

Bertoni, Marco; Brunello, Giorgio; Checchi, Daniele; Rocco, Lorenzo

Working Paper

Where Do I Stand? Assessing Researchers' Beliefs about Their Relative Productivity

IZA Discussion Papers, No. 13637

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Bertoni, Marco; Brunello, Giorgio; Checchi, Daniele; Rocco, Lorenzo (2020) : Where Do I Stand? Assessing Researchers' Beliefs about Their Relative Productivity, IZA Discussion Papers, No. 13637, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/227164>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 13637

**Where Do I Stand? Assessing Researchers'
Beliefs about Their Relative Productivity**

Marco Bertoni
Giorgio Brunello
Daniele Checchi
Lorenzo Rocco

AUGUST 2020

DISCUSSION PAPER SERIES

IZA DP No. 13637

Where Do I Stand? Assessing Researchers' Beliefs about Their Relative Productivity

Marco Bertoni

University of Padova and IZA

Giorgio Brunello

University of Padova and IZA

Daniele Checchi

University of Milan, INPS and IZA

Lorenzo Rocco

University of Padova and IZA

AUGUST 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Where Do I Stand? Assessing Researchers' Beliefs about Their Relative Productivity*

In 2017 the Italian government established the Fund to Finance Basic Research Activities – FFABR – with the purpose of assigning a 3,000 euros research grant to the most productive applicants among eligible assistant and associate professors. We show that, rather surprisingly, many low-productivity researchers applied to the program while many high-productivity ones did not. Our evidence from both a simple structural model of program participation estimated on registry data and a survey of the eligible population suggests that high-productivity researchers under-estimate their own position in the productivity distribution relative to the assignment threshold, while the opposite holds for low-productivity ones.

JEL Classification: I28

Keywords: competitive research funds, overconfidence, Italy

Corresponding author:

Lorenzo Rocco
Department of Economics and Management “Marco Fanno”
University of Padova
Via del Santo 33
35123 Padova
Italy
E-mail: lorenzo.rocco@unipd.it

* We are grateful to Guglielmo Barone, Maria De Paola, Gianni De Fraja and to the audience in Novara (AIEL) and Padova for the comments and suggestions. Although Daniele Checchi was actively involved in the implementation of the policy considered in this paper, the opinions expressed here are entirely personal and do not involve ANVUR. We thank ANVUR for its help with the data. The usual disclaimer applies.

Introduction

Researchers routinely decide whether to apply or not for competitive research grants. These decisions take into account the expected costs and benefits of applying. The costs include the time and effort required to prepare a research project, and the benefits comprise both the financial resources to facilitate scholarly research and the gains in individual reputation.

A crucial ingredient in the decision to apply is the subjective probability of success, which typically depends on the quality of the project as well as on individual productivity relative to the productivity of competitors. Evaluating relative productivity is difficult because the pool of competitors is often unknown and so is the threshold productivity above which grants are awarded.

Due to limited information about the productivity of competitors and the relevant productivity threshold, some researchers under-estimate the threshold and apply, despite a low probability of success, and others over-estimate the threshold and restrain from applying despite having good chances to win the award. Since the efficient allocation of scarce research funds requires that funds should go to the most productive researchers, poor information about relative productivity could produce inefficient outcomes, by altering the composition of applicants with respect to the optimal one.

This paper investigates how accurately researchers assess their relative productivity and rank themselves against their peers, using data on individual participation to a nation-wide funding program launched in Italy in 2017, the Fund to Finance Basic Research Activities, or FFABR hereafter. A peculiarity of FFABR was that applicants were not required to submit a research proposal. Rather, the award of the grant depended exclusively on the attained rank based on (past) productivity. The program was targeted at almost 37,000 assistant and associate professors – employed by Italian public universities – and awarded 3,000 euros in research funds to the top 75 (25) percent of applying assistant (associate) professors.

The individual grant was non-negligible: to give an indication of its magnitude, it was somewhat higher than the net average monthly salary of an associate professor (which ranges in Italian public universities approximately between 2600 and 3300 euro) and significantly higher than the net monthly salary of an assistant professor (between 1600 and 2000 euro).¹ In addition, the application procedure was very simple and required only a small amount of time and effort.

We argue that the observed decisions to apply or not to the program “reveal”, or at least closely reflect, individual beliefs about relative productivity. We characterize the correlation between perceived and actual *relative* productivity, defined as the difference between individual productivity and the productivity threshold above which grants are awarded, using a simple structural model.

We estimate this correlation using the data on the applications to the program by assistant and associate professors in the fields of Economics, Management Sciences and Statistics. We focus on these fields because we are more familiar with their rules and dynamics and because we believe that economists and statisticians are rather familiar with rational and strategic behaviour. We find that the estimated correlation between perceived and actual relative productivity is rather low (about 0.42), and that both the share of applicants with no chance of receiving the award and the share of potential winners who fail to apply are non-negligible.

We further investigate how accurately researchers can predict their (relative) productivity by administering to the target population a web survey, where we ask individuals to indicate to which percentile of the productivity distribution they belong. In spite of the fact that evaluating own productivity relative to average productivity is presumably less complicated than assessing the former relative to the productivity of applicants, we find that low (high) performing researchers tend to over (under)-place themselves, or over (under)-estimate their

¹ Its budget (45 million euro) was about 40 percent of the major funding program for academic research in Italy (PRIN, or *Progetti di rilevante interesse nazionale*).

position in the distribution of productivity. We also suggest that researchers often fail to correctly evaluate not only their relative productivity, but also their absolute productivity.

With many low productivity researchers applying and many high productivity researchers failing to apply, average productivity in the pool of applicants was lower than in the population, and so was the productivity threshold determining whether individuals were awarded the grant or not. Therefore, too many resources were allocated to lower productivity researchers, at the expense of those with higher productivity.

Although our results are derived from a specific program, we think that an inefficient allocation of research funds due to the imperfect assessment of relative productivity is a feature of all competitive funding programs that has been so far overlooked in the literature. Even when the cost of application is high and there is a strong incentive to invest in a reliable assessment of the probability of success, researchers with pessimistic priors might not even consider the option to apply.

To reduce this inefficiency and promote both meritocracy and the take up, comprehensive rankings of individual productivity should be compiled on a regular basis and the results should be publicly available. In the specific case of FFABR, a more effective policy would have ranked the eligible population according to the productivity index and assigned the award to those above the chosen threshold, without requiring any application process.²

Our paper speaks to two strands of economic literature. The first strand investigates the applications to competitive research grants. Cozzi, 2018, for instance, uses data by research centre to study the applications to public research funds in the fields of Social Science and Humanities in Canada and finds that

² To guarantee that public funds are not assigned to researchers who do not benefit from additional research funds, winners were allowed to turn down the award offer. Eligibility requirements also excluded, as in FFABR, the recipients of large international and national awards.

changing the rules to award grants, and consequently the expected success rate, can deter applications.³ This suggests that when reforms reduce the value of past experience, researchers become more uncertain about their probability of success and eventually reduce their applications.

Enger and Castellacci, 2016, consider instead the participation of Norwegian research institutions in the EU Horizon 2020 (H2020) program. They find that the probability of applying to H2020 funding is higher in larger institutions which have participated in previous European funding programs than in institutions which have a higher reputation and are more productive in terms of the weighted number of publications per capita.⁴ These results highlight that researchers may use information from previous rounds to update their probability of success.

The second strand is the growing literature on overconfidence. The situation characterized by individuals over-estimating their relative performance or ability is defined in this literature “over-placement” or overconfidence. Moore and Healy, 2008, classify overconfidence as: i) overestimation of one’s actual performance; ii) excessive precision in one’s beliefs; iii) over-placement of one’s own performance relative to that of others. Over-placement was questioned by Benoit and Dubra, 2011, and more recently re-affirmed by Burks et al, 2013, and, in an experimental setting by Benoit et al, 2015.⁵ Burks et al.,

³ Ley and Hamilton, 2008, consider individual applications to the NIH grants in the US. They note that female scientists apply significantly less than their male counterparts, despite the fact that the success rate is similar across genders. They suggest that females are more attracted by safer career options, such as clinical practice, that do not depend on the success in fund raising

⁴ Lepori et al., 2015, study the probability of application to European Framework Programs (EU-FP) in a large sample of European research institutions and find that participation is concentrated in a small number of large and highly reputed institutions, while country effects do not affect participation. Geuna, 1998, studies the determinants of participation in EU funding programs by European universities in 1992 and concludes that the probability of applying depends primarily on the scientific research productivity of the university. Must, 2010, and Enger, 2018, observe that participation is concentrated in a small network of core institutions which have long traditions of collaboration and high levels of reciprocal trust.

⁵ Benoit and Dubra, 2011, argued that the over-placement (or overconfidence) observed in many empirical situations could be only apparent and entirely consistent with a purely rational Bayesian updating (the so called Bayesian critique). Burks et al. 2013 and Benoit et al. 2015,

2013, analyse a sample of about one thousand trainee truck drivers and argue that over-placement is induced by the desire to send positive signals to others about one's own skill, and can result from either a bias in judgement or from strategic lying.

Our emphasis on researchers' beliefs about their relative performance clearly relates to the work by Arni et al, 2020, who study individual perceptions of own health and find that about 30 percent of the respondents to representative German surveys overestimate their rank in the population health distribution by at least 30 percentiles. This implies, for instance, that these respondents believe to rank at the 60th percentile when they actually rank at the 30th percentile. Similarly, Friehe and Pannenberg, 2019, use representative German survey data to show that people tend to over-estimate their rank in terms of gross wage and that this bias increases with age until about age 50.

