The impact of public research contracts on scientific productivity*

César Alonso-Borrego† Universidad Carlos III

Antonio Romero-Medina Universidad Carlos III

Rocío Sánchez-Mangas Universidad Autónoma de Madrid

Abstract

We analyze a competitive research-oriented public program established in Spain, the Ramon y Cajal Program, intended to offer contracts in public research centers to high-quality researchers. We study the effects of the Program on the ex-post scientific productivity of its recipients, relative to unsuccessful applicants with comparable curricula at the time of application. The full sample results demonstrate that the Program has a positive and significant effect on the scientific impact of the recipients, as measured by the average and the maximum impact factors, but the effect on the number of published papers is not significant. Consequently, receiving a contract does not significantly affect the quantity, but increases the quality, of the contract recipients' publications. This result is primarily driven by the particular relevance of experimental sciences in the Program.

Keywords: Ramon y Cajal Program, Brain Gain, Scientific Productivity, Government Research Programs, Human Capital, Policy Evaluation, Matching.

JEL Classification Numbers: O38, D78, C21, I23, O31, H52.

*We thank the Subdirección General de Formación y Movilidad del Personal Investigador (Spanish Office of Training and Mobility of Researchers) for providing the raw data and Miguel Benavente for assistance in data handling. Research funding from the Ministry of Finance and Competitiveness, Grant Nos. ECO2012-31358 (Alonso-Borrego), ECO2011-25330 (Romero-Medina) and ECO2012-32854 (Sánchez-Mangas) is acknowledged. We thank Raquel Carrasco, Jesús Carro, audiences at the Simposium of the Spanish Economic Association, Vigo, December 2012, the Annual Conference of the European Economic Association, Gothenburg, August 2013, and the Conference of the European Association for Research in Industrial Economics, Evora, September 2013, for their comments.

†Corresponding Author. Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe, Spain. e-mail: cesar.alonso@uc3m.es.

1 Introduction

The Ramon y Cajal Program is a targeted grant-based policy initiative –named after the Spanish neurobiologist and joint winner of the 1906 Nobel Prize in Physiology or Medicine, Santiago Ramón y Cajal– established by the Spanish government in 2001. The Program was designed to attract promising young researchers, both Spanish expatriates and foreign-born, and provide incentives to the Spanish public research centers (PRCs hereinafter) to improve their strategic planning. The grant recipients were offered a well-defined career path, with a 5-year employment contract at a Spanish PRC, and the opportunity to obtain permanent research positions at the end of the contract (see Sanz-Menéndez et al., 2002; Sanz-Menéndez, 2003).

The Program constituted a novel policy measure in the Science and Technology (S&T hereinafter) system, as it shifted the system's focus from training to employability policies (see Cruz-Castro and Sanz-Menéndez, 2005). It has become a relevant S&T instrument, and has attracted attention from the mass media and professional journals (see Bosch, 2001a and 2001b, and Muñoz-Pinedo et al., 2003) since it began.

In this paper, we assess the effect of the Program on the scientific productivity of applicants a few years after application. We exploit data on applications in several calls of the Program, provided by the Direccion General de Investigacion of the former Ministry of Science and Innovation.[1] We complement these data with individual and curricular information on the applicants. We compare successful and unsuccessful applicants using two alternative empirical approaches: linear regression and matching. Our results indicate that contract recipients did not generate a larger number of published contributions, but their scientific impact increased, measured by either the average or the maximum impact of these contributions, with respect to applicants who were not selected by the Program but were comparable in accordance with their curricula at the time of application.

[1] The ministerial arrangements were redesigned after the March 2004 elections, and the Ministry of Science and Technology was eliminated. Its competences were reallocated to the Ministry of Industry, Tourism and Commerce and to the Ministry of Education and Science, which undertook the S&T competences, including the Ramon y Cajal Program. Since the November 2011 elections, these competences have been the responsibility of the Ministry of Finance and Competitiveness.

Our paper belongs to the economic literature on science (Stephan, 1996). This literature initially focused, almost exclusively, on the effect of science on growth through its relationship with technology. Following the development of human capital models in the 1960s, economic research puts its attention on the labor market of scientists. A further line of research that studies reward incentives for scientists was introduced from the sociology of science. Until very recently, little attention has been paid to productivity of scientific research, particularly how scientific outcomes might be affected by resource funding. There are several studies that have analyzed the impact of grant funding on the scientific productivity. The studies differ in the research area under consideration, the measures of scientific productivity (patents, number of publications, citations of published articles, etc.), the unit for which productivity is measured (aggregate research group, individual researchers, head of the research group, etc.), the country and the funding scheme.

Arora et al. (1998) investigate the effects of a public funding program in biotechnology and bio-instrumentation in Italy on the publication records of grants research groups. They find positive, but heterogeneous effects, which are higher the better their past publication performance. Arora and Gambardella (2005) assess the effect of NSF grants in the field of economics on publication output, finding a small impact, except for younger scholars. Chudnovsky et al. (2008) analyze the impact of public grants in Argentina, finding that granted researchers increase the number of publications and their quality, measured by the impact factor. They determine that the grants have been especially relevant for young researchers. Azoulay et al. (2011), focusing on the field of life sciences in the US, show a positive impact of grants on the number of publications of awarded scientists. Similar effects in both the quantity and quality of the publication records are found by Carayol and Lanoe (2013) in France, although the results vary between research areas. However, Jacob and Lefgren (2011), focusing on a program from the National Institute of Health in US show small impacts on both the publications and the citations of granted researchers. Benavente et al. (2012) assess a public research funding program in Chile, finding positive effects on the quantity, but not on the quality of the publications.

