74th Economic Policy Panel Meeting

21-22 October 2022

The Effectiveness of Promotion Incentives for Public Employees: Evidence from Italian Academia

Marco Nieddu (University of Cagliari)
Lorenzo Pandolfi (University of Naples Federico II)

The Effectiveness of Promotion Incentives for Public Employees: Evidence from Italian Academia*

Marco Nieddu[1] and Lorenzo Pandolfi[2]

[1]Department of Economics and Business, University of Cagliari & CRENoS

[2]Department of Economics and Statistics, University of Naples Federico II & CSEF

September 2021

Abstract

We investigate how promotion incentives affect the productivity of a large sample of high-skilled public employees: academics. In a fuzzy regression-discontinuity design, we exploit the three *bibliometric* thresholds of the 2012 National Scientific Qualification (NSQ), the centralized evaluation procedure regulating career advancements in Italian universities. We compare the 2013-2016 research productivity of assistant professors barely qualified for associate professor – whose next goal becomes meeting the higher thresholds for the full professor qualification – with the productivity of candidates who barely miss the qualification – whose goal remains meeting the associate professor thresholds. We find that barely qualified scholars publish significantly more papers than their non-qualified colleagues, in journals of comparable quality. Our results emphasize the importance of promotion incentives as an effective incentivizing tool in the hands of public management.

JEL Classification: I23; J45; M51; O31

Keywords: promotion incentives; public sector; academia; Italy; scientific productivity

*We are grateful to two anonymous reviewers and Roberto Galbiati (the Editor) for their helpful and constructive comments, as well as to Albrecht Glitz and Gianmarco León for numerous insightful discussions and suggestions. This paper greatly benefited from helpful comments by Antonio Ciccone, Niklas Heusch, Tullio Jappelli, Maria Lombardi, Giovanni Mastrobuoni, Roberto Nisticò, Tommaso Oliviero, Marco Pagano, Michela Ponzo, Marta Santamaria, Annalisa Scognamiglio, Vincenzo Scoppa, Saverio Simonelli, Alberto Zazzaro, seminar participants at the Center for Studies in Economics and Finance (CSEF), Universitat Pompeu Fabra, University of Cagliari, and Tor Vergata University of Rome, and participants at the 2018 Petralia Sottana Applied Economic Workshop, the 2018 European Winter Meeting of the Econometric Society, and the 2019 Padova Applied Economics Workshop "Economics of Education, Retirement and the Household".

I Introduction

Public workers account for a large share of total employment in most countries. In OECD economies, approximately one in five workers is employed in the public sector (OECD, 2017). In many cases, public workers are also the main providers of some critically important goods and services, such as healthcare, education, and research. The productivity of public servants is therefore key for economic growth and welfare. Yet, the set of tools available to public management to incentivize employees is typically limited. For instance, pay-for-performance schemes – which are ubiquitous and found to be largely effective in the private sector – are often difficult to implement in the context of the public sector, where compensation schemes are rigid and do not easily allow for the inclusion of discretionary performance-based components (Finan et al., 2017).[1] Public workers' salaries are usually determined in a centralized manner and increase with seniority or career advancements. Promotion-based incentives thus constitute the most widespread motivational lever in public organizations. Nevertheless, there is little evidence on the effectiveness of this type of incentives on the productivity of public servants. Analyzing this important issue is indeed empirically difficult. The major challenge is finding an appropriate setup in which it is possible to measure and compare the productivity of similar public servants exposed to heterogeneous promotion incentives.

In this paper, we study how a quasi-random assignment of different career prospects – which entails a variation in promotion incentives in the form of heterogenous expected future promotion thresholds –, affects the productivity of a large sample of high-skilled public employees, *i.e.*, about 5,000 (already tenured) assistant professors in Italy. The features of the current Italian public university system make it an appropriate laboratory to analyze the role of career-related incentives in the public sector. The Italian public university system is indeed characterized by a clear hierarchical structure and a centralized promotion mechanism based on observable measures of individual scientific productivity. Moreover, salaries are determined at the national level and increase substantially with academic rank, which makes promotions the main source of incentives for scholars

[1] The literature on the effectiveness of financial incentives in the private sector includes, among others, Lazear (2000), Gaynor et al. (2004), Shearer (2004) and Friebel et al. (2017). See also Lazear and Oyer (2013) for an exhaustive review of related studies.

employed in public universities (Haeck and Verboven, 2012). In addition, scientific productivity of academics is a topic of particular importance per se. The production of knowledge is indeed a key driver of economic development and, in many countries, the main provider of research and education is the state. However, while there is some evidence on the determinants of productivity in the scientific production process, little is known about how academics respond to different recruitment and promotion schemes.

To examine the effect of promotion incentives on research productivity, we exploit the introduction, in 2012, of a centralized evaluation procedure awarding the eligibility for career advancements in Italian universities, namely, the *National Scientific Qualification* (henceforth NSQ). We take advantage of a peculiar feature of this procedure: success in the NSQ depends on scholars' research productivity, measured by three *bibliometric* indicators that are required to be above certain observable and well-defined thresholds which vary depending on whether a candidate is applying for the associate or the full professor qualification, and across academic fields. This multidimensional cutoff rule offers two key advantages from an empirical point of view. First, it gives scholars observable and well-defined promotion thresholds that have to be met in order to advance in the academic hierarchy.[2] Second, for scholars in the neighborhood of the frontier, it constitutes a source of quasi-random variation in career prospects, which we exploit in our analysis.

For assistant professors who apply for the associate professor qualification in 2012, earning it sets a new attainable goal to be achieved in the near future: the qualification for the full professorship. Achieving such qualification requires overcoming *bibliometric* thresholds which are typically higher than those in the associate professor qualification. Therefore, in order to fulfill this goal, marginally qualified candidates need to enrich their publication records (articles, citations, and h-index) so as to meet the requirements for the full professor qualification. For non-qualified scholars, instead, the goal remains meeting the thresholds for the associate professor qualification. Failure in 2012 indeed implies that a candidate will have to re-apply for the qualification for an associate professorship

[2] Although the NSQ might not have been explicitly designed to introduce career-based incentives, promotions can constitute an "accidental incentive system" (Lazear and Gibbs, 2014, p.262).

in the future.[3] For the marginally non-qualified applicant, reaching this goal does not require a substantial effort increase since her productivity scores are already very close to the relevant thresholds. In principle, those scholars who do not achieve the associate professor NSQ in 2012 might be forward looking, already targeting the requirements for the full professor qualification. However, as they have to postpone their career progressions, they have weaker incentives to meet this goal in a short period of time. Furthermore, they face more uncertainty, as they have to rely upon the stability of the institutional setting over a longer horizon. Hence, the incentives they have to meet the full professor qualification thresholds in the short run are clearly weaker that those faced by qualified scholars.

Hence, by varying the subsequent career goal of assistant professors applying for the associate professor qualification, the outcome of the NSQ also provides a variation in the future expected promotion thresholds they face. For scholars near the cutoffs, this variation is as good as randomly assigned and we exploit it in a multidimensional regression discontinuity design. Our regression discontinuity estimates show that success in the 2012 NSQ has a positive and significant effect on the number of scientific papers published in the subsequent four years (2013-2016). The marginally qualified scholar publishes on average 6 items more than her marginally non-qualified colleague. This effect is sizable, as it corresponds to a 38% increase with respect to the average number of publications in the entire sample. We find that the average publication quality – proxied by different measures of journals' prestige – does not exhibit any discontinuity at the multidimensional threshold. The increase in the quantity of publications thus does not occur at the expense of the average quality. Additionally, our evidence suggests that qualified scholars tend to expand the co-author network and receive more citations, compared to their non-qualified colleagues. When investigating the heterogeneity of our results depending on candidates' gender, we find that women

[3] Although not explicitly ruled out by the institutional setting, in fact, it is extremely unlikely for candidates who are not already qualified for the associate professorship to attain the full professor qualification. Only 3% of assistant professors who failed in the 2012 associate professor qualification then achieved the full professor NSQ by 2016. Conversely, one-third of the assistant professors who qualified for the associate professorship in 2012 earned the eligibility for a full professorship by 2016. Figure A1 in Appendix A depicts in greater detail the career trajectories of applicants to the 2012 associate professor NSQ. Additionally, in Section VI, we show that assistant professors were much less likely to both apply and succeed in the 2012 NSQ for full professor than associate professors with comparable publication records.

who just comply with the cutoff rule are less likely to achieve the qualification than their male colleagues, which may suggest the presence of gender discrimination in the evaluation procedure. Yet, male and female candidates exhibit similar responses in terms of publications.

Digging deeper into the heterogeneity of treatment effects, we show that the effect of the qualification on the number of publications is heterogeneous, depending on scholars' distance from the (expected) future promotion thresholds. The relation between the increase in productivity and the distance from the minimum number of articles needed to get the full professor qualification is inverted-U shaped: the effect is the strongest for candidates in the middle of the distribution, for whom the minimum requirements to earn the qualification for a full professorship are not too close nor too far. This result is consistent with the view that incentives are mostly effective when "the promotion is possible, but neither too hard to achieve, nor too easy" (Lazear and Gibbs, 2014, p.269). This is in line with several theoretical studies on career incentives, that predict effort to be a non-monotonic function of the distance from the promotion bar faced by workers (Hansen, 2012; Stepanov, 2020): for workers who are either far above or far below the bar, the returns to effort – in terms of the increase in the probability of being promoted – are very low. Instead, workers facing relatively close promotion bars have the highest incentives to exert effort in order to meet this goal.

Besides shedding light on the heterogeneity of treatment effects, this evidence lends support to our empirical strategy. The outcome of the NSQ might in principle affect the research productivity of candidates through channels other than promotion incentives. For instance, succeeding in the NSQ might have a positive motivational effect on qualified scholars, or it might increase their ability to attract research funds and high-quality collaborators. These alternative channels, however, would hardly explain the inverted-U shaped relationship between the increase in publications and the expected future promotion threshold, which is specifically related to the intensity of the promotion incentives implied by the qualification.

To further assess the validity of our results and corroborate their interpretation in light of the proposed promotion incentives mechanism, we provide several additional pieces of evidence. In particular, we take advantage of the longitudinal dimension of our data to show that: i) the observed discontinuity in the post-2012 research productivity is not driven by a decline in performance of discouraged, non-qualified candidates, but rather by an increase in the productivity of qualified scholars; ii) candidates on the two sides of the threshold are comparable to each other in terms of pre-2012 research productivity, as no discontinuity in publication quantity and quality emerges in the years preceding the NSQ; iii) the increase in productivity of qualified candidates occurs already in 2013, prior to the actual promotion to associate professor. This suggests that the variation in teaching duties, research funds, and other aspects of scholars' routine that may change with actual career advancements are not the main driver of our results. In addition, we replicate our analysis within the sample of associate professors applying for the full professor qualification in 2012. Consistent with the fact that promotion incentives vanish once the top ladder of the academic hierarchy is reached, we do not observe any discontinuity in the post-2012 productivity of qualified and non-qualified candidates in this alternative sample. Similarly, when looking at the entire sample of applicants to the associate professor NSQ – thus including also those who are not tenured assistant professors in Italy – we find that the outcome of NSQ has no impact on the productivity of researchers outside Italian academia, who work under different incentive schemes. Lastly, we exploit the heterogeneity in the share of candidates qualified for an associate professorship in each department on the total number of already employed associate professors to show that competition for vacancies among qualified candidates is not driving our findings.

This study contributes mainly to the personnel economics literature and, more precisely, to the stream of studies focusing on the design of incentives in the public sector. As highlighted by Finan et al. (2017), public sector pay schemes are typically flat, with salaries that are mechanically determined by seniority and position and rarely linked to workers' performance.[4] Also because of this, most of the related studies focus on performance-pay programs implemented in the context of randomized control trials. For instance, Muralidharan and Sundararaman (2011) evaluate the impact of a randomized performance-pay program in India and find that linking teachers' pay to

[4]The same argument is also made by Khan et al. (2018) in their study on the efficacy of performance-based job posting schemes for government employees.

students' test scores has a positive effect on learning. In another randomized experiment, Duflo et al. (2012) show that performance-related pay lowers absenteeism among Indian teachers, which in turn translates into better student achievements.[5] This literature also highlights a potential pitfall of implementing performance-pay schemes in the public sector: as public jobs typically involve multiple tasks, financial incentives based on the performance in a specific task can reduce workers' effort in another (Baicker and Jacobson, 2007; Glewwe et al., 2010).

Therefore, while some studies have examined the effectiveness of monetary incentives on the productivity of public employees, the existing empirical literature has overlooked the role of promotion incentives. These constitute a complicated subject of study – even in the private sector – since promotion incentives can hardly be implemented in the context of randomized controlled trials, and quasi-experimental evidence on the topic is rare.[6] To our knowledge, there are few studies that explicitly focus on the relationship between promotion incentives and workers' productivity in the public sector. Among them, Karachiwalla and Park (2017) exploit the Chinese system regulating teachers' career advancement to test the prediction of a tournament model of promotions and show that promotion incentives are associated with higher levels of performance. Bertrand et al. (2020) look at career incentives for Indian bureaucrats, and show that career incentives matter the most for relatively young workers, who have sufficient time to reach the top of the hierarchy before retirement, and more years to spend in top, better paid, positions. Deserranno et al. (2021) instead show that meritocratic promotions increase the productivity of public workers in Sierra Leone, especially when career advancements are expected to entail larger wage increases. Closely related to ours are also the studies by Checchi et al. (2014) – who develop a model of career concerns in academia whose predictions are consistent with data on the publications of Italian scholars from 1990 to 2011 – and Haeck and Verboven (2012) – who analyze the promotion dynamics in a sample of European universities and show that these are characterized by the presence of internal labor markets.[7] Unlike these studies, we exploit quasi-experimental variation in promotion incentives –

[5] Other studies focusing on the role of financial incentives in the public sector are those by Lavy (2002), Gertler and Vermeersch (2013), Dal Bó et al. (2013), Olken et al. (2014), and Khan et al. (2016). See Finan et al. (2017) for an exhaustive review.

[6] Empirical studies on career-based incentives in the private sector include Kwon (2006) and Campbell (2008).

[7] Additionally, Ashraf et al. (2018) study the role of career prospects on the recruitment of workers in the public sector

coming from explicit promotion thresholds implemented within an actual governmental policy – to uncover the causal link between promotion incentives and the productivity of public employees.

More broadly, given our focus on academia, this study also speaks to the literature focusing on the research productivity of scholars. Some studies in this stream of literature investigate the relative importance of human and physical capital in the scientific production process (Waldinger, 2016), the relevance of peer effects in science (Waldinger, 2012; Borjas and Doran, 2015b), and the importance of having access and being exposed to the scientific research frontier (Iaria et al., 2018). Some others focus on the effect of achieving tenure (Faria and McAdam, 2014) or prestigious awards (Azoulay et al., 2014; Borjas and Doran, 2015a) on scholars' research outcomes. We contribute to this literature by uncovering the responsiveness of scholars to hiring and promotion schemes explicitly based on past research performances.

Finally, our paper is also related to the recent literature focusing on the centralized evaluation systems that have been introduced in the last decade in several European countries to regulate access to public university positions. Similar to the Italian NSQ are, among others, the *Acreditación* in Spain and the *Habilitation à diriger des recherches* in France. All the related studies focus on the functioning of the evaluation process and, more specifically, on the role of gender (Bagues et al., 2017; De Paola and Scoppa, 2015; De Paola et al., 2017) or of direct connections between evaluators and candidates (Zinovyeva and Bagues, 2015). None of them examines the potential implications for scholars' productivity, as we do in this study.

The rest of the paper is organized as follows. Section II describes the regulatory framework and the key features of the NSQ. The data used for the empirical analysis and the identification strategy are reported in Sections III and IV, respectively. We then present the first-stage estimates in Section V. Section VI presents the results of our analysis together with some robustness tests. In Section VII, we report the results on the heterogeneity of our treatment effects depending on the intensity of promotion incentives, and discuss potential alternative channels. Finally, Section

(health) in Zambia and on the quality of the service delivered. However, different from ours, this study focuses on the effect of career prospects on the selection of workers rather than on the productivity of workers already employed.

VIII concludes.

II Institutional Setting

In 2010, the Italian Ministry of Education, University and Research (MIUR) deeply reorganized the public university system through the so-called Gelmini reform. This reform introduced a new recruitment and promotion system regulating the access to the two top ranks of the academic hierarchy: the associate professorship and the full professorship.[8] Until that time, the hiring and the promotion processes were decentralized and each academic department had complete discretion over the selection procedure. Since the reform came into force, however, earning an associate or full professorship is conditional on having achieved a qualification – *Abilitazione Scientifica Nazionale* – which is awarded by national committees in a centralized evaluation process. The first round of the NSQ took place in 2012 and was followed by three more rounds, in 2013, 2016-2018, and 2018-2020.[9]

By achieving the NSQ, scholars gain the mere eligibility for associate and full professorships, while actual hirings still occur at the university department level through decentralized public competitions. Coherent with the rationale of the reform, which is to promote research activity and limit local favoritism, the introduction of the NSQ restricts the access to these competitions to candidates whose academic *curriculum vitae* satisfies minimum standards established at the national level. Applicants to the NSQ are evaluated by a committee of five scholars randomly drawn from a list of eligible full professors affiliated with Italian and non-Italian universities. The evaluation criteria vary depending on whether a candidate is applying to the NSQ for the associate or the full professorship and her research field. Academic fields are mapped into 184 different competition sectors, grouped into 14 broader disciplinary areas divided into 2 main groups: *bibliometric* sectors

[8] The hierarchical structure of Italian universities consists of three main ranks: assistant, associate and full professors. Until 2010, the three positions were all tenured and assistant professors were hired under permanent contracts. After the reform, instead, assistant professors are hired under fixed-term contracts. Starting yearly gross salaries for (tenured) assistant, associate, and full professors (as of 2012) are 34,898€, 50,831€, and 72,430€, respectively (source: www.units.it/intranet/tabelle_stipendiali). Thus, career progressions are associated to considerable salary increases.

[9] Differently from the previous two, the 2016-2018 and the 2018-2020 rounds consist of multiple calls which were opened every four months over a two-year time window.

– which include the disciplines of mathematics, physics, chemistry, earth sciences, health sciences, agronomy and veterinary, engineering and architecture, and psychology – and *non-bibliometric* sectors – which include humanities, economics, political sciences, and law.

The committee is in charge of screening the items of each candidate's *curriculum vitae* in order to decide whether to award the NSQ or not. The main aspect that committees must take into account when evaluating a candidate is her publication record, measured by three observable and well-defined productivity indicators. In *bibliometric* sectors, these indicators are i) the number of articles published in scientific journals, ii) the number of citations received, and iii) the scholar's h-index. In *non-bibliometric* sectors, the indicators are i) the number of monographs, ii) the number of book chapters and articles in scientific journals, and iii) the number of articles published in a sector-specific list of A-ranked journals. All indicators are calculated over the ten years prior to the NSQ call and normalized by a candidate's academic age. Other criteria include the participation in national or international projects, editorial activities, fellowships, and awards. Although the MIUR allows committees to decide autonomously the weight assigned to each of the aforementioned elements, it explicitly states that the three productivity indicators should constitute the key criteria.[10]

In particular, the MIUR defines specific minimum thresholds for attaining the qualification in each competition sector. These standards are set by looking at the publication records of associate and full professors already employed in the Italian university system. In order to achieve the associate (full) professor qualification in a *bibliometric* sector, a candidate must score above the median associate (full) professor in her sector in at least two out of the three productivity indicators. A similar one-out-of-three rule is adopted in *non-bibliometric* sectors.[11] These rules represent a (almost) necessary but not sufficient condition to achieve the qualification since committee members

[10]Each committee is composed of four full professors employed in Italian universities and one employed in a university located in a different OECD country. The eligibility requirements for commissioners are similar to those for candidates: when considering the aforementioned productivity indicators, only full professors whose indicators are all above the thresholds for the full professor qualification can be part can be part of the evaluating committee in a given competition sector.