We contribute to these two strands of literature with what we believe is the first study which provides evidence about researchers' biased assessments of their (relative) productivity by using a nation-wide and high-stake policy intervention. We also discuss the implications of our results for the design of competitive research funding programs.

The paper is organized as follows. Section 1 describes the policy and Section 2 introduces the data. The take up of the program is discussed in Section 3 and a model of individual choice is presented and estimated in Section 4. Section 4 presents the results of the web questionnaire. Conclusions follow.

1. The policy: institutional details

FFABR was earmarked to assistant and associate professors with a research performance above a given threshold. Full professors and professors working for private universities were excluded on the grounds that they could raise research money from external sponsors, including the Italian Ministry of

empirically test and reject the Bayesian critique and conclude that over-placement is not only a statistical artefact.

Education and the European Commission. The endowment of the fund was 45 million euros per year, and funding was supposed to cover a five-year period. The policy was first implemented in 2017 and had the following features:

- a) every assistant or associate professor could apply, conditional on being full-time employed (*“regime a tempo pieno”*) and on not being the leading investigator of other major grants (such as ERC, H2020 and the national grant PRIN);
- b) the national agency for the evaluation of university research – ANVUR – ranked applicants according to their productivity over the previous five years;
- c) the top quartile of applying associate professors and the top three quartiles of applying assistant professors⁶ were awarded 3,000 euros to be spent as research funds, with no additional strings attached. The available budget could fund 15,000 researchers per year.

The application procedure consisted of two steps: in the first step (July 2017), applicants were invited to check, update and confirm their publications record as it appeared in the Ministry of University and Research – MIUR – website. In the second step (September 2017), those who completed the first step could view their best publications – selected by ANVUR to compute their personal productivity score – and confirm (or correct) their application.

In a population of 36,935 eligible assistant and associate professors, 21,300 completed the first step in July 2017, but only 17,308 confirmed their application to the program two months later. Therefore, less than half (46.9 percent) of the total pool of assistant and associate professors applied.⁷ Despite

⁶ The quartiles were computed using the distribution of performance within a specific research area (*settore scientifico disciplinare*). All Italian professors are allocated to one of the 371 research areas, which play a crucial role for career development, selection and teaching.

⁷ Details (in Italian) can be found at https://www.anvur.it/wp-content/uploads/2018/04/FFABR_commenti_21122017.pdf.

the differences in the ex-ante probability of being awarded the grant ($\frac{3}{4}$ of applicants for assistants and $\frac{1}{4}$ of applicants for associates), the application rates were quite similar: 48.8 percent for assistants and 44.8 percent for associate professors. There was a significant variation across research areas, ranging from 69.5 percent (61.3 percent) in chemistry to 28.2 percent (28.4 percent) in medicine for assistant (associate) professors.

ANVUR evaluated applicants using the quality of their research output published between 2012 and 2016. Products could be journal articles, chapters in books or monographs. The evaluation procedure allocated these products to four quality groups, taken from the latest National Research Quality Evaluation (VQR), and attributed to each journal article a score equal to 1, 4, 7 or 10. Chapters in books and monographs received instead a score equal to 1 and 10 respectively (at most one monograph is admitted for each scholar). To take multiple authors into account, each score was weighted by the total number of authors (N), with weights equal to $\frac{1}{1+\log_{10}(N-1)}$. For each author, a maximum number M of products was considered, with M varying by research area. When an author had more than M products, the algorithm retained only the products with the highest weighted scores. Finally, individual productivity score P was obtained by summing up all scores.

Researchers in the STEM (Science, Technology, Engineering and Mathematics) area submitted mainly journal articles, with a small share of contributions to edited books. In the SSH (Social Sciences and Humanities), more than half of the submitted contributions were published in collected works and books. In the research area of Economics and Statistics – the focus of the present paper – almost half of the products submitted for evaluation were classified in the lowest group (attributed score: 1), and the remaining half was distributed rather evenly across the remaining three groups. To illustrate, in the research sub-area “Economics” (coded as SECS-P/01), a maximum of six products for each individual was evaluated and the average score attained by assistant (associate)

professors who applied was 26 (32), corresponding to 4 (5) well-published papers.⁸

At the end of the procedure, ANVUR published the list of 9,466 beneficiaries, corresponding to 25.6 percent of the eligible population (36,935 individuals) and the threshold scores for assistant and associate professors within each sub-area. The list of applicants and their score P were not made publicly available. However, since the procedure used by ANVUR to classify and evaluate each publication is public and can be replicated to a very high degree, it is possible to compute a reliable estimate of the score for both applicants and non-applicants.

An important feature of this policy is the determination of the threshold in the distribution of the publication score P above which applicants are awarded the grant. The threshold is not determined ex-ante within each research area, but varies with the quality of applicants. Compared to a baseline situation where everybody applies, it increases (decreases) if low (high) productivity researchers fail to apply.

2. The data

In our empirical analysis, we use data on the population of assistant and associate professors hired in public Italian universities until June 2017, who belong to the research area “Economics and Statistics” – in Italian “*Scienze Economiche e Statistiche*” (“SECS”) according to the classification adopted by the Ministry of Education, University and Research (MIUR). The area consists of 13 research sub-areas for economics and management (SECS-P) and of 6 research sub-areas for statistics (SECS-S)⁹.

⁸ See ANVUR, 2017. Additional information on the evaluation process can be found in the website <https://www.anvur.it/attivita/ffabr/>.

⁹ For economics and management (SECS-P), the sub-areas are: 01: economics; 02: economic policy; 03: public economics; 04: history of economic thought; 05: econometrics; 06: applied economics; 07: business administration; 08: management; 09: corporate finance; 10: business organization; 11: financial intermediation; 12: economic history; 13: commodity science. For statistics (SECS-S) we have: 01: statistics; 02: statistics for experimental and technological

We focus on Economics and Statistics for two reasons: first, as economists we have a comparative advantage in evaluating journals and outlets in this area. We exploit this knowledge in the application of the algorithm used by ANVUR to evaluate the individual productivity of applicants and non-applicants. Second, since most researchers in this area have received some training in game theory and cost-benefit analysis, their behaviour should show relatively small deviations from rational decision-making, an assumption we shall make in the theoretical model discussed below.

We drop from this population those who were not eligible to apply to the program: part-timers (“*a tempo definito*”), those temporary on leave, and principal investigators of competitive national or international funded research project – such as the Italian PRIN or FIRB and the European ERC or H2020 grants.¹⁰ We further exclude two research areas – “History of economic thought – SECS-P/04” and “Statistics for experimental and technological research - SECS-S/02” – because they have very few members.

Our working sample consists of 1,334 associate professors and 1,223 assistant professors (both on open and close-ended contracts). For each individual in this sample, we obtained from the Ministry’s public archives information on rank (assistant or associate professor), the university and area of affiliation, date of birth, gender and tenure in the rank.¹¹ We merge this information with the one provided by ANVUR on both applicants and beneficiaries, which includes the productivity score P assigned by ANVUR, and the area and rank specific threshold T used to identify those awarded the grant.¹²

The productivity score is not available, however, for those who did not apply to the program. In order to obtain a measure of productivity for applicants and non-

research; 03: economic statistics; 04: demography; 05: social statistics; 06: mathematical methods for economics, actuarial sciences and finance.

¹⁰ To exclude those not eligible, we checked individual webpages and CVs.

¹¹ Information on tenure in the rank is missing for roughly 2 percent of the sample. For these subjects, we set this variable to zero and include a dummy indicating a missing value.

¹² The data were merged for us by ANVUR and returned to us in an anonymous format.

applicants, we proceeded as follows. One of the co-authors (Daniele Checchi), who at the time of the preparation of the dataset was a member of ANVUR, had access to the eligible scientific publications of both applicants and non-applicants, available in the Ministry’s scientific catalogue for the period 2012-16. Using this information and the criteria discussed in Section 2, he computed individual scores for both applicants and non-applicants, which he then shared with the other co-authors in an anonymized version. We use these scores to derive an estimate of individual productivity \bar{P} .

We verify whether \bar{P} closely approximate the measure P computed by ANVUR for the sub-sample of applicants by plotting in Figure 1 their densities (Panel A) and scatterplot (Panel B). It turns out that the correlation between P and \bar{P} is equal to 0.95, with marginal variations across areas (0.96 for Economics and Business vs. 0.93 for Statistics).¹³ In the rest of the paper, we shall use the estimated measure of scientific productivity \bar{P} for both applicants and non-applicants.

Table 1 shows the descriptive statistics for our final sample. The table is organized in three columns, one for the full sample and the remaining two for assistant and associate professors. Slightly less than half of the full sample is female, with assistants more gender-balanced than associates. Average age is 47 years, and associate professors are on average 5 years older (49) than assistants (44). Average tenure in the rank is equal to 11 years, and associate professors have longer tenure (13 years) than assistants (9 years). Roughly three-fourths of the scholars in the sample are in economics and management, and one-fourth are statisticians. The majority (close to 40 percent) is employed in a University in Northern Italy, about a quarter in Central Italy and the rest in Southern Italy.

¹³ One reason for the observed discrepancies between the two scores is that, while we followed an automatic procedure, ANVUR could assign a higher score to a product if that product was already peer-review evaluated with a higher score by the national assessment exercise 2011-14. This explains why the actual score P dominates our estimated \bar{P} .

