The Design And The Impact Of Project Funding In Science: Lessons From The ANR Experience*

Nicolas Carayol[a,b,c] and Marianne Lanoë[a,b]

[a] *GREThA, UMR CNRS 5113, France*

[b] *Université de Bordeaux, France*

[c] *Observatoire des Sciences et Techniques, HCERES, France*

April 5, 2019

*Nicolas Carayol is Professor of Economics at the Université de Bordeaux, Research Fellow of the CNRS laboratory GREThA. Email: nicolas.carayol@u-bordeaux.fr. Marianne Lanoë is a PhD student at the Université de Bordeaux and Research Fellow at GREThA (Email: marianne.lanoe@u-bordeaux.fr). We would like to thank G. Imbens, P. David, P. Stephan, R. Veugelers, A. Geuna and the participants to conferences in Brussels, Bordeaux, Barcelona, Paris, Torino, Leiden and Atlanta for their remarks and advices. We acknowledge the support of the teams at OST-HCERES and GREThA and, in particular, C. Mouradian, E. de Turkheim and G. Filliatreau. The project received very limited financial support of the ANR, to avoid conflict of interest. Special thanks go to Philippe Freyssinet in particular, as well as Michael Matlosz and Anne Portier from the ANR, for their advice, fruitful exchanges, and their complete respect of the independence of the authors, who are the only ones responsible for any remaining errors.

Abstract

Competitive allocation of funds to research proposals is a mechanism widely used by government agencies to sustain projects of researchers in universities and other research institutions. However, little is known about how efficient this mechanism is precisely how it affects the recipients' behaviors and how it would be possible to improve the precise design of such funding allocation mechanisms. Relying on empirical evidence stemming from the creation of a French generalist and nationwide research funding agency in 2005, we document the impact of this policy on the grantees. Using data on more than fifty thousand applications over five years, thirty thousand tenured professors and researchers out of which ten thousand applied to the policy, we estimate a 15% impact of funding on citations. We further show that grants awarded by non-directed programs have much larger scientific impact. Directed programs targeting emerging fields do attract and fund scientists who produce more novel research outcomes, but they have no causal effect on the research novelty of the granted.

Keywords: project-based funding, competitive grants, scientific productivity, conditional difference-in-differences, triple difference.

JEL codes: D04, O3, C31.

1 Introduction

Governments financially support research carried out in universities and research organizations via various mechanisms. The competitive funding of research projects has been particularly developped in the US since World War II via federal agencies such as the NIH or the NSF.[1] Presumed advantages of that funding mechanism has led countries and institutions around the globe to develop similar policies. Despite the huge amount of public money at stake worldwide and although the way money reaches research presumably affects efficiency, there is still little large-scale systematic evidence about the impact of such fund allocation schemes. Further, competitive allocation of funds to research proposals actually reveals a significant variability whereas the precise rules and goals of the programs are also likely to affect the outcomes. It is thus important to understand how and why the returns may vary with respect to the specific designs of the funding programs.

This article provides new clues on these issues, relying on the recent French experience. France is the fifth largest scientific nation worldwide in terms of citations. In 2005, the French government created a dedicated agency, the *Agence Nationale de la Recherche* (ANR) to implement project-based research funding in the country. Our study focuses on the first five years of the ANR's existence (2005-2009) since sufficient post-funding time has now elapsed for some of the first consequences of this policy to be observed. Over this period, this institution received about seventy thousand applications and allocated nearly two and a half billion euros to research projects, the total cost of which amounts to approximately ten billion euros. The ANR was set up as a nationwide generalist player welcoming applications from all disciplines. As alternative sources of funding for professors and researchers' projects were rather limited over that period, this experience offers an excellent opportunity for appreciating the impact of fund allocation on a large scale. It further offers interesting forms of variation as regards funding programs, recipients' characteristics, disciplines, etc. that can allow us to appreciate the differential impact of research project funding for precise differences in their design.

The ANR runs two distinct types of funding programs: directed and non-directed programs. Non-directed programs are more neutral as they welcome applications from all fields of science which are examined by single discipline panels. Directed programs target emerging and promising research areas and/or fields that are suspected to have large potential for future applications. Their calls for proposals are designed by panels mixing top-level representatives of large research institutes and R&D performing corporations with well established scientists and the selection of awardees among applicants is made by ad-hoc interdisciplinary panels. The underlying rationale of directed programs is that the traditional academic incentives for

[1] According to the National Science Board (2016), yearly extramural federal funding of US universities and colleges has exceeded fourty billion dollars since 2010. See Stephan, 2012 for a detailed overview.

investigating new or interdisciplinary research areas are not strong enough. It is often argued that risk taking, novelty and interdisciplinarity are under-rewarded because the peer review system would be mainly organised within disciplines and negatively biased toward truely transformative ideas (Braben, 2004; Chubin and Hackett, 1990; Wesseley, 1998). Professors and researchers who respond to incentives (Dasgupta and David, 1994; Stephan, 2012) and who are autonomous in the choice of their research agendas (Carayol and Dalle, 2007; Aghion et al. 2008) may overly refrain from addressing such problems. However nothing garanties non-neutral funding schemes are more efficient. Applicants to directed programs who are already investigating the targetted fields essentially face more limited competition whereas others may make socially inefficient efforts to comply to the specifics of the calls. Further, directed programs are much more complicated to set up and to be efficiently run. Last but not least, the targetted fields may not have larger potential than totally unanticipated new avenues proposed by applicants. As most research funding programs balance between directed and non-directed rationales, we aim to compare the impact of the two types of programs to appreciate which one of the more neutral (non-directed) or the more interventionist (directed) programs is more efficient.

The main methodological issue of estimating the impact of funding on observational data is to disentangle the selection and the funding effect. Indeed, why university professors and researchers apply to the funding agencies, and why evaluators and committees select them, are also often the same reasons why they are likely to be more productive. Confounding factors are thus likely to affect both funding and the scientific outcomes, which would skew estimates in a naive approach. Jacob and Lefgren (2011) use the grades produced in the evaluation process of NIH grants to account for the selection effect in an IV approach. Those grades given by referees and panel members are intended to capture the variation in projects quality which is uncorrelated with the observables.[2] By doing so, the authors aim to measure the impact of funding, holding constant project quality. Though project quality is brought in by applicants, not by the funding agency, this approach captures the average treatment effet only if unfunded applicants can run projects of similar quality than the ones they submitted. Jacob and Lefgren (2011) indeed argue US biomedical sciences are characterized by a variety of available sources of funding. However, unsuccessful applicants often can not undertake the submitted (and evaluated) projects when alternate funding sources are not available. In those circumstances the extra outcome due to the quality differential between the submitted project and the project undertaken when not selected, is obtained thanks to the funding. As such it should be accounted for in the impact analysis.

[2]Li and Agha (2015) find that the grades significantly explain the scientific performance of the recipients of NIH funds, even controlling for the observables (in particular previous performance). Fang et al. (2016) however have reversed conclusions on the same dataset when excluding the projects which got the lowest rates.

As our evidence is characterized by a limited availability of alternative funding sources for projects, the present study thus adopts the conditional difference-in-differences model developed by Abadie (2005) that allows us to control for both the time-invariant individual fixed effects and the selection on observables.[3] In this model, the identification of the impact of the policy relies significantly on the quality of the observables prior to treatment, on which the fund allocation process is modeled. Fortunately, we were able to assemble detailed information that cove scientists' age, their institutions, fine grained research fields, and multidimensional publication profiles, which we can use to model fund allocation. Moreover, information is available for almost the whole of the reference population (not just for applicants) as we match the list of applicants with the list of all professors and researchers associated with a laboratory accredited by the Ministry of Research and Higher Education in France. That represents more than thirty thousand tenured scientists while we restrict ourselves to those fields which are sufficiently covered by the publication database we use (Web of Science).[4]

We can, therefore, estimate the impact of receiving an ANR grant using control groups picked either among unsuccessful applicants to the same program and year, or from the whole reference population. There are good reasons to select controls in each way. On the one hand, all applicants self-select and are thus more "similar". On the other hand, picking individuals in the much larger reference population increases the chance of finding controls that are more similar to the treated in terms of the observables (especially as regards publication profiles and detailed scientific fields). In fact, we consider many ways of defining the selection phase which have advantages and drawbacks. Differences lie either in the chosen list of explaining variables (inclusion of individual, laboratory, or trend variables) or in the exclusion rules (picking controls exclusively in the same program, year, or field, or not). We do not postulate that one design of control groups is preferable to the others but test a number of specifications on a placebo parallel path tests before treatment. The best design of the selection stage only considers applicants as potential controls. The selection turns out to be completely unrelated to pre-treatement trends. Balance diagnosis tests show that those properly weighted controls have very similar observables than the treated when they have similar propensity scores. Productivity divergence between controls and treated only starts two years after funding (most project have a three-year period). These remarks converge convincing us that the chosen controls differentiate from the treated for some reasons that are unrelated to their expected scientific productivity, in the absence of treatment. Those controls and scores are thus used to calculate our reference estimations of the impact of

[3] The previous literature using the IV (or selection approach) also include Carter et al. (1987), Arora et al. (2000), Arora and Gambardella (2005), Benavente et al. (2012), and Gush et al. (2015). The literature using propensity scores include Chudnovsky et al. (2008) and Azoulay et al (2011).

[4] Mainly hard and bio-medical sciences but not exclusively as it also includes some social sciences.

funding, but it turns out that using other variants does not qualitatively affect the results obtained.

Overall, we estimate a 15% impact of ANR funding on citations. This is significantly larger than what Jacob and Lefgren (2011) found for the impact of NIH grants (7% impact on citations) while the average ANR grant is .14 million euros for a total cost of .56 million euros, to be compared with the average NIH grant which is equal to 1.7 million dollars. This difference is likely to be explained by a lower displacement effect due to fewer alternative sources of funding. We attribute to the policy the benefit of recipients working on better quality projects. This is consistent in a context in which alternative sources of funding are scarce. As this value also rests upon funding agencies recruiting good panel members and managing their work efficiently, it is also an outcome of the funding programs that we would like to account for as a component of the policy.

To our knowledge, Azoulay et al (2011) is the only study comparing the impact of different research funding programs. Their focus is different from ours as they are interested in identifying the differential impact of funding persons vs. projects. They compare the impact of a funding program (HHMI grants) which targets young and promising scholars in the medical fields, with the one of NIH early stage career prizes. They do not use information on the applicants to both funding programs arguing that the recipients of NIH early stage career prizes may, in principle, have applied to the HHMI (same age, country and field). Our data are more complete as we do have information on both the awarded and the unsuccessful applicants for the two programs we compare (directed and un-directed). However, the applicants to the two programs are not necessarily the same because directed programs raise barriers to entry. We thus adopt a different estimation strategy than Azoulay et al (2011) to compare the two types of programs. We use a conditional triple difference approach which literaly compares the impacts of the two program which are themselves estimated as double differences.

We find that the impact of directed programs is rather small (about 6% on citations), while the surplus of impact gained by switching to a non-directed program equals 20%. Non directed programs are significantly more efficient. These programs seem to be able to attract and to pick high quality projects. Though we can not exclude that the directed programs may have delayed impact that we can not fully observe, there is no reason to believe that they achieve their specific goals, with the exception that they do attract and fund professors and researchers who write more novel research articles. The same approach is used to compare other dimensions of programs design, such as the age of the applicants, and find that the impact on younger recipients is significantly larger.

The remainder of the article is organized as follows. The data are presented in the second section. Methodology comes next. In the fourth section, we present the selection of controls and the calculation of propensity scores. The fifth section presents our results on the quantification of the average impact of funding. The results concerning the design of funding programs come in the sixth and seventh sections. The last section wraps up and discusses the main results.[5]

2 The Data

Data collection

Data collection starts with a list of all researchers and professors associated with one laboratory accredited by the French Ministry of Higher Education and Research around the year 2010, which contains information on 49,225 persons.[6] All of these persons are tenured, whenever as full or assistant professors, assistant researchers or research directors. Once all individuals for which we do not have full and consistent information (status, institutional employer, laboratory, age, etc.) have been excluded, we are left with 48,328 persons. This list has then been matched to the names of the authors of scientific articles, letters and reviews (on the basis of their surname and first name initials) in the Thomson-Reuters ISI Web of Science, a well-known database which gathers all the documents published in the main scientific journals. The publication period covered in this study goes up to and including year 2012. Thus, the last publication year considered (2012) stands three years after the last funding year (2009) and seven years after the first funding year (2005). We collected more than nine million distinct authorships (listed author$\times$document) which received more than fourty million citations. As these large publication records show, we are faced with a huge homonymy problem due to the absence of any reliably unique identifier of researchers in publication databases. A disambiguation algorithm has thus been developed based on a "seed + expand" methodology (Reijnhoudt et al, 2014). Basically, this algorithm works as follows: in a first step (seed), the algorithm validates articles by imposing strong conditions, particularly on the field and institutions, which should be consistent with what we know for each person. At this stage, the goal is to minimize false positives. In the second step (expand), the algorithm uses the information on the articles already validated in the seed step, to accept other articles which did not fully meet the conditions of the seed step. Typically the information used concerns the co-authors, the references and the keywords. New papers are validated either because, on the top of some of the first-step conditions which are

[5] A Supplementary Material is available online. It provides, in seven appendices, numerous details on the raw data, variables, methodology, tests, results and disambiguation of the publication data.

