sources based on the starting year of the program itself. All PhD programs starting after 1998 are included in the official database of PhD programs maintained by CINECA, the IT service provider of the university ministry. PhD programs starting earlier than 1998 are not included in the database and the information had to be retrieved from the original calls for application.

From 1983 to 1998 (with the exception of 1985 and 1986) the university ministry published on the *Official Gazette of the Italian Republic* one call for applications covering all PhD programs offered in Italy in that year. I have used the OCR software ABBYY FineReader PDF to extract the relevant information and then proceeded to manually check the correctness of each entry.

I categorise PhD programs based on the scientific sectors the university ministry created in 2001 and used until 2015. To retroactively assign these sectors to earlier PhD programs, I have proceeded as follows. Firstly, if a PhD program existed in a university both before and after 2001, I assigned the pre-2001 programs to the same category as the later programs. Then, if a PhD program only existed before 2001 in a university (16.93% of the pre-2001 PhD programs), I assigned it to the same category all other PhD programs on the same subject belong to. As a final step, I manually checked all remaining unassigned PhD program (16.00% of the pre-2001 PhD programs) and assigned them to a specific sector based on the PhD programs on similar subjects.

Based on these ministry-defined subjects, I categorise a PhD program as STEM if it belongs to any area between 1 and 9, that is: Maths, Computer Science, Physics, Chemistry, Earth Sciences, Biology, Medicine, Veterinary and Agricultural Sciences, Industrial and Information Engineering, and Civil Engineering and Architecture. The only exception to this definition is that maths is excluded from STEM in the analysis due to its mostly theoretical nature in the Italian academic environment. Its inclusion does not alter the results.

A.2 PhD Graduates Dataset

The data on PhD graduates come from the online catalogue of the National Central Library of Florence (BNCF). Since universities are required by law to deposit at BNCF each PhD thesis produced by their graduates, BNCF's catalogue comprises the totality of PhD theses produced in Italy since the first PhD cohort.

I obtained the information directly from the website by scraping the results of a search including all PhD dissertations from 1986 on and excluding any dissertation including the name "SISSA". The restriction by year excludes 33 documents which are mostly earlier treaties or books that have been misclassified by the library. The exclusion of SISSA dissertations (2,289 results) ensures I am only selecting people who graduated from PhD programs that are officially recognized by the university ministry.

The information included in the online catalogue for all dissertations includes graduate full name, thesis title, deposit year. University, subject of the program, starting year of the program, year of graduation, and names of the supervisors are included for a subset of dissertations. Since the subject of the PhD program was reported in two different formats (either the name of the program or the scientific sector), I harmonised the dataset by replacing the name of the program with the relevant sector. To do this, I have matched each program name with the PhD programs dataset and replaced it with the scientific sector coming from the PhD programs dataset. Then, I have excluded all entries that did not report the name of the graduate (6,974 entries).

Table A.1 shows the frequency of each piece of information by year of publication.

Table A.1: Availability of Different Types of Information on PhD Graduates

Publication Years	Graduate Name (%)	Dissertation Title (%)	University (%)	Subject (%)
1986-1991	100.00	100.00	77.39	90.08
1992-1996	100.00	100.00	90.05	95.74
1997-2001	100.00	100.00	94.55	97.72
2002-2006	100.00	100.00	40.16	93.20
2007-2011	100.00	100.00	29.00	71.99
2012-2016	100.00	100.00	37.26	77.93
2017-2019	100.00	100.00	52.87	66.55
Total	100.00	100.00	49.13	81.57

Publication Years	Supervisor (%)	Graduation Year (%)	Starting Year (%)	Observations
1986-1991	0.49	18.45	40.96	4878
1992-1996	4.57	5.82	72.77	11984
1997-2001	3.89	3.76	84.36	19718
2002-2006	69.47	20.97	59.79	31822
2007-2011	10.30	0.03	6.24	41941
2012-2016	12.90	0.09	5.34	40591
2017-2019	2.94	0.01	1.16	23284
Total	19.34	5.20	29.52	174218

Notes: This table reports how frequently a specific piece of information is present in the graduates dataset constructed from the PhD theses catalog of the Central National Library of Florence. The information I collect include graduate's name, thesis title, university of PhD program, subject of PhD program, name of supervisor, graduation year, and starting year of the program.

A.3 Matching Applicants to PhD Graduates and Inventors

The first step is matching the applicants data coming from the archives of Italian universities to the PhD graduates dataset from BCNF. First, I restrict the dataset on PhD graduates to all graduates that started their PhD at most in 1998 or finished it at most in 2003, if the starting year is not available. Then, I match applicants to PhD graduates through fuzzy matching on the full name imposing gender to be the same using the *reclink2* STATA package developed by Wasi and Flaaen (2015). I collect the best 10 matches with a matching score higher than 0.8 for each applicant and then apply different rules to determine the success of a match based on whether the applicant enrolled into the PhD program or not. Table A.2 explain in detail the matching rules. It must be noted that the fuzzy matching is necessary in this case due to different naming conventions between sources regarding the order of name and surname, the transliteration of foreign names, and the treatment of secondary names and surnames.

After matching PhD graduates and applicants, I proceed with the match between those and the inventors from the patents dataset. Since different naming conventions were used, I use the *matchit*

STATA package (Raffo, 2015) to match the datasets. Specifically, I use two different matching rules: token-by-token comparison and 3-gram circular comparison. Then, I only keep the matches if the average score is bigger than 0.85 or bigger than 0.80 if the surname includes any preposition.

Finally, I proceeded to manually check each match using Linkedin profiles, academic websites, professional websites, and CVs to establish the correctness of matches. In total, I manually checked 3386 matches, of which 809 were correct.

Table A.2: Rules to Mark a Match between *PhD Applicant* and *PhD Graduate* as Successful

Enrolled	Same University	Same Subject	Same Course	Same Starting Year	Exact Name-Surname	N Matched Names	Minimum Score
X	X				X		0.8000
X	X	X				2	0.8000
X	X			X		2	0.9900
X	X	X		X		1	0.9900
X	X	X	X				0.9844
X	X	X	X	X		1	0.8000
X		X			X		0.8000
X				X	X		0.8000
X		X		X		1	0.9900
X		X	X	X		2	0.9950
		X			X		0.8000
				X	X		0.8000
		X				3	0.8000
		X		X		All-1	0.9998

Notes: This table represents the algorithm that marks the matches coming from *reclink2* as correct or not. Each match need to satisfy at least one of these rules to be marked as correct. The first 10 rules are applied when the PhD applicant has enrolled into that specific PhD program. The remaining 4 rules apply to those who did not enroll. Column 2 is respected if the PhD graduate comes from the university the applicant applies to. Column 3 is respected if the graduate completed a course in the same field the applicant applies to. Column 4 is respected if the name of the course is the same. Column 5 is respected if the starting year of the graduate's PhD program is the year of the call for application of the applicant. Column 6 is respected if all the names and surnames of the applicant are presented in the full name of the graduate. Column 7 is respected if at least the specified number of names and surnames of the applicant are present in the full name of the graduate. Column 8 reports the minimum score attributed by *reclink2* based on the name.

A.4 Matching PhD Graduates to Inventors

I match PhD graduates to inventors using fuzzy matching and calculating a probability of match correctness. The resulting dataset gives me a probability for each inventor-patent couple that the inventor is indeed a PhD graduate.

First, I exclude from the PhD graduates list all graduates included in the applicants dataset, which I have manually matched to patents (see A.3). Then, I use the *matchit* STATA package (Raffo, 2015) to match PhD graduates and inventors. Specifically, I use two different matching

rules: token-by-token comparison and 3-gram circular comparison. Then, I only keep the matches if the average score is bigger than 0.85 or bigger than 0.80 if the surname includes any preposition.

To estimate the correctness probability of each match, I estimate the following logit model using the applicants' manually matched dataset restricted to PhD graduates only:

$$\begin{aligned}\ln\left(\frac{y_i}{1-y_i}\right) =& \alpha_0 + \alpha_1 \text{NgramScore}_i + \alpha_2 \text{TokenScore}_i + \alpha_3 \text{NameFreq}_{n(i)} + \\ &+ \sum_k \beta_k \text{TimeDist}_i^k + \sum_{f,s} \gamma_{fs} \left(\text{PhDField}_i^f \times \text{Patsection}_i^s\right) + \varepsilon_i\end{aligned}$$

The outcome (y_i) is a binary variable equal to one if the match i is correct and the regressors include both fuzzy matching scores, the frequency of a name-surname couple in Italy, and a set of binary variables corresponding to the distance in years between the starting of the PhD and the submission of the patent, and to the interaction between PhD field and patent IPC section. The frequency of names was calculated by scraping the *Pagine Bianche* website[23] and accounts for how many potential namesakes the PhD graduate might have. The binary variables corresponding to the distance between PhD starting year and patent submission account for the evolution of innovativeness throughout the academic career (e.g. a patent submitted 15 years before the start of the PhD is most likely a bad match). The interaction between IPC section and PhD field accounts for the differing degree of closeness between academic subjects and technologies.

Finally, I apply the estimated model to the matches between graduates and inventors the probability that inventor i of patent p was correctly matched to a PhD graduate g, denoted as $P\left(i^p \text{ is } g\right)$.

The resulting aggregate data are calculated using this probability. The probability that a certain patent p has a PhD inventor is:

$$P\left(p \text{ has a PhD inventor}\right) = 1 - \prod_{i \in I^p} \left[1 - P\left(i^p \text{ is } g\right)\right]$$

Where I^p denotes the set of inventors of patent p.

[23] This website includes the owners of all landlines operating in Italy

Given a total of R patents, the probability that a PhD graduate g is an inventor is:

$$P\left(g \text{ is an inventor}\right) = 1 - \prod_{p=1}^{R} \prod_{m=g} \left[1 - P\left(i^p \text{ is } m\right)\right]$$

Where the second product accounts for the probability that the specific PhD graduate g is not matched to any inventor i^p for the specific patent.

A.5 Calculating Aggregate Effects

The total number of additional patents produced thanks to PhD programs between 1986 and 2001 is calculated as follows:

$$\text{PhD Patents} = 118 \cdot \sum_{j=1986}^{2001} \sum_{i>-1} \beta_{i,j} \times N_j \approx 13667$$

Where the $\beta_{i,j}$'s are the distance- and cohort-specific coefficients estimated in Equation 1. N_j is the number of provinces whose first STEM PhD cohort graduates in year j. Through the product between these two elements and the double summation across treatment cohorts and distances in the post-treatment period, I am calculating the total number of patents produced in treated provinces after the first local STEM PhD graduation.

To calculate a percentage change in patent activity, I need to define patent production in a counterfactual scenario with no PhD openings. To do this, I sum the total number of patents produced in treated provinces after the treatment starts and simply subtract PhD Patents. Then, the percentage change is simply the ratio between PhD Patents and the counterfactual:

$$\text{Percentage Change} = \frac{\text{PhD Patents}}{\sum_{j=1986}^{2001} \sum_{t \geq j} patents_{j,t} - \text{PhD Patents}} = \frac{13667}{78442 - 13667} \approx +21.10\%$$

A.6 Effects by Average Novelty of Technologies

Due to the absence of reliable citation data for Italian patents until 2008, I use the importance measure calculated by Kelly et al. (2021) to assess if the effects of PhD programs on patent activity

are concentrated among the least or most novel technologies.