3. *The take up of the program*

Based on the rules designed by the Ministry, the productivity threshold for the selection of beneficiaries is endogenous and depends on the quality of applicants.¹⁴ Although researchers might observe their productivity as measured by ANVUR at a cost, they are uncertain about the relevant threshold. As a consequence, low productivity individuals may apply because they perceive their productivity to be above or at the threshold, and high productivity individuals may fail to apply for the opposite reason.

In our data, we observe that 1,304 out of 2,557 eligible subjects applied for the grant - 51 percent of the population under study.¹⁵ Table 1 breaks down these numbers by rank, showing that – in spite of the differences in the assignment probabilities – the percentage of applicants was similar across the two subsamples, and equal to 54 percent (662/1,223) for assistants and 48 percent (642/1,334) for associates. Consistent with the rules of the game, roughly 75 percent of applying assistant professors (514/662 or 77.6 percent) and 25 percent of associate professors (179/642 or 27.9 percent) who applied were awarded the grant.¹⁶

We investigate the determinants of participation in the program (or take up) by estimating a probit model where the dependent variable is a dummy equal to 1 if the individual applied and to 0 otherwise, and the covariates include a cubic

¹⁴ In the merely illustrative case where individuals have full information on their position in the distribution of productivity, the Nash equilibrium implies that very few apply. To see why, start with the worse candidate: she does not apply because she knows that she will not receive any funding. By so doing she moves (marginally) the threshold upwards. Considering next the $(n - 1)^{th}$ candidate and repeating the same argument, this candidate will not apply either. Assuming rounding to the highest integer, this process ends, in each research area, with only the top performing associate professor and the top three assistant professors applying and obtaining the grant.

¹⁵ This percentage is higher than the national percentage (46.9).

¹⁶ These percentages are slightly higher than those indicated by the law due to ties in ranking and appeals for reconsidering additional research outputs that were discarded by ANVUR. The estimated average productivity score \bar{P} is lower for assistant professors (18.96) than for associate professors (23.41 points), and higher for applicants than for non-applicants.

in the score \bar{P} , a gender dummy, a quadratic in age, tenure in the rank, research area and university fixed effects.

The results are shown in Table 2, which is organized in two columns, one for assistants and the other for associate professors. We find that the probability of applying increases with \bar{P} , but at a decreasing pace. There is also evidence that females are more likely to apply, especially among assistant professors. The application rate increases with age for associate but not for assistant professors. Higher tenure in the rank is instead negatively correlated with the probability of applying. Figure 2 plots the predicted probability of applying to the program as a function of the percentiles of \bar{P} . As one would expect, given the rules of the game that establish a lower threshold for assistants, their predicted probability is generally higher than the one for associate professors, except for the highest percentiles.

Figure 3 shows the actual application rate by rank (assistant and associate professors) and quartile of \bar{P} . For associates, who receive the award only if they belong to the top quartile, we find that more than 20 percent in the bottom quartile and more than 40 percent in the second quartile (from 25th to 50th percentile) applied. On the other hand, less than 70 percent of those in the top quartile applied. A similar pattern is observed for assistants, who received the award if they scored above the bottom quartile. We find that close to 30 percent in that quartile applied, and that between 60 and 70 percent in the third or top quartiles applied.

4. True and perceived productivity.

4.1 The model

In this sub-section, we model the decision to apply and structurally estimate the correlation between perceived and actual relative productivity – defined as the difference between own productivity and the productivity threshold – using the full sample of applicants and non-applicants.

We observe individual productivity \bar{P} , the threshold T and the decision to apply or not for each individual. Conditional on applying, individuals are awarded the grant if the gap between productivity and the threshold – or relative productivity for the purposes of this paper – is positive ($\bar{P} - T \geq 0$), and not awarded if the gap is negative ($\bar{P} - T < 0$). Individuals do not observe $\bar{P} - T$ but estimate the probability of success by relying on their experience and other signals. They could measure their *absolute* productivity using the criteria discussed in Section 2, but this requires time and effort, and may be subject to measurement error. On the other hand, the threshold is only determined at the end of the application process and is unknown at the time of application.

Let $\theta = \bar{P} - T$ and $\hat{\theta}$ be true and estimated relative productivity (the signal). The purpose of this sub-section is to estimate the correlation between θ and $\hat{\theta}$. We assume that θ and $\hat{\theta}$ have a joint normal distribution $f(\theta, \hat{\theta})$, with correlation coefficient equal to ρ , and $Var(\theta) = 1$, an innocuous normalization.

Since information is imperfect, rational individuals decide whether to apply or not using the signal $\hat{\theta}$. Applying to the program is costly. The cost $c(\hat{\theta}) = \gamma(\hat{\theta}) + \Gamma$ consists of: *i*) the deterministic component $\gamma(\hat{\theta})$, which we assume to be increasing in $\hat{\theta}$, because – for given T – individuals with higher estimated \bar{P} are likely to have higher opportunity costs; *ii*) a random component Γ , equal to zero with probability $1 - \eta$ and to $\varepsilon > 0$ with probability η . We assume that Γ is: *i*) independent of $(\theta, \hat{\theta})$; *ii*) observed by individuals before deciding whether to apply or not. Examples of random events are an unexpected professional deadline or a personal occurrence, which increases the cost of applying.

Define $\bar{c}(\hat{\theta}) = \gamma(\hat{\theta}) + \varepsilon$ and $\underline{c}(\hat{\theta}) = \gamma(\hat{\theta})$. Rational researchers apply to the program by comparing the application cost $c(\hat{\theta}) \in \{\underline{c}(\hat{\theta}), \bar{c}(\hat{\theta})\}$ with the expected benefits from applying. Let the subjective probability of being awarded the grant for an individual with perceived $\hat{\theta}$ be $P(\hat{\theta}) =$

$Pr(\theta \geq 0|\hat{\theta}) = 1 - F_{\theta|\hat{\theta}}(0|\hat{\theta})$, where $F_{\theta|\hat{\theta}}(\cdot)$ is the conditional cumulate distribution function of θ given $\hat{\theta}$. This probability is increasing in $\hat{\theta}$.

Utility is linear in income y , which varies across individuals. Since the grant can only be spent for research purposes, its marginal utility k is below one. We specify individual utility conditional on applying as

$$U(y, \hat{\theta}) = \begin{cases} y + 3000k(\hat{\theta}) - c(\hat{\theta}) & \text{if awarded} \\ y - c(\hat{\theta}) & \text{otherwise} \end{cases}$$

where $k(\hat{\theta})$ is assumed to be decreasing in $\hat{\theta}$, as – for given T – researchers with higher expected \bar{P} typically have better funding capacity and therefore already have higher fund endowments.

Individuals apply to the program if their expected utility from doing so is higher than the cost, that is to say if

$$Pr(\theta \geq 0|\hat{\theta})(y + 3000k(\hat{\theta}) - c(\hat{\theta})) + [1 - Pr(\theta > 0|\hat{\theta})](y - c(\hat{\theta})) \geq y \quad (1)$$

which can be written as

$$F_{\theta|\hat{\theta}}(0|\hat{\theta}) \leq 1 - \frac{c(\hat{\theta})}{3000k(\hat{\theta})} \quad (2)$$

Both the left-hand side and the right-hand side of (2) are monotonically decreasing in $\hat{\theta}$. Moreover, the left-hand side tends to 1 and 0 respectively when $\hat{\theta}$ tends to minus and plus infinity. On the other hand, the right-hand side is bounded below 1 when $\hat{\theta}$ goes to minus infinite. Therefore, and assuming single crossing, the right-hand-side crosses the left-hand-side from below.

Consider the case when $c(\hat{\theta}) = \underline{c}(\hat{\theta})$ and let $f(\hat{\theta}|\theta)$ be the conditional density of $\hat{\theta}$ given θ . Single crossing implies that there is a unique threshold value of $\hat{\theta}$, denoted as $\underline{\tau}$, such that all those with $\hat{\theta} \geq \underline{\tau}$ apply and those with $\hat{\theta} < \underline{\tau}$ do not apply the conditional application rate is $\Pi(\theta, \underline{\tau}) = \int_{\underline{\tau}}^{+\infty} f(\hat{\theta}|\theta) d\hat{\theta} = 1 - F_{\hat{\theta}|\theta}(\underline{\tau}|\theta)$, where $F_{\hat{\theta}|\theta}(\cdot)$ is the conditional cumulate density function of $\hat{\theta}$ given

θ . When instead $c(\hat{\theta}) = \bar{c}(\hat{\theta})$, the relevant threshold is $\bar{\tau}$ and the conditional application rate is $\Pi(\theta, \bar{\tau}) = \int_{\bar{\tau}}^{+\infty} f(\hat{\theta}|\theta) d\hat{\theta} = 1 - F_{\hat{\theta}|\theta}(\bar{\tau}|\theta)$.

Since the conditional density $f(\hat{\theta}|\theta)$ is normal with mean $\mu(\theta) = E(\hat{\theta}|\theta)$ and variance $s^2 = \sigma_{\hat{\theta}}^2(1 - \rho^2)$, the application rate can be re-written as

$$\Pi(\theta, \underline{\tau}) = \int_{\underline{\tau}}^{+\infty} f(\hat{\theta}|\theta) d\hat{\theta} = 1 - \Phi\left(\frac{\underline{\tau} - \mu(\theta)}{s}\right) \quad (3)$$

where $\Phi(\cdot)$ is the standard cumulate normal.