[6] Are thus excluded all the tenured researchers and professors who are not associated to a lab, and those associated to laboratories in schools funded only by other ministries (such as the ministries of industry, agriculture or defense), or to laboratories solely associated to national research institutes (such as CNRS or CEA internal labs).

maintained, they have the same co-authors or cite the same references as already validated articles. The program then iterates up to some point. In order to evaluate the quality of this disambiguation process, we have constituted a benchmark of nearly 300 scientists who have created an ORCID number and are thus likely to have disambiguated their own publications. Detailed information on the algorithm and on the quality of the disambiguation are presented in the Supplementary Material, Appendix G. By the end of the disambiguation process, 1.2 million author×documents have been validated (733 thousand distinct articles), approximately 13% of the initial set.

The affiliation of professors and researchers to scientific fields of investigation is based on a fine grained organization of science in France into peer groups called "sections". Such sections are specific to the institutional employer, either a national research institute (such as CNRS or INSERM) or the Ministry of Higher Education and Research for all professors employed in universities and schools. Each section members elect a national committee which usually accredits PhDs for recruitment (or sometimes even recruits directly), evaluates individuals, allocates promotions, etc. Most of the time, sections tend to be organized around specific disciplinary orientations.[7] We computed, for each section, the percentage of professors and researchers for whom no article was found in the database. On the basis of this information, we excluded a long list of sections, mostly in the fields of humanities and social sciences. We suspect that these disciplines are not well covered by the database, either because scientific journal articles are not the main outcomes of their research, or because the principal journals of these disciplines are not well covered by the database. This leaves us with 31,081 persons.

The ANR provided us the list of all applications from 2005 to 2009, comprising 67,812 partners × applications. A project "partner" is defined as an institution which will directly receive the planned funds from the ANR if the application is successful. Each partner has its own scientific coordinator. Multipartner projects have only one project coordinating partner, whose scientific coordinator is the project PI. In multipartner projects, each partner receives its funds directly from the ANR. Each partner coordinator is fully responsible for the engagement of the funds received by his/her institution and thus enjoys significant autonomy. Keeping only the partners × applications emanating from academia and for which the variables of interest are correctly documented (scientific coordinator's surname and first name, the partner, funding decision, amount, and duration), leaves 54,852 partners × applications. The success rate is 30%. The total amount allocated is 2.4 billion euros, but the expected total cost of the funded projects is 9.5 billion euros because the ANR funds only the marginal cost of the projects it supports for public partners.[8] The median fund per partner is 136,000

[7]For a few specialized research institutes, the specialty of the sections is not straightforward, and we had to develop specific strategies. For instance, for INRA (the national research body dedicated to agricultural research), the allocation to disciplines has been performed on an individual basis.

[8]The grants cover the wages of the non-tenured personel hired for the purpose of the project and overheads

euros, while the mean is 138,000 euros. The mean total cost per partner is 545,000 euros.

We next basically matched the list of scientific coordinators of all ANR applications with the personnel list obtained previously. Two types of matching were performed subsequently: an exact matching and a fuzzy one.[9] In the event of homonymy in the full initial list of scientists, a manual check was made, based on the consistency between the discipline of the scientist and the project description, and between the employer of the scientist and the project partner. This matching allowed us to find, in the list of the 31,081 professors and researchers, the scientific coordinators of 46.2% of all applications, 45.5% of the funds, and 46.9% of the total amount of money allocated.

It turns out that more than one third (10,722) of all these persons applied as scientific coordinators of the partners involved in the projects submitted between 2005 and 2009, and that 18.6% (5,831 persons) obtained at least one grant (4,892 applicants were never funded). Therefore, two third (20,498 persons) did not apply. The age distribution of the three populations (reference population, applicants and funded) is similar, though the 35-50 year-olds (in year 2010) are proportionately more numerous among the applicants and the funded. Researchers and full professors are more likely to have applied at least once. Researchers from CNRS and INSERM apply more often and their applications are more likely to be successful. The applicants identified have applied on average 2.4 times over the period (25,364 applications). The distribution of applications is asymmetric, with most professors and researchers not applying or applying only once, while some apply many times. On average, the applicants obtained 1.2 grants over the five years considered (12,757 funds allocated). Like the applications, the funds are also unevenly allocated across the population: More than 75% of the applicants received only one funding, while a few got many. In this study, we will consider only the first funding for those who got multiple grants. There are two types of programs: directed programs that have a specific directed orientation, and non-directed ones which are fully open to any application. While half of the applications go to directed programs and the other half to non-directed programs, directed programs account for 65% of the grants allocated, because these programs have significantly higher rates of success.

When we break down applications by discipline, we observe that the highest rate of application is found for physics (with more than one application per scientist), followed by fundamental biology (.94), chemistry (.91) and applied biology and ecology (.90). The lowest

limited to 4% of the grant. The total costs typically include the grant and all the resources included in the project, in particular the salaries of the tenured researchers and professors paid by the research institutes and universities.

[9] Fuzzy matching authorizes small variations in the surnames and first names and then requires manual verification and cleaning, basically comparing individual information and project information before validation.

rate is found for mathematicians who applied only one-third time on average. The highest average rate of funding can be observed for applied biology and ecology (with .34 funds per scientist). Physics follows immediately (with .32 funds per capita). These two fields differ strongly, however, in terms of supporting programs: physics is most often funded by non-directed programs, whereas nearly two-thirds of the funds allocated to applied biology and ecology come through directed programs. Similarly, fundamental biology, medicine and engineering sciences are mostly funded by directed programs, while the sciences of the universe and mathematics are most often funded by non-directed programs.

Outcome variables

ANR funding is intended to sustain the scientific production and excellence of the awarded. Though this can not be observed directly, different measurements of scientific outcomes can assess its most important dimensions.[10] We build three variables that are labelled respectively Volume, Citations and Impact Factor. Though they are not independent, these indicators are distinct and proxy different dimensions of scientific production and excellence. Volume sums the number of articles published, each being adjusted by the number of co-authors (fractional counts). It relates more to the volume/quantity of scientific production. Impact Factor weights each article by the average number of citations which papers published in the journal that year received on average (again in a three-year window).[11] That variable captures the capacity to publish in well-established journals. Citations weights each article by the number of citations it received (in a three-year time window). As such this indicator captures the impact of each article on the scientific literature and thus corresponds more directly to scientific excellence. All three indicators may be significantly affected by field differences for a number of reasons, but, as these differences are time invariant, they are controled for in the difference-in-differences design.

Publication data also prove to be very helpful in investigating the collaboration behaviors of professors and researchers (Wuchty et al. 2007). We use the number of authors of the article to evidence the size of the research teams, information which is averaged for each given period and person to obtain variable Average Team Size. Collecting all collaborators' names and initials over given time periods and dropping double counts, we also compute the total number of distinct coauthors, labeled Coauthors. This number proxies the size of the collaboration network. We also compare the sets of collaborators between two consecutive

[10] Details on the calculation of all outcome variables are presented in Appendix B of the online Supplementary Material.

[11] This weighting scheme is very close to, but distinct from, the traditional Journal Impact Factor which divides the number of citations received in a given year (thus to articles published that year but also to those published previously) by the number of articles published that year. Therefore our approch is less sensitive to the yearly variations in the average quality or in the number of articles published.

time periods to assess the number of new coauthors they are working with, labeled New Coauthors. The addresses of the authors' institutions can be used to assess the capacity of professors and researchers to extend their collaboration networks at the international level. The variable International Collaborations equals the number of articles that have at least one foreign address.

Descriptive statistics

Some descriptive statistics of the outcome variables measured on the whole data set are presented in Table 1. Mean outcomes are presented separately for the three years before and three years after the year of reference and for the three defined groups, the non-applicants, the not granted applicants and the granted ones. The reference year is the year of first successfull application for the funded. There are exactly 5,831 distinct persons in our dataset that have been funded by the ANR. We do not consider their subsequent successful applications, nor their unsuccessful ones. There are 7,433 distinct applications of the 4,892 persons who applied at least once, but were never funded. We consider all the applications of the never funded applicants, and for each of those applications, the before and after periods are defined according to the application year. As there is no specific reference year for the 20,498 non-applicants, they are considered for each of the five years of the study, and the before and after periods are defined accordingly. Incomplete information about the identity of coauthors before year 2002 prevents us to compute the New Coauthors variable in the period preceding the year of reference.

The means measured in the three-year period after the year of reference are always higher than those in the previous period, but differences in magnitude are observed according to groups. The subset of non-applicants tends to publish more articles after the year of reference (1.64 against 1.42 articles in the previous period, in fractional counts), of a higher cumulated journal impact factor (2.98 against a mean impact factor of 2.43 before) and which receive more citations (6.19 against 5.74 citations before). This positive evolution is also observed for not granted applicants (2.4 against 2.12 articles in the past period), who publish more articles than the non-applicants but less than the granted ones (3.16 against 2.8 articles in the past period). Their publications also received more citations (10.03 against 9.47 citations before) and are associated to higher journal impact factors (4.84 against 3.97 previously), but again in a lesser extent than granted applicants (16.19 against 14.45 citations in the past period and a mean impact factor of 7.14 against 5.77 before).

The collaboration profiles also constrast between groups. The non-applicants are characterized by a larger average team size than the other groups in the past period (10.65 coauthors compared with 7.34 coauthors for the not granted applicants and 7.55 coauthors for the granted applicants). This difference is sharpened in the subsequent period, with an average team size of 20 coauthors for the non-applicants compared to only around 11 coauthors for the other groups. Granted applicants seem however to collaborate more often with different authors, as indicated by the annual number of coauthors. In the subsequent period, they collaborate in average with 81 different coauthors (40.10 individuals in the last period) compared to 65.77 and 57.47 coauthors for the not granted applicants and the non-applicants respectively (34.17 and 31.41 coauthors respectively in the last period). Those numbers of co-authors may look large. They are however consistent with the average team sizes (note the average number of co-authors per paper are always above ten in the "after" period). Moreover, the averages are driven by outliers in those disciplines characterized by very large author teams (the median numbers of coauthors are significantly lower than the means). Granted applicants tend to collaborate more with researchers from abroad (8.22 times in average in the subsequent period against 6.64 in the last period), compared to not granted applicants (6.62 times against 5.34 in the last period) and non-applicants (5.93 times against 4.71 in the last period). In the subsequent period, granted applicants also collaborate more often with partner they never worked with before, with an average of 18.20 new coauthors, compared to 16.05 new coauthors for not granted applicants and only 12.45 new coauthors for the non-applicants. Finally, we observe only little differences between groups in the propensity to address new problems.

3 Identifying the impact of funds: Methodology

Controlling selection on observables

In this paper, as we focus on the effect of receiving an ANR award on successful applicants, we are interested in the so-called average treatment effect on the treated individuals, which is defined as follows:

$$ATT = E\left(Y(1) - Y(0) | T = 1\right), \tag{1}$$

where $Y(1)$ denotes the production when the applicant is funded, while $Y(0)$ refers to the counterfactual, i.e. the production if the applicant had not been funded. The event noted $T = 1$ means treatment occurs. The problem is that the counterfactual outcome is non-observable: either he/she is funded, or is not, but not both.

Propensity scores can help reduce the bias related to the selection on observable characteristics. Rosenbaum and Rubin (1983) show that, under the ignorability condition which states that adjusting for a set of covariates X is sufficient to remove all confounding factors, controlling for the propensity scores is sufficient. The propensity score $P(X)$ is defined as the probability of being "treated" (obtaining a grant in our case) given X: $P(X) = P(T = 1|X)$, with $0 < P(T=1|X) < 1$. The propensity scores are reliably estimated when the conditional independence assumption (CIA) is verified.[12] It states that the potential outcome is independent of the treatment status, conditional on the propensity score. In other words, the treated individuals would have reached the same outcome levels as the controls having the same propensity score, if they had not been assigned to the treatment:

$$E\left(Y(0)|T=1,p(X)\right) = E\left(Y(0)|T=0,p(X)\right) = E\left(Y(0)\right). \tag{2}$$

This equation can be rewritten as:

$$Y(0) \perp T|p(X). \tag{3}$$

Such an assumption, which cannot be tested directly, implies that there is no confounder influencing both the assignment of the treatment and the outcome that is not included in X. Heckman, Ichimura and Todd (1997) show that the non-inclusion of a relevant covariate causes the introduction of a bias in the estimated impact. In other words, the CIA assumption is valid only if all the covariates which influence both the treatment and the outcome variables are included in the set of explanatory covariates used for the estimation of the propensity scores.