[11]After 2012, this rule has changed and candidates in *non-bibliometric* sectors also have to overcome two of the three thresholds to get the qualification.

might deliver a negative judgement even when all of a scholar's indicators surpass the relevant thresholds. Moreover, they also have the right to deviate from the aforementioned rule by awarding the qualification to candidates who do not comply with the productivity requirements. Nonetheless, this latter possibility is allowed only in case of an extremely positive evaluation of the other elements of the *curriculum*.[12]

Figure I depicts the timeline of the first round of the NSQ. It opened between June and July 2012, when both the call for commissioners and that for candidates were published. In August, the MIUR released the sector-specific cutoff values for the each of the three productivity indicators. The deadline for candidates to apply was set for the 20th of November. After this date, candidates' scores, as calculated by the MIUR, the list of commissioners in each sector – randomly drawn from the list of eligible full professors –, and the evaluation criteria adopted by commissioners were made public. Candidates had the right to withdraw their applications in the two weeks following the publication of committee compositions and evaluation criteria. This option was particularly important since a negative assessment by the committee in the 2012 round of the NSQ implied that a candidate could not apply for the qualification in the same competition sector for at least two years. Thus, applicants could decide whether to undergo the evaluation or not after having observed their scores, the cutoff values, the composition of the evaluation committee and the established criteria. Most committees completed their work and published the outcome by February 2014, while in few cases the evaluation process took until mid-2014.[13]

When the NSQ was introduced, rounds were supposed to take place annually. A second call was therefore open in 2013. Given the short time span between the first and second call, much fewer scholars applied to the latter. Some of them had also applied for the NSQ in 2012 but withdrew the application before undergoing the evaluation. In principle, this might introduce additional fuzziness in our design as we could be considering as *untreated* some scholars who, in

[12] Figure A2 in Appendix A depicts the extent of the deviation from the two-out-of-three rule across *bibliometric* competition sectors. On average, fewer than 15% of candidates who did not comply with the two-out-of-three rule achieved the qualification in 2012.

[13] Importantly, as we discuss in Section III, our dataset covers all applicants to the 2012 call at the time of the application deadline (November 2012), thus also including withdrawn applications.

fact, received the treatment during our period of observation. However, candidates who withdrew the application in 2012 and then earned the associate professor qualification in 2013 are only about 100, and excluding them from our analysis does not change our main results. Finally, after the first two calls, the law regulating the NSQ was further modified to give candidates the possibility to submit their applications at any point in time within some given 2-year windows. The first round open under this new regulation was that of 2016-2018, at the end of our sample period, later followed by another round in 2018-2020.

III Data and Sample Description

In this study, we combine several data sources to build a unique and comprehensive dataset containing, for each candidate for the 2012 NSQ, i) the score in each of the three productivity indicators and the outcome of the qualification procedure; ii) the academic position and affiliation at the time of the call; iii) the complete publication record from 2007 to 2016.

The list of applicants to the 2012 round of the NSQ is obtained from the MIUR website. The administrative records include information about each candidate's application(s), that is, the competition sector, the scores in the productivity indicators, the sector-specific cutoffs, and the final outcome of the evaluation procedure. We merge these data with the 2012-2016 professor census, which covers all assistant, associate, and full professors employed in the Italian public university system.[14] This longitudinal database allows us to determine, for each applicant, her position, department of affiliation, academic field as of 2012, and later promotion patterns. Since the NSQ system allows for multiple applications per candidate, in our baseline specification, we consider each candidate's 'best' application in terms of distance from the cutoffs. We do so because, in order to get an associate professorship, an assistant professor does not necessarily need to get the qualification in the same competition sector in which she is employed at the time of the call. Typically, earning the NSQ in any adjacent sector within the same disciplinary area is sufficient for

[14] A small share of professors in the census are employed in a few private Italian universities which however are subject to almost the same regulatory framework that applies to public institutions as regards hiring and promotions. Professors employed in private universities account for 4% of our baseline sample.

her to gain the eligibility for a promotion to associate professor (which in most cases occurs within the same university where she was employed as assistant professor in 2012).[15].

Our measures of research productivity come from the *Scopus* database, the largest repository of peer-reviewed literature. We query the *Scopus* archives in order to retrieve each scholar's complete publication record. For each item, we obtain the cover type (article, conference paper, book chapter or review), author's affiliation, publication date, journal name, and the full list of coauthors. Then, we use this information to build a panel dataset at the scholar-by-year level, which contain measures of the quantity of publications, the quality of the journals in which they are published, and the citations received. The main journal-specific quality indicator is the 2012 *CiteScore index*, which provides a weighted average of the citations received by each journal in a given year. In order to account for the wide heterogeneity between the different academic fields, we look both at the overall *CiteScore index* and at its within-field counterpart, the *CiteScore journal percentile*. Furthermore, we exploit two alternative measures of journals' prestige: the *SJR* (Scimago Journal Rank) and the *SNIP* (Source Normalized Impact per Paper).[16] The citations received by each published paper are counted as of July 2017.

Overall, there were approximately 40 thousand applicants in the first round of the NSQ for both associate and full professorship. These include tenured academics employed in Italian universities, untenured Italian academics (such as post-docs), and also scholars working in foreign universities and non-academic research institutions. To get our final sample, we first discard applicants in *non-bibliometric* sectors, where the cutoff rule is not strictly enforced by most of the committees.[17] In disciplines such as humanities, law, political sciences, and economics, where the number of

[15] As shown in Section VI, results from our analysis do not change if we consider, for each scholar, the outcome of her application in the competition sector in which she is employed at the time of the call or an application at random among the ones she submits.

[16] More precisely, the 2012 *CiteScore* index is computed as the total number of citations received in 2012 by documents published by a journal in the three years before, divided by the total number of documents published over the same period. The *CiteScore journal percentile* ranks the journals belonging to each field according to their *CiteScore* index. The *SJR* and *SNIP* indicators are computed in similar way to the *CiteScore index*, thus making them a weighted average of the citations received in a given year by documents published in the three previous years. However, weighting procedures differ from those used to construct the *CiteScore* index.

[17] For 3000 out of the total number of applicants in the NSQ it was not possible to identify a unique best application and, consequently, a unique sector. Within our baseline sample of assistant professors applying to the associate professor NSQ in *bibliometric* sectors, we cannot identify a unique application only for less than 4% of applicants.

publications per year is typically lower, the thresholds are often very close – or even equal – to zero.[18] Therefore, since more than 90% of the applicants in *non-bibliometric* sectors satisfy the corresponding one-out-of-three rule, compliance with such rule constitutes a very poor proxy for candidates' quality. Moreover, the resulting lack of observations below the cutoff(s) would not allow us to implement our regression discontinuity design in such competition sectors.

Out of the 20 thousand candidates in bibliometrics sectors – including mathematics, physics, chemistry, earth sciences, health sciences, agronomy and veterinary, engineering and architecture, and psychology – about two thirds applied for the associate professor qualification.[19] Given our focus on promotion incentives, our main sample is made of 7000 candidates who are already employed as tenured assistant professors at the time of the deadline. Since these candidates have permanent positions – until 2010, all assistant professors were tenured in Italy – the outcome of the NSQ affects their career progressions but has very limited impact on their incentive to leave the public university system. Attrition in this subsample is indeed very limited, nor it is affected by the outcome of the NSQ (as we show in Section VI). In other words, these candidates participate in an internal labor market in which careers typically evolve vertically within the organization and mobility is very limited. Other applicants to the associate professor NSQ include a few untenured assistant professors hired after 2010, researchers working for non-university institutions in Italy or abroad, academics affiliated with non-Italian universities, and other scholars affiliated with an Italian university, such as post-docs but also graduate students, adjunct professors, and scholars with multiple affiliations.[20] The wide coverage of *Scopus* allows us to detect a unique author identifier for 97% of the candidates in our main sample. For the remaining 3% of the scholars, it could be either the case that none of their publications are recorded in the database or that homonymies and misspelled names result in an unsuccessful merge. Lastly, in our baseline specification we dis-

[18]Given the way cutoffs are established, a threshold equal to zero means that the score of the median associate (or full) professor in that competition sector, for that specific indicator, is equal to zero. This is often the case for the third indicator in *non-bibliometric* sectors, that is, the number of books published.

[19]To perform a placebo exercise which we report in Section VII, we exploit also data on associate professors applying for the full professor qualification in *bibliometric* sectors. The process for selecting this sample of applicants to the full professor qualification (4,866) follows the one for candidates to the associate professor qualification.

[20]In Section VII, we replicate the analysis on the entire sample of applicants to the 2012 associate professor qualification.

card within-sector outliers, and observations belonging to competition sectors with fewer than 30 applicants thus ending up with a final sample of 4920 tenured assistant professors.[21]

A detailed description of this sample is presented in Table I. A significant share of candidates for the associate professor qualification in 2012 were relatively experienced: the average academic age – that is, the number of years since the first publication appears in *Scopus* – was approximately 16 years. Moreover, they published on average 12.40 papers in the four years prior to the NSQ. Slightly less than 60% of candidates achieved the NSQ, whereas two-thirds of them satisfied the two-out-of-three rule. Lastly, the number of collaborations in our sample is relatively high both before and after the NSQ: only 2% of the papers published by the assistant professors in our sample are single-authored, and the average number of coauthors per publication is 10.82 (8.27) in the four years following (prior to) the NSQ.

Figure II provides a visual representation of the compliance with the two-out-of-three rule. To reduce the dimensionality of the problem, the figure is drawn for the subsample of candidates whose h-index is below the cutoff. Each candidate is located in the figure depending on her distances from the relevant cutoffs for the other two indicators (number of articles and citations). Hence, scholars in the upper-right quadrant are those who satisfy the two-out-of-three rule whereas those in the other three quadrants do not. It is evident that the degree of fuzziness is limited as qualified scholars – blue circles in the figure – are concentrated in the upper-right quadrant. This confirms that the two-out-of-three rule constitutes a determinant criterion for awarding the qualification in *bibliometric* competition sectors. Moreover, the figure shows that the mass of observations concentrates around the multidimensional cutoff, and particularly around the intersection of the zero-distance axes. This finding is not surprising since the threshold values are computed by looking at the median associate professor in each competition sector. As discussed in detail in the following section, this particular feature of our data implies that, although local, the effect is estimated in the neighborhood of the representative scholar in each field.

[21] In Section IV, we explain how we determine outliers and why we eliminate them; additionally, in Section VI, we test the robustness of our results to adopting alternative sample restrictions.

IV Empirical Strategy

We exploit the cutoff rule implemented within the NSQ to determine whether a quasi-experimental provision of promotion incentives significantly affects the productivity of a large sample of academics. More precisely, in a regression discontinuity framework, we compare the post-call research productivity of barely successful and unsuccessful assistant professors who participated in the 2012 NSQ call. While the former can achieve the qualification for a full professorship in the subsequent round of the NSQ, the latter first need to re-apply for the associate professor qualification. Although the system does not explicitly prevent assistant professors from applying directly for the full professor qualification, the probability of succeeding without having already obtained the associate professor qualification is *de facto* very low: while almost one-third of assistant professors who qualified for associate professor in 2012 attained also the full professor qualification by 2016, very few (about 70) unsuccessful applicants in 2012 were able to do so.[22] Success or failure in the 2012 NSQ thus varies the expected future promotion threshold faced by scholars and, therefore, the incentives to enrich their own publication records in the short run. Qualified candidates have the incentive to increase their publications so as to meet the higher eligibility requirements for a full professorship in the subsequent round of the NSQ. Conversely, the goal for barely non-qualified scholars remains meeting the associate professor thresholds, which are, by definition, very close. [23]

IV.a A three-dimensional (fuzzy) regression discontinuity design

Our regression discontinuity strategy exploits the discontinuous jump in the probability of obtaining the qualification, arising when two of the three indicators cross the corresponding thresholds. By fully modeling the two-out-of-three rule with three forcing variables – the productivity indicators – we are able to define a three-dimensional cutoff, that is, a hyperplane that is the $\mathbb{R}^3$ equivalent

[22] This is confirmed by the evidence we provide in Section VI, where we show that assistant professors were much less likely to both apply and succeed in the 2012 NSQ for full professor than associate professors with comparable publication records.

[23] By 'non-qualified scholars', we refer to both the applicants who are denied the qualification and those who withdraw their application before the committee evaluates their applications. In our regression discontinuity framework, the 'barely non-qualified' candidates are the ones whose *bibliometric* indicators are 'almost' at the threshold. Thus, the gap they must fill in order to overcome the minimum requirements for the associate professor qualification in one of the later rounds of the NSQ is close to zero.

of the standard single-variable frontier. Therefore, one should think of the discontinuity in the probability of receiving the NSQ around the three-dimensional frontier as a pooled or combined version of three smaller, single-variable, discontinuities. Since the compliance with the two-out-of-three rule alone does not represent a sufficient condition to achieve the qualification, the probability of receiving the treatment will jump by less than 100% when crossing the multidimensional cutoff. Hence, our empirical strategy relies on a fuzzy regression discontinuity design with three running variables and multiple sector-specific cutoffs.

Formally, let us define the assignment variables – number of articles, number of citations and h-index – as x_{i1}, x_{i2} and x_{i3}, respectively. Then, G_{iks} is an indicator function that equals one when score k of candidate i belonging to competition sector s is strictly above the cutoff m, that is

$$G_{iks} = \begin{cases} 0 \text{ if } x_{iks} \leq m_{ks} \\ 1 \text{ if } x_{iks} > m_{ks} \end{cases} \quad \text{for each } k \in \{1, 2, 3\}.$$

The indicator D_{is} thus describes the aforementioned two-out-of-three rule:

$$D_{is} = \begin{cases} 0 \text{ if } \sum_{k=1}^{3} G_{iks} < 2 \\ 1 \text{ if } \sum_{k=1}^{3} G_{iks} \geq 2. \end{cases}$$

Consequently, our first-stage equation is

$$Q_{is} = \alpha_0 + \alpha_1 D_{is} + f(x_{iks} - m_{ks}) + Z_s + \nu_{iks}, \tag{1}$$

where Q_{is} is an indicator that equals one when a candidate achieves the qualification, $f(x_{iks} - m_{ks})$ is a flexible nonlinear function of the distance of the running variables from the threshold(s) (including 2^{nd} order polynomials of the three variables and their interactions), and Z_s are sector-specific fixed effects. Analogously to a 'canonical' RD design – with a single running variable and single cutoff – the coefficient α_1 measures the discontinuous jump in the probability of achieving the qualification that arises when a candidate complies with the cutoff rule. More precisely, α_1 captures a weighted average of the discontinuity in the probability of achieving the qualification when crossing the frontier hyperplane from all the octants below the three-dimensional frontier,.

This discontinuity in the probability of obtaining the qualification is then used as an instrumental variable to estimate our second-stage equation, which is

$$Y_{is} = \gamma_0 + \gamma_1 \hat{Q}_{is} + f(x_{iks} - m_{ks}) + Z_s + \eta_{iks}, \tag{2}$$

where γ_1 is the local average treatment effect (LATE) of achieving the NSQ in 2012 on any of our measures of scientific production Y_{is}, computed in the post-call period. The corresponding reduced form equation is

$$Y_{is} = \beta_0 + \beta_1 D_{is} + f(x_{iks} - m_{ks}) + Z_s + \eta_{iks}, \tag{3}$$

where β_1 measures the intention-to-treat (ITT) effect of complying with the two-out-of-three rule. The interpretation of γ_1 and β_1 in this multidimensional regression discontinuity framework is analogous to that provided for the α_1 coefficient of the first stage: they capture a weighted average of the effect of crossing the three-dimensional frontier from all the neighboring octants.

Of course, for our estimated coefficient γ_1 to be interpreted as a LATE, the monotonicity assumption must hold, *i.e.*, for each candidate, crossing the multidimensional frontier must (weakly) increase the probability of succeeding in the NSQ. As shown in Figure A2, there are candidates who get the NSQ even though not satisfying the two-of-three-rule, as well as candidates who do not get the NSQ even though meeting this requirement. Within this framework, the validity of our estimation strategy thus requires that: i) the former would have succeeded in the NSQ even if they had met the *bibliometric* requirements; ii) the latter, had they not complied with the *bibliometric* rule, would have still been unsuccessful. In other words, while the former must be always-takers the latter must be never-takers. Given the design of the NSQ, this assumption is very unlikely to be violated since, by law, compliance with the two-out-of-three rule is the most important criterion that commissioners must take into account when making their decisions.

To account for the wide heterogeneity in publication standards, thresholds, and committees' strictness among competition sectors, we allow $f(\cdot)$ to be fully flexible across sectors in equations (1), (2), and (3) by interacting each assignment variable centered around its sector-specific cutoffs, their squared values, and their first and second degree interactions, with the competition sector dummies.[24] Because of both the complexity of the framework and the lack of a standard procedure to compute joint bandwidths in a multidimensional regression discontinuity design with multiple cutoffs, our preferred specification is a fully-parametric one. To reduce the weight of potential outliers, we exclude candidates in the top decile or in the bottom percentile of the distribution of the distance from the cutoff. We do so as most outliers are concentrated in the right tail of the distribution of each indicator. Indeed, while the distribution is left-bounded by the fact that indicators cannot take negative values, it is principle unbounded on the right. Yet, as we show in Section VI, our results are robust to considering alternative sample restrictions, and to using a linear specification within an arbitrary range around the zero-distance cutoff(s).

Finally, it is worth remarking that our identification strategy is less vulnerable to the main criticism usually made for regression discontinuity designs, namely, the locality of the estimated effect. The estimated discontinuity is indeed a weighted average of the discontinuities along the three different frontiers, one for each productivity indicator. Furthermore, cutoff values are set by looking at the median associate professor in each competition sector. As a result, a large mass of observations is concentrated around the three-dimensional frontier – as highlighted in Figure II – and the marginal candidate in this setting is a representative scholar in her field.

[24] Figure A3 in Appendix A shows the extent of across-sector heterogeneity in the cutoff values: in many competition sectors in medicine and physics the median number of articles among associate professors, over the 2002-2012 period, is above 40 articles, while it is often below 10 in mathematics and engineering.

IV.b Validity of the RD design

The validity of our regression discontinuity design relies on two main assumptions: 1) the probability of achieving the qualification jumps discontinuously at the multidimensional cutoff describing the two-out-of-three rule; and 2) the joint distribution of the running variables does not exhibit any jump in the neighborhood of the frontier. Furthermore, in a full-parametric, multidimensional regression discontinuity design, special attention should be devoted to possible misspecification issues (3). While the satisfaction of Assumption 1 is discussed in Section V, we address 2 and 3 here.

Testing the validity of Assumption 2 is crucial to discarding two potential threats for our identification strategy: manipulation and sample selection. Regarding the former, the possibility for candidates to manipulate their publication records in order to meet the minimum standards seems remote since both individual scores and thresholds are computed by the MIUR. The MIUR collects candidates' full publication records from their application webpage and cross-validates each research item by querying the two largest databases of peer-reviewed literature: *Scopus* and *Web of Science*.[25] Moreover, because of the short time frame between the publication of the call and the application deadline, it is unlikely that scholars would have the time to adjust their publication records to meet the established requirements. Regarding sample selection, a positive jump in the density could also reveal that scholars who decide to participate in the NSQ without complying with the two-out-three-rule constitute a selected sample. For instance, one potential concern could be that scholars below the cutoff were disincentivized from applying given that a negative evaluation by the committee would have prevented them from participating in the subsequent round of the NSQ. However, it is important to remark that candidates were given the opportunity to withdraw their application after having observed their precise scores and the composition of the committee, and prior to the evaluation itself. Hence, applying to the 2012 NSQ was relatively costless, even for those below the thresholds, and selection concerns should be limited as our sample of candidates is based on the list of applications at the time of the deadline (November 2012) and thus includes withdrawn applications.