As a first step, I restrict the sample of US patents to those filed between 1976 and 2001, corresponding to the period of my analysis. Then, I use the patent-level measure of importance calculated around a 10-year window (q_p^{10}) to estimate the average importance of IPC classes:

$$N_c = \frac{1}{|C|} \sum_{p \in C} q_p^{10}$$

I use this measure as an indicator of how innovative a certain technology was in the years of my analysis. As Table A.3 shows, the top classes mostly pertain to IT, whereas the classes at the bottom of the importance ranking mostly relate to agriculture.

Table A.3: The Most and Least Novel IPC Classes

Most Novel	**Least Novel**
1. G06 - Computing	118. D02 - Yarns
2. G16 - ICT	117. C13 - Sugar Industry
3. H04 - Electronic Communication	116. C10 - Petroleum, Gas, or Coke
4. C12 - Biochemistry	115. C05 - Fertilisers
5. G07 - Checking Devices	114. A24 - Tobacco

Notes: This table shows the most and the least important IPC classes. The measure is based on the average importance of the patents belonging to these classes between 1976 and 2001, based on the methodology of Kelly et al. (2021).

Based on this importance measure, I divide IPC classes into quartiles and estimate a model to assess if the effects from the baseline model are concentrated in the most or least novel technologies:

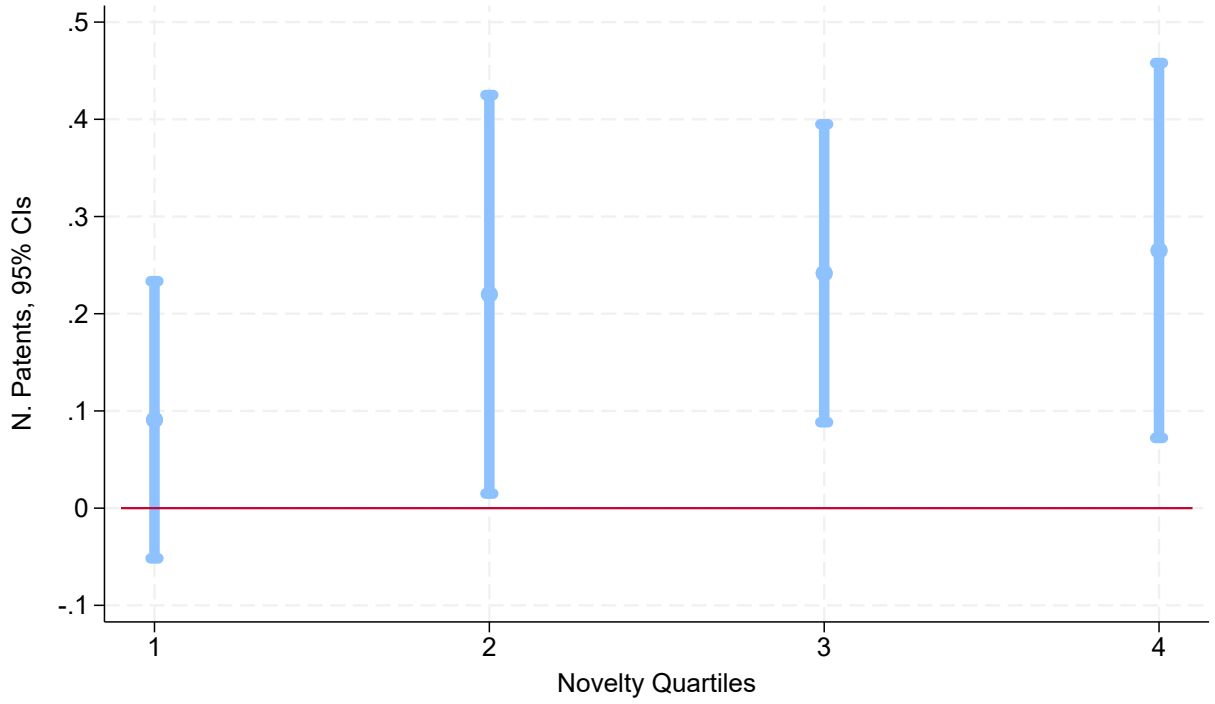
$$\begin{aligned} patents_{p,c,t} &= \alpha_{p,c} + \alpha_{c,t} + \gamma \cdot \alpha_t \times \mathbf{X}_{p,1981} + \\ &+ \sum_{q=1}^{4} \sum_{j=1986}^{2001} \sum_{i \neq -1} \beta_{i,j,q} \cdot 1\left\{Q\left(N_c\right) = q\right\} \times 1\left\{t = g\left(p\right) + i,\ g\left(p\right) = j\right\} + \varepsilon_{p,c,t} \end{aligned} \tag{11}$$

Where $Q\left(N_c\right)$ is the quartile of technology c. The coefficient $\beta_{i,j,q}$ is the average effect of STEM PhD programs on patent activity across technologies belonging to the novelty quartile q.

Figure A.1 shows the average effects by quartile in the post-treatment period. Even though the effects are larger for the most novel technologies, the differences across the quartiles are not

significant. Thus, I conclude that PhD programs did not alter substantially the quality of the patents produced locally.

Figure A.1: Effect of Opening a STEM PhD by Novelty Quartile



Notes: The figure shows the average effect of STEM PhD programs on patent activity by quartile of novelty of the technologies. The standard errors are clustered at the province level.

A.7 Constructing the Knowledge Proximity Measure

I construct a measure of knowledge proximity between PhD fields and IPC classes using US patents' and US PhD theses' texts. The use of US texts ensures that the measure I calculate is exogenous to the Italian context, whereas using Italian texts might create an endogenous measure due to the effect of PhD training on graduates' patenting.

The US patents' information come from the USPTO PatentsView website. I have downloaded the short text description and IPC classes of all US patents submitted for application between 1976 and 2001. From these I have excluded the 8 IPC classes for which I had less than 1000 texts, thus creating a datset with 2.1 million patents across 110 IPC classes.

The US theses' information come from the ProQuest Dissertations & Theses Global dataset. I

have first mapped the 441 ProQuest subjects into the 28 Italian ministry-defined academic fields. Then, I have selected the 1.3 million PhD theses published by 2001.

To calculate the measure of knowledge proximity, I first define a dictionary using words, bigrams, and 3-grams present in the patents and excluding common English stop-words. Since the measure is calculated to estimate how close PhD programs are to technologies, the restriction to patents allows us to focus only on innovation-relevant terms. Then, I refine the dictionary by excluding tokens that are present in less than 10 patents or in more than 50% of patents. This reduces the computational burden and excludes from the dictionary words that are too common to signal knowledge content or too rare to be relevant.

Using this dictionary, I redefine all patents' descriptions and theses' abstracts as Term Frequency (TF) vectors and then multiply them by the Inverse Term Frequency (IDF) of each word across all documents. This way, the weight attributed to each term is scaled by its specificity, as indicated by its frequency across the documents.

Using this TF-IDF vector representation of texts, I can calculate the knowledge proximity of any PhD field f and any IPC class c as the average cosine similarity between their respective theses and patents:

$$\sigma_{c,f} = \frac{1}{|c|\,|f|} \sum_{i \in c} \sum_{j \in f} \frac{\mathbf{patent}_i \cdot \mathbf{thesis}_j}{\|\mathbf{patents}_i\|\,\|\mathbf{thesis}_j\|} \tag{12}$$

As all elements of these vectors are non-negative, the measure is bounded between 0 and 1, where 1 means that a set of patents and a set of theses use the exact same words with the same frequency. If the measure is equal to 0, that means that there is no intersection in word use between patents and theses. Table A.4 shows the most and least similar classes and fields.

Table A.4: The Most and Least Similar IPC Classes and PhD Fields

Most Similar			Least Similar		
	IPC Class	**PhD Field**		**IPC Class**	**PhD Field**
1.	C12 - Biochemistry	Biology	**3080.**	F01 - Machines, Engines	Pedagogy
2.	C12 - Biochemistry	Veterinary Sciences	**3079.**	F15 - Hydraulics	Philosophy
3.	G10 - Musical Instruments	Art	**3078.**	F04 - Pumps for Liquids	Pedagogy
4.	C12 - Biochemistry	Chemistry	**3077.**	D05 - Sewing	Veterinary Sciences
5.	C12 - Biochemistry	Medicine	**3076.**	F22 - Steam Generation	Pedagogy
6.	G06 - Computing	Computer Science	**3075.**	F02 - Combustion Engines	Pedagogy
7.	G06 - Computing	Computer Engineering	**3074.**	F04 - Pumps for Liquids	Philosophy
8.	C12 - Biochemistry	Agricultural Sciences	**3073.**	F02 - Combustion Engines	Philosophy
9.	C30 - Crystal Growth	Physics	**3072.**	D05 - Sewing	Law
10.	A63 - Sports	Sports Sciences	**3071.**	F15 - Hydraulics	Literature

Notes: This table shows the 10 most and least similar couples between IPC classes and PhD fields, based on the average cosine similarity between PhD theses' abstracts and patents' short descriptions.

A.8 Extending Tax Codes Coverage

To analyze spillovers, I need to match patents' assignees to the CERVED dataset in order to recover their headquarters location and industry. This requires firms' tax codes for the entire 1976–2001 period. Since tax codes are available only from 1990 onwards, I extend coverage to earlier years using a supervised matching approach. Specifically, I train a random forest classifier on the 1990–2010 data to identify when two assignee names correspond to the same firm, as verified by their shared tax code. I then apply this model to fuzzy matches of assignee names in the pre-1990 data.

First, I identify all firms with different names sharing the same tax code among the assignees of patents filed between 1990 and 2010. This inconsistency in the way assignees' names are registered does not allow me to straightforward match firms across time. However, thanks to the common tax code that identifies all the different occurrences of the assignee names as being the same firm, I can use this set of correct fuzzy matches to train a matching algorithm. Specifically, I exploit these inconsistencies to estimate 4 different fuzzy matching scores using the *rapidfuzz* Python library, to which I add the raw difference in names' length between the assignees. This gives me for any couple of assignees with different names but a common tax code a set of 5 measures of name similarity: simple ratio, partial ratio, token sort ratio, token set ratio, and length difference.

To train the classification model, I add to the 27,725 correct matches a set of 327,540 randomly

drawn incorrect matches for which I calculate the same measures of name similarity. Then, I use this set of correct and incorrect matches to train a random forest classifier based on the 5 name similarity features. This model gives for any couple of names the probability that they refer to the same firm. This model has an out-of-sample precision (true positives) of 97% and a recall (true negatives) of 93%, resulting in an F-1 score of 95%.

To extend the tax codes from the assignees of post-1990 patents to earlier assignees, I calculate the similarity features for any couple of pre-1990 and post-1990 firms and use the trained model to predict the probability that the match is correct. I take as correct matches all those with a probability higher than 94.74%. I choose this threshold as a rule of thumb to achieve an overall correctness probability of 90%, as multiplying this threshold times the F-1 score returns a value of 0.9. In case of multiple matches fulfilling the condition, I keep the one with the highest probability. This process returns a set of 3,082 matches between more recent and older assignees. Finally, I extend to all matched assignees from before 1990 the tax codes of their matched counterpart from more recent patents.