Therefore, the covariates that are included in the vector X, which is used for estimating the propensity scores, need to be selected with caution. In this study, we use an "agnostic" approach whereby we investigate several specifications of the selection model that we test later.

Matching and weighting

Different methods using the estimated propensity scores can be applied to remove the bias due to the differences between the observed characteristics of the treated and those of the untreated individuals. In this paper, we consider two matching procedures, nearest neighbors matching with replacement, and kernel matching, as well as inverse probability of treatment weighting (IPTW).

In the nearest neighbors matching, each treated individual is assigned its most similar controls (up to five) in terms of propensity score. To improve the quality of the matching, a caliper width is specified, which restricts the selection of the controls within a caliper around the propensity score of the treated individual (to avoid capturing controls that are too distant). The caliper value is calculated in line with Cochran and Rubin (1973), who tested the bias reduction when applying a caliper width $c = a\sqrt{(\sigma_1^2+\sigma_2^2)/2}$, along with σ_1 and σ_2, which are the standard deviations of the propensity scores among the treated individuals and

[12] It is also known as Weak Unconfoundedness for the ATT.

the controls respectively, as well as with a as a positive parameter. Following Rosenbaum and Rubin (1985), we set $a = 0.2$ which is expected to remove around 99% of the bias. Using a caliper condition however reduces the subset of available controls. Note that treated individuals will be excluded from the analysis if no control meets the imposed conditions (the caliper or the common support restriction).

Unlike the nearest neighbor approach which assigns the same weight to all controls of a given treated individual, the kernel matching approach assigns a different weight to each control, which is inversely proportional to the difference between its propensity score and that of the treated individuals. The kernel method provides an interesting solution when the nearest controls have very different propensity scores to those of the treated individuals. Frolich (2004) argues that kernel matching is always preferable to nearest neighbors matching. We exclude observations with extreme propensity score values. Following Imbens and Wooldridge (2008), we remove all individuals i, such that $p(x_i) > .9$ or $p(x_i) < .1$. We also apply the common support restriction, which implies that we do not consider controls with a lower propensity score than the lowest score among the treated individuals (Dehejia, 1999).

Robins, Hernan and Brumback (2000) and Hirano and Imbens (2001) argue that the controls with higher probabilities of being treated are likely to be under-represented in the control population (because they are likely to have been treated), whereas the controls with lower propensity scores are likely to be over-represented. To correct for this bias, the authors suggest weighting the controls by the inverse of the probability of being treated. The weights allow under-represented controls (because they are likely to have been treated) to have a more important role in the analysis as compared to the controls who have a low probability of receiving the treatment (who are thus likely to be over-represented). Hirano, Imbens and Ridder (2003) argue that this approach is more efficient.

Conditional difference-in-differences

So far we have considered that observed heterogeneity was sufficient for explaining the selection into treatment. However, in the applicants' CVs or in their project proposals, the selection committees and the external solicited referees can find relevant information that cannot be observed in our data, but which reveals their ability to perform in science. If this occurs, and if it influences the selection, then propensity scores are not sufficient for identification. However, if these unobserved variables are time-invariant, such as personal fixed-effects, then time differentiation can be used to solve the problem. The relevant approach is the so-called difference-in-differences methodology, which basically compares the variation in the performances of the treated individuals and the controls, before and after treatment. The outcomes variables are calculated by pooling together the information on the three years before and the three years after the year of funding.[13] Therefore, the publication outcomes issued in the year of funding are not considered. The three-year window ensures that the post-funding publication period considered is complete, even for the last funding year considered (2009) because publication data are available until 2012.

In the context of our application, we conceptualize scientific outcome as given by:

$$Y_{i,t} = T_{i,t} \times \delta + \eta_t + \theta_i + \mu_{i,t}, \tag{4}$$

where i refers to the professors or researchers, $t = 1, 2$ denotes the time period (pre vs. post treatment), $T_{i,t}$ is the treatment dummy, η_t is a time dummy equal to one in the post treatment period $t = 2$, θ_i is a fixed individual effect, and $\mu_{i,t}$ is the error term. The term δ is the impact of funding that we aim to estimate. Assuming we remove confoundedness ($cov\,(T_{i,t}, \mu_{i,t}) = 0$), δ can be properly estimated in a difference-in-differences approach.

To do so, we use the conditional difference-in-differences model (Abadie, 2005) that combines a treatment selection model based on the estimation of the propensity scores with the difference-in-differences method. The estimation of the impact can be calculated as follows:

$$\hat{\delta} = \frac{1}{|N_T|} \sum_{i \in N_T} \omega_i \left(Y_{i,1} - Y_{i,0}\right) - \frac{1}{|N_C|} \sum_{j \in N_{NT}} \omega_j \left(Y_{j,1} - Y_{j,0}\right), \tag{5}$$

where N_T denotes the set of treated individuals and N_C the set of controls. $Y_{i,t}$ is the outcome variable observed in period t, with $t = 1$ in the period after the treatment assignment, and $t = 0$ in the period before treatment. The weights ω_j are defined according to the chosen matching or weighting method. When the nearest neighbors or the kernel methods are chosen, the treated individuals have a unitary weight ($\omega_i = 1$) when included and the controls have a total weight which is accumulated over the treated individuals to which they are associated: $\omega_j = \sum_{i \in N_T} \dfrac{1}{|M(i)|} \omega_{j,i}$, with $\omega_{j,i}$ the weight of control j vis-à-vis treated agent i, and with $M(i)$ the set of controls for treated agent i. With the IPTW approach, the weights are calculated following a slightly different logic as controls are no longer specifically associated with given treated individuals. They are calculated as follows: $\omega_i = T_i + \frac{(1-T_i)p(x_i)}{1-p(x_i)}, \forall i \in N_T \cup N_C$, with $T_i = 1_{\{i \in N_T\}}$ the treatment dummy and $p(x_i)$ the propensity score of agent i (cf. Robins, Hernan and Brumback, 2000; Hirano and Imbens, 2001).

4 Selection on observables

In this section we first provide descripive statistics on the variable used to model the selection stage, before presenting the models and the tests.

[13]In principle, this analysis could be done on a yearly basis. However, we follow Bertrand et al. (2004) who show that using only two periods is preferable because it reduces serial auto-correlation.

Descriptive statistics

Some descriptive statistics of the selection variables are presented in Table 2. The three population sets (non-applicants, not-granted applicants and granted applicants) are exposed separately. In Table 2, we also distinguish, among the not granted applicants, the ones who applied to directed vs. to non-directed programs. The same is done for the granted. This distinction makes no sense for the non-applicants. Out of the 5,831 professors and researchers who have been funded by the ANR, 3,385 got their first application thanks to a directed program, and 2,446 thanks to the non-directed program. There are a few cases for which a first grant from the directed and from the non-directed programs occur simulaneousy. When this happens, one is randomly selected while others are excluded from the sample. We do not consider the subsequent successful applications of the granted, nor their unsuccessful ones. The never successfull applicants are considered for each of their applications, 4,085 to directed programs, and 5,567 to non-directed ones. The non-applicants are considered five times, that is, once for each of the funding years. As all presented statistics are time-variant, they are calculated for each considered year.

As age is likely to explain both the probability to apply and the probability of being granted, we consider age at the time of application (Age). The different subpopulations however do not differ significantly in terms of their average or median age, but the applicants (granted or not) to directed programs who are two-to-three years older in average. The number of articles (fractionned to account for co-authorship) published in the previous three years is intended to capture recent research intensity (Articles). It is significantly larger for the applicants than for the non-applicants. Among applicants, it is 20% larger for the granted than for the not granted. We use the number of citations to those articles (keeping the fractionnal counting) received in a three-year period after publication to account for the scientific impact of recent research (Citations). Similar differences are found between the three groups (the granted perform better than not granted applicants, who perform better than non-applicants). However, we now also find that applicants (both granted or not-granted) to non-directed programs have larger citations records than their counterparts in the directed programs. The total number of citations received over their career, recorded since 1999, accounts for long-run reputation (Total Citations). It may affect both self-selection (applying or not) and the odds of passing the formal selection process. We observe similar differences as for Citations. Note that if non-applicants have significantly lower scores in average, this is mainly explained by a large proportion of low performing individuals in this population as the median equals 9, which is only 11% to 16% of the median in the different groups of applicants. The largest Impact Factor of the journals which published their three previous years papers accounts for the capacity to publish in well established journals (Max Impact Factor). We observe a neat difference between the non-applicants and the applicants on this variable (three times greater), but differences between sorts of applicants are rather limited.

A series of variables are employed to capture the characteristics of the research environment. The average per capita number of articles in the laboratory accounts for the intensity of research activity in the close professionnal environment (Av. Lab Articles). To account for the presence of one particularly reputed member of the lab, which could affect the probability to apply and to be selected, we use the maximum number of citations reached among lab members (Max Lab Citations). We find that applicants have more intense research environments than non-applicants and are more likely have a star in their lab. Larger labs often have larger supporting staff which may affect the probability to apply and the quality of the project proposal. We thus use the number of faculty members to capture the size of the lab (Lab Size). There are however limited differences between the various sorts of applicants in these respects. The average laboratory size is pretty similar between groups of applicants and with the non-applicants. It is however larger for the applicants (treated or not) to non-directed programs.

Selection models and tests

Eight different designs have been retained and tested for selecting potential controls and calculating propensity scores. Table 3 synthezises the different models.[14] All logit estimations regress the treatment dummy on individual variables. The latter include age which may affect the odds of being granted as well as the various variables discussed above on individual past publication profiles (number of publications in the last three years, number of citations to articles published over the same period, highest impact factor, and total number of citations received over their career so far). The designs for calculating propensity scores differ, however, in several respects. Publication trend variables in the years preceding treatment are included in some logit regressions so as to capture the recent dynamics of scientific production before treatment. Some designs exclude all non-applicants, while others select controls from within the reference population as a whole. Some, however, require the controls to be in the same section as the treated individuals,[15] while others do not. Let's recall that the section allows to control for both the detailed scientific field and the employer (a specific national research institute or any university). Since the quality of the research environment is one of the selection criteria of the ANR, we have also considered the inclusion of laboratory variables

[14]For each design, the three weighting methods (five nearest neighbors, kernel and inverse probability of treatment weighting) have been tested. That makes 24 estimations. Details are presented in the online Supplementary Material (Appendix C)

[15]Then propensity scores are computed separately for each section. This implies that controls are exactly in the same field of study than the treated and have broadly the same status (researcher or professor) and the same employer, that is the Ministry of higher education if the treated is professor, or a specific research institute if the treated is a researcher.

among the regressors,[16] such as the ones presented above: the average research intensity of the lab, the size of the lab, and the presence of an outstanding reputation member. Last but not least, in some designs the directed and non-directed programs are considered jointly while in others, logit regressions are performed by program type, basically assuming that the selection mechanisms of directed and non-directed programs are distinct, based on different weights given to the observables, and even on different observables.

The difference-in-differences identification relies on the parallel path hypothesis, that is, the treated individuals would have had production paths parallel to the ones of their controls if they had not been treated. This hypothesis cannot be tested comparing before and after treatment outcomes, since the counterfactual is not available after treatment. However, the parallel path hypothesis can be tested between different periods before treatment (Imbens, 2004; Abaddie, 2005). Therefore, we estimate a hypothetical impact of the treatment on the treated individuals, between two distinct periods before treatment: $t_0 - 3$ and $t_0 - 1$, where t_0 stands for the year of funding. The goal is to verify that the variations in outcomes of the treated individuals before treatment are not significantly different from those of the controls. Performing such a test on all estimations, we find that the predicted impact of the treatment is always very small and never significant. A similar test is performed between $t_0 - 3$ and t_0. This placebo test is more helpful in sorting out candidate estimations, and thus in selecting our reference propensity score estimations. The propensity scores that lead to the most parallel paths are those for which the controls are applicants exclusively; the estimations are performed distinctly between program types (directed vs. non-directed); the laboratory variables are not included; and production trend variables before treatment are included. The variables used in the two models (for directed and for non-directed programs) are presented in Tables 4 and 5.