We repeat the same procedure when $c(\hat{\theta}) = \bar{c}(\hat{\theta})$. Averaging over the two values of $c(\hat{\theta})$, we obtain the average application rate as function of θ

$$\Pi(\theta) = \eta\Pi(\theta, \bar{\tau}) + (1 - \eta)\Pi(\theta, \underline{\tau}) = 1 - \left[\eta\Phi\left(\frac{\underline{\tau} - \mu(\theta)}{s}\right) + (1 - \eta)\Phi\left(\frac{\bar{\tau} - \mu(\theta)}{s}\right)\right] \quad (4)$$

Result: the correlation between true and estimated relative productivity is a monotonic transformation of the slope of the average application rate $\frac{\partial \Pi(\theta)}{\partial \theta}$

$$\rho = \sqrt{\frac{2\pi\left(\frac{\partial \Pi(\theta)}{\partial \theta}\right)^2}{2\pi\left(\frac{\partial \Pi(\theta)}{\partial \theta}\right)^2 + 1}} \quad (5)$$

Proof: using a first order Taylor expansion around 0,¹⁷ we have $\Phi\left(\frac{\underline{\tau} - \mu(\theta)}{s}\right) \cong$

$\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \frac{\underline{\tau} - \mu(\theta)}{s}$ and $\Phi\left(\frac{\bar{\tau} - \mu(\theta)}{s}\right) \cong \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \frac{\bar{\tau} - \mu(\theta)}{s}$ and Eq.(4) can be written as

$$\left[\frac{1}{2} + \frac{1}{s\sqrt{2\pi}}\left(\eta\bar{\tau} + (1 - \eta)\underline{\tau} - \mu(\theta)\right)\right] = 1 - \Pi(\theta), \text{ and after rearranging terms}$$

we get that

$$\mu(\theta) = [\eta\bar{\tau} + (1 - \eta)\underline{\tau}] - s\sqrt{2\pi}\left(\frac{1}{2} - \Pi(\theta)\right) \quad (6)$$

Because of normality and the independence of Γ with respect to $[\theta, \hat{\theta}]$, we have that $E(\hat{\theta}|\theta, \Gamma) = E(\hat{\theta}|\theta) = \mu(\theta) = a + b\theta$, where $a = E(\hat{\theta}) - bE(\theta)$ and

¹⁷ Since the derivative of the normal density at the value zero is also zero, this is also the second order Taylor expansion.

$b = cov(\hat{\theta}, \theta)$. Taking the derivatives of both sides of (6) with respect to θ and using $\frac{\partial \mu(\theta)}{\partial \theta} = cov(\hat{\theta}, \theta)$ we obtain

$$cov(\hat{\theta}, \theta) = s\sqrt{2\pi} \frac{\partial \Pi(\theta)}{\partial \theta} \quad (7)$$

Dividing both sides of (7) by s , the conditional standard deviation of $\hat{\theta}$, and recalling that the standard deviation of θ is equal to 1, we get

$$\frac{cov(\hat{\theta}, \theta)}{s} = \frac{\rho}{\sqrt{1-\rho^2}} = \sqrt{2\pi} \frac{\partial \Pi(\theta)}{\partial \theta} \quad (8)$$

which yields (5). Equation (8) also implies that $\frac{\partial \Pi(\theta)}{\partial \theta}$ needs to be constant. QED

A distinguishing feature of our data is that, since we observe both θ and the average application rate Π , we can estimate the function $\Pi(\theta)$ and derive an estimate of the correlation coefficient ρ . The higher this correlation, the smaller the discrepancy between perceived and true relative productivity.

4.2 Empirical application

Figure 4 shows that the distribution of $\theta = \bar{P} - T$ is approximately normal. We divide $\bar{P} - T$ in 50 quantile-spaced bins, each containing about 50 observations and plot in Figure 5 the application rate as function of each bin, the linear fit and the locally weighted scatterplot smoothing. The figure suggests that $\Pi(\bar{P} - T)$ is approximately linear, consistent with $\frac{\partial \Pi(\theta)}{\partial \theta}$ being constant, an implication of the model presented in the previous sub-section.

We estimate the linear probability model

$$E(D_i | (\bar{P} - T)_i) = \alpha_0 + \alpha_1(\bar{P} - T)_i + \alpha_2 X_i + \varepsilon_i \quad (9)$$

where D_i is a binary variable taking value 1 if individual i applied to the program and 0 otherwise and X is a vector of field-by-rank dummies (with standard errors clustered at the same level). We find that $\hat{\alpha}_1 = \frac{\partial \Pi(\theta_i)}{\partial \theta_i} = 0.185$ (95% confidence

interval equal to [0.143, 0.227]) and that $\hat{\rho}$, the correlation between $\hat{\theta}$ and $\bar{P} - T$ is 0.421 (95% confidence interval=[0.340,0.494], a relatively low value.

Since $\hat{\theta} = E(\hat{\theta}|\theta) + \omega$, where ω is a measurement error with zero mean and variance σ_ω^2 , and the variance of $\theta = \bar{P} - T$ is normalized to 1, the correlation coefficient ρ is equal to $\frac{1}{\sqrt{1+\sigma_\omega^2}}$, suggesting that the lower is ρ the higher is the

variance of measurement error. With $\hat{\rho}=0.421$, we get that $\widehat{\sigma_\omega^2}=4.64$, a very large value considering that $Var(\theta) = 1$. Such a high variance σ_ω^2 implies that in our sample the discrepancy between $\hat{\theta}$ and θ can be high. Therefore, we are likely to observe researchers with high $\hat{\theta}$ but low θ – who apply to the program but do not receive the grant – as well as researchers with low $\hat{\theta}$ but high θ – who fail to apply but would have won the grant had they applied.

The correlation between perceived and true relative productivity can vary across population groups. Older researchers, for instance, could have a more precise prior because a longer experience cumulates signals and expands the information set. To test this hypothesis, we split our sample between those aged above and below median age (45) and estimate Eq. (9) separately for the two groups.¹⁸ As expected, the estimated $\hat{\rho}$ is 0.244 [confidence interval: 0.165, 0.218] for the younger age group and 0.472 [confidence interval: 0.374, 0.555] for the older group.

Researchers learn about their relative productivity by comparing themselves with their proximate colleagues. If the level of productivity is similar, such comparison is not particularly informative. Conversely, researchers who work in more heterogeneous environments have a better chance of learning about their relative productivity by observing the performances of others. We compute the standard deviation of \bar{P} , $SD(\bar{P})$, by university, scientific field (economics,

¹⁸ To meet the requirements of the model, we always standardize $(\bar{P} - T)$ and check normality in each subsample.

business and statistics), and rank¹⁹ and estimate Eq. (8) separately for researchers who work in an homogeneous ($SD(\bar{P})$ below the median) and heterogeneous ($SD(\bar{P})$ at or above the median) environment.²⁰ We find that $\hat{\rho}$ is larger for the former than for the latter group – and equal to 0.500 [confidence interval: 0.428,0.563] and 0.414 [confidence interval: 0.321, 0.497] respectively – but the difference is not statistically significant.

Researchers can also learn about their relative productivity by consulting publicly available rankings such as RePEc (Research Papers in Economics), which rely on publications and citations and are therefore correlated with the ranking provided by ANVUR to award the grant. Since these rankings are not available for statisticians and business economists, we expect that the correlation between perceived and true relative productivity is higher for economists than for business economists and statisticians. In line with this, we find that the estimated $\hat{\rho}$ is 0.467 [confidence interval: 0.373, 0.547] for the former and ranges between 0.338 [confidence interval: 0.185,0.469] and 0.461 [confidence interval: 0.304, 0.585] for the latter.

Interestingly, we find that $\hat{\rho} = 0.527$ among researchers who work in Southern universities [confidence interval: 0.443,0.598], much higher than $\hat{\rho} = 0.373$ in the Centre-North [confidence interval: 0.278,0.457]. A candidate interpretation of this difference is that researchers in the South have fewer opportunities of obtaining research funds than those in the rest of the country. Therefore, they value more the 3000 euro offered by FFABR (a higher k in the parlance of the model), and have stronger incentives to have an accurate estimate of the probability of success.²¹

¹⁹ The threshold T is approximately constant within each scientific field and rank.

²⁰ We drop cells (i.e individuals in the same university/field/rank) with less than five observations.

²¹ We have also estimated ρ by gender and rank. We obtain $\rho=0.432$ for males (95%CI [0.335,0.515]); $\rho=0.421$ for females (95%CI [0.339,0.497]); $\rho=0.382$ for associates (95%CI [0.271,0.479]) and $\rho=0.328$ for assistants (95%CI [0.237, 0.411]).

5. *The questionnaire*

Our finding that the correlation between perceived and true relative productivity is low helps explaining why we have found that many relatively low performers applied to the program, while many relatively high performers did not apply. The observed low correlation is partially due to the fact that the assignment threshold is difficult to predict because it refers to the population of applicants rather than to the population itself.

In order to assess whether researchers over or under-place themselves relative to the relevant population as well, and also whether they manage to reliably assess their own absolute productivity, we used a web survey to ask them to report their expected position – or percentile – in the distribution of productivity by rank and sector. In December 2019 we sent to the population of 2,557 researchers a short questionnaire, asking their perceived productivity percentile, $\widehat{pct(\bar{P})}$, whether they applied to the program in 2017, and inquiring about the possible reasons for applying or not applying.

The response rate was 26.5 percent, but only 20.3 percent gave us permission to use their personal details to link them to their productivity \bar{P} . Since the sample of respondents who gave us permission to use their personal records is clearly a non-random sample of the relevant population, we reweight it to guarantee that observables are balanced between respondents and non-respondents. We do this by using entropy balancing, a data pre-processing procedure that reweights the dataset to obtain that the covariate distributions in the reweighted data satisfy a set of specified moment conditions (see Hainmueller 2012 for details).