[25] More precisely, a ministerial agency (ANVUR) computes both the individual scores and the thresholds.

To formally test whether the distribution of candidates is discontinuous around the cutoff we follow Cattaneo et al. (2017). In Figure III, we report the frequencies, as well as the density and confidence intervals, separately for each of the three forcing variables centered around the cutoff. Because of the criteria used to define the cutoffs – which are tied to median associate professor in each competition sector – the distribution of candidates is concentrated around the zero-distance cutoffs. However, none of the running variables exhibits a significant (at the 10% level) discontinuous jump in the density in the neighborhood of the cutoff (number of articles: T=1.32, p-val=0.19; number of citations T=0.45, p-val=0.65; h-index: T=-1.26, p-val.=0.21).[26] Further evidence in support of the assumption that scholars do not endogenously sort or select around the threshold comes from the continuity tests that we discuss in Section VI, where we show that marginally qualified and non-qualified scholars are statistically indistinguishable in terms of quantity and quality of publications, number of collaborations and coauthor network size in the period 2009-2011.

Finally, our fuzzy regression discontinuity design also relies on the assumption that scoring above the median professor in two out of the three *bibliometric* indicators should have no impact on future scientific productivity other than that passing through the achievement of the NSQ. It seems, however, extremely unrealistic that other confounding factors or policies could drive the observed jumps at such particular cutoffs.

V First Stage

A crucial condition must hold to implement our empirical strategy: overcoming the *bibliometric* thresholds and satisfying the two-out-of-three rule must result in a discrete jump in candidates' probability of achieving the qualification. In this section, we show that this is indeed the case. In Table II, we report both the estimates of the first-stage equation when considering each of the three *bibliometric* indicators and the corresponding cutoffs, separately (Columns 1 to 6), and when exploiting the three running variables simultaneously, as formalized by Equation (1) (Column 7).

[26] We report the robust bias-corrected manipulation tests obtained using the Stata package described in Cattaneo et al. (2018).

The estimated coefficient from this three-dimensional-RDD – that is, our preferred specification – shows that compliance with the *bibliometric* two-out-of-three rule discontinuously increases the probability of achieving the qualification for an associate professorship by approximately 30 percentage points. The magnitude of the first stage confirms that commissioners attribute a strong weight to the compliance with the two-out-of-three rule when making their decisions.[27]

The single-RDD estimates are also positive and significant in all specifications, consistent with the graphical evidence in Figure IV. In Columns (1), (3), and (6), we estimate the discontinuity in the probability of achieving the qualification when passing each of the three *bibliometric* threshold – the number of articles, the citations, and the h-index – , assuming a quadratic functional form on the entire support and including both competition sector fixed effects and sector-specific interactions. In Columns (2), (4), and (6), we replicate the same estimates assuming a linear functional form within the MSE-optimal bandwidths. In this case, we use as running variables the relative distances from each threshold, that is, the original running variable divided by the threshold itself. By doing so, we are also able to compute three optimal bandwidths, expressed in relative terms, which can be used across the different fields.[28] The estimation results are very close to their fully parametric counterparts. Of course, the magnitude of each single-RDD coefficient is lower than that resulting from the three-dimensional-RDD estimation since the former measures the discontinuous jumps around each single threshold regardless of whether the specific indicator is pivotal for the compliance with the two-out-of-three rule. Hence, estimating three standard, single-forcing variables RDD would not account for the compliance (or defiance) with the other two requirements, thus increasing the degree of fuzziness. This is precisely the reason why we adopt a three-dimensional-RD design, in which the α_1 coefficient of Equation (1) should be interpreted as a combined version of three smaller discontinuities.

[27]The corresponding estimates for the sample of candidates to the NSQ for full professorship are presented in Table A1 in Appendix A.

[28]Specifically, the optimal-MSE bandwidths are computed following Calonico et al. (2014) for each of the three relative distances, separately.

VI Results

VI.a Quantity of publications

Table III reports the main result of our empirical analysis: achieving the qualification for an associate professorship in 2012 – thus being provided with higher promotion incentives – has a positive impact on the number of papers published in the subsequent years. The local average treatment effect (LATE) of achieving the qualification on the number of scientific publications over the 2013-2016 period corresponds to 6.5 publications and is 3.25 times larger than the intention-to-treat (ITT) effect of complying with the two-out-of-three rule (which is equal to 2 publications). Both the LATE and the ITT coefficients are statistically significant at the 1% level. The estimated LATE is sizable and corresponds to approximately 40% of the sample average number of publications over the same period. By looking at the different publication types, we find the effect to be driven mostly by an increment in the number of published articles and, to a smaller extent, reviews and conference papers.

In principle, the estimated effect could be due not only to the increased productivity of barely qualified scholars but also to a decline in publications by narrowly non-qualified candidates. This latter group of scholars might indeed become frustrated and discouraged or could revise their research production function after missing the qualification. In order to disentangle these two hypotheses – the discontinuity being driven by marginal successful or unsuccessful applicants – we first test whether our results are driven by differential attrition at the threshold. Results are given in Table IV and show that this is not the case as the outcome of the NSQ does not affect the probability of exiting the professor census nor that of not publishing any article in the post-NSQ period. This is not surprising given our focus on tenured assistant professors who, having permanent positions, have very little incentives to leave their jobs in case of an unsuccessful application. In fact, attrition in general is very low in our sample: less than 3% of assistant professors in our sample exit the professor census in the period 2013-2016.

We then exploit the panel dimension of our dataset and replicate our baseline estimation using the yearly number of publications before and after the first call of the NSQ as the dependent variable of interest. Figure V, Panel A, reports the estimated LATE of the qualification on the number of publications for each year between 2007 and 2016 (the estimates are also reported in Appendix Table A2). This effect is persistent, large in magnitude, and significant for the whole post-call period, with the only exception of 2015 when the discontinuity is still positive but the larger variance in the data lowers its significance. In Panel B, instead, we depict the predicted number of publications per year of barely qualified and non-qualified candidates (blue diamonds and red circles, respectively).[29] This figure shows that the productivity of barely non-qualified applicants remains relatively constant in the post-NSQ, further confirming that the effect on publications in mostly driven by the increase in the productivity of barely qualified candidates. The evidence in Figure V also lend strong support to our identification strategy. For the entire pre-2012 period, the difference between treated and non-treated individuals is close to a precise zero. Hence, the results are not driven by *ex ante* differences between candidates on the two sides of the frontier, nor by a possible misspecification of the functional form assumed when estimating the relation between the treatment and outcomes.

Although the estimated effect is not statistically distinguishable from zero, a divergence in productivity between barely qualified and barely non-qualified assistant professors seems to emerge already in 2012, before the results of the NSQ are made public. This can be rationalized by some anticipation effects. In late 2012, applicants know both the field-specific cutoffs and their score in each *bibliometric* indicator. As a result, they know whether they meet the *bibliometric* cutoffs or not. It is therefore reasonable that applicants barely above the cutoffs anticipate the higher likelihood of passing the NSQ relative to candidates barely below the cutoffs, and respond to the resulting variation in career patterns and related promotion incentives accordingly.

Of course, this implies that candidates barely above the thresholds manage to increase their publications in a relatively short time period. This is not surprising given that the publication

[29] The former are calculated as the predicted publications of scholars at zero-distances from the cutoffs, and therefore corresponds to a weighted average of sector-specific fixed effects. The latter are the sum of such weighted average and the LATE coefficient in each year.

process in most of the *bibliometric* fields in our sample is typically fast, especially if compared to *non-bibliometric* fields such as economics and business. In chemistry, engineering, biomedicine, physics and earth science, the average time from submission to acceptance (publication) is less than 6 (12) months, much shorter than that in *non-bibliometric* fields such as social sciences, arts and humanities, and business and economics (Björk and Solomon, 2013).[30] In some fields, even top journals have very fast publication times. Sebo et al. (2019) show that, in a sample of randomly selected articles published in 2016 by the highest impact journals of general medicine, the average submission-to-acceptance time was only 123 days, the minimum being 24 days. Additionally, when we look at the year-by-year effect of achieving the NSQ on articles, conference papers, and reviews, separately, we find that almost 40% of the overall increase in publications in 2013 is due to an increase in conference papers published by barely qualified scholars, which are typically faster to produce than journal articles. (estimates are given in Appendix Table A2).

Additionally, we investigate whether the effect of passing the NSQ in 2012 is heterogeneous depending on candidates' gender and disciplinary area. The estimates are reported in Table V and Table VI, respectively. When looking at gender heterogeneity, we find the LATE to be similar for female and male candidates: the promotion incentives associated with the achievement of the qualification are equally effective, regardless of gender. However, we find a negative and significant coefficient for the interaction between the female indicator and the one for compliance with the *bibliometric* rule when estimating the first-stage equation. Hence, women who satisfy the two-out-of-three rule are less likely to achieve the qualification than men with comparable publication records. This result is consistent with the evidence provided by Bagues et al. (2017) and De Paola and Scoppa (2015) – who also document that female candidates have lower success rates in the Italian qualification procedure – and could be due to gender discrimination.

We find moderate evidence of between-disciplines heterogeneity. To ease comparison across fields, in this analysis we express the outcome variable in units of discipline-specific standard deviations. The coefficients of the interactions between our treatment and the dummies identifying the

[30] Björk and Solomon (2013) show that, for instance, the average submission-to-publication delay in chemistry is 9 months, half than that in economics and business (18 months).

ten disciplinary areas are positive (with the only exception of psychology) even though heterogenous in magnitude and not always statistically different from zero. Yet, this evidence shows that our main effect is not driven by a few peculiar fields. Indeed, when excluding one discipline at a time our results remain substantially unchanged (results are given in Appendix Table A3 where we show that estimated LATE ranges between 4.8 publications – when excluding Physics from the sample – to 8.5 – excluding Medicine – being significant at the 5% level in all cases).

VI.b Robustness checks

Our three-dimensional regression discontinuity model is an extended version of the regression discontinuity with multiple assignment variable proposed, among others, by Papay et al. (2011) and Papay et al. (2014). In particular, it is close to what the latter define as the 'Response-Surface RD'. These models depend heavily on a correct specification of the parametric functional form, as the gain in both efficiency and power resulting from multidimensionality comes at the expenses of lower flexibility.[31] Moreover, as for any full-parametric approach, the presence of (within-sector) outliers can bias the estimated coefficients, as all observations are assigned an equal weight irrespective of their distance from the cutoff.

To account for this issue, in our baseline specification, we exclude observations in the top decile and the bottom percentile of the sector-specific distribution of distances from the cutoffs. In this section, we show that our main results are robust to adopting alternative sample restrictions. More specifically, we replicate our analysis varying the lower and the upper bounds of the distribution of distances from the cutoffs, thus progressively excluding candidates whose scores lie outside specified interpercentile ranges. Results from this test – presented in Figure VI Panel A – show that considering a broader or narrower sample does not deliver significantly different estimates for the LATE, unless we include observations in the far right tail of the distribution of the productivity indicators (the top 2%). In this case, the estimated LATE is still positive but lower in magnitude and in some cases barely below the statistical significance level. This is not surprising as the dis-

[31] Since we want to estimate the average treatment effect along the multidimensional borders, we cannot include a two- or three-dimensional spline since, by doing so, we would estimate a very local effect at the intersection of all cutoffs.

tribution of distances has a lower bound (no candidate can score less than zero in each indicator) but virtually no upper bound. Hence, in the very top percentiles of the distribution of distances there are candidates who are really far away from the thresholds, whose inclusion strongly affects the estimated functional form. Furthermore, in our baseline analysis, we exclude small competition sectors with fewer than 30 applicants, in which the field-specific functional form is likely to be imprecisely estimated because of the sparse distribution of applicants. Yet, in VI, Panel B, we show that considering alternative thresholds, in terms of minimum number of applicants, does not alter our main results.[32]

To further address possible concerns owing to the functional form assumed in our baseline estimation, we also replicate our analysis assuming a linear specification in the neighborhood of the thresholds. More precisely, we first normalize each running variable by dividing it by the corresponding cutoff value and then select three different bandwidths, one for each running variable. Finally, we re-estimate Equation (2) on the sample of scholars whose productivity indicators lie within the resulting multidimensional joint bandwidth, assuming a linear specification.[33] Table A4 in Appendix A reports the result of this robustness check and a comparison with our baseline results. The point estimates resulting from this local linear approach are very close in magnitude to those obtained assuming a second-degree polynomial form over the entire support. However, they are less precise, as standard errors are larger. Additionally, we replicate our analysis estimating a series of two-dimensional, nonparametric regression discontinuity designs following the procedure in Cattaneo et al. (2020). For each pair of the running variables, we estimate the effect of crossing the corresponding two-dimensional frontier in different points around the intersection of the zero-distance axes. Results and further details are given in Appendix B. As with the local linear specification, estimates tend to be noisier but in line with those obtained with our preferred, parametric specification. Our main specification with sector-specific interactions indeed allows us to better estimate the effect of complying with the two-out-of-three rule accounting for both

[32]In any case, we have to exclude sectors with fewer than 14 applicants since, in these fields, we have too few observations to estimate a quadratic, field-specific functional form.

[33]The bandwidths for the three productivity indicator are the MSE-optimal bandwidths computed separately for each running variable, following Calonico et al. (2014)

the heterogeneity in the distribution of the productivity indicators across competition sectors and candidates' score in all of the three indicators.

The concern that our findings could be driven by any specification or sorting issue is also ruled out by the evidence on the yearly publications in the pre-NSQ period (reported in Figure V). In Table VII, we show also that marginal applicants on the two sides of the cutoff do not differ in terms of quality of their publications or number of collaborations when these measures are computed over the entire 2009-2011 period. As a further test, we apply a perturbation to each sector-specific threshold. We expect the magnitude of both our estimated first-stage and ITT coefficients to decline and the associate confidence intervals to broaden the farther we get from the original cutoff(s). Specifically, we reshuffle the cutoff values by adding a randomly generated error component $\epsilon \sim N(0, \sigma)$, which is defined as a percentage of the original cutoff.[34] The resulting perturbation, which we impose to lie within plus and minus the 100% of the original cutoff value, then has a different intensity depending on the standard deviation (σ) of the error. We then estimate the LATE from our baseline regression for increasing values of σ, replicating this exercise for 30 different draws from the ϵ distribution. We show in Figure A4 that the magnitude of the effect is the highest in the zero-perturbation case – that is, when using the true threshold values – and decreases in the variance of the perturbation.

As an additional robustness check, we test whether our results hold when using different approaches to deal with multiple applications. Since the rules of the NSQ allow candidates to apply for the qualification in different competition sectors, in our baseline specification, we consider for each candidate her 'best' application, that is, the one in which she scores the highest in terms of distance from the relevant thresholds. Here, we first replicate our analysis considering for each applicant the indicators, the cutoffs, and the qualification outcome in the competition sector to which she already belongs as an assistant professor at the time of the application. Table A5 in Appendix A shows that the effect of achieving the qualification on the number of articles published between 2013 and 2016

[34] We first generate the error $\epsilon \sim N(0, \sigma)$ and then draw from the ϵ distribution in order to assign a different perturbation to the cutoff value of each sector. We do this to account for the between-sector heterogeneity in each of the three productivity indicators.

is still positive and significant under this alternative specification. Coefficients are slightly lower in magnitude, consistent with the fact that, in this case, barely non-qualified candidates might have succeeded in another competition sector. Therefore, a significant share of the candidates below the multidimensional cutoff are actually qualified and consequently exposed to promotion incentives, which makes the discontinuity in terms of post-call productivity smaller. Additionally, we replicate our analysis picking, for each candidate, an application at random among the ones she submits. In Figure A5 we report the results thus obtained under 50 different random draws. Coefficients of both the first stage and the second stage are fairly stable across draws, and all of them are statistically significant at the 5% level (the vast majority being significant also a the 1% level).

Finally, when motivating our empirical strategy we argue that the outcome of the NSQ changes the relevant promotion thresholds because unsuccessful assistant professors need first to re-apply to the associate professor NSQ. In other words, the chances to get the full professor qualification without having already achieved the associate professor one are *de facto* very low. This is consistent with the evidence we provide in Figure A1, where we depict the career patterns of assistant professors participating in the 2012 NSQ. While 30% of those who succeed in the 2012 associate professor qualification earn also the full professor qualification by 2016, only 3% (about 70 candidates) of candidates who fail the associate professor qualification in 2012 do the same. Here, we further corroborate this hypothesis by showing that assistant professors are less likely to succeed in the 2012 full professor NSQ than comparable associate professors. This is true also for the marginal candidate: crossing the multidimensional frontier increases assistant professors' chances of earning the full professor NSQ less than it does for associate professors. When considering the entire sample of potential applicants (thus including also assistant professors who might have applied to the full professor NSQ but did not), we find that being an assistant rather than an associate professor dramatically reduces the probability of success both on average and for marginal applicants (results are given in Appendix Table A6).

VI.c Additional results

After analyzing the impact of passing the NSQ on the quantity of published items, we explore in this section whether it also affects other dimensions of the research activity of the academics in our sample. In particular, we investigate whether any significant discontinuity between (barely) qualified and non-qualified candidates emerges in terms of citations, publication quality and academic network size.

Citations. By replicating our baseline specification using the post-2012 citations received by each scholar as the dependent variable, we find that passing the qualification for the associate professorship also affects scholars' citations. The results in Table VIII show that for papers published from 2013 to 2016, barely successful candidates receive on average 44 citations more than their barely non-qualified colleagues (Column 1). This result can be attributed to both the increased number of publications of qualified scholars and the increase in the average number of citations *per* paper (Column 2). The probability of publishing an article with more than 50 citations (Column 3) or a non-cited article (Column 4) does not exhibit any jump, however.

Thus, scholars who are provided with higher promotion incentives in 2012 not only increase their publications but also manage to improve on another dimension that is taken into account in the qualification procedure: the number of citations received. This effect is in part simply driven by the increased research productivity of qualified scholars but could also reflect an augmented effort to promote and disseminate scientific works, greater visibility following a promotion, or an increase in the average publication quality. This last aspect seems of particular importance and is therefore the next dimension on which we focus.

Average publication quality. The publication quality does not directly enter among the productivity indicators considered in the NSQ but could be indirectly affected by qualified scholars' incentives to maximize both citations and publications in a direction that is *a priori* ambiguous. On the one hand, publishing in better, more prestigious journals can increase a scholar's citations and H-index. On the other hand, there is a potential tension between the quantity and quality of publications, as submissions to prestigious journals are costly, especially in in terms of time, owing to the higher standards required and the more selective review processes. This trade-off could induce qualified scholars to sacrifice the quality dimension in order to minimize publication times and quickly increase their publication records.

We test these hypotheses by replicating our analysis using as dependent variables several alternative measures of a journal's quality and prestige. Specifically, we consider the *CiteScore*, the *Sjr* and *Snip* indexes, and the within-field *CiteScore* ranking – that is, a measure grouping journals according to their position in the field-specific distribution of the *CiteScore* index. According to the results reported in Table IX, Columns (2) to (5), no significant discontinuity in the average publication quality emerges, as all coefficients are not statistically different from zero. Additionally, we test whether the probability of publishing in a journal ranking in the top percentile of the CiteScore index (Column 1) or in a journal with no available measures of quality in the *Scopus* database (Column 6) changes discontinuously at the multidimensional threshold and find that this is not the case.

Hence, the documented increase in publications and citations by barely qualified scholars is not associated with a contemporaneous change in their average publication quality. Importantly, the large increase in the number of publications induced by the provision of promotion incentives does not appear to come at the expense of the average quality.