The resulting dataset matches 31,369 patents to limited liability firms. Using this match, I link patents to the assignee's headquarters and to their industry. I use the headquarters' location as a proxy for the location of a firm's R&D deartment to assign patents to a province, under the assumption that a firm is treated if their R&D department is located where a PhD cohort has graduated in the past. The link to the industry allows me to classify firms based on their sector's R&D intensity and propensity to hire graduates from different university courses.

A.9 Estimating Patent Value

I estimate the patent value of Italian patents published by listed firms using the method developed by Kogan et al. (2017). This method consists of linking each patent's assignee to their stock market value in the days following the announcement of a patent and estimating the value of the patent as the escess return of the stocks. This method relies on the assumption that financial markets can predict the discounted flow of profits deriving from the patents published by firms.

As a first step, I match the historical data on stock market values from the Milan stock exchange from the LSEG Datastream database to the firms' tax code through the Bureau Van Dijk's AIDA (BvD) database. Whereas the LSEG Datastream databases contain information on stock exchanges throughout the world, the BvD data provide information on Italian firms linking their tax code to their International Securities Identification Number.

The matching between these databases was made first through the ISIN codes, then through the firms' names. Since the BvD data do not cover the whole period analyzed, not all firms from the patent data were present in the AIDA dataset.

To match these remaining firms, I first ran a fuzzy matching algorithm using the *rapidfuzz* Python library to estimate the partial ratio and token sort ratio between any couple of firms from the patent dataset and Refinitiv LSEG. Then, I manually checked the correctness of each couple match that had achieved a score of at least 80 for both measures.

The final dataset has matched 159 listed firms from Italy out of 667 listings on the Milan stock market between 1973 and 2001. I restrict the patent dataset to include only patents from treated provinces after the start of treatment with a single assignee, thus restricting the sample to 3,718 patents.

Matching these patents to the stock market value of their assignees around the publication date, I estimate their monetary value as:

$$\xi_j = \frac{1}{N_j} E\left[v_j | R_j\right] M_j \tag{13}$$

Where ξ_j is the value of patent j. N_j is the total number of patents granted on the same day as

patent j to j's assignee. M_j is the assignee's market capitalization the day before the patent is granted. $E\left[v_j|R_j\right]$ is the expected value of the patent conditional on the realized excess return on the stocks of the assignee in the three days following the announcement of the patent.

As the IPO was only controlling the bureaucratic correctness of patent applications and and not their novelty until 2008, I am setting the ex-ante probability of having a successful patent application equal to 1 ($\bar{\pi} = 1$). Given the formula from Kogan et al. (2017), this assumption has the effect of lowering the estimates and returning a lower bound for the true value of the patent.

The conditional expected value of patents is defined as:

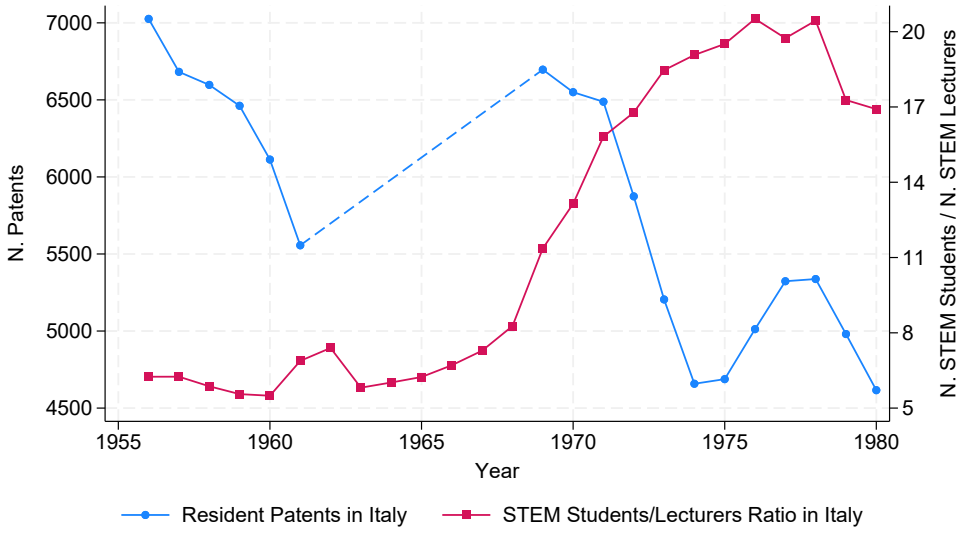
$$E\left[v_j|R_j\right] = \delta R_j + \sqrt{\delta}\sigma_{\varepsilon ft}\frac{\phi\left(-\sqrt{\delta}\frac{R_j}{\sigma_{\varepsilon ft}}\right)}{1-\Phi\left(-\sqrt{\delta}\frac{R_j}{\sigma_{\varepsilon ft}}\right)}$$

Where R_j is the excess stock market returns from the stock of patent j's assignee. $\sigma_{\varepsilon ft}$ is the the sum of squared market-adjusted returns estimated separately for each year. δ is the signal-to-noise ratio which I take directly from the calibration of Kogan et al. (2017): it is equal to 0.0145 and the same across time and firms.

Finally, I calculate the monetary value of 3,718 patents published by firms located in treated areas using the definition of ξ_j, and apply it to estimate the within-technology average value of patents.

B Appendix Figures

Figure B.1: STEM Student-Lecturer Ratio and Patents in Italy (1950-1980)



Notes: The figure shows in red squares the ratio between students and lecturers in STEM subjects in Italian universities. The blue circles represent the 3-year moving average of the number of granted patents submitted by Italian residents each year. Solid lines represent actual data, dashed lines are linear projection covering periods with missing data.

Figure B.2: Timeline of the centralized PhD allocation process

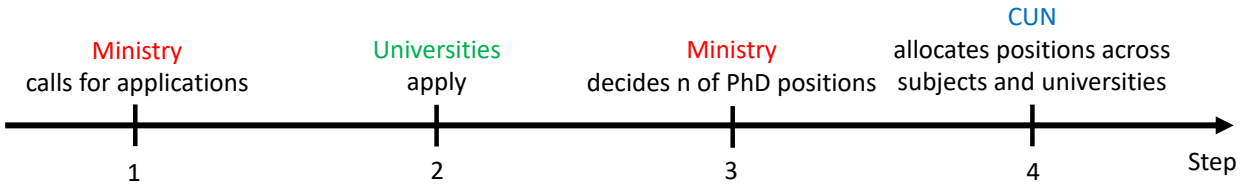
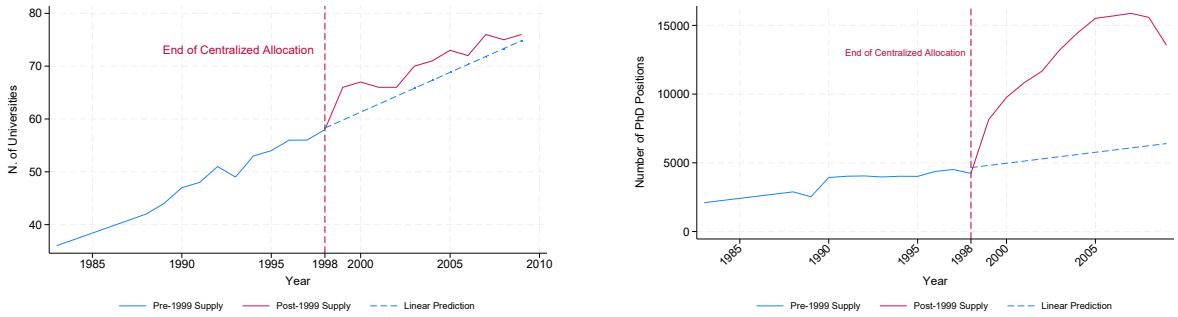


Figure B.3: Supply of PhD Programs over Time

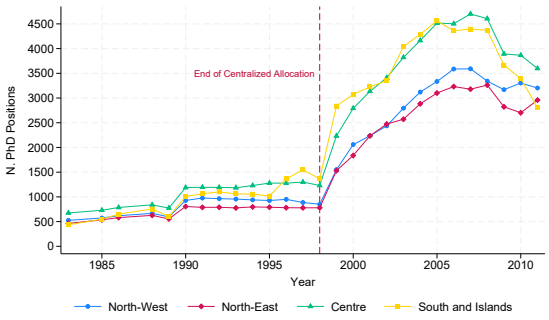


(a) Number of Universities with PhD Programs

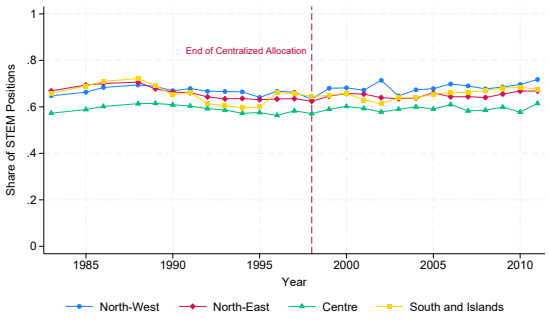
(b) Number of PhD Positions

Notes: The figures show a comparison between actual data and linear predictions of the supply of PhD programs in Italy. Solid lines represent actual data, dashed lines are linear predictions based on the data of the centralized allocation period. Panel A shows the number of universities offering PhD programs each year; panel B reports the total number of funded PhD positions available at PhD programs each year.

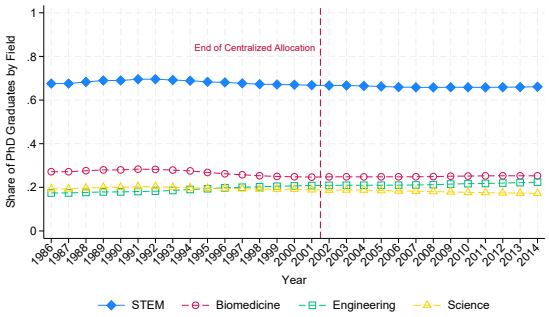
Figure B.4: Distribution of PhD Programs over Time



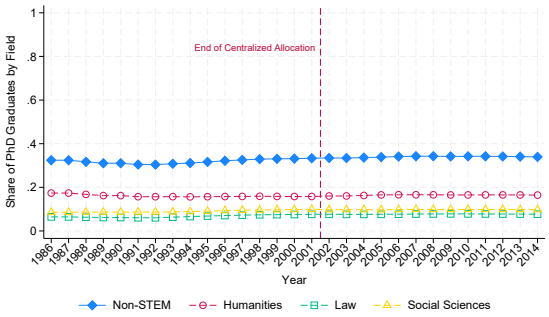
(a) PhD Positions by Area



(b) Share of STEM PhD Positions by Area



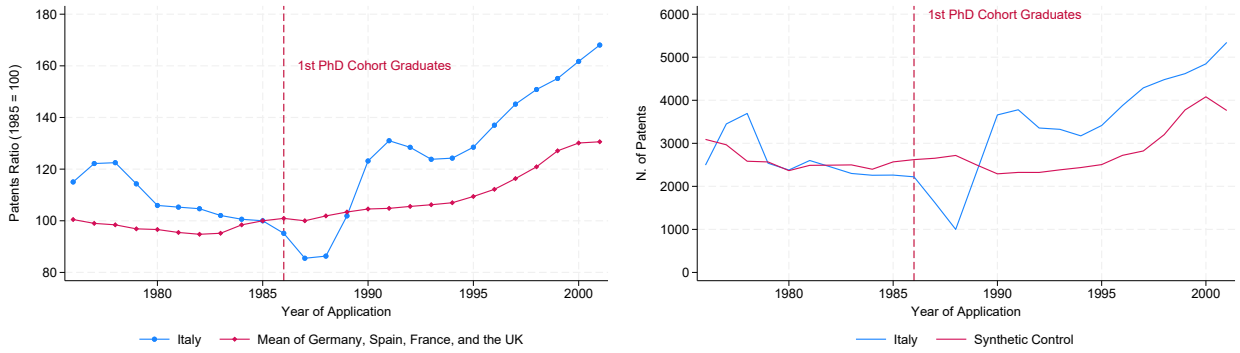
(c) Share of STEM PhD Graduates



(d) Share of Non-STEM PhD Graduates

Notes: The figures in panels A and B show the number of total PhD positions and the share of STEM PhD positions opened between 1983 and 2011 by geographical area. Panels C and D show the composition of the stock of PhD graduates by field between 1986 and 2014. The categorisation of PhD programs as STEM closely follows the classification of the university ministry and is described in A.1. The definition of geographical areas comes from ISTAT.