Table 3 shows the mean and variance of observables for respondents and non-respondents before and after reweighting. When we compare the two samples, we find that, before reweighting, the former includes a lower share of applicants and has a lower average perceived productivity percentile $\widehat{pct(\bar{P})}$ than the latter.

After reweighting, however, these two key variables have the same mean and variance.

Using the reweighted sample, we find that the estimated correlation between $\widehat{pct(\bar{P})}$ and the percentile of $pct(\bar{P})$ is below 0.5 (0.449). We plot in Figure 6 – separately for assistant and associate professors – a local polynomial smooth of $\widehat{pct(\bar{P})}$ on $pct(\bar{P})$. The fitted line (with shaded confidence intervals) shows that the predicted prior lies above $pct(\bar{P})$ when the latter is below the 6th decile, and below $pct(\bar{P})$ for higher deciles, suggesting that low (high) performing researchers tend to over (under)-place themselves in the distribution of productivity. These findings suggest that those researchers who apply but fail to win the award may do so not only because they under-estimate the assignment threshold, but also because they over-estimate their relative productivity. On the other hand, those who fail to apply in spite of having a high productivity may do so not only because they over-state the assignment threshold but also because they under-estimate their relative productivity.

If we assume that, within each rank and sector, actual productivity \bar{P} (in level) and perceived productivity \hat{P} (in level) are identically and jointly normally distributed, $corr(\bar{P}, \hat{P})$ can be obtained from $corr(pct(\bar{P}), \widehat{pct(\bar{P})})$ and turns out to be equal to 0.466.²² This result implies that not only researchers have on average an imperfect perception of their relative productivity, however defined, but also often fail to correctly predict their absolute productivity.²³

²² Under joint normality there exists a close relationship between Spearman correlation ρ_s and Pearson correlation ρ_p . The latter is approximatively $\rho_p = 2 \sin(\pi/6 * \rho_s)$ (see Boudt et al. 2012).

²³ Survey respondents include those who applied to the program, who knew at the time of the survey whether they received the grant or not. This information could have been used to improve their perception of their relative productivity, thereby reducing observed over- or under-confidence. There is no obvious relationship between $corr(\bar{P}, \hat{P})$ and ρ , the parameter that we have estimated in the previous section. We recall that ρ is the correlation between $\bar{P} - T$, the researcher's actual productivity relative to the threshold T , and the signal $\hat{\theta}$. By letting $\hat{\theta} = \hat{P} - \hat{T}$, where \hat{T} is the perceived threshold, we have $\rho = corr(\bar{P}, \hat{P}) \frac{\sigma_{\bar{P}} \sigma_{\hat{P}}}{\sigma_{\bar{P} - \hat{T}}} + corr(T, \hat{T}) \frac{\sigma_T \sigma_{\hat{T}}}{\sigma_{\bar{P} - \hat{T}}} -$

We use the results of the web survey to investigate the reasons given by respondents for applying and not applying to the program. Since each respondent could provide at most three answers, total frequencies do not add up to unity. Starting from the reasons for applying, we consider the following options: a) the probability of being awarded the grant was good; b) it did not cost to try; c) I needed the money; d) many colleagues in my Department applied; e) I was hoping that few participated; f) other.

Table 4 shows the frequencies of each answer by rank and quartile of productivity \bar{P} . We find that the view that the probability of getting the award was good is the modal answer, selected by close to 90 percent of assistants and associates in the top quartile of \bar{P} . Surprisingly, option a) was chosen by 66.8 (64.4) percent of associate professors with an estimated productivity in the bottom (second) quartile of the distribution, who could only get the award if those in the higher quartiles did not apply. Similarly, 43.4 percent of assistants in the bottom quartile, who were unlikely to get the grant, replied that they judged their probability of being awarded as good.

Second only to a) is option c) “I needed the money”, which attracted between 37 and 56 percent of the answers. Many (between 23 and 39 percent) selected option b), confirming that the cost of going through the procedure in order to apply was negligible. There is instead little evidence that applicants were influenced by the behaviour of their peers and that they behaved strategically (applying when they expected a lower threshold because few people with a better score participated).²⁴

$corr(\bar{P}, \hat{T}) \frac{\sigma_{\bar{P}\hat{T}}}{\sigma_{\bar{P}-\hat{T}}} - corr(\hat{P}, T) \frac{\sigma_{\hat{P}T}}{\sigma_{\hat{P}-T}}$. We cannot establish a priori whether ρ is larger or smaller than $corr(\bar{P}, \hat{P})$. Much depends on 1) the correlations between \bar{P} and \hat{T} and 2) the correlation between \hat{P} and T .

²⁴ We have investigated peer effects in program participation by including in the probit model of Table 2 the average productivity of scholars in the same university, rank and area (economics, management or statistics). We find that peers' productivity has no effect on the individual decision to apply. Results are available from the authors.

Turning to the reasons for not applying (Table 5), we consider the following: a) I did not know about the program; b) I forgot to apply; c) I had no chance of getting the award; d) I had no time to apply; e) I was afraid that results could become public; f) few colleagues in my Department applied; g) the rules of the game excluded me; h) other.

We find that about one third of the respondents selected option a) and about 20 percent chose option b). In either case, we do not observe a clear relationship with productivity. Both results are perhaps to be expected for those in the lower quartiles, who had a low probability of getting the grant, but not for those with a relatively high performance, who were likely to win. The fact that many report lack of information as a reason is surprising given the effort spent by ANVUR and universities to inform.²⁵ The importance of lapses of memory is confirmed also by the comparison between the number of researchers who started their application procedure in July 2017 (1,576)²⁶ and the number who finalized it in September 2017 (1,304).²⁷

These lapses could be due to several reasons, including that academic research is not the priority for many assistant and associate professors, either because of their teaching and administrative obligations or because of their involvement in non-academic activities (such as private consulting), political opposition to the

²⁵ The model in Section 4 assumes that the choices of applying and non-applying are deliberate and excludes the possibility that someone could be unaware of FFABR. We check whether our results change when we exclude the uninformed from the population. Given that we know the identity of the uninformed only for those who answered the web survey, we proceed by adopting a rather crude imputation strategy. Taking at face value the responses to the web survey, we assume that one third of non-applicants are uninformed and we keep this proportion constant at all levels of θ . We split θ in 50 bins of equal size, compute $\Pi(\theta)$ within each bin, and re-scale it as $\tilde{\Pi}(\theta) = \frac{\Pi(\theta)}{\Pi(\theta) + 2/3(1 - \Pi(\theta))}$. Finally, we alternatively regress $\tilde{\Pi}(\theta)$ and $\Pi(\theta)$ on θ with the usual controls and clustering standard errors by bin. We find that the coefficients associated with θ are very similar to the baseline. We conclude that our findings do not depend on the assumption that all researchers are aware of FFABR.

²⁶ Only six among those who failed to complete were able to prove that they could not complete the application because of technical problems. They were eventually restored in the procedure.

²⁷ The total number of researchers, across all fields, who started and finalized the procedure in July was 21,300 and 17,308 respectively.

program²⁸ and inertia – or failure to take advantage of the opportunities that arise, a non-rational behaviour already documented in the literature (see Ornaghi and Tonin, 2018).

As expected, researchers in the lower quartiles of the distribution of performance were more likely to choose option c), and to admit that they had a small chance of winning the grant. Virtually none of the respondents did not apply because of the fear that results could have been made public, and very few were discouraged by the fact that few colleagues in their Department applied.

Conclusions

In 2017 the Italian government established the Fund to Finance Basic Research Activities – FFABR – with the purpose of assigning a 3,000 euros research grant to the best performers among the assistant and associate professors who applied. Unique features of this program were that funds were assigned exclusively on the basis of past productivity, and that the application procedure was almost costless. Applicants were ranked by a productivity index, using an algorithm that was made public in advance. Assistant professors ranked in the top 75 percent and associate professors ranked in the top 25 percent were awarded the grant.

We have used data on applications and measured productivity to estimate the correlation between perceived and actual relative productivity for the population of assistant and associate professors in the fields of Economics, Management and Statistics. We have estimated that the correlation between perceived and actual relative productivity is about 0.42, suggesting that these scholars have a rather poor assessment of their own relative productivity. This low correlation is consistent with the fact that many low performers applied and many high

²⁸ The program was criticised for its failure to ensure a minimum grant to everyone and for being excessively meritocratic. In addition, political opponents of the government could have intentionally discarded this opportunity. Yet, unless these opponents were concentrated in specific parts of the distribution, our results would still hold.

performers did not apply to the program. Using the results of a web survey, we have also shown that the observed behaviour is consistent with many low (high) performers over (under)-placing themselves relative to the relevant population, and with many individuals suffering of perception bias.

These results have important implications for the design of competitive funding policies, which are increasingly adopted to provide resources for scientific research. They suggest that –over or under-confidence affect participation to competitive grants, with implications for the efficient allocation of resources. We believe that one way to address this problem is that timely and reliable information on relative productivity should be offered to researchers.

In the specific experience of FFABR, an alternative design would have considered the entire population, ranked it by productivity and awarded the grant to those performing above a given threshold, without requiring an explicit application process. The current and the proposed design would be equivalent if everybody applied under the original policy. Yet we have shown that, in spite of the trivial application costs and the non-negligible size of the grant, many did not apply even among the more productive. Because of this, the average productivity of the pool of applicants turned out to be lower than the one prevailing in the population, and the productivity threshold determining whether individuals were awarded or not was also below the one associated with full application.