Co-author network. Finally, we study whether the outcome of the 2012 NSQ has any effect on the number of collaborations or on the size of scholars' co-authors network. In Table X, we report the estimated coefficients from our ITT and LATE equations, using as dependent variables i) the mean and the median number of authors per paper, ii) the probability of publishing a single-authored paper, and iii) the number of distinct co-authors. The three variables are computed for the 2013-2016 period. While the first two outcomes measure how each research paper is produced – that is, whether scholars tend to publish more or less coauthored works – the third proxies for the size of the academic network. We find suggestive evidence of a positive effect of achieving the associate professor qualification on scholars' co-authoring decisions, although the only significant (at the 10% level) coefficient is that for the median number of co-authors. Specifically, the estimated LATE in Column (2) shows that the median paper published by a barely qualified scholar has 2.2 more coauthors than that published by a barely non-qualified scholar. These findings suggest that although some scholars might strategically expand their academic network in order to meet the thresholds for the full professor qualification more quickly, this behavior is not the main driver of the increase in productivity documented in the previous sections.

VII Heterogeneous Treatment Effects and Alternative Mechanisms

VII.a Heterogeneity of Treatment Effects and Incentive Intensity

In this section, we dig deeper into the promotion incentive mechanism and explore whether scholars' reaction to achieving the NSQ is heterogeneous depending on the intensity of the incentives. Thus, we first exploit across-individual differences in the distance between the productivity indicator and the full professor thresholds in 2012, which is the best estimate that a candidate can have about the future thresholds she will face. This distance measures the size of the gap a scholar needs to fill in order to pass the (future) full professor threshold and therefore proxies for her chances of meeting the promotion thresholds in a relatively short time interval. According to the theoretical literature on career incentives, the effort that workers exert in order to clear a given promotion bar should be non-monotonic in their distance from such bar (Hansen, 2012; Stepanov, 2020). This happens because the increase in the chances of meeting the threshold induced by an increase in effort is very small when a worker is either very close to (or even above) the promotion threshold or far away from it. As a consequence, her incentive to exert extra effort is very low, especially when the cost of effort is convex. Instead, workers who are relatively close to the promotion threshold have the highest incentive to increase effort since this largely increases the chances of meeting the bar and thus get promoted. In other words, promotion incentives are most effective when the promotion is "neither too hard to achieve, nor too easy" (Lazear and Gibbs, 2014, p.269).

In our setting, we therefore expect incentives to be low when the probability of obtaining the full professor qualification in the short or middle run is close to zero or to one, that is, when the gap that scholars have to fill is either too large or too small. To reduce the dimensionality of the problem – and consistent with the dependent variable in our main regression (the number of publications post-2012) – we focus on the distance between the first *bibliometric* indicator – the number of articles published over the 2002-2012 period – and its (sector-specific) full professor cutoff. Moreover, in order to account for the heterogeneity in pre-2012 research productivity across sectors, we normalize this distance dividing it by each candidate's number of publications as of 2012. The resulting index therefore varies across candidates and measures the relative increase in publications that a candidate has to produce in order to reach the first full-professor cutoff. It is worth remarking that this index has two main sources of variation: i) differences across fields in the distance between the associate and the full professor cutoffs for the first indicator, *i.e.*, marginal applicants at zero-distance from the first cutoff (the number of publications) may need to fill a larger or smaller gap to reach also the cutoff for the full professor qualification depending on the competition sector; ii) differences in the number of publications of marginal candidates who are at zero-distance from either of the two other cutoffs, even within the same academic field, *i.e.*, marginal candidates in the neighborhood of the other two cutoffs (defined on citations and h-index) can be further or closer to the first full professor threshold depending on their score in the first indicator.[35]

Table XI reports, in its upper panel, the estimates from regressing the number of post-2012 publications on our treatment, interacted with a categorical variable grouping observations in quintiles defined on the above-specified index. In order to ease the interpretation of the magnitude of the quantile-specific effects, the average effect is reported in Column (1). The estimated coefficients of the interaction terms (Columns 2 to 6) show that the relationship between the increase in productivity and the number of publications needed to meet the full professor promotion threshold is inverted-U shaped. The estimated LATE is the lowest in magnitude – especially if compared to

[35] More precisely, the index is defined as $dist_{i,1,s} = \frac{m_{1,s}^{full} - x_{i,1}}{x_{i,1}}$, where $m_{1,s}^{full}$ is the sector-specific cutoff for the first indicator (number of articles) in the full professor NSQ and $x_{i,1}$ is candidate's i score in the same indicator.

the quintile-specific mean of the dependent variable – and is not significantly different from zero in the groups of assistant professors who, in 2012, were either too close or too far from meeting the publication requirements for a full professor qualification. On the contrary, the estimated LATE is much larger and statistically significant in the groups of candidates located in the middle of the relative-distance distribution.[36]

In order to fill the gap with the full professor cutoff, scholars in the third quartile would have to increase their stock of publications by approximately 25%.[37] This goal is realistic in a short- or middle-run horizon. Conversely, scholars in the last quintile would need to almost triple their stocks of publications, a target that is much more difficult to meet in a relatively short time interval. Importantly, this heterogeneity in the effect is not driven by differences across quintiles in the pre-2012 research productivity of candidates (see Table A7 in Appendix A).

In Figure VII, we complement this evidence by providing a graphical representation of the relationship between the effect of achieving the 2012 NSQ and the distance from the full professor qualification threshold. In this case, we exploit also the time dimension of our data. Each point in the figure is the LATE effect of achieving the associate professor NSQ on the productivity in a given year, both before and after 2012, in each of the above-defined quantiles. For all years with the only exception of 2016, coefficients tend to be larger for intermediate values of the distance from full-professor cutoffs, while being close to zero and statistically insignificant when such distance is very large or very small. Moreover, the evidence in Figure VII further validates our previous empirical findings by showing that no significant effect nor heterogeneity emerges in the years prior to the 2012 NSQ.

Finally, the bottom panel of Table XI reports the estimated effect of achieving the associate

[36] Testing the hypothesis that the treatment effect in the third quintile is greater than that in the fifth quintile (*vs.* the null that the latter is greater or equal than the former) delivers a p-value of 0.098 . Similarly, the hypothesis that the coefficient in the fifth quintile is greater or equal than that in the fourth one is also rejected (the p-value is 0.089). However, not all pairwise comparisons deliver results that are significant at the conventional level, mostly because the standard errors of some of the interaction coefficients are large. In particular, testing that coefficients in the third and fourth quintiles are larger than that in the first one delivers p-values of 0.262 and 0.363, respectively, even though coefficient in the first quintile is very close in magnitude to that in the last one.

[37] The stock of publications is computed in the ten years prior to the NSQ.

professor qualification in 2012 on the probability of achieving the full professor qualification by 2016, that is in either of the two following rounds of the NSQ. Column (1) shows that passing the qualification for associate professor in 2012 increases the probability of attaining the full professor NSQ in either 2013 or 2016 from almost zero (0.7%) to about 18%. This result thus supports the view that the outcome of the 2012 represents a quasi-random variation in career prospects. Additionally, the estimates reported in Columns (2) to (6) show that the probability of actually achieving the qualification for a full professorship by 2016 is heterogeneous across the above-defined quintiles. Candidates who were already very close to the full professor cutoff and those who increase their publication records the most after achieving the associate professor qualification are also those who are more likely to effectively achieve the qualification for a full professorship by the end of 2016. Conversely, candidates in the last two quintiles have a much lower likelihood to succeed in the full professor NSQ in one of the following rounds.

Furthermore, according to Bertrand et al. (2020), career incentives should be stronger for younger civil servants, who have higher chances to reach the senior-most positions within the organization, and more time to spend in top, better paid positions.[38] Here we bring this prediction to our data. Since we do not observe candidates' age, we split the sample into older and younger assistant professors, defined on the basis of two proxies, and then look at the heterogeneity of the treatment effect across these two groups. First, we compare assistant professors who were in the professor census since 2000 (the first year covered by the census) and those who got hired afterwards. Second, we compare assistant professor with a below- and above-average academic age, defined as the distance in years between the first publication in Scoups and 2012. Results are given in Table XII. The coefficient of the interaction between our treatment and the dummy identifying more senior assistant professors is negative in both Column (2) and Column (4), even though not statistically different from zero. Although suggestive, this evidence is in line with the prediction in Bertrand et al. (2020).

Taken together, these results show that candidates provided with the strongest incentives are

[38] The same argument is made also in Schneider (2005) in the context of judges.

also those who increase their post-2012 productivity the most. These results not only shed light on the distributional consequences of the promotion incentives induced by the qualification process but also lend important support to the promotion incentives channel. Most of the alternative mechanisms through which earning the qualification could affect the productivity of scholars – for instance, qualified scholars obtaining different teaching duties or easier access to research funds – would hardly explain the observed heterogeneity of the effect depending on the variation in the intensity of the incentives.

VII.b Promotion incentives *vs.* alternative channels

Results from our analysis document that scholars who attain the NSQ in 2012 increase the quantity of publications in the four years following the call. We argue that the increased productivity of qualified scholars is mostly due to the provision of promotion incentives: gaining the eligibility for an associate professorship 'unlocks' the possibility to achieve also the qualification for a full professorship. Thus, it incentivizes scholars to enrich their publication records so as to meet the requirements for the full professor qualification in the subsequent round.

Of course, the variation in the future expected promotion thresholds may not be the only driver of our main results. For instance, obtaining the qualification could have a motivational effect, thus enhancing productivity, if it is perceived as a reward for past effort. Also, passing the qualification could induce substantial changes in scholars' daily life, as career advancements in academia are possibly associated with different teaching or bureaucratic duties, better access to research funds or broader networks. Still, these hypotheses seems inconsistent with the observed timing of the effect. Scholars' productivity begins rising already in 2013, before any actual promotion occurs (for more than 75% of qualified candidates, promotions did not take place earlier than 2015). To further disentangle between the effect of promotion incentives and that of other potential mechanisms, we estimate our baseline equation for the sample of associate professors who apply for the full professor qualification in 2012. The NSQ indeed regulates both the access to associate and full professor positions, but candidates for this latter rank will have vanishing career incentives once the goal is achieved since no further advancements are possible. In other words, earning the full professor qualification does not entail an increase in the future expected promotion thresholds faced by scholars.

The estimates reported in Table XIII show that applicants who barely earn or barely miss the eligibility for the top academic position do not exhibit any significant difference in terms of later research productivity. This zero (or even negative) effect clashes with several alternative explanations for our main result. It shows that the effect of achieving the qualification is specific to the group of academics (assistant professors) facing further promotion thresholds. In other words, consistent with our identification hypothesis, achieving the NSQ when this does not involve an increase in future expected promotion thresholds does not boost scholars' research productivity.[39]

Additionally, we look also at the sample of applicants to the 2012 NSQ who are not tenured assistant professors. This residual sample includes non-tenured academics but also scholars working in non-Italian universities and researchers working in other institutions either in Italy or abroad.[40] In principle, the incentive channel should apply to untenured Italian academics as well, while it is *a priori* unclear whether applicants outside Italian academia – who work under different incentive schemes – should exhibit similar responses. Thus, we categorize applicants to the 2012 associate professor NSQ into five groups – 1) tenured assistant professors (our baseline sample); 2) untenured assistant professors; 3) other applicants affiliated with an Italian university; 4) non-academic researchers in Italy; 5) applicants with a foreign affiliation – and look at the heterogeneity of the effect across these groups. Results are given in Table XIV and show that not only tenured assistant professors increase their publications in response to success in the NSQ, but also untenured assistant professors do so, even though the latter group is very small.[41]. Of course, this effect may not only be due to the variation in promotion thresholds but also to the achievement of the first tenured contract (getting tenure can have an effect on productivity *per se* (Faria and McAdam,

[39]It is worth remarking that this does not imply that achieving the full professor qualification impair associate professor's incentives to publish. Yet, the main goal for associate professors who earn the full professor qualification becomes the actual promotion to a full professorship, which is appointed at the local level rather than through a centralized procedure and typically not tied to specific *bibliometric* requirements.

[40]We retrieve from Scopus information about the employment status of most of these candidates by exploiting the affiliation reported in their 2011-2012 publications.

[41]Untenured assistant professors were introduced only in 2010 by the Gelmini reform, so they are 295 in our sample

2014)). A positive, yet non-significant, effect is found also within the residual sample of candidates with an affiliation to an Italian university. However, this sample does not only include post-docs but also PhD students, scholars with multiple affiliations, adjunct professors and others, for whom incentives may be different. We find the coefficients to be much lower for non-academic researchers in Italy (especially if compared to the average of the outcome variable within this sample) and even negative in the group of scholars working abroad. In both cases, coefficients are not statistically different from zero. All in all, this evidence lends further support to our interpretation of the results in light of the promotion incentive channel. Our effects are driven by the sample of candidates already employed by the Italian university system, for whom career progressions represent the main source of incentives. Instead, scholars outside Italian academia who work under different incentive schemes do not exhibit the same response to the outcome of the NSQ.

Finally, we test whether the observed increase in publication is due to the competition at the decentralized stage, where associate professorships are actually awarded. Achieving the NSQ might indeed incentivize qualified scholars to publish more in order to maximize their chances of obtaining an associate professor position as soon as a job vacancy opens rather than to meet the future full professor thresholds. Data on scholars' promotion patterns suggest that there is limited across- and within-department competition: approximately two-thirds of eligible candidates in our sample obtained an associate professorship within three years from achieving the NSQ; additionally, 97% of them obtained a promotion within the same university where they were employed in 2012. Nonetheless, it could still be the case that the productivity jump is driven by departments with few vacancies and many qualified scholars competing for a position.

Hence, we exploit the discipline-by-university heterogeneity in the degree of internal competition for being promoted to associate professor and test whether the effect of promotion incentives is actually stronger when there is more competition at the decentralized stage. Since we do not observe the actual number of vacancies but rather the equilibrium outcome, we use the ratio between the number of assistant professors who succeeded in the NSQ in 2012 and the number of existing associate professors in each university and discipline at the end of 2012 as a proxy for the degree of competition at the academic-department level. This ratio indeed measures the ease of access to an associate professor position conditional on having achieved the qualification. Departments in which there is a large mass of qualified candidates and few associate professors on staff are indeed likely to have lower turnover rates and therefore fewer vacancies, which make them relatively more competitive than those with a relatively low share of qualified scholars. Table XV shows that the effect of passing the NSQ on the subsequent research productivity does not vary systematically across quintiles, defined based on the distribution of the degree of competition at the academic department level. This evidence, together with the heterogeneous effects discussed in the previous section and with the zero-effect found in the sample of candidates to the full professor NSQ, as well as in that of scholars outside Italian academia, strongly supports the fact that promotion incentives are the main mechanism at work.

VIII Conclusion

This paper studies the effectiveness of promotion incentives for high-skilled public employees. For a sample of 5,000 tenured assistant professors participating to the first round of the Italian NSQ – the centralized evaluation procedure awarding the eligibility for career advancements – we find that scholars exposed to a quasi-random variation in promotion incentives in 2012, owing to a success in the NSQ, increase the number of publications by almost 40% over the 2013-2016 period. Additionally, we find that the effect the incentives is the strongest for those scholars who are neither too far nor too close from the relevant future promotion thresholds. That is, promotion incentives are most effective when the promotion is "neither too hard to achieve, nor too easy" (Lazear and Gibbs, 2014, p.269). When exploring additional aspects of scholars' research activity, we find that qualified candidates receive more citations after the achievement of the NSQ and tend to expand the number of collaborations. The average publication quality – proxied by several measures of the journal's prestige – remains constant.

Several robustness tests and placebo exercises confirm the validity of our three-dimensional fuzzy regression discontinuity design. Further, consistent with the identification hypothesis according to which achieving the qualification affects productivity through an increase in promotion incentives, we do not find a similar effect in the sample of associate professors applying for a full professorship. Once the top ladder of the academic hierarchy is reached, achieving the qualification does not provide any further incentive to meet specific and pre-determined *bibliometric* requirements. Finally, we provide evidence that our results are not driven by possible changes in scholars' routine associated with a promotion nor by local within-department competition among qualified candidates.

Although our study focuses on Italian academia, our results can be generalized to other contexts as well. First, centralized evaluation criteria regulate promotions within the public university system in many other countries, such as France, Germany, Portugal, and Spain. In all of these countries, advancements along the academic ladder are conditional on achieving a national qualification which is typically awarded on the basis of publication records. Our findings highlight that these systems can be an important tool to enhance the research productivity of scholars – especially when the qualification criteria are based on some observable and well-defined productivity indicators – and thus to foster the production of knowledge. This is particularly important for countries in which the higher-education system is predominantly public, which constitute the majority of OECD countries (OECD, 2019).

More broadly, our results can apply to other public organizations with internal labor markets, and especially those where the individual output of workers is recognizable and measurable, and workers have some discretion in the choice of effort. Typically, these features are common to high-skilled jobs such as university professors, school teachers, judges, and public managers in general. Due to their importance, promoting productivity in these contexts is a key concern for policy makers. Yet, legislative and budget constraints – and in some cases also political concerns, as these workers are paid by taxpayers – make it hard to implement pay-for-performance schemes such as those used by private organizations.

Career-based incentives thus stand out as the main motivational lever in the hands of policy makers. Because of this, many countries recently introduced evaluation-based promotion criteria in these contexts. For instance, meritocratic promotion mechanisms are used to govern the career advancements of school teachers in China, (Karachiwalla and Park, 2017), and those of civil-law judges in Italy and Germany (Schneider, 2005). However, how to design these evaluation procedures – whether they should be discretionary or strictly output-based – is an open and key policy issue. For instance, discretionary performance evaluations in which criteria are not set and communicated in advance may be less effective in promoting productivity than those based on objective criteria (consistent with the evidence in Meyer-Sahling et al. (2018)). Our study contributes to this discussion by showing that meritocratic promotion mechanisms based on transparent and output-based promotion thresholds can significantly boost the performance of public workers. Of course, promotion bars must be set so that they are neither too easy nor too hard to be met by the majority of workers.

References

Ashraf, Nava, Oriana Bandiera, and Scott Lee, "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," Working Paper March 2018.

Azoulay, Pierre, Toby Stuart, and Yanbo Wang, "Matthew: Effect or Fable?," *Management Science*, January 2014, *60* (1), 92–109.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva, "Does the Gender Composition of Scientific Committees Matter?," *American Economic Review*, 2017, *107* (4), 1207–1238.

Baicker, Katherine and Mireille Jacobson, " Finders Keepers: Forfeiture Laws, Policing Incentives, and Local Budgets," *Journal of Public Economics*, 2007, *91* (11), 2113–2136.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu, "The Glittering Prizes: Career Incentives and Bureaucrat Performance," *The Review of Economic Studies*, 05 2020, *87* (2), 626–655.

Björk, Bo-Christer and David Solomon, "The publishing delay in scholarly peer-reviewed journals," *Journal of Informetrics*, 2013, *7* (4), 914–923.

Borjas, George J. and Kirk Doran, "Prizes and Productivity: How Winning the Fields Medal Affects Scientific Output," *Journal of Human Resources*, 2015, *50* (3), 728–758.

_ and _ , "Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors," *The Review of Economics and Statistics*, 2015, *97* (5), 1104–1117.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , Max H. Farrell, and Rocìo Titiunik, "Rdrobust: Software for Regression-Discontinuity Designs," *Stata Journal*, June 2017, *17* (2), 372–404.

Campbell, Dennis, "Nonfinancial Performance Measures and Promotion-Based Incentives," *Journal of Accounting Research*, 2008, *46* (2), 297–332.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," Working Paper, University of Michigan 2017.

_ , _ , and _ , "Manipulation testing based on density discontinuity," *Stata Journal*, March 2018, *18* (1), 234–261.

_ , Roc??o Titiunik, and Gonzalo Vazquez-Bare, "Analysis of regression-discontinuity designs with multiple cutoffs or multiple scores," *The Stata Journal*, 2020, *20* (4), 866–891.

Checchi, Daniele, Gianni De Fraja, and Stefano Verzillo, "Publish or Perish: An Analysis of the Academic Job Market in Italy," Discussion Papers 14/04, University of Nottingham, School of Economics July 2014.