Following the recent literature we will use the IPTW weighting scheme as our reference estimations but consider the other weighting methods. In fact, the results with the nearest-neighbors and the kernel matching methods remain essentially the same as the ones obtained with the IPTW. The results of the parallel path test for our preferred propensity score specification are presented in Table 6. Appendix D of the online Supplementary Material presents the same tests parallell path tests for for the other specifications.

The conditional independence assumption on which our identification strategy is based implies that the treatment dummy is independent of the variables included in the logit model, conditional on the propensity score. Therefore, if the estimated propensity scores are

[16] A potential issue is that laboratory variables are observed after funding (approximately around year 2010). Therefore, if the consequence of funding is a mobility in a different lab (potentially of a higher quality), the impact of the funding may actually be underestimated (because the treated would be basically compared to controls in higher quality labs). Our results show, however, that the estimations retaining lab variables do not differ significantly and thus that lab variables do not lead to an underestimation.

correct, we expect that controls and treated individuals do not significantly differ with respect to the explanatory variables when they have similar propensity scores. Dehejia and Wahba (1999) suggest a balance test that builds upon this property. It consists in a comparison of weighted means between treated individuals and controls for each variable included in the logit regression.[17] All logit specifications pass this balance test. Due to space constraint, we present that test for our preferred propensity score estimation and weighting scheme in Figure 1. Those balance diagnosis tests (standardized differences of the means tests for each covariate of the logit regression) are also performed within seven distinct strata of the propensity score (separately for the non-directed and directed programs because the preferred propensity score estimations are also performed separately), and pass that test.[18] Further, the treated and controls sets are also well balanced according to the propensity sores, when obervations are weighted according to the IPTW method. Figures 2 and 3 show that matched samples are pretty similar in terms of propensity scores distribution for both directed and non-directed programs.

5 The impact of project funding

Figure 4 shows that ANR funds persons who have an increasing publication trend, which starts before the year of first funding (at $t = 0$) and subsequently expands. Figure 4 also reports the properly weighted performances of their controls (dashed blue line). Note those data are expurged of time trends and yearly shocks. We see that the performances of the controls are slightly lower than those of the treated individuals. This is because the pre-treatment difference-in-differences placebo test has given priority to the similarity in trends with the treated individuals. Other sets of controls which were more similar in outcome levels have been discarded because they do not satisfy all placebo parallel paths tests as well as the preferred ones.

And as expected, non-funded applicants have very similar trends to the funded agents until the year of funding (included). In fact, it turns out that the trends diverge only starting from the second year after funding. It is sometimes claimed that researchers have often nearly completed their project when applying. If these projects were also more likely to be funded, then a positive impact could be partially driven by this phenomenon. However, in that case, divergence should occur early after the funding date, something that is not observed here. This does not mean that anticipated projects are not more likely to be funded but that the conditions we imposed for the selection of controls seem to have sorted out such an effect.

[17] When possible, we use the procedure proposed by Becker et Ichino (2002).

[18] Appendix E of the Online Supplementary Material is dedicated to balancing tests.

Publication outcomes The main conditional difference-in-differences results are shown in Table 7. We find that receiving an ANR fund increases publications by 3.5% according to the preferred estimation. When the impact factor of the scientific journals in which articles are published is taken into consideration, receiving an ANR fund increases production by 8.3%. The impact of funding is strongest when citations are considered: a 15.2% increase is found. Impact is thus much stronger on indicators that capture the quality dimension of scientific output. The research project of the granted seem to attract more attention from the scientific community which more likely cites their work.

As we have seen that age can influence scientific productivity, we are worried that these results may be slightly biased by age differences between treated individuals and controls. Moreover, the literature has long emphasized that age plays a significant role in scientific outcomes,[19] as an inversed-U shape of scientific productivity has been found in most fields of science. However, age differences between funded individuals and controls at the time of application are very limited in each funding program (see Table 2). Unreported regressions that are similar to the main ones but controlling for age and age squared, exhibit no significant change in the results.

These results are larger than those obtained for NIH grants (7% impact on citations) by Jacob and Lefgren (2011), though the mean amount of the funds allocated in our sample is far less than the average NIH N01 grant. Jacob and Lefgren (2011) report a 1.7-million-US-dollar NIH N01 grant on average as compared to an average ANR grant of less than .14 million euros, and an average total cost of .56 million euros. This difference may be due to the specificity of the biomedical sciences in the US for which the availability of funds and the variety of funding sources may induce a displacement effect (as the authors themselves argue). Such an effect occurs if the funded individuals expend less energy in obtaining more funds than the unsuccessful applicants taken as controls. The plausibility of that explanation is reinforced by the fact that alternative sources of project funding than the ANR at the national level were relatively limited at the time of the study.[20] Note that our results are quite similar to those obtained in Gush et al. (2015), who use a different methodology, and data from a different country.

Collaboration patterns The literature has recently documented a long-run increase in the size of research teams proxied by the number of co-authors of the articles (Wuchty et al. 2007). We now document a hypothetical impact of project funding on team size. Coordinators may

[19]To name a few: Lehman (1953), Zuckerman and Merton (1972), McDowell (1982), Levin and Stephan (1991).

[20]At the European level, the ERC was launched in 2007. It had however a limited budget in period 2007-2009: less than 1.7 billion euros for the whole of Europe. We matched the PIs of ERC grants in this period with our list of French professors and researchers, but found only a few scientists in the two lists.

have incentives to delegate research tasks because they experience rising time constraints and because they have more financial resources to staff their teams. We find (see Table 7) a positive but limited impact of funding on the average number of authors per paper (2.2%). However, the impact of ANR funding on the total number of co-authors is significantly larger (9.8%). Thus, project-based funding increases the network of collaborators of the funded individuals more than it does the size of their research teams. This increase seems essentially due to the turnover of coauthors, as treated individuals have 6.7% more new collaborators than controls. This could be due to a higher capacity to hire PhD students or postdocs that eventually become coauthors on specific projects. It could also indicate that the funded individuals become more attractive as coauthors on the academic "collaboration market". To disantangle the two effects we would need to characterize further the collaborators of the treated individuals and controls, which is very difficult because of data limitations. We can however proxy the international span of their individual networks by counting the number of publications for which the authors gave at least one professional address outside France. Funding is found to increase the number of such articles by 4.2%, a result which is positive and significant albeit below the impact of funding on publication volume. This supports the idea that the two effects are at play: the funded individuals increase their networks by hiring, and also by collaborating more with independent colleagues. Moreover, this shows that ANR funding, which is mainly organized on a national basis, does not decrease the internationalization of collaborations but increases it, though to a limited extent.

6 The impact of directed vs. undirected programs

We now exploit variations in the program characteristics to uncover which funding design has larger impact on scientific outcomes. Over the considered period, the ANR ran two main types of programs: the directed and the undirected funding programs. The non directed programs are standard programs, open to any fields of science and managed by disciplinary based panels. The directed programs correspond to specific calls for project proposals in new fields of research for which the agency has diagnosed a specific need or opportunity for its financial support. The proposals are selected by transdisciplinary panels. Because these calls are targeted, only subsets of possible recipients can apply in practice. Therefore, non-directed programs are likely to be characterized by a higher degree of competition. Observed success rates are consistent with this statement: 37% in the directed programs vs. 25% in the non-directed ones. Therefore, self-selection is also likely to be more pronounced in the non-directed programs, and indeed (see Table 2) average applicants and funded via non-directed programs outperform on average applicants and funded to directed programs, when articles are weighted by their citations. Which of the two types of programs should be more efficient, in the sense that it has a larger impact on scientific outcomes? On the one hand, we expect that directed programs may make a big difference on targeted fields. If, as intended by the policy, it encourages the investigation of promissing emerging research areas, it should lead to more path breaking research, leading to more cited papers published in well established journals. On the other hand, non-directed programs may have a larger impact because, thanks to a stronger competition and to their openness toward ideas heading in unspecified direction, they should be able to pick unexpectedly high quality projects.

Conditional triple difference model Our identification strategy builds upon the basic conditional difference-in-differences model by introducing a supplementary level of differentiation. As this basically differentiates double-differences, this estimation is called conditional triple-difference. We here shortly explain the model, before presenting results on the differentiated impacts according to the two types of programs launched, the directed (a non-neutral funding design) vs. the non-directed (a more standard and neutral funding design). For instance, the ATT differential of being treated by the non directed program as compared to being treated by a directed program is given by:

$$\begin{aligned}\hat{\delta}_{N-D} = &\frac{1}{\left|N_T^N\right|}\sum_{i\in N_T^N}\omega_i\left(Y_{i,1}-Y_{i,0}\right)-\frac{1}{\left|N_C^N\right|}\sum_{j\in N_C^N}\omega_j\left(Y_{j,1}-Y_{j,0}\right)\\ &-\left(\frac{1}{\left|N_T^D\right|}\sum_{i\in N_T^D}\omega_i\left(Y_{i,1}-Y_{i,0}\right)-\frac{1}{\left|N_C^D\right|}\sum_{j\in N_C^D}\omega_j\left(Y_{j,1}-Y_{j,0}\right)\right),\end{aligned} \tag{6}$$

where N stands for "non directed" or "neutral", D stands for "directed", N_T^p is the set of persons who received funding of type $p\in\{N,D\}$ and N_C^p is the set of controls for the funded individuals of type p. $Y_{i,t}$ is the outcome variable observed in period t, with $t=1$ in the period after the treatment assignment, and $t=0$ in the period before treatment. The weights ω_j are defined as in Equation 5. The first part of the right side of the equation refers to the difference between the treated and control groups of non directed programs, whereas the second part is the same difference for directed programs. The differential ATT of non directed program over directed program is simply equal to the difference between those two terms. It is estimated using a similar regression as Equation 4, but now considering the coefficient of a term to be added, formed of a triple interaction between a post-funding dummy, a treatment dummy and a non directed program dummy.

The impact of directed vs. non-directed programs Figure 5 shows the publication records of the granted and the (properly weighted) applicants to the two programs at different years before and after the application year. Red solid (blue dashed) lines stand for the granted (controls). Circles (crosses) denote directed (non-directed) programs. Controls in each program have publication trends that are very similar to the granted before granting date. For non directed programs, treated and controls match nearly perfectly in levels as well before treatment. Applicants and granted to the non-directed program have larger publication records than their counterparts in the directed programs which is consistent with the idea of a higher level of competition in this program. As a first sign of a presumed superior efficiency of non-directed programs, we observe an increasing spread between granted and controls posterior to the application year, for this program only when publications are weighted by citations or by the journal Impact Factor.

The precise impact analysis is reported in Table 8. Directed and non-directed programs barely differ in their impact on the volume of scientific production: a 2.8% difference in favor of non-directed programs, only significant at the 10% level. However, non-directed programs turn out to be significantly more efficient when the impact factor of the journals or the number of direct citations are taken into account. Directed programs have a treatment effect on the treated (baseline ATT in the table) of 3.1% when articles are weighted by the journal Impact Factor, while switching to a non-directed program increases that outcome by 11.1%. The difference between directed and non-directed programs is even sharper when articles are weighted by citations: the baseline treatment effect of directed programs is 5.9%, while switching to a non-directed program raises output by a 20.3%. These differences between program types are even larger than the overall impact of ANR funding.[21]

These results strongly support the idea that non-directed programs are very efficient, while directed programs have a limited impact on scientific outcomes. However, we find that directed programs are much more stimulating coauthor turnover as Table 8 shows the recipients of directed program funds have 17% more new collaborators than the recipients of non-directed programs. More new collaborations could be an early sign of professors and researchers granted on directed programs investigating more new and original problems.

Novelty Peer review procedures have been repreatedly criticized as being negatively biased toward really groundbreaking and innovative projects (Braben, 2004; Chubin and Hackett, 1990; Wesseley, 1998). Boudreau et al. (2016) show that highly novel projects are associated with lower ratings in a field experiment. Azoulay et al. (2011) show that scientists supported on a program specifically funding researchers (vs. projects) explore more novel research lines. The authors interpret their finding arguing researchers granted on projects are bound to their project proposals whereas others can more easily redesign their goals. Agencies could also face more difficulties in inducing (often disciplinary based) committees to support risky research projects rather than to fund researchers willing to take such risks. As we

[21]It may sound surprising that summing the treatment effect of the baseline directed programs with the effect of switching to non-directed programs, we obtain a larger number than the overall 15.2% effect seen in the previous section. The explanation is that projects granted via non-directed programs are less numerous overall (42% vs. 58% for directed programs).

have access to two different types of funding programs, we can specifically look how both programs deal with novelty. Directed programs focus primarily on new and promising areas of science. If successful in their explicit goals, they should attract and fund more often professors and researchers who investigate new research problems. However, the non-directed programs, which are open to any field of science, may as well attract and select unanticipated pathbreaking research proposals.