Figure B.5: Aggregate Trends in Patenting in Italy and other European Countries

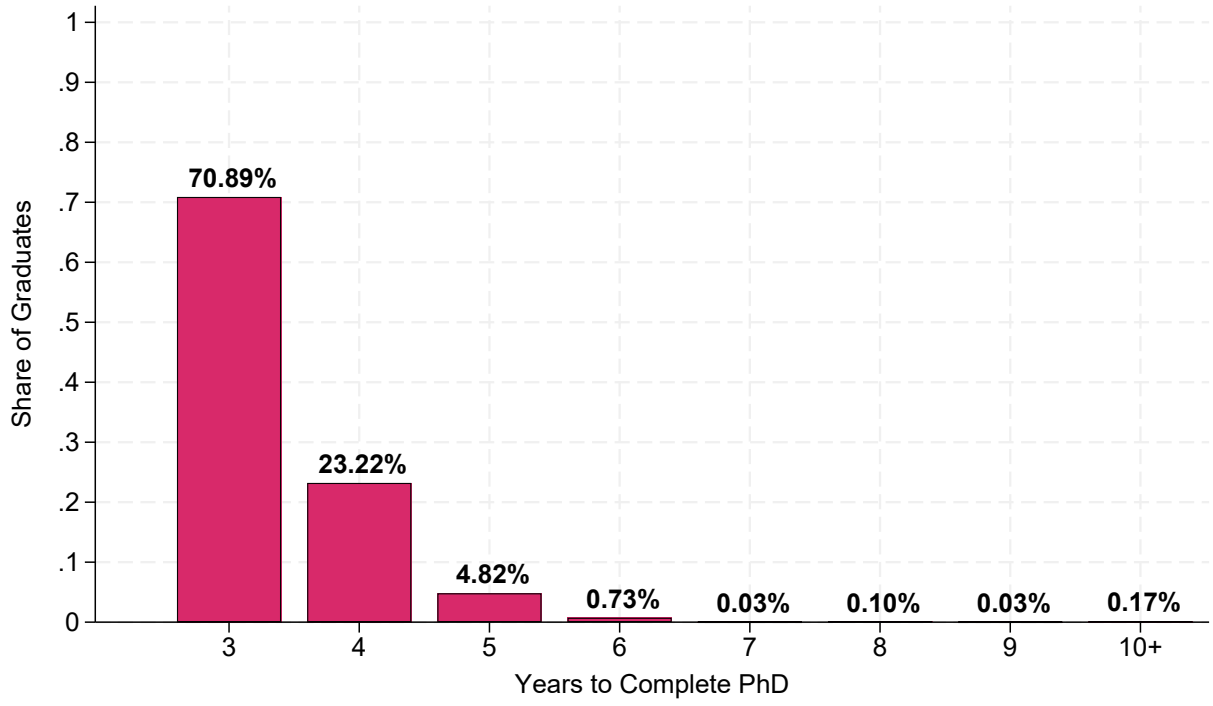


(a) Patenting in European Countries

(b) Italy versus Synthetic Contro

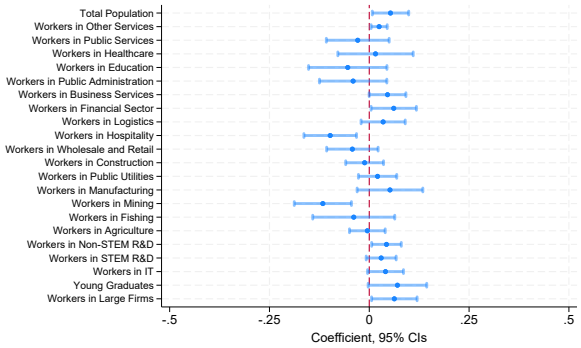
Notes: The figures show the trend in yearly patent activity in Italy and other European countries. Panel (a) plots the annual production for Italy and the other 4 major European countries (Germany, Spain, France, and the United Kingdom) as 3-year moving averages, setting 1985 as the baseline year. Panel (b) shows the annual production for Italy and a synthetic control estimated using the method of Arkhangelsky et al. (2021). The donor pool includes all European OECD countries for which WIPO data are available for the period examined: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Norway, Poland, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

Figure B.6: Distribution of PhD Duration

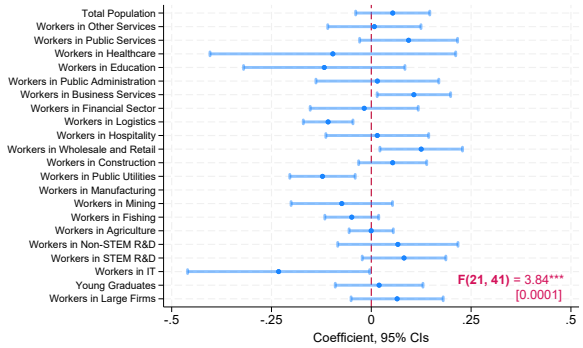


Notes: The figure shows the distribution of the duration of PhD studies by graduate. The years to complete the PhD are calculated as the difference between the year of graduation of the thesis and the starting year of the PhD program. This graph is based on the 2900 graduates for which both starting and graduation years were available.

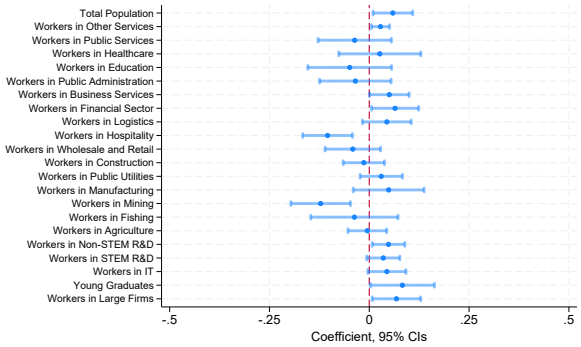
Figure B.7: Treatment Status versus Local Characteristics



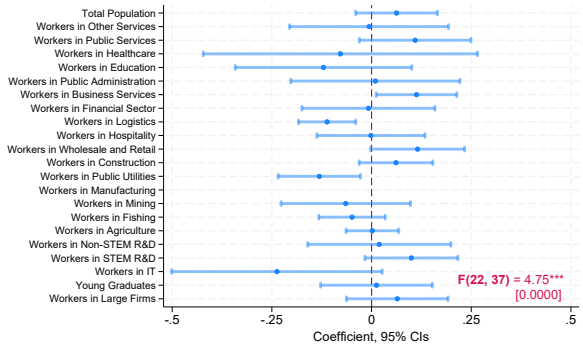
(a) Any PhD - Univariate Regressions



(b) Any PhD - Multivariate Regression



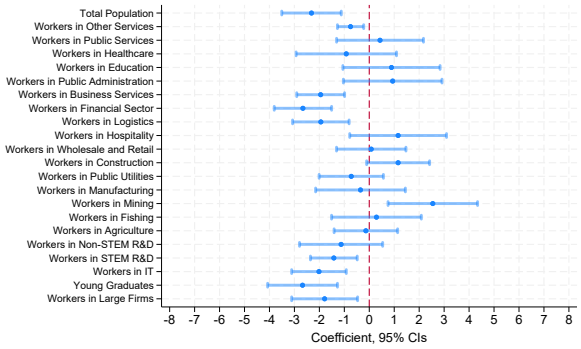
(c) STEM PhD - Univariate Regressions



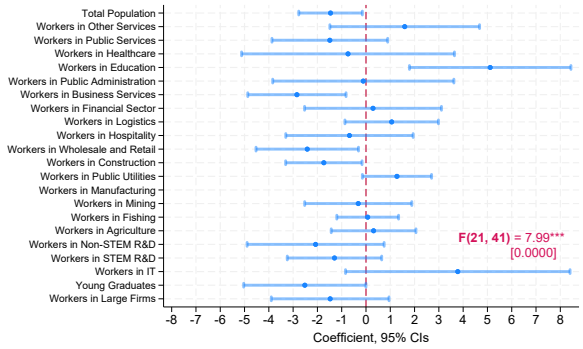
(d) STEM PhD - Multivariate Regression

Notes: This figure reports the results of regressing a dummy variable indicating treatment status on a set of standardised observable characteristics of provinces before the institution of PhD programs. In panels B.7a and B.7b, the treatment is defined as having opened the first local PhD program by 1998. In the bottom panels B.7c and B.7d, the treatment is restricted to the provinces that have opened the first local STEM doctoral program by 1998. I report the point estimates and the 95% confidence intervals. The variables included are share of university graduates aged between 25 and 34 in 1981, share of workers employed in R&D institutions in 1981, share of workers employed in IT firms in 1981, share of workers employed in firms with 500 or more employees in 1981, total population in 1981, and industrial composition in 1981.