Dal Bó, Ernesto, Frederico Finan, and Martín Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.

De Paola, Maria and Vincenzo Scoppa, "Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia," *Economica*, 2015, *82* (325), 162–188.

_ , Michela Ponzo, and Vincenzo Scoppa, "Gender Differences in the Propensity to Apply for Promotion: Evidence from the Italian Scientific Qualification," *Oxford Economic Papers*, 2017, *69* (4), 986–1009.

Deserranno, Erika, Philipp Kastrau, and Gianmarco León-Ciliotta, "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector," Working Papers 1239, Barcelona Graduate School of Economics February 2021.

Duflo, Esther, Rema Hanna, and Stephen Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 2012, *102* (4), 1241–78.

Faria, J Ricardo and Peter McAdam, "Does Tenure Make Researchers Less Productive? The Case of the "Specialist"," School of Economics Discussion Papers 0514, School of Economics, University of Surrey 2014.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande, "The Personnel Economics of the Developing State," *Handbook of Economic Field Experiments*, 2017, *2*, 467–514.

Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov, "Team Incentives and Performance: Evidence from a Retail Chain," *American Economic Review*, 2017, *107* (8), 2168–2203.

Gaynor, Martin, James B. Rebitzer, and Lowell J. Taylor, "Physician Incentives in Health Maintenance Organizations," *Journal of Political Economy*, 2004, *112* (4), 915–931.

Gertler, Paul and Christel Vermeersch, "Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes," NBER Working Papers 19046, National Bureau of Economic Research, Inc 2013.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 205–27.

Haeck, Catherine and Frank Verboven, "The Internal Economics of a University: Evidence from Personnel Data," *Journal of Labor Economics*, 2012, *30* (3), 591 – 626.

Hansen, Stephen E., "Performance Feedback with Career Concerns," *The Journal of Law, Economics, and Organization*, 10 2012, *29* (6), 1279–1316.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger, "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science," *The Quarterly Journal of Economics*, 2018, *133* (2), 927–991.

Karachiwalla, Naureen and Albert Park, "Promotion Incentives in the Public Sector: Evidence from Chinese Schools," *Journal of Public Economics*, 2017, *146* (C), 109–128.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken, "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics*, 2016, *131* (1), 219–271.

_ **, Asim Ijaz Khwaja, and Benjamin A. Olken**, "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 2018, *forthcoming.*

Kwon, Illoong, "Incentives, Wages, and Promotions: Theory and Evidence," *The RAND Journal of Economics*, 2006, *37* (1), 100–120.

Lavy, Victor, "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, 2002, *110* (6), 1286–1317.

Lazear, Edward P., "Performance Pay and Productivity," *American Economic Review*, December 2000, *90* (5), 1346–1361.

_ **and Mike Gibbs**, *Personnel Economics in Practice*, John Wiley & Sons, 2014.

_ **and Paul Oyer**, "Personnel Economics," in Robert Gibbons and John Roberts, eds., *The Handbook of Organizational Economics*, Princeton University Press, 2013.

Meyer-Sahling, Jan-Hinrik, Christian Schuster, and Kim Sass Mikkelsen, "Civil service management in developing countries: what works?: Evidence from a survey with 23.000 civil servants in Africa, Asia, Eastern Europe, and Latin America," Working Papers, University of Nottingham 2018.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39 – 77.

OECD, *Government at a Glance 2017*, OECD Publishing, 2017.

_ , *Benchmarking Higher Education System Performance*, OECD Publishing, 2019.

Olken, Benjamin A., Junko Onishi, and Susan Wong, "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia," *American Economic Journal: Applied Economics*, 2014, *6* (4), 1–34.

Papay, John P., Richard J. Murnane, and John B. Willett, "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables," *Journal of Econometrics*, 2011, *161* (2), 203–207.

_ , _ , and _ , "High-School Exit Examinations and the Schooling Decisions of Teenagers: Evidence From Regression-Discontinuity Approaches," *Journal of Research on Educational Effectiveness*, 2014, *7* (1), 1–27.

Schneider, Martin R., "Judicial Career Incentives and Court Performance: An Empirical Study of the German Labour Courts of Appeal," *European Journal of Law and Economics*, 2005, *20* (2), 127–144.

Sebo, Paul, Jean Pascal Fournier, Claire Ragot, Pierre-Henri Gorioux, François R. Herrmann, and Hubert Maisonneuve, "Factors associated with publication speed in general medical journals: a retrospective study of bibliometric data," *Scientometrics*, May 2019, *119* (2), 1037–1058.

Shearer, Bruce, "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *The Review of Economic Studies*, 2004, *71* (2), 513–534.

Stepanov, Sergey, "Biased performance evaluation in a model of career concerns: incentives versus ex-post optimality," *Journal of Economic Behavior & Organization*, 2020, *179*, 589–607.

Waldinger, Fabian, "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, 2012, *79* (2), 838–861.

_ , "Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge," *The Review of Economics and Statistics*, December 2016, *98* (5), 811–831.

Zinovyeva, Natalia and Manuel Bagues, "The Role of Connections in Academic Promotions," *American Economic Journal: Applied Economics*, 2015, *7* (2), 264–292.

Figures

Figure I: Timeline of the 2012 NSQ

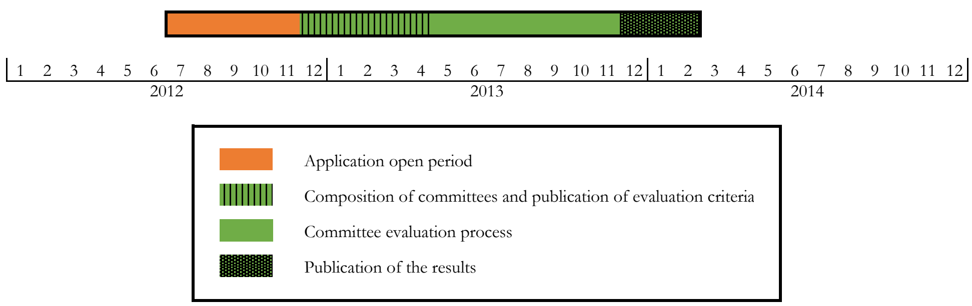
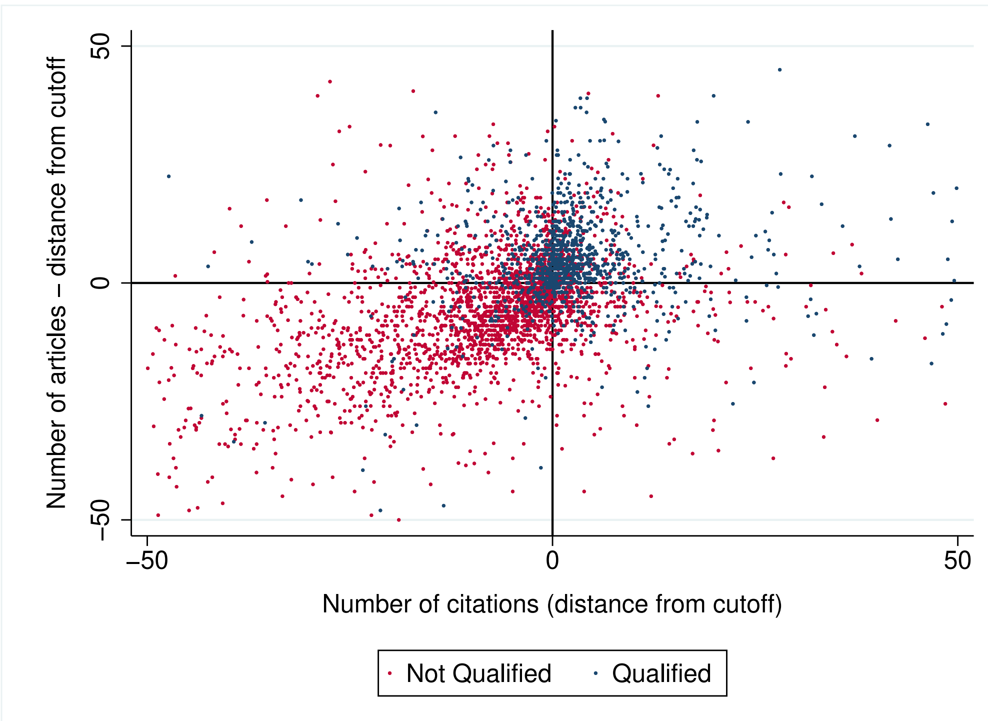


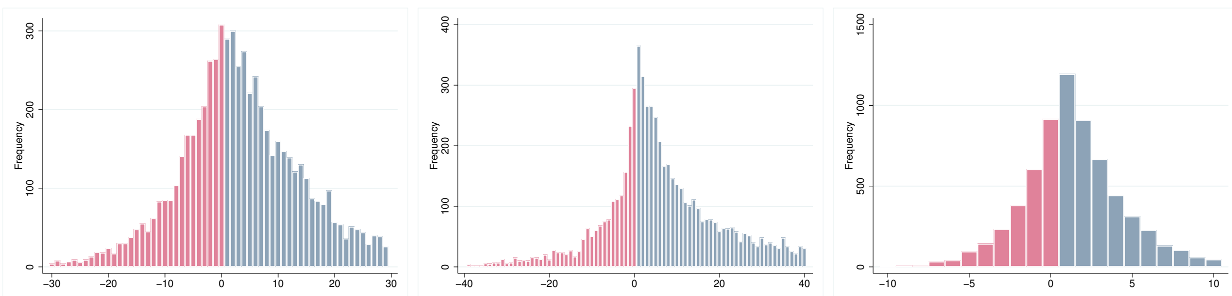
Figure II: Compliance with the two-out-of-three rule and outcome of the evaluation



NOTES. This figure depicts the distribution of candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors, depending on the distances between the first two *bibliometric* indicators and the corresponding cutoffs. The distance between the number of articles and the cutoff is on the y-axis, while the distance between the number of citations and the cutoff is on the x-axis. The figure is drawn for the subsample of applicants whose H-Index is below the cutoff, and whose distances from the other two indicators ranges between -50 and +50. Therefore, circles in the upper-right quadrant correspond to candidates complying with the two-out-of-three rule. Blue circles indicate qualified candidates, while red circles indicate non-qualified candidates.

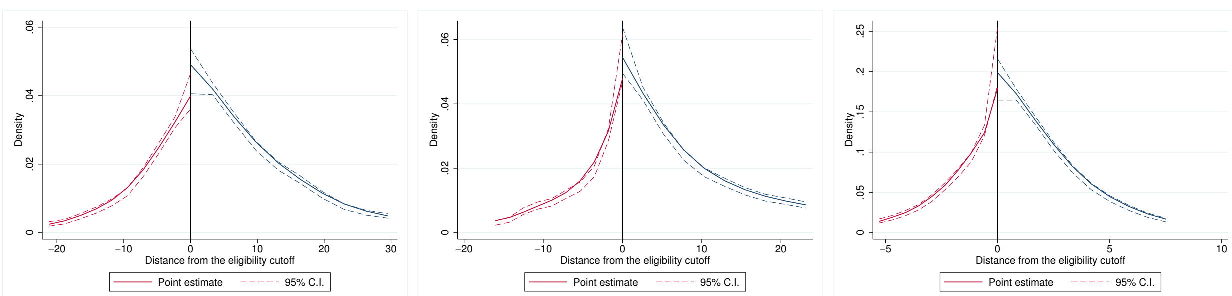
Figure III: Frequency distribution and manipulation test

Panel A. Frequency Distribution



Number of articles | Number of citations | H-index

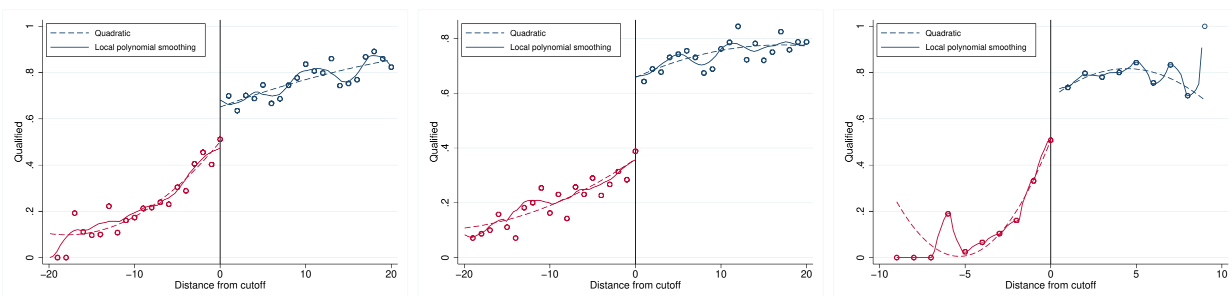
Panel B. Manipulation Test



Number of articles | Number of citations | H-index

Notes. This figure depicts the frequency distribution (Panel A) and the local polynomial density estimation of the distribution (Panel B) of candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors, depending on their distance from each of the three *bibliometric* cutoffs. The frequency distributions in the left, center, and right panel are constructed within the intervals $[-30, 30]$, $[-40, 40]$, $[-10, 10]$ for indicators 1, 2 and 3, respectively. In all panels, the bin width is equal to 1. The number of articles is rounded at the nearest integer. The local polynomial density is estimated following Cattaneo et al. (2017) and using the companion Stata package described in Cattaneo et al. (2018). When estimating the density of the distribution of candidates' distance from each sector-specific cutoff, we exclude observations in the top and the bottom 1%.

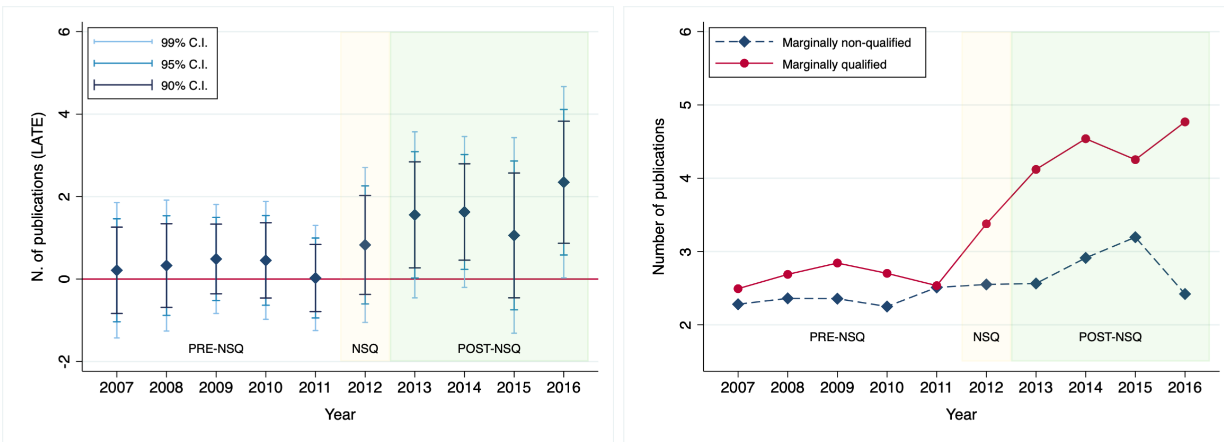
Figure IV: First stage



Number of articles | Number of citations | H-index

Notes. This figure depicts the discontinuous jumps in the probability of achieving the qualification arising when each of the three indicators overcomes the corresponding cutoff. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. Each circle represents the average probability of achieving the NSQ within each unit-sized bin. The running variables for the three indicators are the distances from the sector-specific median centered at zero. The dependent variable in the quadratic and local polynomial smoothing regression is an indicator that equals one when a candidate achieves the qualification. Both the quadratic and the local polynomial smoothing regressions are estimated within a $[-20, 20]$ interval of the distance from the thresholds for the first two indicators (Panel A and B), and within a $[-10, 10]$ interval of the distance from the threshold of the third indicator (Panel C). Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations.

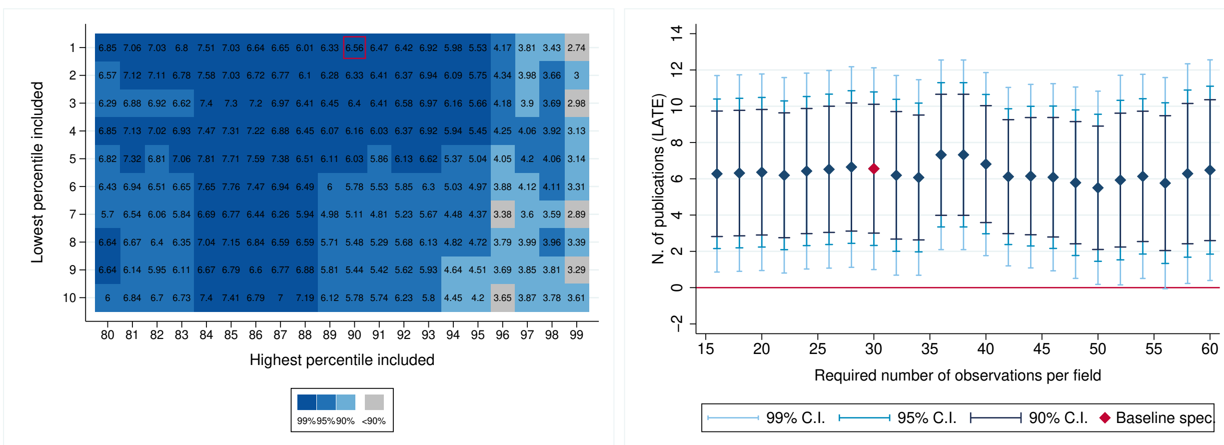
Figure V: Year-by-year effect of the NSQ on the number of publications



a. LATE *b.* Predicted publications

Notes. Panel A of this figure depicts the estimated LATE of achieving the qualification for associate professor on research productivity in the sample of assistant professors applying for the associate professor qualification in 2012 in *bibliometric* competition sectors. Coefficients and the corresponding confidence intervals are obtained by running a set of year-specific regressions (for each year between 2007 and 2016) where the dependent variable is the number of publications (including articles, conference papers, reviews and other items) of each candidate in a given year, and the main independent variable is the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator for compliance with the two-ouf-of-three rule (first-stage estimates are reported in Column (7) of Table II). Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions. Panel B instead depicts the estimated publications of the marginal candidates at the two sides of the multidimensional threshold in each year between 2007 and 2016. Blue diamonds in the figure are the predicted publications in each year of marginally non-qualified candidates. These are computed as the weighted average of the field-specific fixed effects, which correspond to the predicted number of publications conditional on being at a zero-distance from the three cutoffs. Red diamonds, instead, are the predicted publications of marginally successful scholars, computed as the sum of the publications of barely unsuccessful candidates and the LATE coefficients reported in Panel A.