To address this issue, we need to proxy the novelty of the research articles of funded and non successful applicants to the directed and non-directed programs, before and after funding date. As we need to look at this dimension in the longer run, we perform a supplementary extraction of WoS data up to year 2015 that is up to six years after the last funding year of funding (2009). Article novelty is calculated using the frequencies of pairwise combinations of Author Keywords as introduced in Carayol et al. (2018). This measurement of novelty is intended to identify the originality of the research directions, the very problem addressed by research articles. Carayol et al. (2018) show, on more than ten million research articles published by journals indexed in the Web of Science (WoS), that Pairwise Author Keywords Novelty is a very good predictor of citations and highly cited articles, even in the relatively short run.[22] We compute yearly average and maximum pairwise author keywords novelty to appreciate to what extent their research is novel over time, before and after the application year.

In Figure 6, as in previous figures, we use residuals obtained after yearly scores are first regressed on year dummies. We find that directed programs indeed attract and fund professors and researchers whose research is more novel in average than non-directed programs. Differences between programs are more pronounced when we look at maximum novelty rather than at average novelty. It is interesting to observe that granted professors and researchers on non-directed programs perform less novel reseach than unsuccessfull applicants. This difference however shrinks when considering maximum novelty. There is no post-treatement tendency of the granted from directed programs to specifically undertake more novel research. If significant, the impact would rather be negative but Table 9 confirms there is no overall significant impact of funding on novelty, and no significant differentiation between programs in this respect.

These results lead to the conclusion that directed programs are more successful in attracting and funding researchers and professors who produce more novel science. However, both directed and undirected programs are ineffective in incentivizing the funded toward addressing more novel research lines than they did before treatment, even in the longer run.

[22] Carayol et al. (2018) show that relying on the frequencies of pairs of author keywords is key to this result, as using either keyword frequencies, predefined keywords or journal co-citations does not lead to the same results.

7 Designing funding programs: more results

We want to shed light on the conditions under which project funding turns out to be more efficient, and to what extent precisely.[23]

Impacts along the career path Estimating the impact of fund allocation at different career stages is an important policy issue. We thus ran estimations similar to the preceding ones, allowing us to differentiate the impact on younger scientific coordinators (equal to or less than 43 years old, the median age) from that on older ones. The impact of choosing a younger coordinator is then estimated by interacting post-funding dummy with treatment dummy in a fixed effect regression using two time periods' panel data, where observations are weighted according to the chosen method.Results are reported in Table 10. We find non significant differences in the volume of publications and when articles are weighted by the journal's average impact factor. However, an important and significant difference is found in terms of citations: the impact on younger coordinators is 9.5% higher than that on older ones. This implies that the impact in terms of citations for younger coordinators is more than twice that observed among older scientists. This result is pretty strong and has significant policy implications. Further, no significant differentiated effect on collaborations is observed. Funding only increases the team size of the older scientists slightly more than that of the younger scientists (2.8%, significant at the 10% level only).

We now differentiate the impact according to the publication profiles of the treated individuals at the time of funding. Our goal is to investigate whether some publication profiles are more likely to be positively impacted by the funding policy than others. Treated individuals and controls are ranked within each discipline according to the number of citations received by their articles published in the preceding three years,[24] and are categorized in either one of the four largest deciles or in the remaining six deciles. In the triple difference approach, the performances of the top 10% are taken into reference. It is found that the treated individuals who are in the top 10% are never those on which the impact is the largest. Largest impacts are found in terms of publication volume when the treated individuals are in the second to the fourth deciles only, which are significantly larger than those of the first decile (from 8.2% to 11.8% larger). Similar statements can be made in terms of impact factor and citations, though coefficients are less significant. This can be explained by the fact that the top professors and researchers may have access to other sources of funds. Though the committees should select applicants who have strong publication records, the impact is not likely to be the largest when the funds are targeted to those who can obtain funds elsewhere,

[23] The main results are presented in this section whereas their associated tables are to be found in Appendix F of the online Supplementary Material.

[24] We have used alternative performance variables to rank them, such as the number of articles, or even when such articles are weighted by the journal impact factor. Results are qualitatively similar.

at the European level, for instance. Note that this statement is in terms of elasticities, not in absolute terms (number of citations for instance). A lower impact in terms of elasticity on top-10% performers may well correspond to a larger impact in absolute outcomes. On the other side of the distribution, when the treated individuals are not in the four largest deciles, the impact is likely to be significantly lower, not on the volume of publications, but both when the impact factor of the journal is considered and for citations. When, for instance, in the six lower deciles, the treated individuals have an average impact in terms of citations reduced by 9%, that is no longer significantly different from zero.

PI or not PI Project variables are also available. In particular we have information on the role each person plays in the project: is she/he scientific coordinator of the whole project (the PI of the project), or only scientific coordinator of one institutional partner in a multi-partner project. As the design of the ANR grant system provides each partner's scientific coordinator with a significant level of autonomy (in particular financial), we have chosen the partner level of analysis rather than the project level. However, the project PI role is specific, often not a desirable one to play and one that keeps busy with administration and coordination tasks. We thus keep track of the status of each partner's scientific coordinator in the project with a dummy labeled PI, which will allow us to check whether PIs are compensated for their efforts by increased scientific productivity and/or collaborations.

In a project, do partners free-ride on the PI who bears most of the between-partners coordination costs? Or, conversely, does the project PI free-ride on the partners' scientific coordinators, using their labor force to increase his or her scientific production? We find no significant difference according to the status of the treated individuals in the project, who can be either PI or partner scientific coordinator. Thus it seems that the benefits and costs of coordinating multi-partner projects counterbalance each other. Gains of assuming the PI role are also not observed in collaborations. Unreported estimates show that the PI role has no effect on team size, number of coauthors and number of new coauthors. These results highlight how burdensome the PI role is. At the time of the proposal, assembling together all partners' contributions. At the time of the project, coordinating the work of all partners. The specific rules of the ANR, which give broad autonomy to the institutional partners and thus less power to the PI of the projects, probably does not help reduce such coordination costs. Another explanation is that it is still complex (though not impossible) in France to use project funds to reduce coordination costs or at least buy back teaching time, for instance.

Year effect No significant difference is found according to the year of funding. This result may seem surprising, bearing in mind that the agency was created in 2005. We guess that the agency has significantly increased its capabilities over the time period considered. We also know that the level of competition has been fairly different across years. For instance, the rate of success of the first year was much higher than that of the second year (48% as compared to 26%). In a sense, the fact that we find no significant difference between years is reassuring vis-à-vis our estimation methodology - tending to show that appropriate controls have been found for each year.

Scientific field effect When interaction with the scientific discipline is considered, we find that the impact of receiving funds is never significantly larger than in the life sciences, which is the reference. The only exception applies to Information and Communication Sciences and Technologies, where the impact is greater by 8.8% on citations and by 6.2% on the number of articles. Note that significance levels are however low (in particular for citations) and should thus be treated with caution.

8 Conclusions and discussion

In this article we have taken advantage of the recent French experience in which a new institution for project-based funding was created in 2005. This institution operates on a large scale, having distributed funds to research projects whose accumulated total costs approach ten billion euros over the five years covered by the study. Moreover, a certain level of variation in programs' rules and recipients' characteristics allows us to investigate the relative efficiency of variants of project funding. The results are not specific to one field of science, as all disciplines of hard and natural sciences are concerned (as well as some social sciences).

We identify the impact of receiving a research grant essentially by comparing the research production trajectories of the scientific coordinators of the funded projects with those of control groups. The controls are selected and weighted thanks to propensity scores that model the treatment on observables. Because the data on the whole reference population (not only on applicants) as well as several useful variables potentially explaining selection are available, we can compare how various sets of controls pass parallel paths tests. The "best" set of controls according to those tests picks controls among applicants exclusively, models treatment by program types, and includes past publication performances at the time of treatment as well as recent trends. This suggests future studies should have similar information to obtain satisfactory control sets.

Concerning the global efficiency of project-based funding, our study concludes that a grant increases the number of publications weighted by citations by about 15%. That result is larger than what was previously observed in Jacob and Lefgren (2011). However, as our study is not limited to a specific scientific field and as few alternative opportunities for project-based funding were available at the time of the study in France, our results are less prone to be affected by a displacement effect (negative bias). This suggests that our quantification of project-based funding is the closest to the real effect.

Further, we also find that funding has a positive effect on the size of collaborators' network and on the turnover of collaborators. Although the agency under investigation operates on a national basis, it does increase international collaborations. Funding thus has a significant and positive impact on the scope of collaboration networks. One concern remains, however, since project funding does not affect the novelty of the research problems that are tackled by the funded individuals. This is a serious issue often raised by funding agencies themselves which would need further investigation.

Some of the most striking practical results of our study concern the differentiated impacts with respect to the types of program. We find that when programs have no specific direction, so that they are open to wider competition, they have a much larger impact. Directed programs have a significant but rather small impact, while the surplus of impact of non-directed programs is quite large, even larger than the average impact of funding. This nominal advantage of non-directed programs is not counterbalanced by any sign of increased novelty of the research performed by recipients of directed grants. However, the directed programs prove successful in attracting and funding professors and researchers who develop (essentially before the funding date) more novel research than non-directed programs. Last but not least, the funds allocated to younger applicants have much larger impacts than those allocated to older applicants. This strongly supports the idea that project-based funding should keep a large door open to younger applicants. If confirmed by other studies, these results may provide some guidelines for improving project funding in science.

References

Abadie, A., 2005. Semiparametric difference-in-differences estimators, The Review of Economic Studies 72(1), 1-19.

Aghion, P.,Dewatripont, M., Stein, J.C., 2008. Academic Freedom, Private Sector Focus, and the Process of Innovation. RAND Journal of Economics 39, 617-635.

Arora, A., David, P. A., Gambardella, A., 2000. Reputation and competence in publicly funded science: estimating the effects on research group productivity, 141-176: in Encaoua, D., Hall B., Laisnay F., Mairesse, J., The Economics and Econometrics of Innovation, Springer US.

Arora, A., Gambardella, A. 2005. The impact of NSF support for basic research in economics, Annales d'Economie et de Statistique 79-80, 91-117.

Austin, P. C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies, Multivariate Behavioral Research 46(3), 399-424.

Azoulay, P., Graff Zivin, JS., Manso, G., 2011. Incentives and Creativity: Evidence from the Academic Life Sciences, RAND Journal of Economics 42 (3), 527-554.

Becker, S., Ichino, A., 2002. Estimation of average treatment effects based on propensity scores, The Stata Journal 2(4), 358-377.

Benavente, J. M., Crespi, G., Garone, L. F., & Maffioli, A., 2012. The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT. Research Policy, 41(8), 1461-1475.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates?, Quarterly Journal of Economics 119, 249-275.

Boudreau, Kevin J., Guinan, Eva C., Lakhani, Karim R. and Riedl, Christoph. 2016. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science, Management Science 62(10), 2765-2783.

Braben, D.W., 2004. Pioneering research: A risk worth taking. Hoboken, NJ: Wiley- Interscience.

Carayol, N., Llopis, O., Lahatte, A., 2018, The Right Job and the Job Right: Novelty, Impact and Journal Stratification in Science, mimeo.

Carayol, N., J.-M. Dalle., 2007. Sequential problem choice and the reward system in Open Science, Structural Change and Economic Dynamics 18, 167–191.

Carayol N., Lanoë, M., 2013. The impact of blue sky project based funding of academic research, mimeo.

Carter, G.M., Winkler, J.D., Biddle A.K., 1987. An evaluation of the NIH Research Career Development Award. Santa Monica, Calif, Rand Corporation.

Chubin, D.E., Hackett, E.J., 1990. Peerless science: Peer review and U.S. science policy. Stony Brook, NY: State University of New York Press.

Chudnovsky, D., López, A., Rossi, M. A., & Ubfal, D., 2008. Money for science? The impact of research grants on academic output. Fiscal Studies, 29(1), 75-87.

Cochran, W. G., Chambers, S. P., 1965. The planning of observational studies of human populations, Journal of the Royal Statistical Society Series A (General), 234-266.

Cochran, W. G., Rubin, D. B., 1973. Controlling bias in observational studies: A review. Sankhyā: The Indian Journal of Statistics, Series A, 417-446.

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy 23 (5), 487-521.

Dehejia, R., Wahba, S., 1999. Causal effects in non experimental studies: reevaluating the evaluation of training programs, Journal of Statistical American Association 94(448), 1053-1062.

Fang, F.C., Bowen, A., Casadevall, A., 2016. NIH peer review percentile scores are poorly predictive of grant productivity, eLife 5, e13323.

Frölich, M., 2004. Finite-sample properties of propensity-score matching and weighting estimators, The Review of Economics and Statistics 86(1), 77-90.