Figures 7 and 8 illustrate, for assistant and associate professors, the distribution of the productivity score \bar{P} for both applicants and non-applicants. Each figure includes two thresholds: the actual one (dashed vertical line) and the potential one, that would have been obtained had everybody applied (solid vertical line). For both groups, the actual threshold is lower than the potential one, the gap being larger for associate professors. The actual threshold is lower because of the many who did not apply in spite of their high productivity, and of the many who applied even though their productivity was low.

Compared to a policy where grants are assigned by the relevant agency by ranking the population rather than applicants, the current policy based on the application by candidates and an endogenous threshold has two drawbacks: first, since the endogenous threshold was lower than the given threshold, the average productivity of funded individuals was also lower. Second, since the application rate was well below 100 percent, the share of awarded individuals was much lower than expected. Therefore, not all the available resources were distributed.

The failure to use the available resources during the first year led to the discontinuation of the policy. After all, if so many grants could not be awarded, perhaps this money was not needed. The policy was considered by many opponents as too meritocratic, for instance because, rather than simply adding up publications, it relied on weighing each publication with weights based on the ranking of journals. We believe that the alternative policy – without applications and with a given threshold – would have been a better tool to foster the incentives to produce high quality research, especially if maintained for five years, as originally planned.

References

- ANVUR, 2017. Analisi degli esiti delle procedure per il finanziamento delle attività base di ricerca. mimeo
- ANVUR, 2018. Rapporto Biennale sullo Stato del Sistema Universitario e della Ricerca 2018. Roma
- Arni, P, Dragone, D, Goette, L and Ziebarth, N., 2020, Biased Health Perceptions and Risky Health Behaviors: Theory and Evidence, IZA Discussion Paper n. 13308
- Auranen, O and Nieminen, M, 2010, University Research Funding and Publication Performance – an International Comparison, *Research Policy*, 39 822-834.
- Benoit, JP, and Dubra, J., 2011, Apparent Overconfidence, *Econometrica*, 79(5) :1591–1625
- Benoit, JP, Dubra, J and Moore, D, 2015, Does the Better than Average Effect Show that People are Overconfident? Two Experiments, *Journal of the European Economic Association*, 13, 293-329
- Boudt, K., Cornelissen, J., Croux, C., 2012, The Gaussian rank correlation estimator: robustness properties, *Statistics and Computing*, 22:471-483
- Burks, S., Carpender, J, Goette, L and Rustichini, A., 2013, Overconfidence and Social Signalling, *Review of Economic Studies*, 80, 949-83
- Cozzi, M., 2018, Public funding of research in the social sciences: are Canadian academics discouraged? University of Victoria, Economics Department Discussion Paper DDP1809
- Enger, S.G. and Castellacci, F., 2016, Who gets Horizon 2020 research grants? Propensity to apply and probability to succeed in a two-step analysis. *Scientometrics* (2016) 109:1611–1638
- Enger, S.G., 2018, Closed clubs: network centrality and participation in Horizon 2020. *Science and Public Policy*, 45(6), 884–89
- Friehe, T and Pannenberg, M, 2019, Overconfidence over the lifespan: Evidence from Germany, *Journal of Economic Psychology*, 74
- Geuna, A., 1998, Determinants of university participation in EU-funded R&D cooperative projects. *Research Policy*, 26(6), 677–687.
- Hainmueller, H., 2012, Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies, *Political Analysis*, 20, 25-46.
- Lepori, B., Veglio, V., Heller-Schuh, B., Scherngell, T., & Barber, M., 2015. Participations to European Framework Programs of higher education

institutions and their association with organizational characteristics. *Scientometrics*, 105(3), 2149–2178. doi:10.1007/s11192-015-1768-2.

Ley, T.J. and Hamilton, B.H., 2008, The gender gap in NIH grant applications. *Science*, 322:1472-74

Must, U., 2010, Collaboration in EU Framework programmes: the case of the social sciences and humanities. *Innovation: The European Journal of Social Science Research*, 23(1), 79–83.

Ornaghi, F and Tonin, M, 2018, Water tariffs and cosumer's inaction, IZA Discussion Paper, n.11458.

Tables and Figures.

Table 1. Descriptive statistics

	(1) Full sample	(2) Assistant professors	(3) Associate professors
Female	0.46	0.50	0.42
Age	47.03	44.30	49.53
Tenure in the rank (years)	11.17	9.06	13.10
Tenure in the rank missing	0.02	<0.01	0.03
Economics and management sub-area	0.73	0.74	0.72
Statistics sub-area	0.27	0.26	0.28
Employed in a University in Northern Italy	0.42	0.38	0.46
Employed in a University in Central Italy	0.24	0.25	0.24
Employed in a University in Southern Italy	0.34	0.37	0.30
Applied to the program	0.51	0.54	0.48
Awarded the grant	0.27	0.42	0.13
Awarded the grant conditional on applying	0.53	0.78	0.28
Estimated score \bar{P}	21.28	18.96	23.41
Estimated score \bar{P} conditional on applying	25.51	22.44	28.68
Observations	2,557	1,223	1,334

Table 2. Probability of applying to the program. Marginal effects, probit model. Assistant and associate professors.

	Assistants	Associates
Estimated score	0.039*** (0.059)	0.042*** (0.006)
Estimated score ²	-0.064*** (0.020)	-0.074*** (0.016)
Estimated score ³	0.002 (0.002)	0.004*** (0.001)
Female	0.091*** (0.032)	0.058* (0.032)
Age	-0.18 (0.026)	0.056** (0.028)
Age ²	0.013 (0.026)	-0.062** (0.027)
Tenure in the rank	-0.012** (0.005)	-0.017*** (0.006)
Observations	1,220	1,334
University fixed effects	Yes	Yes
Research sub-field fixed effects	Yes	Yes

Notes: *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$. Robust standard errors within parentheses. Each regression includes also a dummy equal to 1 if tenure in the rank is missing and to 0 otherwise. We lose three observations in the first column because this dummy cannot be estimated.

Table 3. Rebalancing the sub-sample of respondents using entropy balancing.

	Respondents Mean	Respondents Variance	Non- respondents Mean	Non- respondents variance
<i>Before re-balancing</i>				
Score \bar{P} percentile	50.67	833.90	61.03	713.60
Applied to the program	0.51	0.25	0.71	0.21
Rank (1: associate; 0: assistant)	0.52	0.25	0.51	0.25
<i>After rebalancing</i>				
Score \bar{P} percentile	50.67	833.90	50.69	834. 0
Applied to the program	0.51	0.25	0.51	0.25
Rank (1: associate; 0: assistant)	0.52	0.25	0.52	0.25

Table 4. Reasons for applying to the grant. At most three choices per respondent. By quartile of productivity score \bar{P} .

Score quartile	1 st	2 nd	3 rd	4 th
<i>Assistant professors</i>				
Good probability of getting the grant	0.434	0.686	0.803	0.944
It did not cost trying	0.362	0.326	0.251	0.252
I needed the money	0.501	0.423	0.512	0.447
Many colleagues in my Department applied	0.099	0.083	0.038	0
I was hoping few with a better score participated	0.025	0	0	0
Other reasons	0	0	0	0.029
<i>Associate professors</i>				
Good probability of getting the grant	0.668	0.644	0.732	0.931
It did not cost trying	0.396	0.351	0.307	0.237
I needed the money	0.445	0.566	0.426	0.373
Many colleagues in my Department applied	0.032	0.046	0.024	0.068
I was hoping few with a better score participated	0	0.032	0	0
Other reasons	0.074	0	0.042	0

Source: replies to questionnaire. Columns do not add to 100 because multiple answers were allowed.

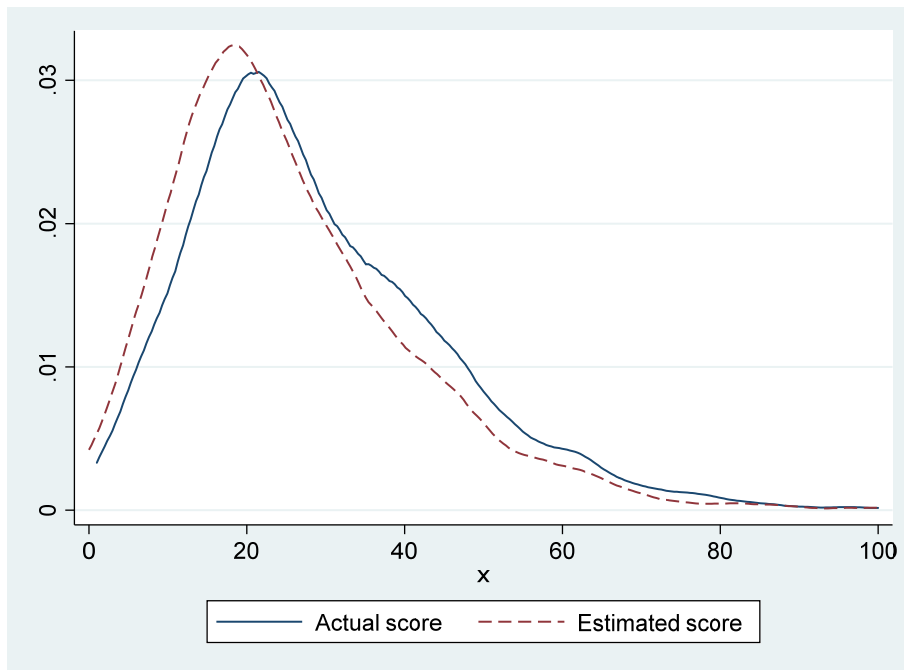
Table 5. Reasons for not applying to the grant. At most three choices per respondent. By quartile of score \bar{P} .