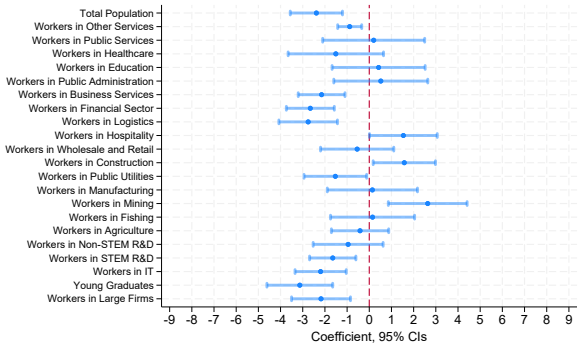
Figure B.8: Treatment Timing versus Local Characteristics



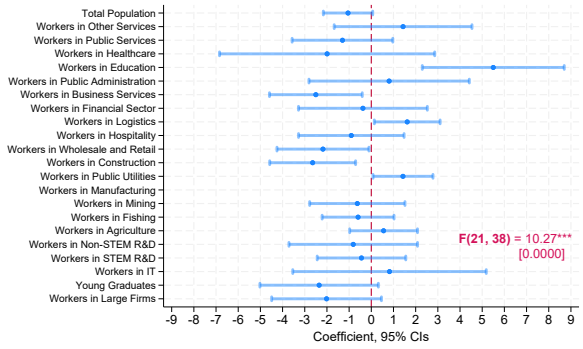
(a) Any PhD - Univariate Regressions



(b) Any PhD - Multivariate Regression



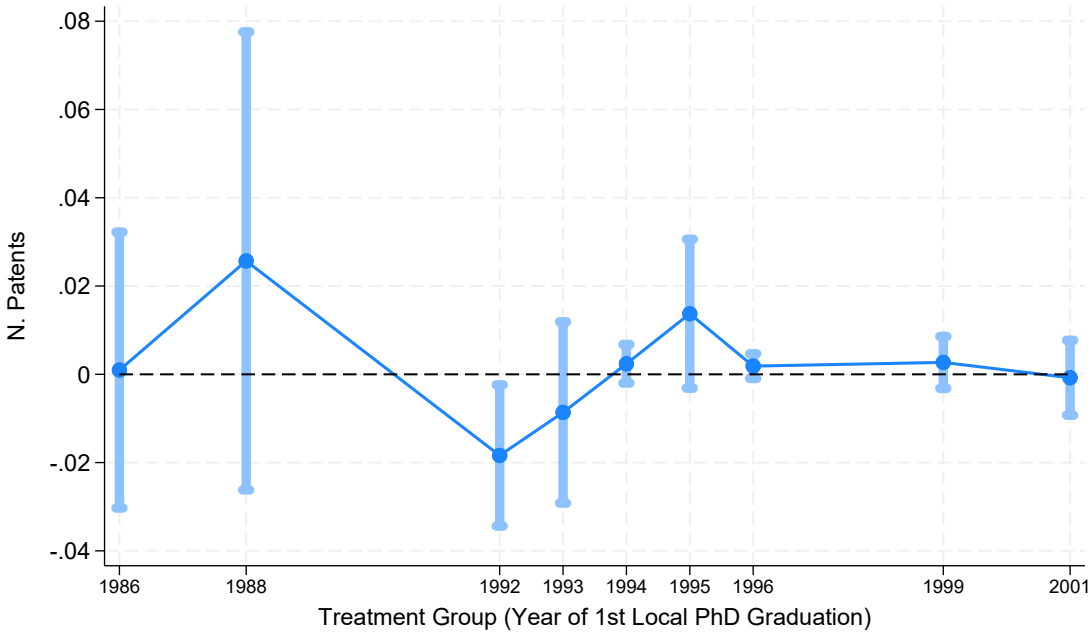
(c) STEM PhD - Univariate Regressions



(d) STEM PhD - Multivariate Regression

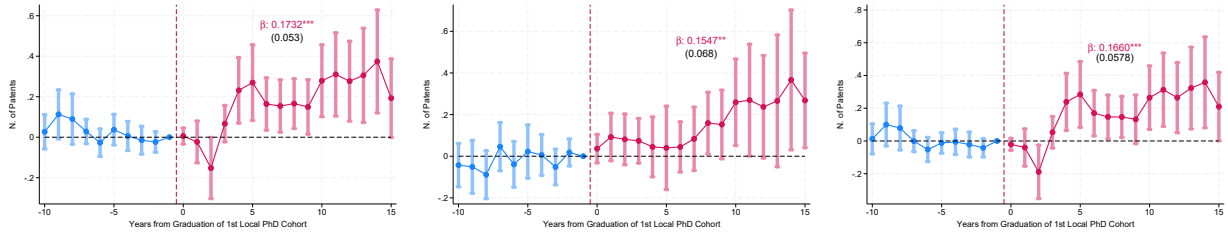
Notes: This figure reports the results of regressing the starting year of the first local PhD on a set of standardized observable characteristics of provinces before the institution of PhD programs. In panels B.8a and B.8b, the treatment is defined as having opened the first local PhD program by 1998. In the bottom panels B.8c and B.8d, the treatment is restricted to the provinces that have opened the first local STEM doctoral program by 1998. I report the point estimates and the 95% confidence intervals. The variables included are share of university graduates aged between 25 and 34 in 1981, share of workers employed in R&D institutions in 1981, share of workers employed in IT firms in 1981, share of workers employed in firms with 500 or more employees in 1981, total population in 1981, and industrial composition in 1981.

Figure B.9: Heterogeneous Effects by Treatment Group



Notes: The figure shows the estimation of the effects as in Equation 1 at the time of graduation of the first PhD cohort. The Y axis shows the point estimate of the effects; the X axis reports the treatment group those estimates belong to. Each line represents coefficients estimated for the same distance from the opening of the 1st local PhD program.

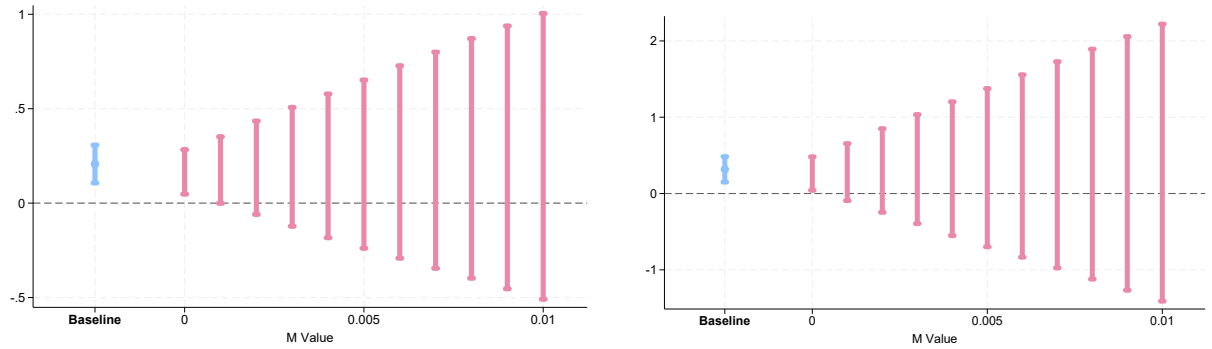
Figure B.10: Dynamic Effects of Additional Results in Table 2



(a) All PhDs - Only FE (b) All PhDs - Full Controls (c) STEM PhDs - Only FE

Notes: The figure shows the effects of PhD programs on local patent activity as estimated in the models presented in columns (1) to (3) of Table 2. Panel B.10a shows the effects estimated for the baseline model when all PhD programs are included in the definition of treatment and without the full set of control variables. In Panel B.10b, I include the full set of control variables to the previous specification. Panel B.10c shows the effects when the treatment only includes STEM PhD programs and I do not use the set of control variables.

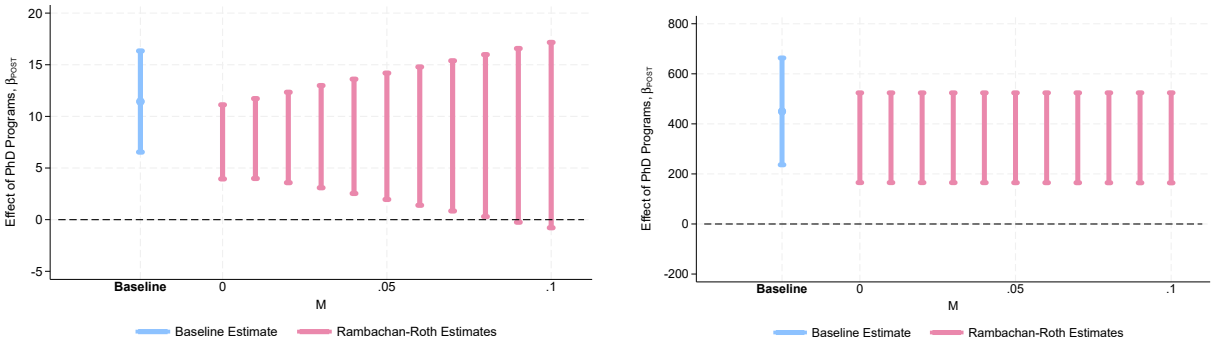
Figure B.11: Robustness to Violations of Parallel Trends for Patenting



(a) Extended Pretrends for β_{POST} (b) Extended Pretrends for β_{15}

Notes: The figure shows the results of the Rambachan and Roth (2023) tests checking the robustness of the baseline estimates to potential violations of the parallel trends assumption. Panel B.11a shows the results of the test for β_{POST}. Panel B.11a shows β_{15}. M indicates the degree of non-linearity of the pre-trends estimated from the realized pre-treatment DiD coefficients. When M is 0, the pre-trend is taken as linear. As M grows larger, the estimated pre-trend is allowed to get farther from linearity.

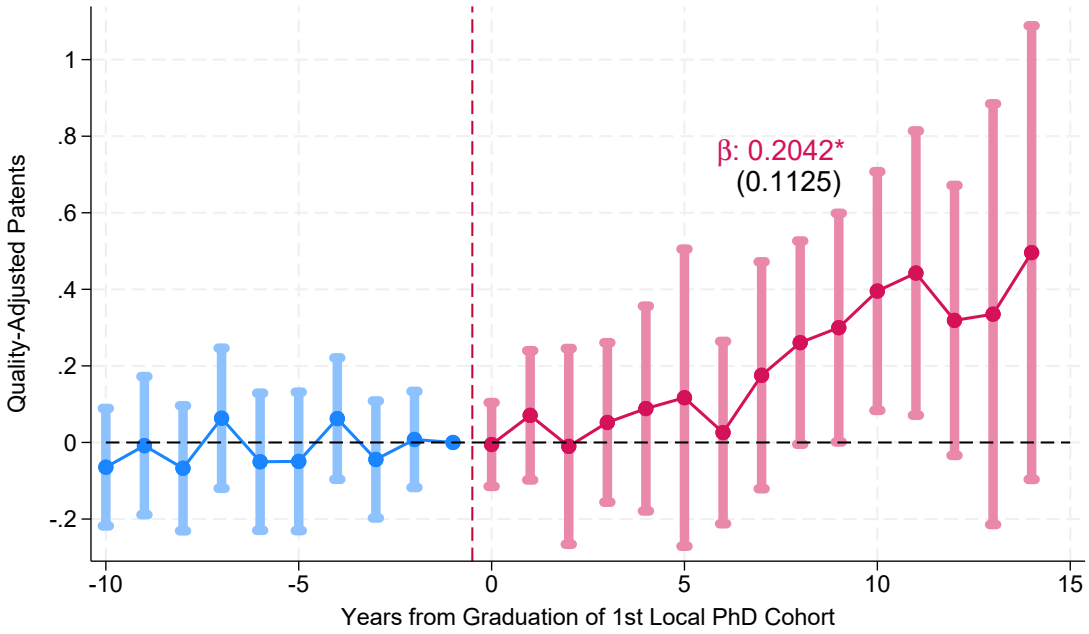
Figure B.12: Robustness to Violations of Parallel Trends for Academic Output