Gush, J., Jaffe, A.B., Larsen, V., Laws, A., 2015. The effect of public funding on research output: the New Zealand Marden Fund, NBER Working Paper #21652.

Heckman, J. J., Ichimura, H., Todd, P. E., 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. The review of economic studies, 64(4), 605-654.

Hirano, K., Imbens, G. W., 2001. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcomes Research Methodology 2(3-4), 259-278.

Hirano, K., Imbens, G. W., Ridder, G., 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, Econometrica 71(4), 1161-1189.

Imbens, G. W., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review, Review of Economics and statistics 86(1), 4-29.

Imbens, G. W., Wooldridge, J., 2008. Recent Developments in the Econometrics of Program Evaluation, National Bureau of Economic Research w14251.

Jacob, B. A., Lefgren L., 2011. The Impact of Research Grant Funding on Scientific Productivity, Journal of Public Economics 95(9-10), 1168-1177.

Lehman, H.C., 1953. Age and Achievement. Princeton University Press, Princeton.

Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: evidence for academic scientists. American Economic Review 81, 114–132.

Li, D., Agha, L., 2015. Research funding. Big names or big ideas: do peer-review panels select the best science proposals?, Science 348(6233), 434-8.

McDowell, J.M., 1982. Obsolescence of knowledge and career publication profiles: some evidence of differences among fields in costs of interrupted careers. American Economic Review 72, 752–768.

National Science Board, 2016. Science & Engineering Indicators 2016, National Science Foundation.

Reijnhoudt, L., Costas, R., Noyons, Ed, Börner, K., Scharnhorst, A., 2014. 'Seed + expand': a general methodology for detecting publication oeuvres of individual researchers, Scientometrics 101, 1403-1417.

Robins, J. M., Hernan, M. A., Brumback, B., 2000. Marginal structural models and causal inference in epidemiology, Epidemiology 11(5), 550-560.

Rosenbaum, P. R., Rubin, D. B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, The American Statistician, 39(1), 33-38.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects, Biometrika, 70(1), 41-55.

Rosenbaum, P. R., 2002. Observational Studies. New York: Springer.

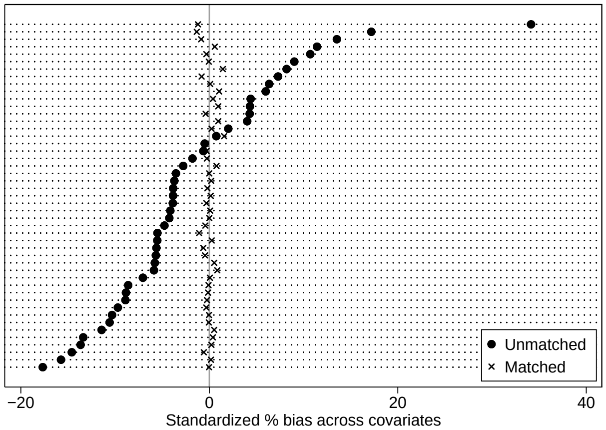
Stephan, P., 2012. How Economics Shapes Science, Harvard University Press.

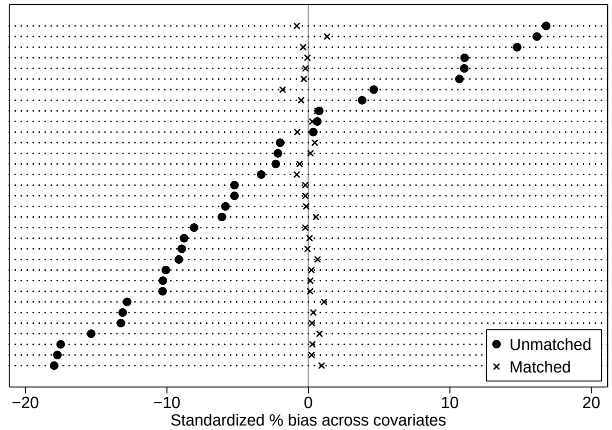
Wesseley, S., 1998. Peer review of grant applications: What do we know?, Lancet 352(9124), 301-305.

Wuchty, S., Jones, B. F, Uzzi, B., 2007. The Increasing Dominance of Teams in the Production of Knowledge, Science 316(5827), 1036-1039.

Zuckerman, H.A., Merton, R.K., 1972. Age, aging, and age structure in science. In: Riley, M.R., Johnson, M., Foner, A. (Eds.), A Sociology of Age Stratification: Aging and Society, vol. 3, Russel Sage foundation, New York, Reprinted in: Storer, N.W. (Ed.), 1973. The Sociology of Science: Collected Papers of R.K. Merton, Chicago University, Chicago Press, 497-559.

Figure 1: Standardized bias (in %) associated with each explanatory covariate in the original unmatched sample and in the weighted sample for the directed (left graph) and non-directed (right graph) programs, using the estimated inverse probability of treatment weights.





Notes: Each dotted line represents an explanatory covariate included in the propensity score logit model (the X vector). The variables used in the propensity score model for non-directed programs are presented in Table 4, and those used for the non-directed programs are presented in Table 5.

Figure 2: Density and box plot of the estimated propensity scores before and after weighting by the inverse probability of treatment weights for the directed programs.

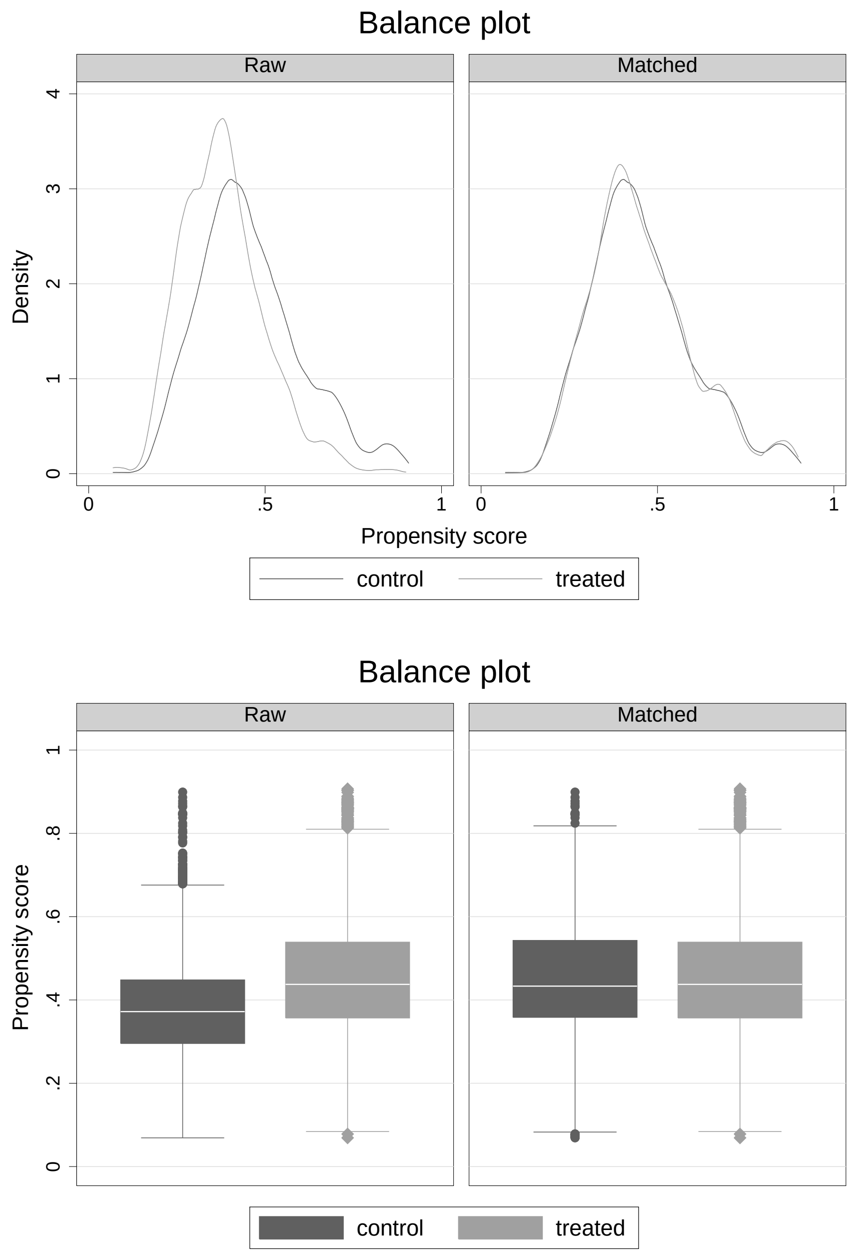


Figure 3: Density and box plot of the estimated propensity scores before and after weighting by the inverse probability of treatment weights for the non-directed programs.