Score quartile	1 st	2 nd	3 rd	4 th
<i>Assistant professors</i>				
Did not know about the grant	0.309	0.249	0.295	0.427
Forgot to apply	0.095	0.401	0.292	0.136
I had no chance of getting the grant	0.414	0.173	0.096	0
I had no time to apply	0.036	0.062	0.207	0.297
I was afraid results could be public	0.036	0	0	0
Few colleagues in my Department applied	0	0	0	0
The rules excluded me	0.036	0	0.107	0.139
Other reasons	0.086	0.067	0.094	0.138
<i>Associate professors</i>				
Did not know about the grant	0.316	0.175	0.353	0.349
Forgot to apply	0.126	0.294	0.102	0.167
I had no chance of getting the grant	0.393	0.214	0.163	0.080
I had no time to apply	0.029	0.119	0.224	0.158
I was afraid results could be public	0	0	0	0
Few colleagues in my Department applied	0.034	0	0	0
The rules excluded me	0.145	0.100	0.211	0.171
Other reasons	0	0.193	0	0.157

Source: replies to questionnaire. Columns do not add to 100 because multiple answers were allowed.

Figure 1. Densities and scatterplot of actual and estimated scores P and \bar{P} . Sub-sample of applicants.

Panel A. Densities



Panel B. Scatterplot

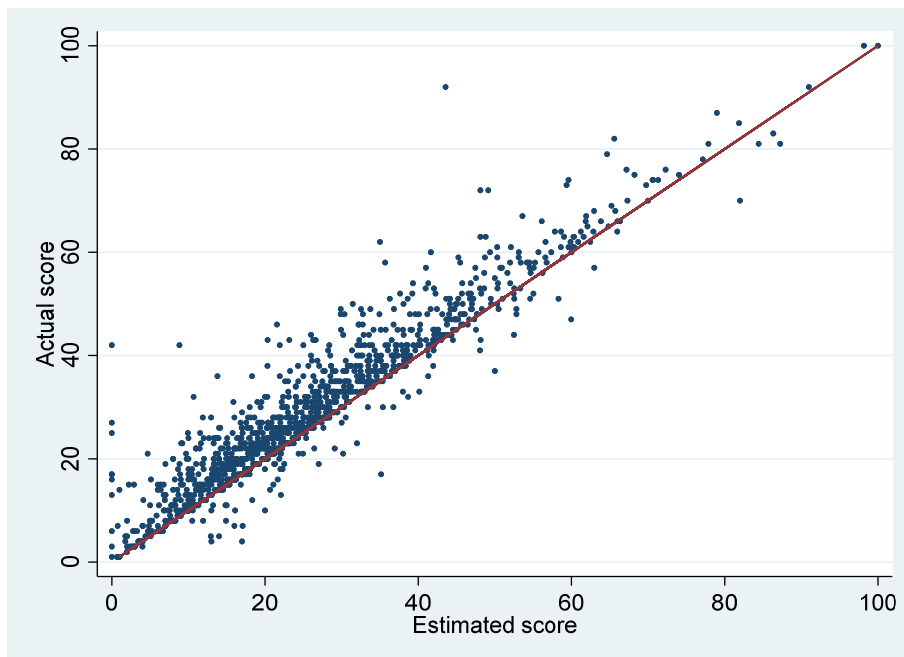


Figure 2. Predicted probability of applying to the program as function of the score \bar{P}