(a) Publications (b) Citations

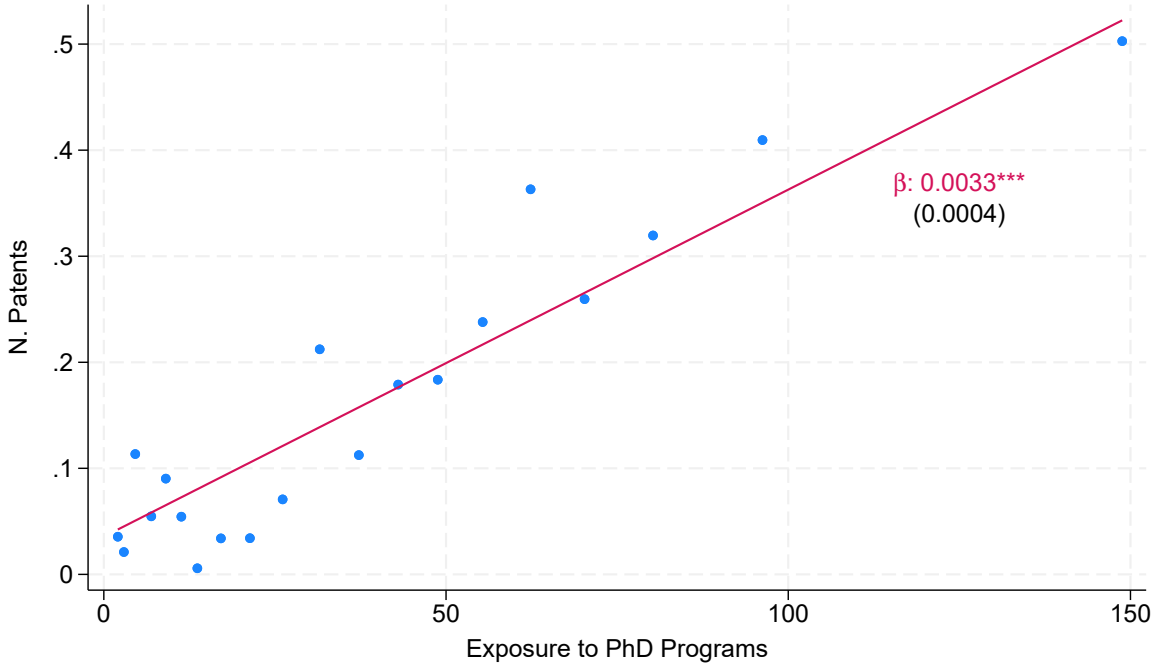
Notes: The figure shows the results of the Rambachan and Roth (2023) tests checking the robustness of the findings for publications and citations. Panel B.12a shows the results of the test for publications. Panel B.12b shows the results for citations. M indicates the degree of non-linearity of the pre-trends estimated from the realized pre-treatment DiD coefficients. When M is 0, the pre-trend is taken as linear. As M grows larger, the estimated pre-trend is allowed to get farther from linearity.

Figure B.13: Effects of PhD Programs on Quality-Adjusted Patenting



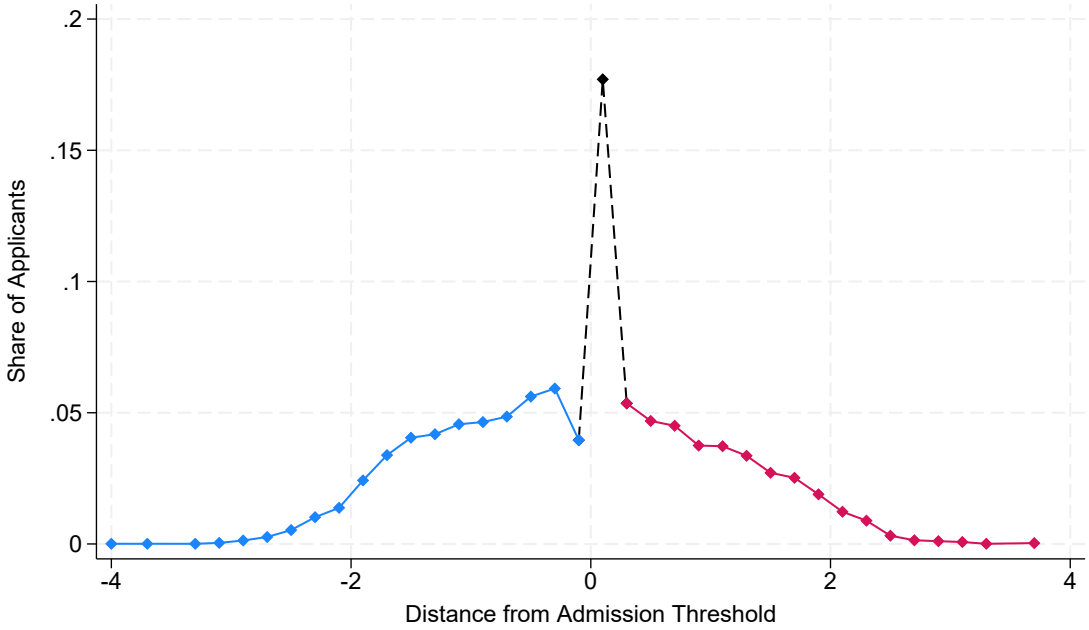
Notes: The figure shows the effects of PhD programs on local quality-adjusted patent activity. This measure of patenting is constructed using the fixed-effect approach outlined by Hall et al. (2001). Standard errors are clustered at the province level.

Figure B.14: Correlation Between Average Exposure and Average Treatment Effect



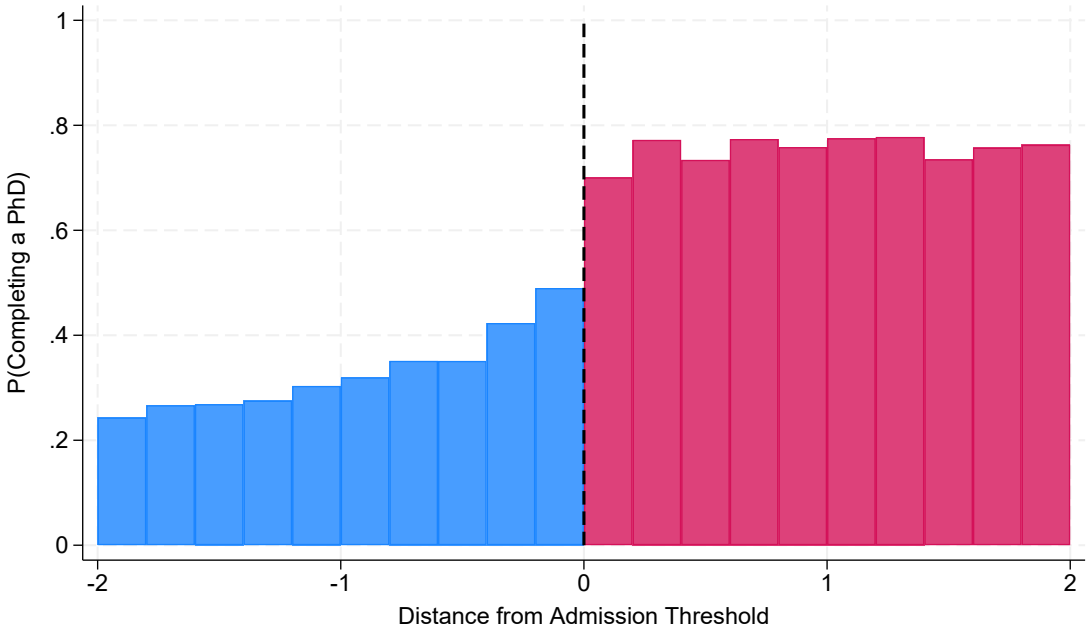
Notes: The figure shows in a *binscatter* plot the correlation between average exposure and average treatment effects across technology–year cells, where the effects are calculated by interacting the treatment variable with technology–year binary variables in the baseline model.

Figure B.15: Distribution of Applicants by Distance from the Cutoff



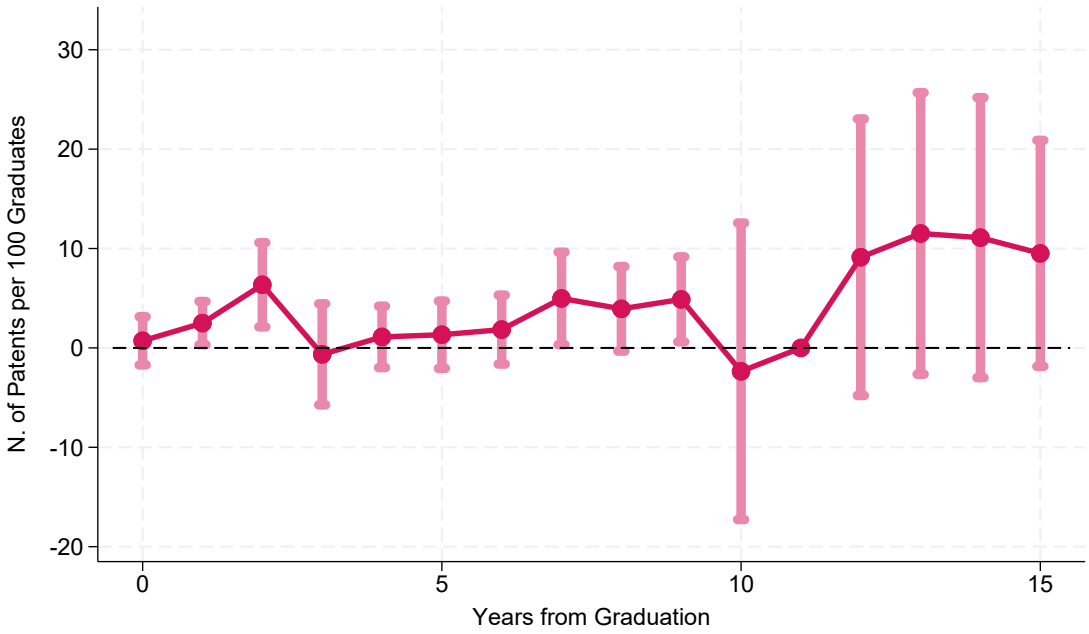
Notes: The figure shows the distribution of applicants by their normalized score. The discontinuity at 0 is caused by the way the measure is constructed, which creates a mass point at zero with an absolute frequency larger or equal to the number of PhD admission exams.