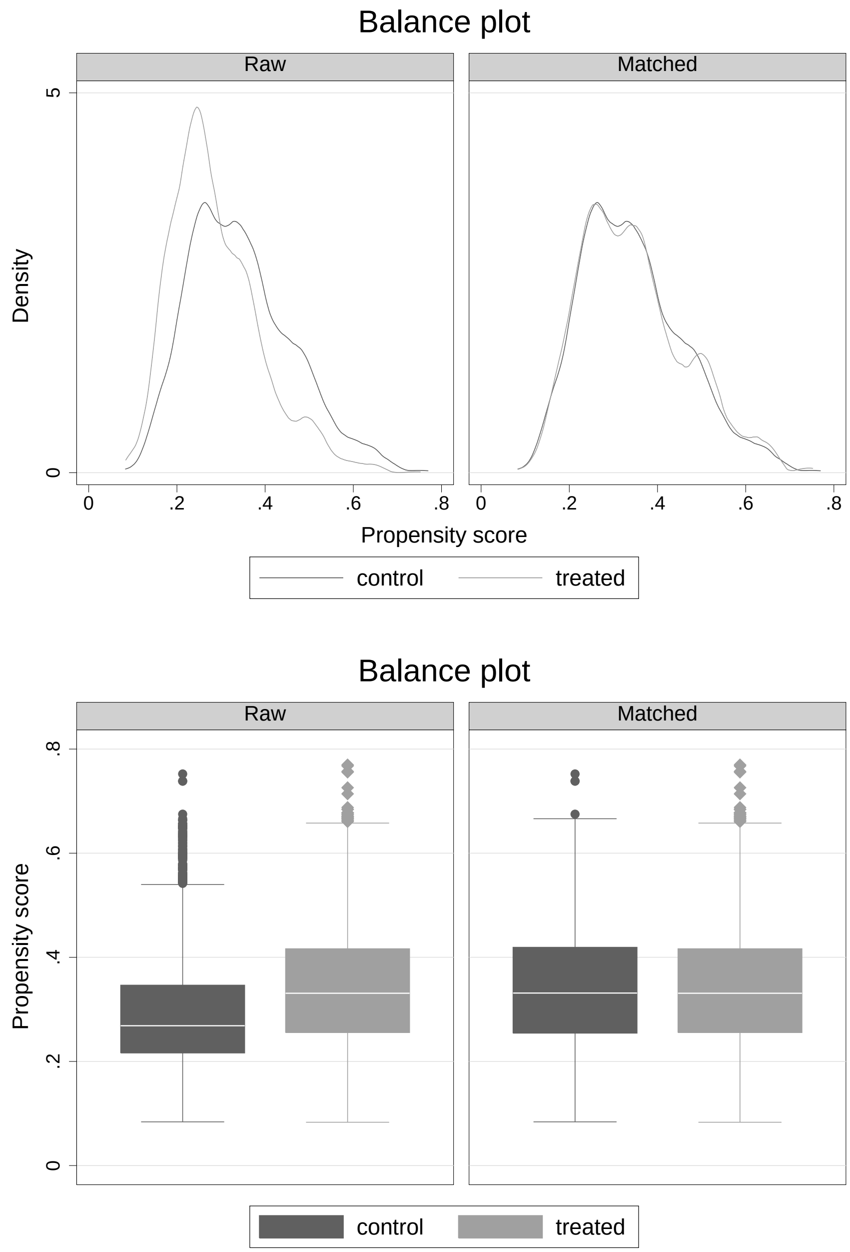
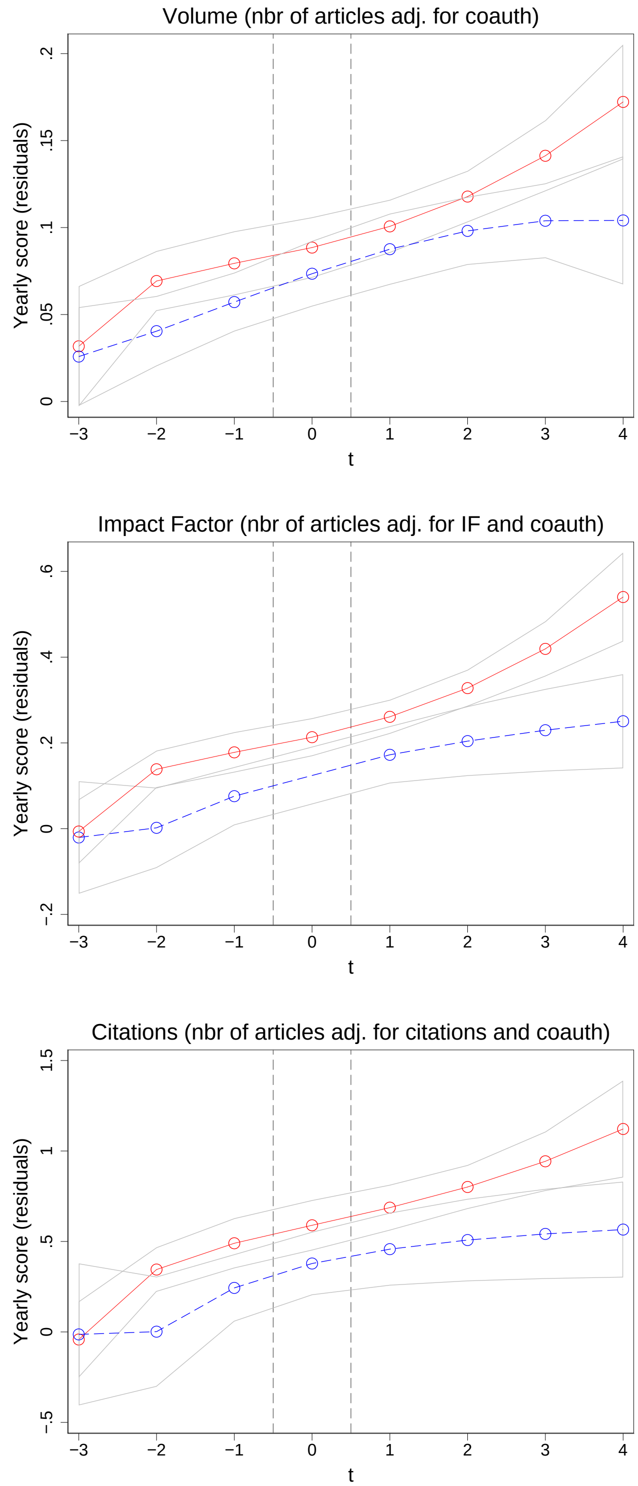
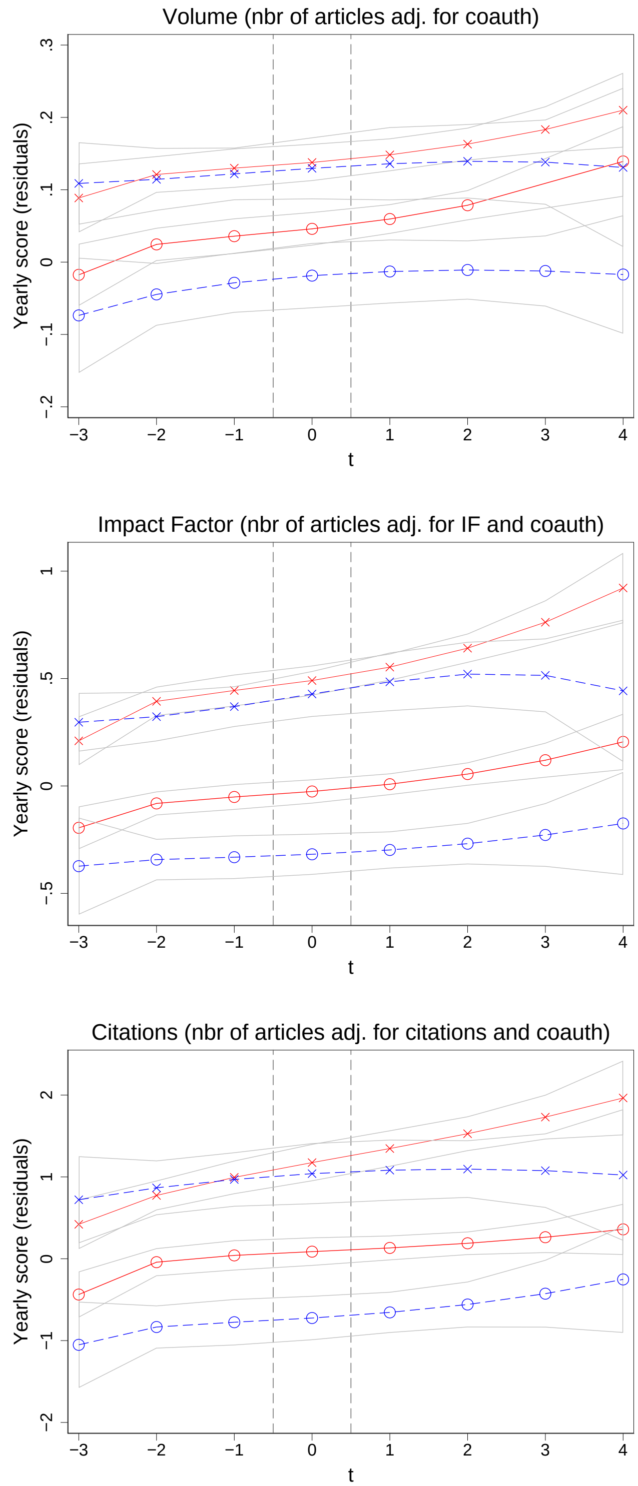


Figure 4: Yearly scientific outcomes of the funded professors and researchers (red solid line) and of their controls (blue dashed line) with respect to the funding year ($t = 0$).



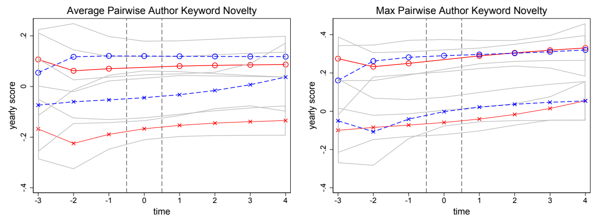
Notes: The red solid line stands for the granted and the blue dashed line stands for the unsuccessful applicants. Mean and 95% fractional polynomial confidence intervals are presented. The first year of funding occurs at $t = 0$. *For each variable considered (Volume, Impact factor and Citations), we present the residuals obtained after regressing yearly scores on year dummies (absorbing potential year shocks and trends). Observations are weighted according to the inverse probability of treatment. The variables used in the propensity score models are reported in Tables 4 and 5.*

Figure 5: Yearly scientific outcomes of the funded professors and researchers (red solid line) and of their controls (blue dashed line) who applied to the two funding schemes: directed (o marks) and non-directed ($\times$ marks) with respect to the funding year ($t = 0$).



Notes: The red solid line stands for the granted and the blue dashed line stands for the unsuccessful applicants. The circle points correspond todirected programs while the crosses stand for the non-directed programs. Mean and 95% fractional polynomial confidence intervals are presented. The first year of funding occurs at $t = 0$. For each variable considered (Volume, Impact factor and Citations), we present the residuals obtained after regressing yearly scores on year dummies (absorbing potential year shocks and trends). Observations are weighted according to the inverse probability of treatment. The variables used in the propensity score models are reported in Tables 4 and 5.

Figure 6: Yearly average (left graph) and maximum (right graph) Pairwise Author Keyword Novelty of professors and researchers research articles. Red solid lines (blue dashed line) stand for the funded (unsuccessful applicants). Applicants to directed programs (non directed programs) have o marks ($\times$ marks).



Notes: The red solid line stands for the granted and the blue dashed line stands for the unsuccessful applicants. The circle points correspond to non-directed programs while the crosses stand for the directed programs. Mean and 95% fractional polynomial confidence intervals are presented. The year of first funding occurs at $t = 0$. *The included data points go up to year 2015 included, that is up to six year after the last funding year of funding (2009). For each variable considered (mean and maximum article novelty in the considered year), we present the residuals obtained after regressing yearly scores on year dummies (absorbing potential year shocks and trends). Observations are weighted according to the inverse probability of treatment. The variables used in the propensity score models are reported in Tables 4 and 5.*

Table 1: Descriptive statistics on outcome variables among non-applicants, unsuccessful applicants and granted ones, before and after the reference year.

Variables	*stat.*	Non-Applicants Before	Non-Applicants After	Not Granted Applicants Before	Not Granted Applicants After	Granted Applicants Before	Granted Applicants After
	mean	1.42	1.64	2.33	2.58	2.80	3.16
Volume	*med.*	0.53	0.68	1.45	1.58	1.83	2.05
	s.d.	(2.66)	(3.01)	(3.04)	(3.40)	(3.35)	(3.72)
	mean	2.43	2.98	4.69	5.34	5.77	7.14
Impact Factor	*med.*	0.59	0.88	2.53	2.89	3.14	3.93
	s.d.	(5.41)	(6.40)	(7.02)	(7.86)	(8.19)	(9.95)
	mean	5.74	6.19	11.14	10.58	14.45	16.19
Citations	*med.*	0.92	1.12	4.90	4.27	6.68	7.13
	s.d.	(15.16)	(15.99)	(20.19)	(20.24)	(23.97)	(27.45)
	mean	10.65	19.93	7.24	10.29	7.55	11.34
Av. team Size	*med.*	5.14	5.60	5.50	6.00	5.29	5.91
	s.d.	(57.67)	(152.21)	(21.80)	(70.70)	(27.47)	(83.95)
	mean	31.41	57.47	37.37	71.32	40.10	81.82
Coauthors	*med.*	13	20	22.00	37.00	24	44
	s.d.	(51.25)	(107.04)	(49.56)	(104.06)	(50.85)	(116.56)
	mean	4.71	5.93	5.92	7.29	6.64	8.22
Internat. Collab	*med.*	1	2	3	3	3	4
	s.d.	(10.65)	(14.77)	(10.49)	(13.35)	(11.04)	(13.94)
	mean		12.45		17.34		18.20
New Coauthors	*med.*		8		15		16
	s.d.		(11.55)		(12.55)		(12.84)
	mean	10.58	10.84	10.78	11.01	10.73	10.99
New problems	*med.*	10.73	10.96	10.85	11.10	10.79	11.07
	s.d.	(0.86)	(0.85)	(0.78)	(0.77)	(0.75)	(0.74)
Nb. of observations		102,490		9,652		5,831	

Notes: The "before" ("after") columns refer to the three years which precede (follow) the year of reference. It is the year of first successfull application for the funded. There are exactly 5,831 distinct persons in our dataset that have been funded by the ANR. We do not consider their subsequent successful applications, nor their unsuccessful ones. There are 9,652 distinct applications of the 4,892 persons who applied at least once, but were never funded. We thus consider all the applications of the non funded applicants, and for each of those applications, the before and after periods are defined according to the application year. As there is no specific reference year for the 20,498 non-applicants, they are considered for each of the five years of the study (102,490 observations), and the before and after periods are defined accordingly. Incomplete information about the identity of coauthors before year 2002 prevents us to compute the New Coauthors variable in the period preceding the year of reference.

Table 2: Descriptive statistics on selection variables for non-applicants, not granted applicants and granted ones, by progam type (directed or non-directed).