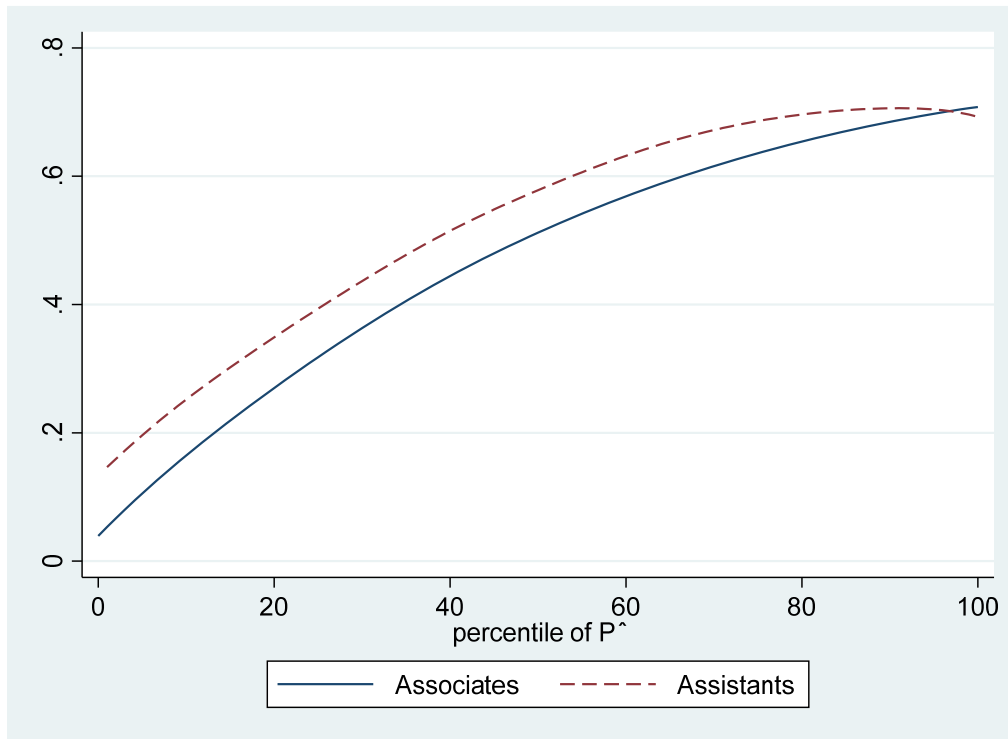


Figure 3. Actual application rate, by quartile of productivity \bar{P}

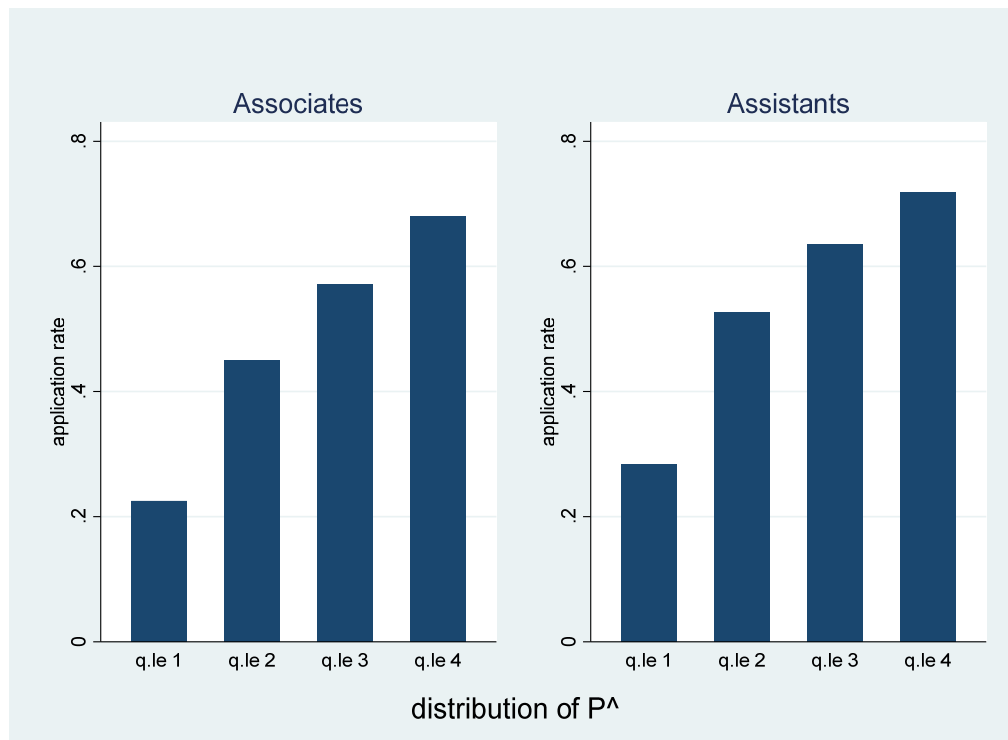


Figure 4. The distribution of $\bar{P} - T$ in the data

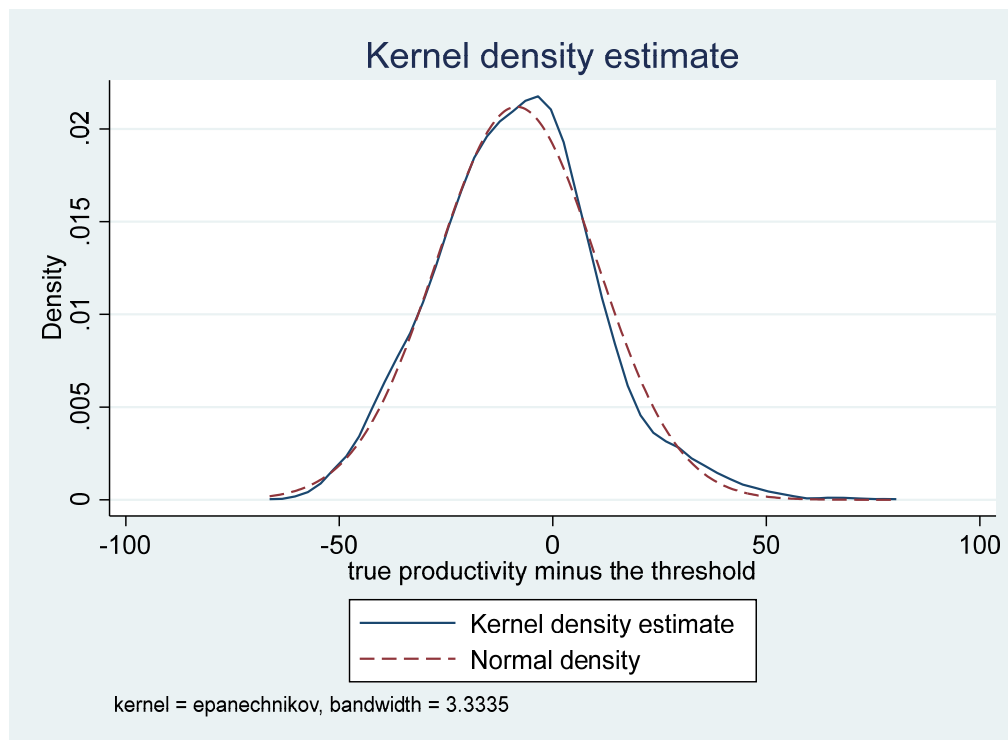
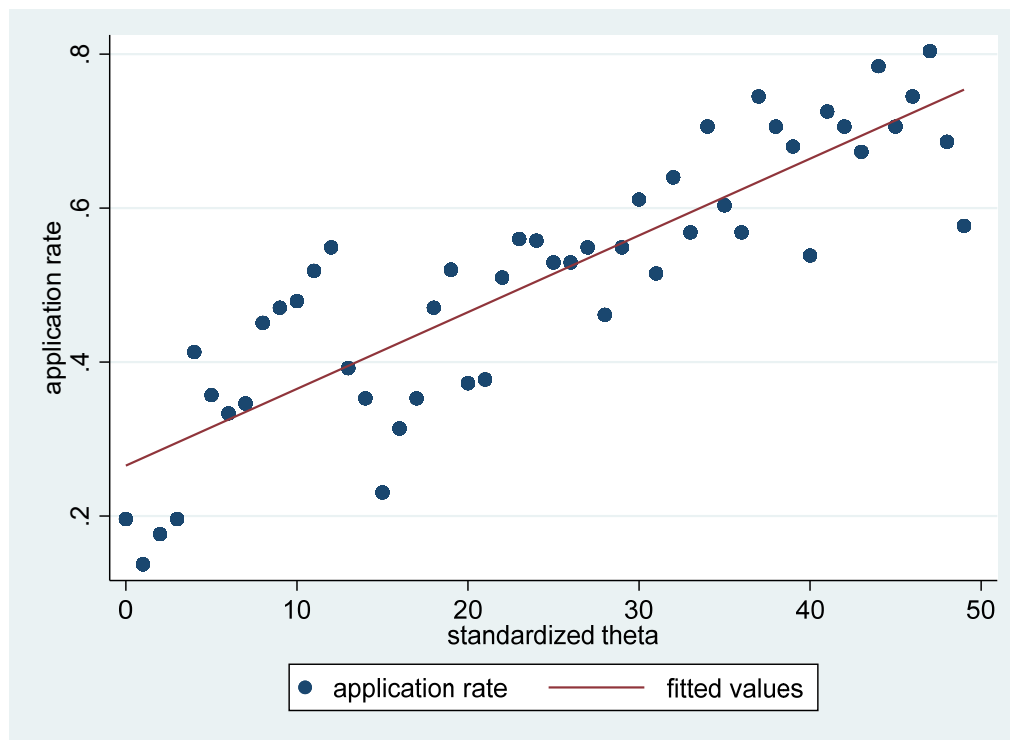


Fig 5. The application probability $\Pi(\theta)$ as a function of the bins of θ



Note: each bin includes about 50 observations. Average θ by bin is reported on the horizontal axis. The lowest and the highest bins are trimmed out.

Figure 6. Local polynomial smooth of the productivity prior \tilde{P} (percentiles) on true productivity \bar{P} (percentiles). By rank.

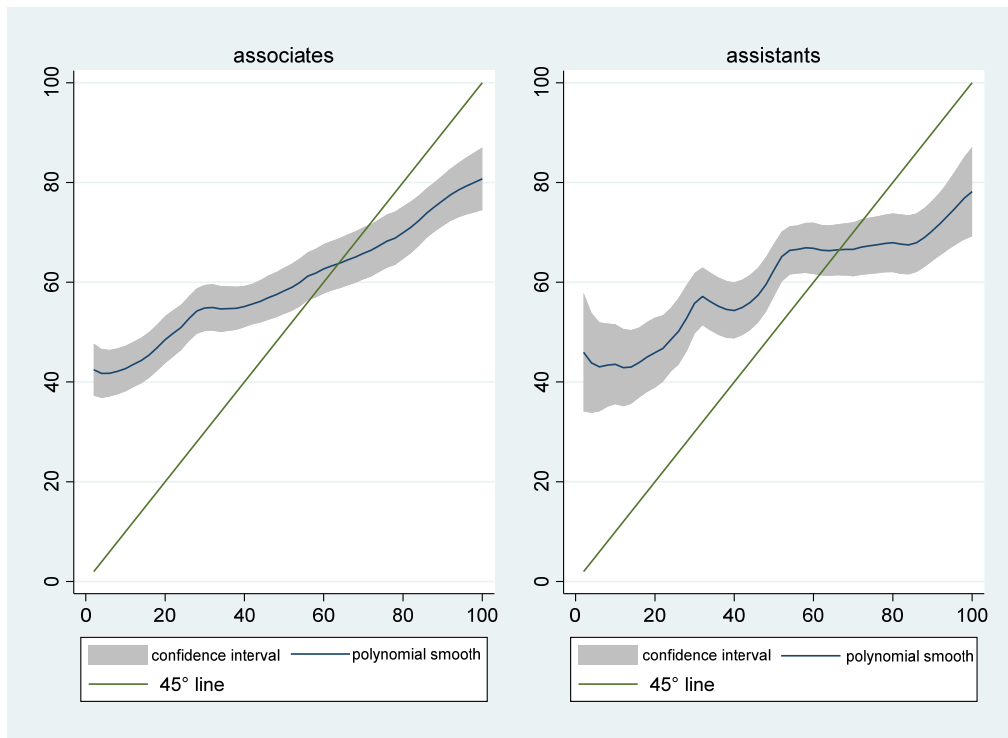


Figure 7. Density of the productivity score \bar{P} (percentiles). Applicants and non-applicants.

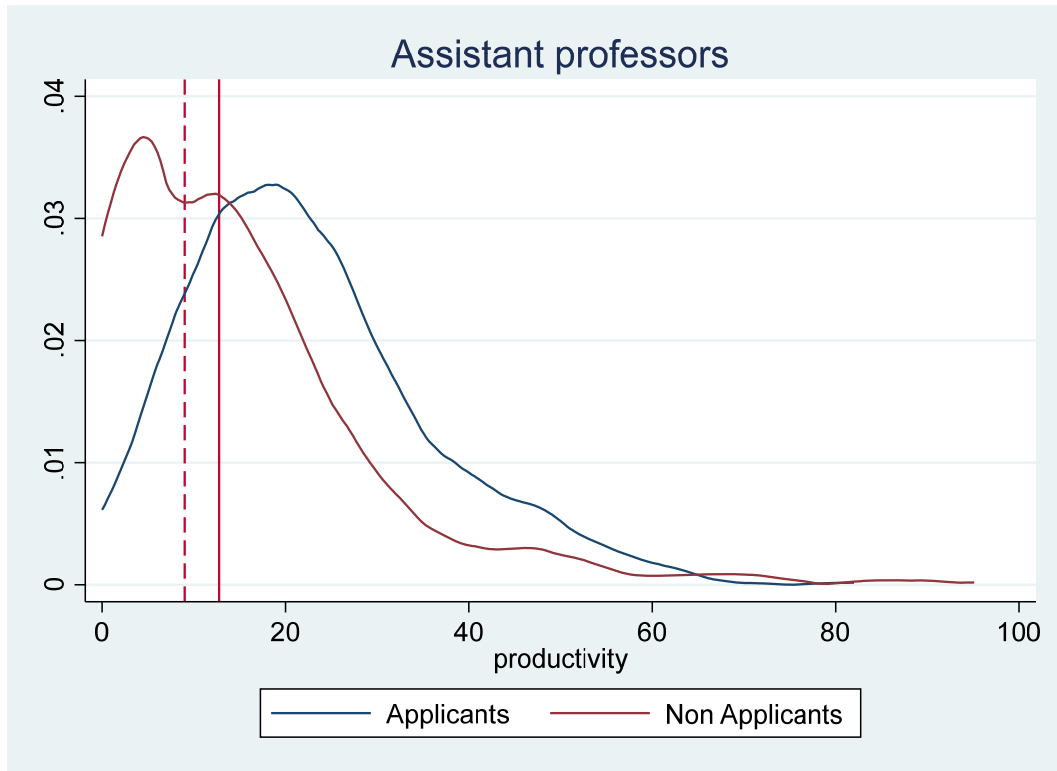


Figure 8. Density of the productivity score \bar{P} (percentiles). Applicants and non-applicants.