Figure VI: Robustness to alternative sample restrictions

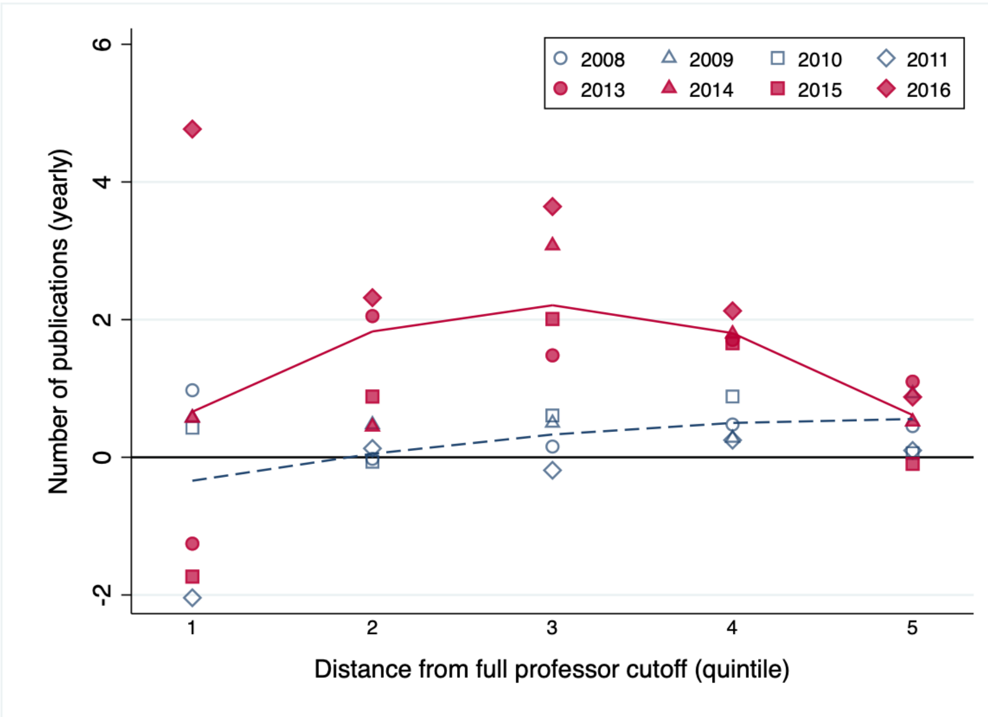


a. Interpercentile range

b. Sector size

NOTES. This figure depicts the effect of achieving the qualification on the quantity of publications over the 2013-2016 period under different sample restrictions. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. Each number in the left panel corresponds to the LATE coefficient of a regression estimated within the specified interpercentile range defined based on candidates' distance from the relevant cutoffs. For instance, when estimating the regression within the interpercentile range 5- 90 – the lower and the higher limit are reported on the y- and the x-axis, respectively – we exclude all candidates belonging to the bottom 5% and the top 10% of the pool of applicants in the same competition sector for any of the three indicators considered. Square color intensity indicates whether the coefficient is significant at the 1%, 5%, or 10% level. In each regression, the dependent variable is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period. The reported LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. In the right panel, the figure reports the LATE coefficients estimated under different restrictions on the minimum size (n_s) of the competition sectors. For instance, when estimating the regression for n_s>20, we exclude competition sectors with fewer than 20 observations. The red square (diamond) in the left (right) panel indicates our baseline specification. In both panels, we exclude sectors with more than 90% successful candidates. In the left panel we also exclude competition sector with fewer than 30 observations, while in the right panel we exclude observations in the top 10% and the bottom 1% of the within-sector distribution of the distances. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.

Figure VII: Heterogeneity in the distance from full professor thresholds



Notes. This figure depicts the heterogeneity of the effect of achieving the associate professor qualification depending on the distance from the full professor threshold on the yearly number of publications. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* fields, grouped by quintiles of the distribution of distances from the full professor threshold for the first indicator (number of articles). Each marker correspond to the estimated LATE within each quintile, that is, the 2SLS coefficient of the indicator for achieving the qualification interacted with the corresponding quintile-specific dummy. Full markers indicate point estimates for the post-NSQ period (2013-2016), while hollow markers the pre-NSQ one (2008-2011). The solid (dashed) line indicates the predicted average number of publications per year over the period 2013-2016 (2008-2011), obtained from a regression where the distance from the full professor threshold, grouped in quintiles, and its square product are interacted with the 2SLS coefficient of the indicator for achieving the qualification. The distance from the full professor threshold is expressed as a percentage of the initial stock of articles, thus defined as $dist_{i,1,s} = \frac{m^{full}_{1,s} - x_{i,1}}{x_{i,1}}$, where $m^{full}_{1,s}$ is the field-specific cutoff for the full professor NSQ and $x_{i,1}$ is the professor's score in the same indicator. The dependent variable is the total number of papers (including articles, conference papers, reviews and other items) published in each year prior (2008-2011) and after (2013-2016) the 2012 NSQ. Within each academic field, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We also exclude the fields with more than 90% successful candidates, those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes field-specific dummies and interactions.

Tables

Table I: Descriptive statistics

Panel A: applicants' characteristics				
	Mean		Sd	
Academic age	15.87		7.65	
Female	0.42		0.49	
Number of applications	1.26		0.65	
Qualified	0.58		0.49	
Above 2/3 cutoffs	0.66		0.47	
Above cutoff 1 (number of articles)	0.62		0.49	
Above cutoff 2 (number of citations)	0.68		0.47	
Above cutoff 3 (h-index)	0.58		0.49	
Distance from sector median 1 (n. of articles)	4.43		18.26	
Distance from sector median 2 (n. of citations)	9.62		38.36	
Distance from sector median 3 (h-index)	0.92		3.55	
Panel B: research activity				
	2008-2011		2013-2016	
	Mean	Sd	Mean	Sd
Number of publications	12.40	10.37	17.01	20.41
Number of articles	8.67	7.42	12.75	17.82
Number of conference papers	2.29	5.37	2.31	5.72
Number of reviews	0.68	1.55	0.92	2.09
Number of scholars	4920			
Panel C: publication characteristics				
	2008-2011		2013-2016	
	Mean	Sd	Mean	Sd
Top 5% journal	0.16	0.37	0.18	0.39
CiteScore (percentile)	72.18	24.79	75.44	23.19
CiteScore	2.69	2.26	2.80	2.13
Sjr	1.48	1.68	1.52	1.58
Snip	1.27	0.94	1.31	0.91
Journal unlisted	0.14	0.35	0.13	0.34
Number of citations received	17.19	38.81	5.34	15.00
Single-authored	0.02	0.15	0.02	0.12
Number of authors	8.27	12.03	10.82	17.09
Number of publications	43810		83670	

NOTES. This table reports the baseline characteristics of candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors together with the summary statistics about their research activity over the period 2008-2016. The unit of analysis is the single candidate in Panels (A) and (B), and the single publication in Panel (C). The variables 'Top 1% journal', 'Journal unlisted', and 'Single-authored' are binary indicators that take the value of one if the publication appears in journals scoring in the top 1% of the distribution of the 2012 CiteScore journal percentile index, it appears in journals not classified in the *Scopus* database, or it has a single author, respectively. All statistics are computed on the baseline sample used in our main analysis, in which, within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations.

Table II: First stage

Dependent variable: Qualified	Single RD (Articles)		Single RD (Citations)		Single RD (H-Index)		Triple RD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Quadratic	LLR	Quadratic	LLR	Quadratic	LLR	Quadratic
Above cutoff	0.131***	0.122***	0.210***	0.161***	0.153***	0.162**	
	(0.040)	(0.046)	(0.041)	(0.057)	(0.046)	(0.066)	
Above 2/3 cutoffs							0.306***
							(0.043)
Competition sector FE	Yes	No	Yes	No	Yes	No	Yes
Sector specific interactions	Yes	No	Yes	No	Yes	No	Yes
Mean dep. var.	0.577	0.573	0.577	0.536	0.577	0.564	0.577
Sd dep. var.	0.494	0.495	0.494	0.499	0.494	0.496	0.494
BW (MSE)		0.396		0.364		0.304	
N of clusters	82	89	82	89	82	89	82
Observations	4920	2752	4920	1798	4920	3034	4920

Notes. This table reports the OLS coefficients of overcoming the sector-specific *bibliometric* cutoffs on the probability of success in the 2012 NSQ. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. In all columns, the dependent variable is an indicator that equals one when a candidate achieves the qualification, and zero otherwise. In Columns (1) to (6), the main independent variable is an indicator that equals one when a candidate overcomes the relevant cutoff for either the number of articles (Columns 1 and 2), the number of citations (Columns 3 and 4) and the h-index (Columns 5 and 6). In Column (7), the main independent variable is an indicator that equals one when a candidate complies with the two-out-of-three rule, that is, when her scores in at least two indicators are above the relevant cutoffs. Regressions in Columns (1), (3), (5) and (6) are estimated using a quadratic specification, which includes competition sector-specific dummies and interactions, over the entire support after excluding observations in the top 10% and the bottom 1% of the within-sector distribution of the distances from each cutoff. In Columns (2), (4) and (6), we replicate the estimates in (1), (3), and (5), performing local linear regressions (LLR) within the MSE-optimal bandwidths computed following Calonico et al. (2014) – using the companion Stata package described in Calonico et al. (2017) – after normalizing each distance from the cutoff by dividing it by the cutoff itself. We exclude competition sectors with more than 90% successful candidates and those with fewer than 30 observations.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table III: The effect of achieving the NSQ on the number of papers published

	Publications	Articles	Conf. Papers	Reviews
	(1)	(2)	(3)	(4)
ITT	2.003***	1.254**	0.540	0.198*
	(0.701)	(0.477)	(0.336)	(0.102)
LATE	6.557***	4.105***	1.769*	0.648**
	(2.159)	(1.485)	(0.956)	(0.290)
Mean dependent variable	17.006	12.747	2.313	0.916
Sd dependent variable	20.410	17.820	5.719	2.088
Number of clusters	82	82	82	82
Observations	4920	4920	4920	4920

Notes. This table reports the ITT and the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variable in Column (1) is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period; the dependent variables in Columns (2), (3) and (4) are the total number of articles, conference papers and reviews published during the 2013-2016 period, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p< 0.01$, ** $p<0.05$, and *$p<0.10$.

Table IV: The effect of achieving the NSQ on the probability of leaving academia

	Left academia	Zero publications
	(1)	(2)
ITT	-0.007	-0.006
	(0.015)	(0.007)
LATE	-0.024	-0.019
	(0.043)	(0.019)
Mean dependent variable	0.028	0.017
Sd dependent variable	0.166	0.128
Number of clusters	82	82
Observations	4920	4920

Notes. This table reports the ITT and the LATE of achieving the qualification on the probability of leaving the Italian academia or not publishing any paper during the 2013-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variable in Columns (1) and (2) are an indicator that equals one if a candidate is not longer registered in the Italian professor census as of 2016 and an indicator that equals one when a candidate does not publish any paper during the 2013-2016 period, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses.
*** p< 0.01, ** p<0.05, and *p<0.10.

Table V: Gender heterogeneity

Dependent variable: Number of publications		
	First stage	LATE
Female	0.005	-1.148
	(0.020)	(0.875)
Above 2/3 cutoffs	0.329***	
	(0.046)	
Above 2/3 cutoffs × Female	-0.057*	
	(0.030)	
Qualified		6.742***
		(2.253)
Qualified × Female		-0.684
		(1.563)
Mean Dep. Var.	0.577	17.006
Standard dev.	0.494	20.410
N. of clusters	82	82
Observations	4920	4920

NOTES. This table reports the gender heterogeneity of the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period, and the corresponding first-stage estimates. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. Column (2) reports the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification and its interaction with the *female* dummy. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p< 0.01$, ** $p<0.05$, and * $p<0.10$.

Table VI: Disciplinary area heterogeneity

Dependent variable: Number of publications (std)										
	MATH	PHYS	CHEM	EARTH	BIOL	MED	AGRO	ENG	ARCH	PSY
Panel A: LATE										
Qualified × DA	0.121	0.883***	1.227*	1.015**	0.010	0.227	2.008	2.025*	1.262	-0.296
	(0.251)	(0.157)	(0.681)	(0.486)	(0.161)	(0.157)	(2.005)	(1.128)	(2.005)	(0.467)
Panel B: First Stage										
Above 2/3 cutoffs × DA	0.372***	0.143***	0.242***	0.507**	0.313***	0.394***	0.186*	0.252	0.114	0.532**
	(0.109)	(0.051)	(0.064)	(0.243)	(0.074)	(0.096)	(0.105)	(0.206)	(0.140)	(0.215)

Notes. This table reports the disciplinary area heterogeneity of the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period, and the corresponding first-stage estimates. The quantity of publication is expressed in within-disciplinary area standard deviations. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. In each column, the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification interacted with the corresponding discipline-specific dummy (DA). Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p< 0.01$, ** $p<0.05$, and *$p<0.10$.
Legend: MATH=Mathematics; PHYS=Physics; CHEM=Chemistry; EARTH=Earth Sciences; BIOL=Biology; MED=Health Sciences; AGRO=Agronomy and Veterinary; ENG=Engineering; ARCH=Architecture; PSY=Psychology.

Table VII: Continuity test (2009-2011 measures)

	(1) Publications	(2) Zero publ.	(3) % Unlisted	(4) Top 5%	(5) CiteScore	(6) Coauthors	(7) Single-author	(8) Network
Above 2/3 cutoffs	0.392	-0.002	0.028	0.173	0.010	0.310	-0.013	1.372
	(0.467)	(0.007)	(0.034)	(1.143)	(0.010)	(0.315)	(0.025)	(3.608)
Mean dep. var.	11.658	0.026	0.580	70.999	0.122	6.860	0.143	41.859
Sd dep. var.	9.598	0.159	0.494	15.860	0.196	5.766	0.351	53.641
Number of clusters	82	82	82	82	82	82	82	82
Observations	4920	4920	4755	4755	4792	4763	4861	4785

NOTES. This table reports the OLS coefficients of overcoming two out of the three sector-specific *bibliometric* cutoffs on the quantity of publications, their quality and the number of collaborations computed over the period 2009-2011. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variables are: the total number of papers – including articles, conference papers, reviews and other items – (Column 1); an indicator that equals one when a candidate does not publish any paper (Column 2); the share of publications in journals not classified in the *Scopus* database (Column 3); the share of articles published in journals scoring in the top 5% according to the 2012 *CiteScore journal percentile* (Column 4); the average *CiteScore journal percentile* (Column 5); the average number of co-authors per publication (Column 6); the share of single-authored publications (Column 7); and the total number of distinct co-authors (Column 8). In Column (6) the sample includes only scholars with at least one record in the *Scopus* database during the period 2013-2016. In Columns (3) to (5) the sample is further limited to scholars with at least one publication in a journal classified in the *Scopus* database (with a non-missing score) during the period 2009-2011. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p< 0.01$, ** $p<0.05$, and *$p<0.10$.

Table VIII: The effect of achieving the NSQ on the number of citations received

	Total citations	Cit. per paper	% Papers with cit.$\geq$ 50	% Papers with 0 cit.
	(1)	(2)	(3)	(4)
ITT	13.526**	0.519*	0.002	0.012
	(6.372)	(0.273)	(0.002)	(0.012)
LATE	43.934**	1.687**	0.006	0.039
	(18.087)	(0.748)	(0.004)	(0.034)
Mean dependent variable	92.371	4.417	0.005	0.356
Sd dependent variable	200.847	4.184	0.028	0.211
Number of clusters	82	82	82	82
Observations	4838	4838	4838	4838

NOTES. This table reports the ITT and the LATE of achieving the qualification on the citations received by papers published over the 2013-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors with at least one record in the *Scopus* database over the 2013-2016 period. The dependent variables in Columns (1) and (2) are the total and the average number of citations received by papers published during the 2013-2016 period, respectively. The dependent variable in Columns (3) and (4) are the share of papers published during the 2013-2016 period with at least 50 citations and with zero citations, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table IX: The effect of achieving the NSQ on the quality of publications

	Top 5%	CiteScore (pct)	CiteScore	Sjr	Snip	% Unlisted
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	0.038	-0.110	0.063	0.052	0.040	0.000
	(0.030)	(1.087)	(0.069)	(0.059)	(0.032)	(0.011)
LATE	0.126	-0.360	0.206	0.170	0.130	0.001
	(0.083)	(3.065)	(0.192)	(0.168)	(0.086)	(0.029)
Mean dependent variable	0.655	72.403	2.509	1.336	1.239	0.126
Sd dependent variable	0.475	14.773	1.343	0.843	0.456	0.180
Number of clusters	82	82	82	82	82	82
Observations	4809	4809	4809	4808	4809	4838

NOTES. This table reports the ITT and the LATE of achieving the qualification on the quality of publications over the 2013-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variable in Column (1) is the share of articles published in journals scoring in the top 5% according to the 2015 *CiteScore journal percentile* during the 2013-2016 period; the dependent variables in Columns (2), (3), (4) and (5) are the average *CiteScore journal percentile*, the average *CiteScore* index, the average *Sjr* index, the average *Snip* index of papers published during the 2013-2016 period, respectively; the dependent variable In Column (6) is the share of publications in journals not classified in the *Scopus* database during the period 2013-2016. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. In Column (6) the sample includes only scholars with at least one record in the *Scopus* database during the period 2013-2016. In Columns (2) to (5) the sample is further limited to scholars with at least one publication in a journal classified in the *Scopus* database of journals (with a non-missing score) during the period 2013-2016. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table X: The effect of achieving the NSQ on the number of collaborations

	(1) Coauthors (mean)	(2) Coauthors (median)	(3) Single-author	(4) Network size
ITT	0.241	0.686	0.016	1.748
	(0.459)	(0.419)	(0.029)	(5.334)
LATE	0.785	2.236*	0.053	5.722
	(1.309)	(1.292)	(0.082)	(15.061)
Mean dependent variable	8.185	7.255	0.137	62.016
Sd dependent variable	8.245	8.217	0.344	84.882
Number of clusters	82	82	82	82
Observations	4801	4801	4801	4806

NOTES. This table reports the ITT and the LATE of achieving the qualification on the number of co-authorships over the 2013-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors with at least one record in the *Scopus* database over the 2013-2016 period. The dependent variables are the average number of coauthors per publication (Column 1), the maximum number of co-authors per publication (Column 2), the share of single-authored publications (Column 3) and the total number of distinct coauthors (Column 4). The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table XI: Heterogeneity in the distance from the full professor thresholds

Panel A - *Dependent variable:* Number of publications 2013-2016

	Average effect	By quintile of the distance				
	(1)	(2)	(3)	(4)	(5)	(6)
		Q1	Q2	Q3	Q4	Q5
Qualified	6.279***					
	(2.175)					
Qualified × Quintile		2.550	5.652	10.100***	7.245***	2.366
		(12.905)	(4.173)	(3.526)	(2.618)	(4.174)
Mean dep. var.	17.079	25.680	19.533	17.027	14.160	9.202
Sd dep. var.	20.452	35.700	16.372	13.253	11.493	10.102
Observations	4887	968	986	975	954	1004

Panel B - *Dependent variable:* Probability of achieving the full professor NSQ by 2016

	Average effect	By quintile of the distance				
	(1)	(2)	(3)	(4)	(5)	(6)
		Q1	Q2	Q3	Q4	Q5
Qualified	0.179**					
	(0.071)					
Qualified × Quintile		0.648*	0.195	0.309***	0.130*	0.027
		(0.336)	(0.123)	(0.120)	(0.073)	(0.079)
Mean dep. var.	0.177	0.309	0.259	0.190	0.108	0.022
Sd dep. var.	0.382	0.462	0.438	0.392	0.311	0.146
Observations	4887	968	986	975	954	1004

Notes. This table depicts the heterogeneity of the effect of achieving the associate professor qualification depending on the distance from the full professor thresholds on the quantity of publications over the 2013-2016 period (Panel A), and on the probability of achieving the full professor qualification by 2016 (Panel B). The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors, grouped by quintiles of the distribution of distances from the full professor threshold for the first indicator (number of articles). Column (1) in both panels reports the (average) estimated LATE, that is, the 2SLS coefficient of the indicator that equals one when a candidate achieves the associate professor qualification instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. Columns (2) to (6) report the estimated LATE within each quintile, that is, the 2SLS coefficient of the indicator for achieving the qualification interacted with the corresponding quintile-specific dummy. The distance from the full professor threshold is expressed as a percentage of the initial stock of articles, thus defined as $dist_{i,1,s} = \frac{m_{1,s}^{full} - x_{i,1}}{x_{i,1}}$, where $m_{1,s}^{full}$ is the sector-specific cutoff for the full professor NSQ and $x_{i,1}$ the professor's score in the same indicator. The dependent variable in the upper panel is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period, while it is an indicator that equals one when a candidate achieves the full professor qualification by 2016 in the bottom panel. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates, those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table XII: Age heterogeneity

	Years of tenure		Academic age	
	(1)	(2)	(3)	(4)
	First stage	Publications	First stage	Publications
Senior	-0.033	-1.113	-0.012	-2.218***
	(0.028)	(0.750)	(0.030)	(0.808)
Above 2/3 cutoffs	0.298***		0.316***	
	(0.045)		(0.048)	
Above 2/3 cutoffs × Senior	0.032		-0.025	
	(0.039)		(0.038)	
Qualified		6.820***		6.493***
		(2.139)		(2.243)
Qualified × Senior		-1.657		-1.028
		(1.190)		(1.356)
Mean Dep. Var.	0.577	17.006	0.577	17.006
Standard dev.	0.494	20.410	0.494	20.410
N. of clusters	82	82	82	82
Observations	4920	4920	4920	4920

Notes. This table reports the age heterogeneity of the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period, and the corresponding first-stage estimates. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. In Columns (1) and (2) *senior* is an indicator that equals one for candidates who were tenured assistant professors in 2000 (the first available year of the professor census). In Columns (3) and (4) it is defined as an indicator that equals one when a candidate's first recorded publication in Scopus lies below the sample average (1997). Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p < 0.01$, ** $p<0.05$, and *$p<0.10$.