Figure B.16: Share of PhD Graduates by Distance from the Cutoff



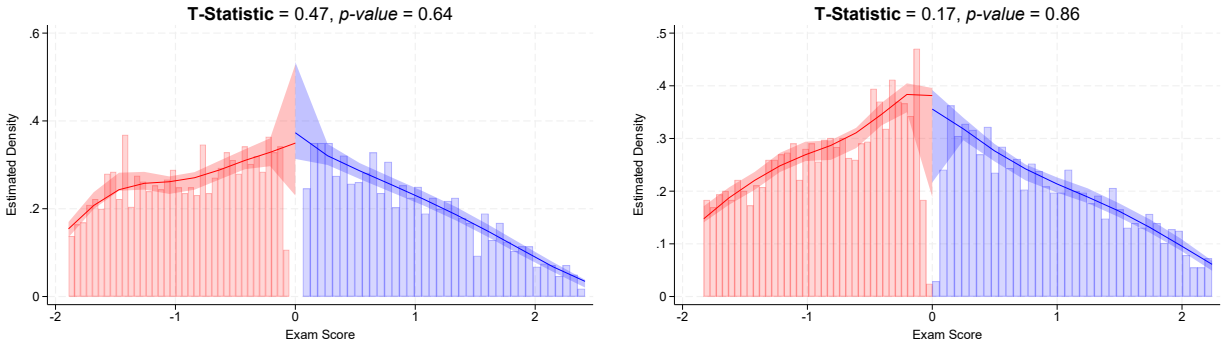
Notes: The figure shows the share of applicants completing a PhD in Italy between 1986 and 2001 by normalized distance from the cutoff. Data have been censored below -2 and above 2 due to low numerosity.

Figure B.17: Dynamic Effects of PhD Programs on Individual Patenting



Notes: The figure shows the fuzzy RDD estimate of Equation 8 for applied STEM PhD programs, separately for each event-time from graduation. Standard errors are clustered at the individual level.

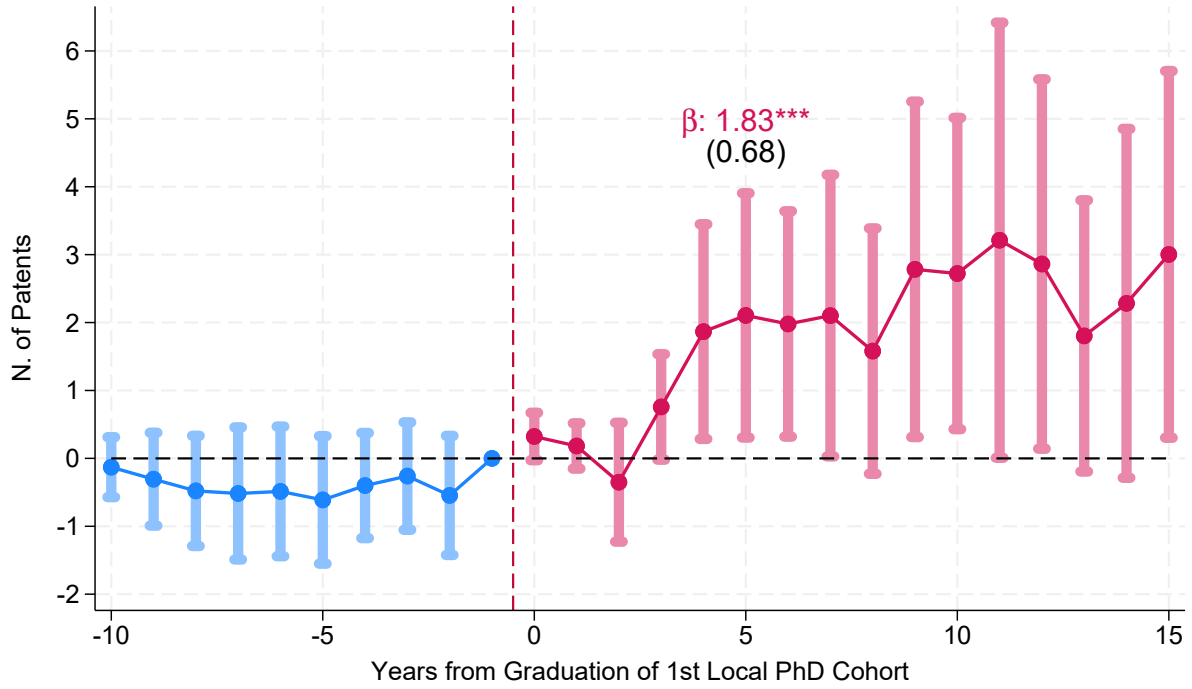
Figure B.18: Manipulation Test on The Distance from the Cutoff



(a) **Applied STEM** (b) **Non-Applied STEM**

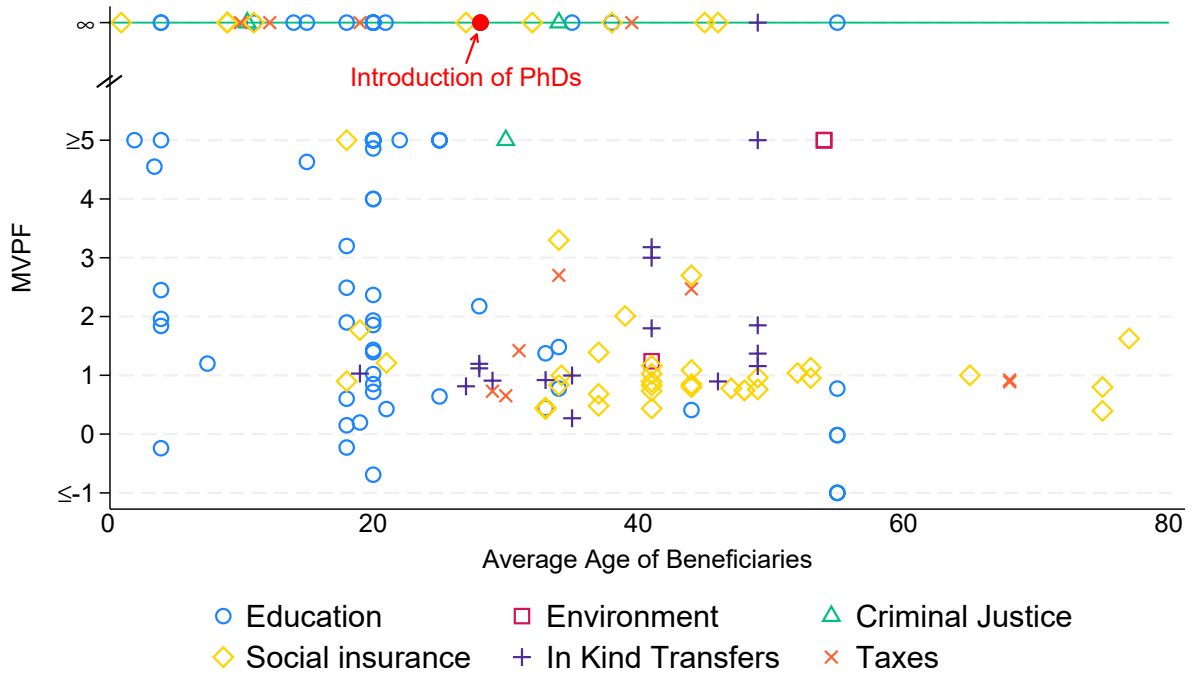
Notes: This figure shows the results of the McCrary tests on the running variable separately for each PhD category. The tests have been made using the *rddensity* command (Cattaneo et al., 2018) and the reported t-statistic and p-value are bias-corrected.

Figure B.19: Spillovers to High R&D Intensity and High Exposure Industries



Notes: The figure shows the spillover effects of PhD programs on local firms. The plot presents the results of a regression in which treatment is defined as belonging to the high R&D intensity and high exposure group and province-year fixed effects are included. Standard errors are clustered at the province level.

Figure B.20: MVPF for Different Policies



Notes: The figure shows the MVPF of various policies as reported by the *Policy Impacts* website. Different policy types have been plotted using different symbols and colors. The MVPF for the introduction of PhD programs is represented by the only solid symbol, as a solid red circle.

C Appendix Tables

Table C.1: High School Track Requirements by University Course

	High School Track							
University Course	Classics	Sciences	Teaching	Arts	Agriculture	Commerce	Industrial	Nautical
Law	X							
Humanities	X							
Pol. Sci.	X	X						
Medicine	X	X						
Engineering	X	X						
Sciences	X	X						
Industrial Chemistry	X	X						
Pharmacology	X	X						
Veterinary Med.	X	X						
Agricultural Sci.	X	X			X			
Economics	X	X			X	X	X	X
Statistics	X	X			X	X	X	X
Naples' Nautical Institute	X	X			X	X	X	X
Naples' Oriental Institute	X	X	X		X	X	X	X
Architecture	X	X		X				

Notes: The table shows the possibilities to enroll into university courses for high school graduates based on their high school track of choice before the 1961 reform.

Table C.2: Descriptive Statistics on PhD Programs

Starting Year	1983	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
N. PhD Positions	2100	2887	2528	3904	3956	3982	3253	3973	3969	4328	4496	4226
N. Programs	488	722	798	993	1090	1164	842	1110	1228	1429	1559	1663
N. Universities	36	42	44	47	48	51	51	53	54	56	56	58
Share of PhD Position by Field (%)												
Computer Science	2.29	2.22	2.18	2.10	2.28	2.34	2.89	2.62	2.57	2.89	2.98	2.44
Maths	4.52	2.98	3.28	3.28	2.78	3.31	4.24	3.42	3.43	3.05	3.05	3.17
Physics	5.14	5.02	5.06	4.46	4.25	4.47	5.35	4.53	4.51	4.39	4.31	4.14
Chemistry	7.05	7.14	7.32	6.4	6.14	6.18	6.33	5.89	5.92	5.89	5.74	4.90
Earth Sciences	3.33	3.33	3.68	2.92	2.83	2.91	3.44	2.74	2.65	2.96	2.80	2.77
Biology	10.67	10.29	9.93	9.99	9.83	9.42	7.78	8.91	8.94	9.17	9.76	10.15
Medicine	15.90	17.56	16.73	15.93	15.57	12.96	4.8	12.26	11.82	11.92	12.41	11.59
Agricultural Sciences	4.00	5.72	4.67	4.82	4.95	4.75	5.29	4.46	4.41	5.38	5.76	5.11
Veterinary Medicine	1.29	1.84	1.58	1.82	1.62	1.68	1.69	1.91	1.59	1.80	1.94	1.66
Civil Eng. & Architecture	5.43	6.30	6.61	6.94	7.61	7.89	9.07	8.63	8.31	8.32	8.23	8.83
Industrial Eng.	10.52	10.84	10.76	11.63	12.21	12.15	13.80	12.01	12.50	12.78	12.10	12.47
Classics, Linguistics & Art	9.33	8.21	8.62	9.71	9.30	9.49	9.93	9.11	8.97	8.76	8.79	9.28
History, Philosophy & Psychology	8.05	6.41	7.00	6.33	6.24	6.88	7.10	7.27	7.68	6.7	6.67	6.79
Law	6.52	5.82	6.09	7.38	7.96	8.34	9.71	8.66	9.07	8.43	8.12	8.61
Economics & Statistics	5.52	6.23	5.97	6.48	7.43	8.26	9.01	7.95	8.16	8.29	8.19	8.52
Political Science & Sociology	3.14	2.70	2.97	2.84	2.55	2.46	2.37	2.89	2.77	2.70	2.65	2.63

Notes: The table reports descriptive statistics for the PhD programs opened between 1983 and 1998, by starting year of the programs.