Variables		Non-Applicants	Not Granted Applicants		Granted Applicants	
	stat.		Thema	Non-Thema	Thema	Non-Thema
	mean	42.87	45.30	42.73	44.44	42.50
Age	*med.*	41.00	45.00	41.00	44.00	41.00
	s.d.	(10.25)	(8.09)	(8.55)	(8.16)	(8.35)
	mean	6.49	10.92	10.26	12.21	12.17
Articles	*med.*	2.00	7.00	6.00	8.00	8.00
	s.d.	(12.61)	(14.64)	(13.16)	(14.98)	(14.19)
	mean	37.50	58.25	64.45	75.48	80.94
Citations	*med.*	4.00	23.00	26.00	30.00	37.00
	s.d.	(117.74)	(111.97)	(129.73)	(133.83)	(138.06)
	mean	78.34	132.60	148.28	153.85	164.80
Total Citations	*med.*	9.00	57.00	67.00	63.00	79.00
	s.d.	(231.82)	(253.57)	(272.00)	(276.58)	(271.08)
	mean	2.53	3.88	4.47	4.46	4.95
Max Impact Factor	*med.*	1.26	3.25	3.65	3.36	3.74
	s.d.	(3.42)	(3.71)	(4.11)	(4.57)	(4.53)
	mean	7.68	9.08	9.36	8.97	8.57
Av. Lab Articles	*med.*	6.64	8.38	8.68	8.18	8.14
	s.d.	(5.96)	(6.39)	(5.65)	(5.97)	(4.88)
	mean	369.09	409.39	461.10	408.82	447.05
Max Lab Citations	*med.*	247.00	305.00	358.00	321.00	333.50
	s.d.	(422.83)	(421.27)	(421.91)	(409.88)	(445.70)
	mean	52.01	51.58	51.99	54.12	56.20
Lab Size	*med.*	42.00	41.00	45.00	42.00	48.00
	s.d.	(40.41)	(41.38)	(36.98)	(43.91)	(39.12)
	mean	2007.00	2007.13	2007.95	2006.64	2006.68
Application Year	*med.*	2007.00	2007.00	2009.00	2006.00	2006.00
	s.d.	(1.41)	(1.18)	(1.20)	(1.32)	(1.39)
Nb. of observations		102,490	4,085	5,567	3,385	2,446

Notes: The "Directed" ("Non-Directed") distinction does not make sense for the 20,498 non-applicants. They are considered in this table as five distinct potential controls, one for each of the five years of the study, and the statistics are computed accordingly. Out of the 5,831 who have been funded by the ANR, 3,385 got their first application thanks to a directed program, 2,446 thanks to the non-directed program. We do not consider their subsequent successful applications, nor their unsuccessful ones. There are 7,433 distinct applications of the 4,892 persons who applied at least once, but were never funded. 4,085 applications to directed programs, and 5,567 to non-directed ones. Articles is the number of articles (fractionnal counts) published in the previous three years. Citations is the number of citations to those articles (fractionnal counts) in a three-year period after publication. Total Citations is the number of citations to articles published since 1999. Max Impact Factor records the largest impact factor among the journals in which these articles were published. Av. Lab Articles is the average number of articles (fractionnal counts) among all professors and researchers affiliated to the same laboratory. Max Lab Citations is the number of citations (fractional counts, to articles published in the previous three years) received by the professor or researcher who got the maximum number of such citations in the lab. Application Year is self explanatory. As there is formally no year of application for non-applicants controls, for them the current year is considered.

Table 3: Synthesis of the eight specifications of the propensity score model.

	1	2	3	4	5	6	7	8
Restriction on the controls								
All the reference population	X	X						
Only applicants			X	X	X	X	X	X
Exact matching								
Section (detailed field & employer)	X	X						
Field & research institute			X	X		X	X	
Program type					X			X
Covariates explaining the treatment								
Individual covariates	X	X	X	X	X	X	X	X
Laboratory covariates		X		X			X	
Trend covariates						X	X	X

Notes: The reference selection model is the eighth.

Table 4: List of covariates used for the propensity score estimation in the reference model, for directed programs.

Variable Description
Age at the time of application
Number of publications in the previous 3 years
Number of citations to papers published in the previous 3 years
Maximum Impact Factor in the previous 3 years
Total number of citations to papers published since 1999
The specific directed program
Year of the application
Interaction between the specific directed program and the application year
Variation in absolute terms in the number of publications $(t-3, t)$
Variation in percentage points in the number of citations $(t-3, t-1)$

Notes: All outcome variables (apart from the total number of citations) are adjusted for co-authorship (fractionnal counting) and categorized in four classes (four dummies are created): top 10%, next 20%, next 30% and last 40% . The distribution of the variables is restricted to researchers of the same scientific field (31 disciplines used).

Table 5: List of covariates used for the propensity score estimation in the reference model, for non-directed programs.

Variable Description
Age at the time of application
Number of publications in the previous 3 years
Number of citations to papers published in the previous 3 years
Maximum Impact Factor in the previous 3 years
Total number of citations to papers published since 1999
Large scientific disciplines dummies
Dummies when an university or a specific research institute is the employer
Variation in absolute terms in the maximum Impact Factor $(t-3, t)$
Variation in percentage points in the maximum Impact Factor $(t-3, t-1)$

Notes: All outcome variables (apart from the total number of citations) are adjusted for co-authorship (fractionnal counting) and categorized in four classes (four dummies are created): top 10%, next 20%, next 30% and last 40%. The distribution of the variables is restricted to researchers of the same scientific field (31 disciplines used). The large scientific disciplines dummies are Life sciences, Medicine, Chemistry, Physics, Science of the Universe, Engineering, Mathematics, Information science, Human & social sciences. The specific research institutes are: CNRS, INRA, INRIA, IRD, and INSERM.

Table 6: Parallel path test: Difference-in-differences estimates of the mean effect of treatment on various production variables with the reference specification of the selection stage (calculated from $t-3$ to $t-1$ and from $t-3$ to t).

	from $t-3$ to $t-1$			from $t-3$ to t		
	5 nn	kernel	iptw	5 nn	kernel	iptw
Volume	-.00954	-.00809	-.00727	.00078	-.00101	-.00139
	(-1.39)	(-1.27)	(-1.13)	(0.11)	(-0.15)	(-0.21)
Citations	-.00454	-.0012	-.00268	-.01868	-.02018	-.02246
	(-0.3)	(-0.09)	(-0.19)	(-1.23)	(-1.45)	(-1.58)
Impact Factor	-.00515	-.00383	-.00588	-.00626	-.0040	-.00484
	(-0.6)	(-0.47)	(-0.72)	(-0.69)	(-0.47)	(-0.56)

*Notes: Conditional difference-in-differences results. Dependent variables in Log. Robust standard errors in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. Observations are weighted according to the inverse probability of treatment.*

Table 7: Average treatment effect of receiving an ANR grant on publication outcomes and collaboration behaviors (the three years after treatment against the three years before).

Volume	Impact Factor	Citations
0.0350***	0.0825***	0.1525***
(4.46)	(7.53)	(9.30)
Av. Team Size	**Coauthors**	**Internat. Collab.**
0.0218***	0.0981***	0.0418***
(2.71)	(7.02)	(2.82)
New Coauthors[a]		
0.0668***		
(3.03)		

*Notes: Conditional difference-in-differences results. Coefficients and standard errors of the interaction term between the post-funding period dummy and the treatment dummy in fixed effect regressions. Observations are weighted according to the inverse probability of treatment. Dependent variables in Log. Robust standard errors in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. The variables used in the propensity score models are reported in Tables 4 and 5.*

[a] *Conditional differences results only as this variable counts the new items in the post-treatment period as compared to the pre-treatment period for treated and control units.*

Table 8: Differentiated effects of receiving an ANR grant on outcomes according to non-directed versus directed funding schemes (the three years after treatment against the three years before). Baseline average treatment effect for directed programs are in italics.

	Volume	Impact Factor	Citations
Non-Directed vs. Directed programs	0.0277* (1.77)	0.1111*** (5.10)	0.2028*** (6.26)
Baseline ATT of Directed programs	*0.022*** *(1.99)*	*0.0314*** *(2.04)*	*0.059*** *(2.54)*

	Av. Team Size	Coauthors	Internat. Collab.
Non-Directed vs. Directed programs	-0.0011 (-0.07)	0.0201 (0.72)	0.0286 (0.97)
Baseline ATT of Directed programs	*0.0223*** *(2.36)*	*0.0885**** *(4.48)*	*0.0283* *(1.32)*

	New Coauthors[a]
Non-Directed vs. Directed programs	-0.049 (-1.12)
Baseline ATT of Directed programs	*0.0901**** *(2.91)*

*Notes: "Non-Directed vs. Directed programs" lines report conditional difference-in-difference-in-differences results. Coefficients and standard errors of the triple interaction term between the post-funding period dummy, the treatment dummy and the non-directed-program dummy, in fixed effect regressions. "Baseline ATT of Directed programs" lines report the estimation of Conditional difference-in-differences for the directed programs only. Coefficients and standard errors of the interaction term between the post-funding period dummy and the treatment dummy in fixed effect regressions. All observations are weighted according to the inverse probability of treatment. Dependent variables are in log. Robust t-stats in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. The variables used in the propensity score models are reported in Tables 4 and 5.*

[a] *Conditional differences results only as this variable counts the new items in the post-treatment period as compared to the pre-treatment period for treated and control units.*

Table 9: Differentiated effects of receiving an ANR grant on the average and maximum Pairwise Author Keywords Novelty according to non-directed versus directed funding schemes (the three years after treatment against the three years before). Average treatment effect for the baseline (all projects and directed programs) in italics.

	Average Pairwise Author Keywords Novelty	Maximum Pairwise Author Keywords Novelty
Non-Directed vs. Directed programs	-.0071 (-0.56)	-.0097 (-0.76)
Baseline ATT	*-.0038* *(-0.69)*	*-.005* *(-0.83)*
Baseline ATT of Directed programs	.0027 (0.29)	.0019 (0.21)

*Notes: "Non-Directed vs. Directed programs" lines report conditional difference-in-difference-in-differences results. Coefficients and standard errors of the triple interaction term between the post-funding period dummy, the treatment dummy and the non-directed-program dummy, in fixed effect regressions. "Baseline ATT of Directed programs" lines report the estimation of Conditional difference-in-differences for the directed programs only. Coefficients and standard errors of the interaction term between the post-funding period dummy and the treatment dummy in fixed effect regressions. All observations are weighted according to the inverse probability of treatment. Dependent variables are in log. Robust t-stats in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. The variables used in the propensity score models are reported in Tables 4 and 5.*

Table 10: Differentiated effects of receiving an ANR grant on outcomes according to age dummy: below the median age (43) versus over the median age (the three years after treatment against the three years before).

	Volume	Impact Factor	Citations
Young (below median age) vs. Older	0.0221 (1.41)	0.0266 (1.29)	0.0952*** (3.09)
Baseline ATT on the Older	0.0227** (2.00)	0.0679*** (4.49)	0.1008*** (4.45)

	Av. Team Size	Coauthors	Internat. Collab.
Young (below median age) vs. Older	-0.0279* (-1.80)	-0.0034 (-0.13)	0.0280 (0.95)
Baseline ATT on the Older	0.0367*** (3,31)	0,0993*** (5,02)	0,0266 (1,27)

	New Coauthors[a]
Young (below median age) vs. Older	-0.0568 (-1.37)
Baseline ATT on the Older	0.0994*** (3.17)

*Notes: "Young (below median age) vs. Older" lines report conditional difference-in-difference-in-differences results. Coefficients and standard errors of the triple interaction term between the post-funding period dummy, the treatment dummy and the below median age dummy, in fixed effect regressions. "Baseline ATT on the older" lines report the estimation of Conditional difference-in-differences for the directed programs only. Coefficients and standard errors of the interaction term between the post-funding period dummy and the treatment dummy in fixed effect regressions. All observations are weighted according to the inverse probability of treatment. Dependent variables are in log. Robust t-stats in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. The variables used in the propensity score models are reported in Tables 4 and 5.*

[a] *Conditional differences results only as this variable counts the new items in the post-treatment period as compared to the pre-treatment period for treated and control units.*

Table 11: Differentiated effects of receiving an ANR grant on publication outcomes according to the position in the citation distribution at the time of funding (the three years after treatment against the three years before).

	Volume	Impact Factor	Citations
Baseline Top-10% publication performance	-0.0474**	-0.0340	0.1037**
	(-2.27)	(1.15)	(2.29)
Top-10-to-20% publication performance (vs. top-10%)	0.0823***	0.0608	0.0989*
	(3.06)	(1.62)	(1.80)
Top-20-to-30% publication performance (vs. top-10%)	0.106***	0.0675*	0.0716
	(4.05)	(1.87)	(1.33)
Top-30-to-40% publication performance (vs. top-10%)	0.118***	0.0880**	0.0921*
	(4.42)	(2.39)	(1.68)
Bottom-60% publication performance (vs. top-10%)	0.0632**	-0.0188	-0.0898*
	(2.38)	(-0.55)	(-1.70)

*Notes: "Baseline Top-10% publication performance" report the estimation of Conditional difference-in-differences for the top 10% publishing professors and researchers only. Coefficients and standard errors of the interaction term between the post-funding period dummy and the treatment dummy in fixed effect regressions. The other lines report the conditional difference-in-difference-in-differences results. Coefficients and standard errors of the triple interaction term between the post-funding period dummy, the treatment dummy and the percentile-class-of-the-citations-volume-prior-to-application dummy (mentioned at the right of each line, the top-10% being in reference), in fixed effect regressions. All observations are weighted according to the inverse probability of treatment. Dependent variables are in log. Robust t-stats in parentheses, clustered at the project level. Significance levels: 0.01: ***, 0.05: **, 0.10: *. The variables used in the propensity score models are reported in Tables 4 and 5.*

[a] *Conditional differences results only as this variable counts the new items in the post-treatment period as compared to the pre-treatment period for treated and control units.*