Table XIII: The effect of achieving the full professor NSQ on the number of papers published

	Publications	Articles	Conf. Papers	Reviews
	(1)	(2)	(3)	(4)
ITT	-0.349	-0.251	0.195	-0.105
	(1.195)	(1.009)	(0.221)	(0.189)
LATE	-0.836	-0.601	0.468	-0.251
	(2.408)	(2.037)	(0.467)	(0.382)
Mean dependent variable	22.592	16.776	2.727	1.431
Sd dependent variable	24.987	21.383	6.982	2.758
Number of clusters	51	51	51	51
Observations	2746	2746	2746	2746

NOTES. This table reports the ITT and the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period. The sample includes the candidates for the full professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variable in Column (1) is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period; the dependent variables in Columns (2), (3) and (4) are the total number of articles, conference papers and reviews published during the 2013-2016 period, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table XIV: The effect of achieving the NSQ among all applicants

Panel A - *Dependent variable:* Number of publications 2013-2016	Average effect	By 2012 employment				
	(1)	(2) Tenured assistant prof.	(3) Untenured assistant prof.	(4) Others in Italian univ.	(5) Researchers in non-university inst.	(6) Researchers Abroad
Qualified	3.553*					
	(2.075)					
Qualified × 2012 Employment		4.592**	9.521**	2.324	1.198	-4.999
		(1.947)	(4.451)	(2.812)	(2.885)	(4.188)
Mean dep. var. (Group)	16.871	17.006	21.329	14.405	18.896	16.940
Sd dep. var. (Group)	18.516	20.410	19.255	13.672	17.558	16.295
Observations (Group)	9071	4920	295	1993	1433	430

Panel B - Dependent variable: Probability of achieving the NSQ	Average effect	By 2012 employment				
	(1)	(2) Tenured assistant prof.	(3) Untenured assistant prof.	(4) Others in Italian univ.	(5) Researchers in non-university inst.	(6) Researchers Abroad
Above 2/3 cutoffs	0.278***					
	(0.033)					
Above 2/3 cutoffs × 2012 Employment		0.337***	0.198***	0.183***	0.245***	0.178***
		(0.034)	(0.070)	(0.040)	(0.040)	(0.059)
Mean dep. var. (Group)	0.497	0.577	0.614	0.346	0.426	0.444
Sd dep. var. (Group)	0.500	0.494	0.488	0.476	0.495	0.497
Observations (Group)	9071	4920	295	1993	1433	430

NOTES. This table depicts the heterogeneity of the effect of achieving the associate professor qualification on quantity of publications over the 2013-2016 period (Panel A) and the probability of achieving the qualification (Panel B) depending on the applicant's employment at the time of the NSQ. The sample includes the candidates for the associate professor NSQ in *bibliometric* competition sectors. Besides applicants who at the time of the NSQ are tenured assistant professors (our baseline sample), the sample in this table also include untenured assistant professors (*ricercatori a tempo determinato*), other applicants with an Italian university affiliation, non-academic researchers in Italy, applicants with a foreign affiliation. In Panel A, Column (1) reports the (average) estimated LATE in this extended sample, that is, the 2SLS coefficient of the indicator that equals one when a candidate achieves the associate professor qualification instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. Columns (2) to (6) report the estimated LATE within each employment group, that is, the 2SLS coefficient of the indicator for achieving the qualification interacted with the corresponding employment group-dummy. In Panel B, Column (1) reports the (average) first-stage estimates, while Columns (2) to (6) reports the estimated first-stage coefficient within each employment group, that is, the coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule interacted with the corresponding employment group-dummy. The dependent variable in the upper panel is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period, while in the bottom panel it is an indicator that equals one when a candidate achieves the qualification. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates, those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table XV: Heterogeneity in the degree of internal competition for promotions

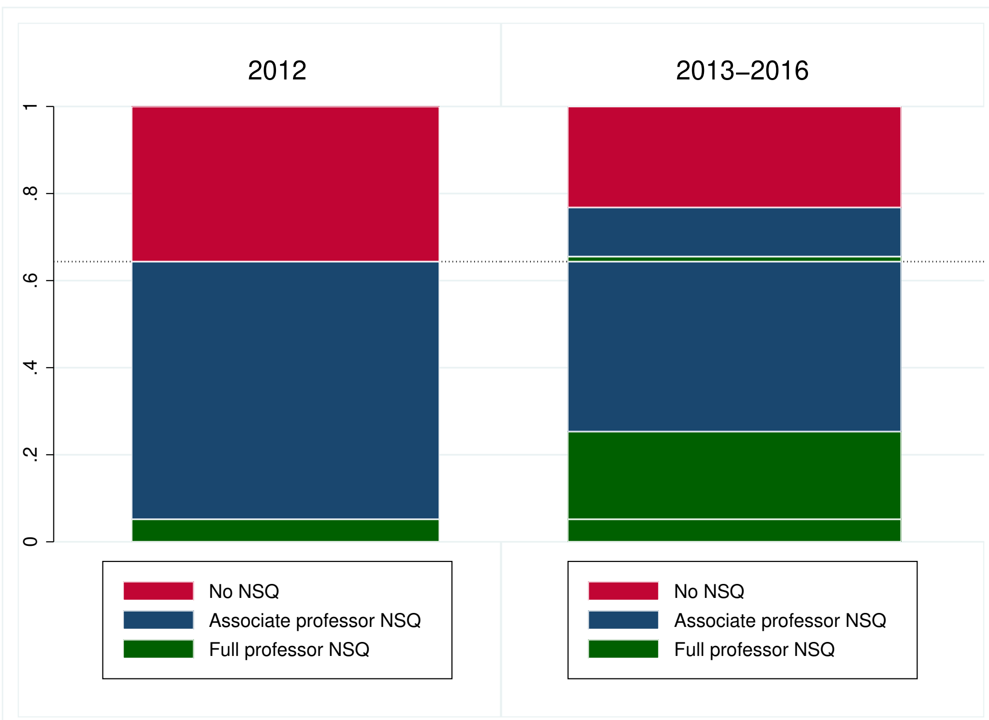
Panel A - Dependent variable: Number of publications					
	Q1	Q2	Q3	Q4	Q5
Qualified × Quintile	5.920**	6.237***	4.048	7.614***	8.062**
	(2.600)	(1.871)	(2.560)	(2.723)	(3.829)
Mean dep. var. (Quintile)	14.323	16.647	16.957	16.807	19.794
Sd dep. var. (Quintile)	12.883	18.816	15.053	22.199	28.099
Observations (Quintile)	827	1095	998	909	1057
Panel B - Dependent variable: Probability of achieving the NSQ					
	Q1	Q2	Q3	Q4	Q5
Above 2/3 cutoffs × Quintile	0.269***	0.328***	0.329***	0.293***	0.311***
	(0.052)	(0.055)	(0.050)	(0.055)	(0.053)
Mean dep. var. (Quintile)	0.418	0.519	0.596	0.644	0.696
Sd dep. var. (Quintile)	0.494	0.500	0.491	0.479	0.460
Observations (Quintile)	827	1095	998	909	1057

NOTES. This table depicts the heterogeneity of the effect of achieving the associate professor qualification depending on the degree of competition for vacancies at the department level on the quantity of publications over the 2013-2016 period (Panel A), and on the probability of achieving the qualification (Panel B). The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors, grouped by quintiles of the distribution of the degree of competition at the decentralized stage. This is defined as the ratio between the number of assistant professors employed in a given university who qualify for associate professor in a given disciplinary area on the number of associate professors already employed in the same university and belonging to the same disciplinary area. The reported coefficients correspond to the estimated ITT effect within each quintile, obtained by multiplying an indicator that equals one when a candidate complies with the two-out-of-three rule with a set of quintile-specific indicators. The dependent variable in the upper panel is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period, while in the bottom panel it is an indicator that equals one when a candidate achieves the qualification. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p< 0.01$, ** $p<0.05$, and *$p<0.10$.

Appendix A

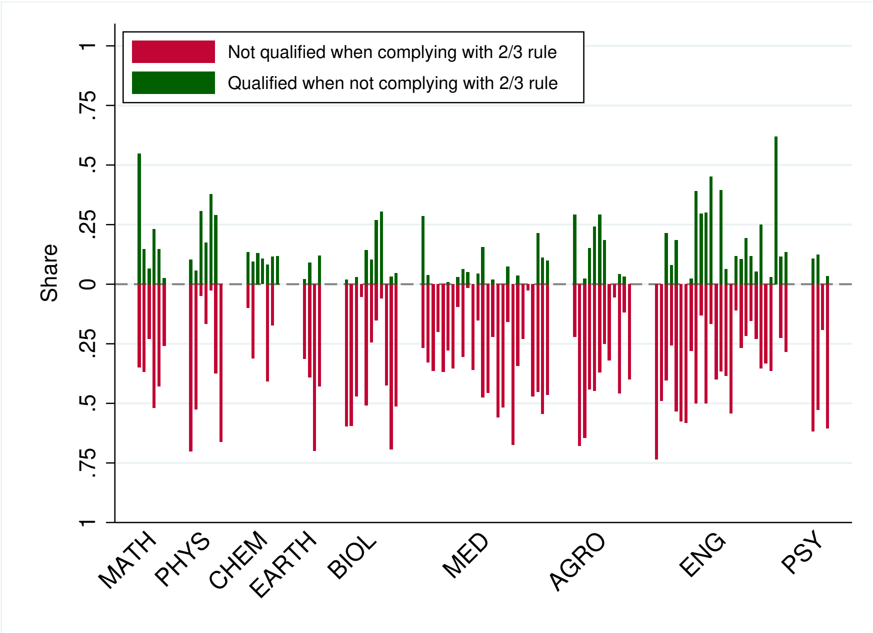
Additional Figures and Tables

Figure A1: NSQ trajectories 2012-2016



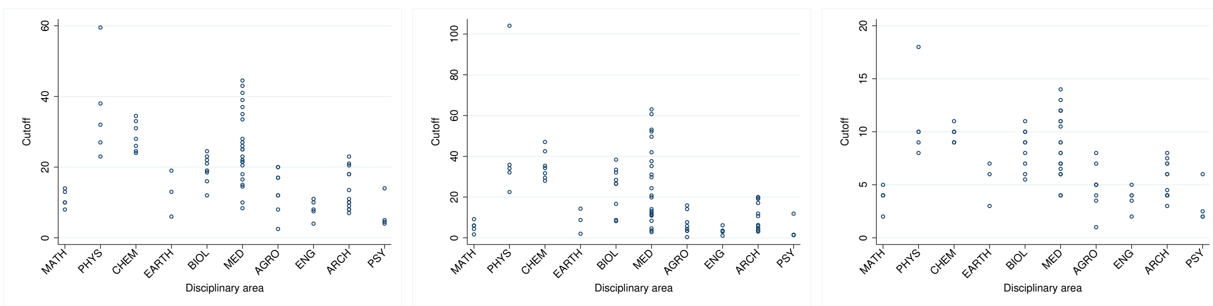
Notes. This figure depicts the share of qualified and non-qualified assistant professors applying to the 2012 NSQ and their later NSQ trajectories. The left bar reports the share of candidates who miss the qualification (in red), who qualify for associate professor (in blue), and who qualify for full professor (in green). In the right bar, we report, for each of the three groups, the share of those who hold an associate or full professor qualification (or none of the two) as of December 2016, that is, after the 2013 and 2016 rounds of the NSQ.

Figure A2: Compliance with the two-out-of-three rule



NOTES. This figure depicts the share of candidates for the associate professor NSQ in 2012 in each *bibliometric* competition sectors who achieve the qualification when not complying with the two-out-of-three rule (green bars) and the share of those who do not achieve the qualification even if complying with the rule (red bars). We exclude competition sectors with more than 90% successful candidates and those with fewer than 30 observations. Competition sectors are grouped by disciplinary area.
Legend: MATH=Mathematics; PHYS=Physics; CHEM=Chemistry; EARTH=Earth Sciences; BIOL=Biology; MED=Health Sciences; AGRO=Agronomy and Veterinary; ENG=Engineering; ARCH=Architecture; PSY=Psychology.

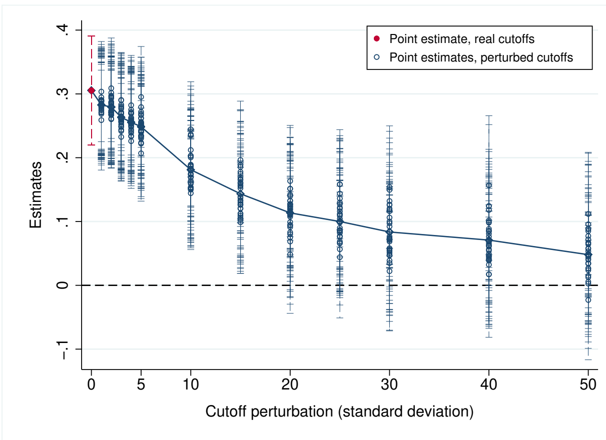
Figure A3: Heterogeneity in sector-specific cutoffs



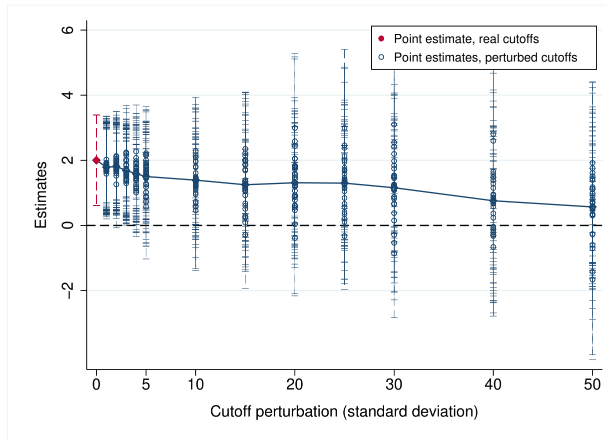
Number of articles | Number of citations | H-index

NOTES. This figure depicts the value of the cutoffs for each *bibliometric* indicator and competition sector. We exclude sectors with more than 90% successful candidates and those with fewer than 30 observations. Competition sectors are grouped by disciplinary area.
Legend: MATH=Mathematics; PHYS=Physics; CHEM=Chemistry; EARTH=Earth Sciences; BIOL=Biology; MED=Health Sciences; AGRO=Agronomy and Veterinary; ENG=Engineering; ARCH=Architecture; PSY=Psychology.

Figure A4: Cutoff perturbation



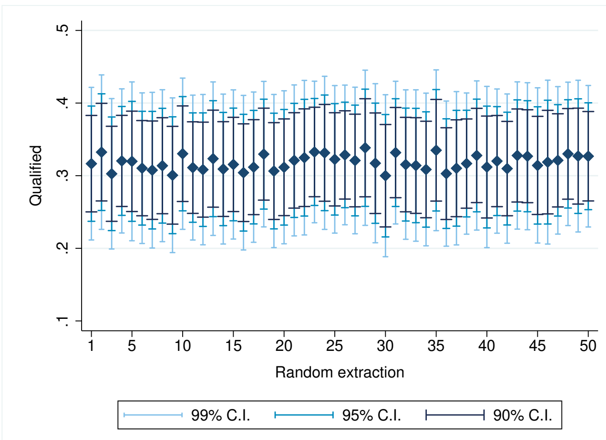
a. First stage



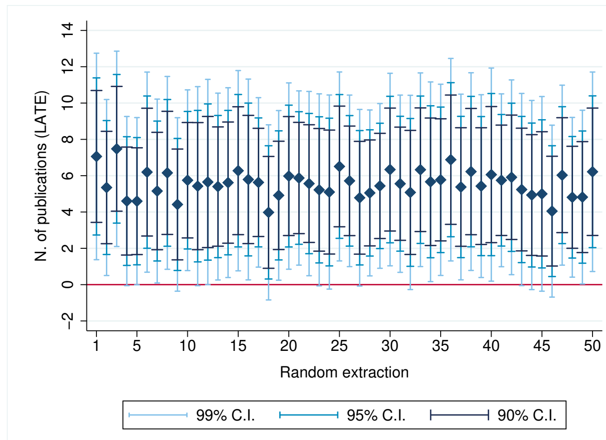
b. Number of publications 2013-2016 (ITT)

NOTES. This figure depicts the OLS coefficients of the indicator for the compliance with the two-out-of-three rule on the outcome of the 2012 NSQ (left panel) and on the quantity of publications over the 2013-2016 period (right panel) applying different perturbations to the cutoff values. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. Each dot corresponds to the coefficient of an indicator that equals one when a candidate complies with the two-out-of-three rule on the depended variable considered, under a different perturbation of the cutoff values. The dependent variables in the left and right panel are an indicator that equals one when a candidate achieves the qualification and the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period, respectively. The permutations of the cutoff values are obtained by adding a randomly generated error component $\epsilon \sim N(0, \sigma)$, where the standard deviation (σ) of the error determines the intensity of the reshuffling. For each value of σ, we estimate 30 separate regressions for different realizations of ϵ. We apply the same reshuffling to the three *bibliometric* cutoffs, and we force the perturbation to lie within within - and +100% of the original cutoff values. In each regression, Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.

Figure A5: Robustness to random application selection



a. First stage



b. Number of publications 2013-2016 (LATE)

Notes. This figure depicts the OLS coefficients of the indicator for the compliance with the two-out-of-three rule on the outcome of the 2012 NSQ (left panel) and the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period (right panel) for different random draws from a applicant's application list. The sample, includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. Each dot in the left (right) panel corresponds to the first stage (LATE) coefficient from a regression where we select, for each applicant, an application at random among the ones she submits. In the left panel, the dependent variable is an indicator that equals one when a candidate achieves the qualification, while in the right panel it is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period. The reported LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. The different random draws (50) are reported on the x-axis. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.