Table C.3: Descriptive Statistics on Patent Activity

Year Interval	1976-1980	1981-1985	1986-1991	1992-1996	1997-2001
N. Patents	25079	22389	16930	27598	40113
Share of Patents by IPC section (%):					
A: Human Necessities	20.60	22.02	23.38	24.31	24.61
B: Performing Operations	32.28	32.46	32.91	35.94	34.05
C: Chemistry & Metallurgy	10.05	12.45	12.36	11.63	8.62
D: Textiles & Paper	4.31	4.72	5.45	5.05	4.51
E: Fixed Constructions	9.00	7.64	7.29	7.76	7.93
F: Mechanical Engineering	16.08	14.15	13.43	12.42	12.84
G: Physics	8.05	8.19	9.39	9.41	9.52
H: Electricity	6.90	7.06	8.34	8.21	7.97

Notes: The table reports descriptive statistics for the patents published between 1976 and 2001 by year of application. The years have been grouped into 5 groups for clarity's sake.

Table C.4: Descriptive Statistics on PhD Applicants

Field	Biomedicine	Civil Eng. & Architecture	Industrial Eng. & CS	Sciences	Non STEM
N. Applicants	4766	1321	1960	4327	5587
N. Positions	3176	795	1446	2439	2920
Share of Female Applicants (%)	59.27	37.12	19.32	38.66	49.87
Share of PhD Graduates (%)	55.23	49.48	55.79	59.29	49.39
Share of Inventors (%)	1.77	0.72	4.57	3.46	0.21
Avg. N. Patents (per 100 People)	3.56	1.17	7.74	7.82	0.25
Share of Applicants by Uni (%):					
Rome La Sapienza	43.28	84.48	52.52	25.99	29.95
Milan La Statale	30.16	0.00	5.74	28.81	32.31
Padua	15.67	7.72	28.48	23.44	23.57
Catania	8.95	7.49	2.93	9.76	9.39
Genoa	1.94	0.31	10.33	12.00	4.79

Notes: The table presents some descriptive statistics on the applicants to PhD programs. When an individual was present multiple times, a weight equal to the inverse of their number of applications was applied.

Table C.5: Data Sources

Source	**Information**
Main datasets:	
Official Gazette of the Italian Republic	PhD programs' university, subject, number of spots *before 1999*
University Ministry *Cerca Università* website	PhD programs' university, subject, number of spots *from 1999*
Bianchi and Giorcelli (2019)	Italian patents' inventors, assignees, locations, and IPC classes
Italian Patent Office (IPO)	Italian patents' inventors
European Patent Office (EPO)	Italian patents IPC classes, citations from foreign patents
National Central Library of Florence (BCNF)	PhD graduates' name, subject, enrollment year, graduation year
Universities' PhD archives	Names, subject, score, ranking of applicants in PhD entry exams
ISTAT - 1981 General Population Census	Total population, % graduates among people aged 25-34
ISTAT - 1981 Census of Industry	Industrial composition, % workers in firms with 500+ w.s, % STEM R&D workers, % non-STEM R&D workers, % workers in IT
Additional datasets:	
SCOPUS Database	Aggregate local and university departments' publications and citations
ProQuest Dissertations & Theses Global	US PhD dissertatons' abstract and subject
United States Patent and Trademark Office (USPTO)	Patents' texts and IPC classes
Kelly et al. (2021)	US patents' novelty
CERVED	Italian limited liability firms' name, location, industry
ISTAT Survey on Graduates' Employment (1989-2004)	University graduates' studies and employment outcome
ISTAT Yearbook on University (1983-1998)	Number of graduates for each course and university
ISTAT 1982 Survey on Scientific Research	Public R&D expenditure by topic, R&D expenditure and workforce by industry
Law: D.M. 11 settembre 1998, MURST	Sholarship amount
ISTAT Historical Series' on R&D	National university expenditure and workforce for research, aggregate R&D expenditure
Sapienza University of Rome Balance Sheets	Administrative cost of PhD programs
LSEG Datastream	Italian publicly listed firms' market capitalization and price
Laws: D.P.R. 29 settembre 1973, n. 597; D.P.R. 22 dicembre 1986, n. 917	Labor tax schedule
Laws: D.P.R. 29 settembre 1973, n. 598; D.P.R. 22 dicembre 1986, n. 917; L. 23 dicembre 2000, n. 388; L. 23 dicembre 2000, n. 388; D.L. 12 dicembre 2003, n. 344; L. 24 dicembre 2007, n. 244	Corporate tax schedule

Notes: This table lists all data sources used in the paper and the specific information used in the analysis.

Table C.6: Effects of PhD Programs on Local Innovation

Method	Dynamic TWFE	Baseline Method		SDID
	(1)	(2)	(3)	(4)
Treatment	STEM PhDs	STEM PhDs	STEM PhDs	STEM PhDs
Effects of PhD Programs:				
β_{POST}	0.1195	21.51**	0.1952***	0.1265***
	(0.0782)	(9.65)	(0.0763)	(0.0297)
β_{+15}	0.2841*	38.44*	0.3171**	0.1987***
	(0.1467)	(19.74)	(0.1314)	(0.0615)
Mean of DV, Treat=1	0.7232	71.06	0.7793	0.7232
Control Variables:				
Province FE		X		
Year FE		X		
Province-Tech FE	X		X	X
Tech-Year FE	X		X	X
Full Controls	X	X	X	X
Obs	177,944	1,508	162,604	177,944

Notes: Column (1) is an OLS estimate of a TWFE model. Columns (2) and (3) are OLS estimates of Equation 1. Column (4) uses the method of Arkhangelsky et al. (2021). One observation is a province-technology-year cell in columns (1), (2), and (4); a provnce-year cell in column (3). The dependent variable is the number of patents produced in a given cell. In column(3), all treated provinces from the imputed treatment cohort 1988 are excluded. All columns control for local characteristics in 1981 interacted with year fixed effects: industrial composition using the 17 one-digit 1991 ATECO industries, local population, share of workers employed in firms with 500 or more employees, share of workers in IT, share of workers in STEM RD, share of workers in non-STEM RD, and share of university graduates among people aged between 25 and 34. Standard errors are clustered at the province level and reported in parentheses. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

Table C.7: Effects of PhD Programs on Local Innovation

Type of Analysis	Baseline	Baseline	No Big Cities	No New Universities	No Neighbours
Dependent Var.	Patents per 1M People	Patents per 1K Firms	N. Patents	N. Patents	N. Patents
	(1)	(2)	(3)	(4)	(5)
Treatment	STEM PhDs	STEM PhDs	STEM PhDs	STEM PhDs	STEM PhDs
Effects of PhD Programs:					
β_{POST}	0.2976***	0.0400***	0.1941***	0.3287***	0.1927***
	(0.0932)	(0.0114)	(0.0642)	(0.0790)	(0.0534)
β_{+15}	0.4724***	0.0630***	0.2741**	0.3797***	0.3094*
	(0.1556)	(0.0186)	(0.1116)	(0.1391)	(0.1637)
Mean of DV, Treat=1	0.7294	0.1041	0.4140	0.7987	0.8179
Control Variables:					
Province-Tech FE	X	X	X	X	X
Tech-Year FE	X	X	X	X	X
Full Controls	X	X	X	X	X
Obs	177,944	177,944	165,672	144,196	119,652

Notes: All columns present OLS estimations of Equation 1. Column (1) and (2) use patents per 1 million people and patents per 1 thousand firms as outcome instead of the number of patents, respectively. In column (3) the 4 biggest Italian cities (Milan, Naples, Rome, and Turin) are excluded from the analysis. In column (4), all treated provinces with provinces founded between 1983 and 2001 are excluded. In column (5), all treated provinces neighboring provinces in the control group are excluded from the analysis. Standard errors are clustered at the province level and reported in parentheses. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

Table C.8: Birthplace of Applicants

Birthplace	Same Province	Other Treated	Control Province	Other
Sapienza Applicant	944.5	601.5	58	195
	[0.53]	[0.33]	[0.03]	[0.11]
Statale Applicant	469	484	10	129
	[0.43]	[0.44]	[0.01]	[0.12]
Total	1413.5	1085.5	68	324
	[0.49]	[0.38]	[0.02]	[0.11]

Notes: The table shows the birthplace of the subset of applicants to PhDs offered by Sapienza University of Rome and Statale University of Milan for whom this information was available. The share of applicants born in a specific area is reported within square brackets. Other includes both foreign-born applicants and individuals born in provinces excluded from the analysis. Observations has been weighted inversely according to the number of total applications made by individuals across the two universities.

Table C.9: Effects of PhD Programs on Individuals' Innovation

PhD Category	Applied	Non-Applied
	(1)	(2)
Effects on $P\left(Inventor_i\right) : \hat{\tau}_I$	0.0680**	-0.0348
	(0.0340)	(0.0627)
Effects on $Patents_i$: $\hat{\tau}_P$	0.1902***	-0.4627
	(0.0704)	(0.2832)
First Stage: $\hat{\delta}$	0.2303***	0.1701***
	(0.0368)	(0.0395)
Control Variables:		
University FE	X	X
Field FE	X	X
Gender FE	X	X
Main Bandwidth	0.3346	0.2212
Bias Bandwidth	1.0724	0.6800
Effective Obs.	1688	1410
Total Obs.	5440	6599

Notes: All columns present estimations of Equation 8. Each column shows the results of completing a PhD program in an applied or non-applied STEM subject. A subject is defined as non-applied if the relative majority of public research funds directed to the subject are used for theoretical research. The models have been estimated using the *rdrobust* package and the bandwidth have been selected through the *rdbwselect* (Calonico et al., 2017). The plotted coefficients are bias-corrected and the reported confidence intervals are calculated using robust bias-corrected standard errors clustered at the individual level. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).

Table C.10: Balance Test by STEM PhD Category

Covariate	**Applied**	**Non-Applied**	**Covariate**	**Applied**	**Non-Applied**
Sapienza Applicant	0.02	0.05	Biology	-0.05*	0.04
	(0.05)	(0.05)		(0.03)	(0.03)
Statale Applicant	0.05	-0.08**	Medicine	0.17***	0.15***
	(0.04)	(0.04)		(0.04)	(0.02)
Genova Applicant	-0.01	-0.01	Agricultural Sci.	-0.01	0.00
	(0.02)	(0.02)		(0.02)	(0.01)
Catania Applicant	0.03*	0.04**	Veterinary Med.	0.02***	.
	(0.02)	(0.02)		(0.01)	.
Padova Applicant	-0.04	-0.04	Civil Eng. & Architecture	0.00	0.02
	(0.04)	(0.03)		(0.03)	(0.02)
Female Applicant	0.08***	0.06*	Industrial Eng.	-0.03	0.00
	(0.03)	(0.03)		(0.03)	(0.01)
Physics Applicant	-0.02	-0.21***	Electronic Eng.	-0.07***	0.02*
	(0.02)	(0.04)		(0.03)	(0.01)
Chemistry Applicant	0.02***	0.00	Computer Science	-0.02	0.00
	(0.01)	(0.02)		(0.01)	(0.01)
Geology Applicant	-	-0.01			
	-	(0.02)			

Notes: The table shows the results of the balancing tests for each of the covariate used in the RDD. The models have been estimated using the *rdrobust* package and the bandwidth have been selected through the *rdbwselect* (Calonico et al., 2017). No coefficient is reported for geology and applied PhDs, and veterinary medicine and non-applied because no program from that field was present in the category. The standard errors reported in parentheses are robust bias-corrected standard errors clustered at the individual level. (*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$).