Table A1: First stage - full professor NSQ

Dependent variable: Qualified	Single RD (Articles)		Single RD (Citations)		Single RD (H-Index)		Triple RD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Quadratic	LLR	Quadratic	LLR	Quadratic	LLR	Quadratic
Above cutoff	0.141*	0.175***	0.242***	0.174**	0.265***	0.312***	
	(0.072)	(0.064)	(0.075)	(0.081)	(0.086)	(0.108)	
Above 2/3 cutoffs							0.420***
							(0.067)
Competition sector FE	Yes	No	Yes	No	Yes	No	Yes
Sector specific interactions	Yes	No	Yes	No	Yes	No	Yes
Mean dep. var.	0.566	0.581	0.566	0.547	0.566	0.543	0.566
Sd dep. var.	0.496	0.494	0.496	0.498	0.496	0.498	0.496
BW (MSE)		0.298		0.270		0.182	
N of clusters	47	89	47	89	47	89	47
Observations	2369	1039	2369	675	2369	905	2369

NOTES. This table reports the OLS coefficients of overcoming the sector-specific *bibliometric* cutoffs on the probability of success in the 2012 NSQ. The sample includes the candidates for the full professor NSQ in 2012 in *bibliometric* competition sectors. In all columns, the dependent variable is an indicator that equals one when a candidate achieves the qualification, and zero otherwise. In Columns (1) to (6), the main independent variable is an indicator that equals one when a candidate overcomes the relevant cutoff for either the number of articles (Columns 1 and 2), the number of citations (Columns 3 and 4) and the h-index (Columns 5 and 6). In Column (7), the main independent variable is an indicator that equals one when a candidate complies with the two-out-of-three rule, that is, when her scores in at least two indicators are above the relevant cutoffs. Regressions in Columns (1), (3), (5) and (6) are estimated using a quadratic specification, which includes sector-specific dummies and interactions, over the entire support after excluding observations in the top 10% and the bottom 1% of the within-sector distribution of the distances from each cutoff. In Columns (2), (4) and (6), we replicate the estimates in (1), (3), and (5), performing local linear regressions (LLR) within the MSE-optimal bandwidths computed following Calonico et al. (2014) – using the companion Stata package described in Calonico et al. (2017) – after normalizing each distance from the cutoff by dividing it by the cutoff itself. We exclude the competition sectors with more than 90% successful candidates and those with fewer than 30 observations.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table A2: The effect of achieving the NSQ on the number of publications, per year

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Panel A - Dependent variable: Number of publications										
ITT	0.064	0.099	0.149	0.138	0.008	0.253	0.474**	0.496**	0.323	0.717***
	(0.194)	(0.184)	(0.154)	(0.167)	(0.153)	(0.222)	(0.236)	(0.211)	(0.273)	(0.254)
LATE	0.212	0.327	0.486	0.452	0.025	0.827	1.556**	1.626***	1.058	2.348***
	(0.551)	(0.534)	(0.446)	(0.482)	(0.430)	(0.634)	(0.679)	(0.618)	(0.801)	(0.783)
Mean Dep. Var.	2.706	2.825	2.940	2.962	3.146	3.533	3.820	4.159	4.518	4.574
Standard dev.	2.655	2.668	3.015	3.019	3.141	3.793	3.862	4.956	7.547	8.446
Observations	4743	4795	4827	4843	4855	4861	4869	4898	4914	4919
Panel B - Dependent variable: Number of articles										
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
ITT	0.112	0.122	0.056	0.097	0.079	0.169	0.231	0.342*	0.160	0.530***
	(0.124)	(0.115)	(0.120)	(0.117)	(0.114)	(0.165)	(0.161)	(0.174)	(0.204)	(0.186)
LATE	0.369	0.403	0.183	0.316	0.258	0.555	0.758	1.121**	0.526	1.735***
	(0.351)	(0.334)	(0.338)	(0.336)	(0.326)	(0.474)	(0.468)	(0.514)	(0.590)	(0.567)
Mean Dep. Var.	1.888	1.981	2.084	2.041	2.206	2.468	2.734	3.054	3.408	3.598
Standard dev.	1.903	1.928	2.339	2.229	2.327	2.979	2.849	4.183	6.854	7.807
Observations	4743	4795	4827	4843	4855	4861	4869	4898	4914	4919
Panel C - Dependent variable: Number of conference papers										
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
ITT	-0.024	0.005	0.113	0.077	-0.029	0.064	0.173*	0.098	0.114	0.147
	(0.140)	(0.109)	(0.077)	(0.084)	(0.068)	(0.100)	(0.100)	(0.090)	(0.111)	(0.109)
LATE	-0.081	0.017	0.368	0.251	-0.094	0.209	0.569**	0.322	0.373	0.483
	(0.393)	(0.308)	(0.229)	(0.243)	(0.192)	(0.277)	(0.283)	(0.256)	(0.321)	(0.304)
Mean Dep. Var.	0.565	0.551	0.562	0.589	0.575	0.596	0.605	0.610	0.617	0.490
Standard dev.	1.564	1.455	1.508	1.594	1.577	1.615	1.690	1.699	1.770	1.508
Observations	4743	4795	4827	4843	4855	4861	4869	4898	4914	4919
Panel D - Dependent variable: Number of reviews										
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
ITT	-0.024	-0.009	0.029	-0.009	-0.026	0.005	0.073*	0.021	0.041	0.064
	(0.033)	(0.031)	(0.032)	(0.033)	(0.031)	(0.038)	(0.037)	(0.036)	(0.047)	(0.054)
LATE	-0.080	-0.029	0.096	-0.029	-0.087	0.018	0.241**	0.070	0.134	0.210
	(0.093)	(0.086)	(0.090)	(0.094)	(0.085)	(0.107)	(0.107)	(0.100)	(0.129)	(0.157)
Mean Dep. Var.	0.128	0.134	0.126	0.158	0.167	0.236	0.222	0.238	0.209	0.251
Standard dev.	0.449	0.467	0.478	0.570	0.513	0.681	0.676	0.708	0.692	0.743
Observations	4743	4795	4827	4843	4855	4861	4869	4898	4914	4919

Notes. This table reports the ITT and the LATE of achieving the qualification on the quantity of publications in each year over the 2007-2016 period. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variables in Panels A, B, C, and D are the numbers of publications, articles, conference papers and reviews, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. First-stage estimates are reported in Column (7) of Table II. In all panels, the sample includes only scholars that were 'active' in the year considered, that is, scholars whose first publication is not later than that year. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Table A3: Robustness to the exclusion of single disciplinary areas

Dependent variable: Number of publications										
	MATH	PHYS	CHEM	EARTH	BIOL	MED	AGRO	ENG	ARCH	PSY
ITT	2.218***	1.529**	1.881**	1.957***	2.351***	2.264**	1.895**	1.928***	1.985***	2.124***
	(0.780)	(0.705)	(0.746)	(0.714)	(0.814)	(0.873)	(0.725)	(0.712)	(0.637)	(0.711)
LATE	7.491***	4.772**	6.033***	6.500***	7.729***	8.497***	6.039***	6.284***	6.058***	7.130***
	(2.472)	(1.953)	(2.214)	(2.232)	(2.566)	(3.266)	(2.140)	(2.178)	(1.891)	(2.263)
Mean Dep. Var.	17.631	15.352	16.935	17.104	17.838	17.147	17.318	17.126	16.458	17.227
Standard dev.	21.149	13.166	21.118	20.650	21.577	22.167	20.924	20.750	20.952	20.709
N. of clusters	77	77	75	79	73	59	74	77	69	78
Observations	4406	4587	4476	4782	4260	3402	4600	4706	4345	4716

NOTES. This table reports how the ITT and the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period vary when excluding from the sample a disciplinary area (reported in the column header). The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule, interacted with the corresponding discipline-specific dummy (DA); the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, interacted with the corresponding discipline-specific dummy. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p < 0.01$, ** $p<0.05$, and *$p<0.10$.
Legend: MATH=Mathematics; PHYS=Physics; CHEM=Chemistry; EARTH=Earth Sciences; BIOL=Biology; MED=Health Sciences; AGRO=Agronomy and Veterinary; ENG=Engineering; ARCH=Architecture; PSY=Psychology.

Table A4: Local linear specification

Dependent variable: Number of publications		
	(1)	(2)
	Full Parametric	LLR
ITT	2.003***	1.559*
	(0.701)	(0.815)
LATE	6.557***	6.620*
	(2.159)	(3.718)
First Stage	0.306***	0.236***
	(0.043)	(0.056)
Competition sector FE	Yes	Yes
Sector specific interactions	Yes	No
Mean dependent variable	17.006	15.392
Sd dependent variable	20.410	11.950
Bandwidth (MSE) - Number of articles		0.349
Bandwidth (MSE) - Number of citations		0.465
Bandwidth (MSE) - H-Index		0.230
Number of clusters	82	84
Observations	4920	931

Notes. This table reports the ITT and the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period, and the corresponding first-stage estimates. In Column (1) we use our baseline full-parametric specification over the entire support. In Column (2), we use a linear approach in the neighborhood of the threshold. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. The dependent variable in both columns is the total number of papers (including articles, conference papers, reviews and other items) published during the 2013-2016 period. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate achieves the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. The polynomial (quadratic) specification in Column (1) is estimated over the entire support after excluding observations in the top 10% and the bottom 1% of the distribution of the distances. The local linear specification in Column (2) is estimated within a joint three-dimensional bandwidth. The bandwidth for each productivity indicator is the MSE-optimal bandwidth computed following Calonico et al. (2014) and using the companion Stata package described in Calonico et al. (2017). In both columns, we exclude the competition sectors with more than 90% successful candidates and those with fewer than 30 observations.
Standard errors, clustered at the competition sector level, in parentheses.
*** p< 0.01, ** p<0.05, and *p<0.10.

Table A5: Candidates applying to the competition sector they belong to as of 2012.

	Publications	Articles	Conf. Papers	Reviews
	(1)	(2)	(3)	(4)
ITT	1.520**	1.230**	0.218	0.045
	(0.761)	(0.533)	(0.290)	(0.109)
LATE	4.238**	3.431**	0.609	0.127
	(1.951)	(1.376)	(0.709)	(0.263)
Mean dependent variable	15.374	11.285	2.216	0.895
Sd dependent variable	13.728	10.111	5.719	2.048
Number of clusters	83	83	83	83
Observations	5024	5024	5024	5024

NOTES. This table reports the Intention-to-treat (ITT) and the Local Average Treatment Effect (LATE) of achieving the qualification on the quantity of publications over the period 2013-2016. The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors. In case of multiple applications, we consider the one to the competition sector to which the scholar already belongs as an assistant professor as of December 2012. The dependent variable in Column (1) is the total number of papers (including articles, conference papers, reviews and other items) published during the period 2013-2016; the dependent variables in Columns (2), (3) and (4) are the total number of articles, conference papers and reviews published during the period 2013-2016, respectively. The ITT is the OLS coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule; the LATE is the 2SLS coefficient of the indicator that equals one when a candidate gets the qualification, instrumented by the indicator that equals one when a candidate complies with the two-out-of-three rule. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions. Standard errors, clustered at the competition sector level, in parentheses. *** $p < 0.01$, ** $p<0.05$, and *$p<0.10$.

Table A6: Academic rank and success in the full professor NSQ

Panel A: Real applicants	All applications		Own field applications	
	(1)	(2)	(3)	(4)
	Qualified	Qualified	Qualified	Qualified
Assistant professor	-0.248***	-0.055**	-0.259***	-0.038
	(0.030)	(0.025)	(0.041)	(0.034)
Above 2/3 cutoffs		0.399***		0.453***
		(0.041)		(0.050)
Above 2/3 cutoffs × Assistant professor		-0.250***		-0.284***
		(0.039)		(0.054)
Mean Dep. Var.	0.503	0.503	0.564	0.564
Standard dev.	0.500	0.500	0.496	0.496
N. of clusters	90	90	74	74
Observations	6411	6411	4512	4512
Panel B: Potential applicants				
	All applications		Own field applications	
	(1)	(2)	(3)	(4)
	Qualified	Qualified	Qualified	Qualified
Assistant professor	-0.411***	-0.115***	-0.436***	-0.118***
	(0.016)	(0.017)	(0.018)	(0.020)
Above 2/3 cutoffs		0.482***		0.528***
		(0.025)		(0.029)
Above 2/3 cutoffs × Assistant professor		-0.469***		-0.534***
		(0.021)		(0.021)
Mean Dep. Var.	0.233	0.233	0.264	0.264
Standard dev.	0.423	0.423	0.441	0.441
N. of clusters	113	113	108	108
Observations	15085	15085	11470	11470

NOTES. This table depicts the effect of being an assistant professor, rather than an associate professor, on the probability of success in the 2012 full professor NSQ. In Panel A, the unit of analysis is an application to the full professor NSQ in 2012, and the dependent variable is an indicator that equals one in case of a successful application and zero otherwise. Columns (1) and (2) refer to the whole sample of applications in *bibliometric* competition sectors, while Columns (3) and (4) consider only applications to the candidate's competition sector (the competition sector to which she already belongs as an assistant professor at the time of the application). The main independent variable is an indicator that equals one for candidates who are assistant professor at the time of the 2012 NSQ, and zero for those who are associate professor, an indicator that equals one when a candidate complies with the two-out-of-three rule, and their interaction. Panel B is the analogous of Panel A for potential applications to the full professor NSQ in 2012. These are defined based on the list of applications to the associate professor NSQ in the same year, and thus include also assistant professors who apply to the associate professor NSQ but not to the full professor one. Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude also sectors with more than 90% successful candidates and those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A7: Heterogeneity in the distance from full professor thresholds (II)

Panel A - Dependent variable: Probability of achieving the NSQ					
	(1)	(2)	(3)	(4)	(5)
	Q1	Q2	Q3	Q4	Q5
Above 2/3 cutoffs × Quintile	0.148	0.333***	0.283***	0.328***	0.343***
	(0.099)	(0.072)	(0.057)	(0.045)	(0.064)
Mean dep. var. (Quintile)	0.812	0.754	0.673	0.498	0.178
Sd dep. var. (Quintile)	0.391	0.431	0.469	0.500	0.383
Observations (Quintile)	968	986	975	954	1004
Panel B - Dependent variable: Number of publications pre-NSQ (2008-2011)					
	(1)	(2)	(3)	(4)	(5)
	Q1	Q2	Q3	Q4	Q5
Above 2/3 cutoffs × Quintile	-0.342	-0.128	0.214	0.717	0.950
	(0.900)	(0.806)	(0.685)	(0.556)	(1.089)
Mean dep. var. (Quintile)	14.235	14.889	12.925	10.363	6.348
Sd dep. var. (Quintile)	10.611	9.838	8.674	7.089	8.686
Observations (Quintile)	968	986	975	954	1004

NOTES. This table depicts the heterogeneity of the effect of achieving the associate professor qualification depending on the distance from the full professor thresholds on the probability of achieving the qualification (Panel A) and on the quantity of publications over the pre-NSQ period (Panel B). The sample includes the candidates for the associate professor NSQ in 2012 in *bibliometric* competition sectors, grouped by quintiles of the distribution of distances from the full professor threshold for the first indicator (number of articles). Each column reports the estimated effect within each quintile, that is, the coefficient of the indicator that equals one when a candidate complies with the two-out-of-three rule interacted with the corresponding quintile-specific dummy. The distance from the full professor threshold is expressed as a percentage of the initial stock of articles, thus defined as $dist_{i,1,s} = \frac{m_{1,s}^{full} - x_{i,1}}{x_{i,1}}$, where $m_{1,s}^{full}$ is the sector-specific cutoff for the full professor NSQ and $x_{i,1}$ the professor's score in the same indicator. The dependent variable in the upper panel is an indicator that equals one when a candidate achieves the qualification, while in the bottom panel it is the total number of papers (including articles, conference papers, reviews and other items) published during the three years before the NSQ (2008-2011). Within each competition sector, we exclude observations in the top 10% and the bottom 1% of the distribution of the distances. We exclude competition sectors with more than 90% successful candidates, those with fewer than 30 observations. All regressions are estimated using a quadratic specification over the entire support that includes sector-specific dummies and interactions.
Standard errors, clustered at the competition sector level, in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.

Appendix B

Two-dimensional RDDs

Here, we replicate our analysis implementing a series of non-parametric, two-dimensional RDDs, estimated using the `rdms` command developed by Cattaneo et al. (2020). In particular, for each pair of running variables, we report the effect of passing the corresponding two-dimensional frontier in 5 different points in the neighborhood of the intersection between the zero-distance axes (where most candidates are located). First stage estimates are consistent with those obtained with our main specification: crossing two cutoffs increases the probability of getting the NSQ by 9 to 29 percentage points, depending on the specific region of the frontier that we consider. On average, the estimated increase in the probability of getting the NSQ when crossing these two-dimensional frontiers is therefore lower than the one we get with our preferred, three-dimensional specification (31 percentage points). Indeed, considering two indicators at the time does not take into account where a candidate is located in the third, omitted dimension, thus increasing the fuzziness of the design. The estimated LATE on the number of 2013-2016 publications exhibits a substantial heterogeneity depending on the region of the frontier where the treatment effect is estimated. Also, they tend to be more volatile and less precise: both the standard errors of the coefficients and the optimal bandwidths are large. That is because, within this setting, we cannot take into account the wide cross-field heterogeneity as we do in our baseline regression, in which we allow the functional form to vary across competition sectors. Still, the average of the estimated coefficients (6.55), reported at the bottom of the table, is actually very close to our baseline one (6.56).

Table B1: Two-dimensional RDDs

	First Stage			N. of publications (LATE)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Art.; Cit.	Cit.; H-Index	Art.; H-Index	Art.; Cit.	Cit.; H-Index	Art.; H-Index
(0;0)	0.294***	0.246***	0.126***	0.596	10.592**	-0.368
	(0.037)	(0.030)	(0.040)	(3.753)	(4.809)	(7.601)
BW	19.113	13.714	10.026	13.790	7.696	9.750
Within-BW Observations	4108	4146	3937	3282	3033	3871
	Art.; Cit.	Cit.; H-Index	Art.; H-Index	Art.; Cit.	Cit.; H-Index	Art.; H-Index
(1;0)	0.233***	0.203***	0.120***	-0.661	11.557**	-8.766
	(0.040)	(0.028)	(0.042)	(3.115)	(4.856)	(9.548)
BW	15.183	13.109	8.306	18.309	8.847	7.616
Within-BW Observations	3575	4128	3606	4031	3357	3377
	Art.; Cit.	Cit.; H-Index	Art.; H-Index	Art.; Cit.	Cit.; H-Index	Art.; H-Index
(0;1)	0.269***	0.205***	0.176***	1.505	20.132***	3.455
	(0.037)	(0.031)	(0.039)	(4.259)	(6.011)	(7.216)
BW	17.988	11.801	10.045	13.672	7.328	7.316
Within-BW Observations	3991	3850	3953	3313	2931	3258
	Art.; Cit.	Cit.; H-Index	Art.; H-Index	Art.; Cit.	Cit.; H-Index	Art.; H-Index
(2;0)	0.170***	0.168***	0.107**	-2.287	5.020	2.928
	(0.043)	(0.028)	(0.044)	(3.437)	(7.852)	(9.013)
BW	12.504	12.435	7.587	19.034	7.411	7.621
Within-BW Observations	3100	4056	3351	4151	3087	3399
	Art.; Cit.	Cit.; H-Index	Art.; H-Index	Art.; Cit.	Cit.; H-Index	Art.; H-Index
(0;2)	0.246***	0.156***	0.086*	1.137	19.261***	34.257*
	(0.037)	(0.035)	(0.048)	(4.689)	(6.174)	(20.053)
BW	17.523	11.517	7.549	13.981	10.748	6.392
Within-BW Observations	3969	3790	3221	3378	3660	2818
Average effect		.187			6.557	

NOTES. This table reports the LATE of achieving the qualification on the quantity of publications over the 2013-2016 period, and the corresponding first-stage estimates, obtained following the procedure in Cattaneo et al. (2020) to deal with multiple running variables. In particular, each LATE coefficient in the table is the effect of crossing a two-dimensional frontier in a given point. The two-dimensional frontier is defined on the number of articles and citations in Columns (1) and (4), the number of citations and the H-Index in Columns (2) and (5), and the number of articles and the H-index in Columns (3) and (6). In each row, the numbers in parentheses indicate the specific point along the frontier where the treatment effect is estimated (all running variables are expressed in terms of distances from the corresponding cutoff). Estimates are obtained using the `rdms` package described in Cattaneo et al. (2020). All regressions are estimated within a MSE-optimal bandwidth and include as a covariate an indicator that equals one when a candidate overcomes the third, omitted cutoff. The bottom row reports the average first stage and LATE coefficients, defined as the average of the corresponding cutoff-specific estimated effects. In all columns, we exclude the competition sectors with more than 90% successful candidates.
Robust standard errors in parentheses. *** p< 0.01, ** p<0.05, and *p<0.10.