Our work contributes to this literature by evaluating the impact on the individual scientific productivity of a hiring policy, instead of project-based research funding, targeted at young researchers in any research area. The fact that the hiring policy is aimed at young researchers avoids confounding effects associated with differences in scientific maturity and productivity in the life cycle. We also propose a measure of scientific impact that permits comparison between different areas.

To the best of our knowledge, our paper is the first evaluation of the Program's impact on the performance of contract recipients. However, several previous studies have analyzed other aspects of the Program. Sanz-Menéndez et al. (2002) undertook a descriptive study of the first call for applications, finding that researchers who were already in the Spanish S&T system obtained 60 percent of the contracts, while the remaining 40 percent of contracts were awarded to researchers external to the Spanish S&T system (two thirds of whom were Spanish). Cruz-Castro and Sanz-Menéndez (2005) use data from the first three calls to analyze the Program's impact on the spread of information regarding the quality of researchers and PRCs in Spain, and ascertain how the Program affected the organizational strategies of the PRCs. They concluded, "in the second call, the Program has earned a solid reputation abroad" and increased the chances of those PRCs with a good reputation to attract high-quality young researchers.

Ten years after the first call, the Ministry of Science and Innovation published a Results Report entitled "10 años del Programa Ramón y Cajal" ("Ten years of the Ramon y Cajal Program", DGI, 2010). The report summarized the results of a survey of the recipients and PRCs involved in the Program, and concluded, confirming previous findings, that the evaluation process is generally acknowledged to be objective and transparent, and that the Program has proven to be a good instrument to attract expatriate Spanish researchers.

According to Cruz-Castro and Sanz-Menéndez (2005), the Ramon y Cajal Program provided short-term relief for the key problems of the system that it was intended to address. Particularly it improved the employment opportunities, working conditions and academic career prospects of PhDs. On the supply side, the effects of the Program "have been pressurizing the PRCs to develop strategies for human resource recruitment by research field, and organizing their priorities in terms of competitive research capabilities"-. Overall, the professional community has claimed that the Program "offers a rare opportunity for young scientists trying to gain a foothold in the rigid Spanish academic system" (see Schiermeier, 2004). Our results allow us to conclude that the Ramon y Cajal recipients were able to achieve a quality level above that of comparable applicants not selected by the Program.

The remainder of the paper is organized as follows. In section 2, we describe the Program and the institutional context in which it was implemented. In section 3, we introduce the main data set of applications, and the complementary data set on the applicants' curricular information and preliminary results. In section 4, we present our empirical approach. In section 5, we evaluate the effectiveness of the Program with respect to the scientific productivity of successful applicants. In section 6, we summarize the major results and discuss their policy implications, and conclude.

2 The Ramon y Cajal Program

The Spanish Government implemented the Ramon y Cajal Program in 2001 to meet the specific needs of the Spanish S&T system. At the time it was created, the low level of R&D investment and the scarcity of researchers were considered two of the central problems in the Spanish S&T system. In 2001, the share of gross domestic expenditures in R&D relative to GDP in Spain was 0.91 percent, which contrasts with the averages of 1.76 percent in the EU-27 and 2.27 percent in the OECD. In that same year, the number of researchers as a share of total employment in Spain amounted to 4.7 per thousand, below the EU-27 average of 5.3 and the OECD average of 6.8 per thousand (see OECD, 2007). However, R&D personnel in Spain had been increasing rapidly: Six years before, in 1995, the share of R&D personnel in total employment was only 3.5 per thousand.[2] As there was no corresponding increase in R&D funding, the growth in research personnel was primarily achieved through the creation of precarious jobs.

To understand the emergence of this trend, we need to consider the 1970s. At that time, the training of new PhDs through doctoral programs was a political objective that

[2] According to Cruz-Castro and Sanz-Menéndez (2005), part of this increase might be due to statistical adjustment. Since 2000, doctoral and post-doctoral personnel with fellowships (but not contracts) are counted as researchers.

became governmental policy (Fernández-Esquinas, 2002). Since the mid-1980s, there was a steady increase in the availability of four-year public grants to fund doctoral studies (see Sanz-Menéndez, 1997). As a consequence of this policy, in 2000, there were more than 60,000 PhD students in Spain, and approximately 6,000 students received their PhD every year, these figures being three times those at the beginning of the 1980s (see Cruz-Castro and Sanz-Menéndez, 2005).

However, this increase in the supply of PhDs was not accompanied by a similar increase in job positions for researchers. At the end of the 1990s, access to a permanent research position or a promotion became more difficult than had been previously. The labor market for graduate students and experienced PhDs lacked a career path. Fellowships (typically tied to project funding) became the regular labor relationship. Therefore, the Spanish labor market for researchers in the late 1990s was characterized by both a very high proportion of temporary jobs and a low expenditure level per researcher (Cruz-Castro and Sanz-Menéndez, 2005). The National R&D and Innovation Plan 2000-2003 was conceived with the intention to improve R&D spending and broad ideas regarding the creation of 2,000 new "five-year contracts for PhDs in PRCs". This latter objective was implemented in 2001 through a new policy instrument, the Ramon y Cajal Program.

The Program was designed to undertake two objectives. First, to ameliorate the working conditions and long-term employment prospects of a sizeable stock of postdoctoral researchers, within the S&T system, lacking well-defined career paths. Second, to attract numerous Spanish PhD graduates with high-quality scientific records who, at that time, were working abroad. The Program provided postdoctoral researchers with a point of entry to the Spanish S&T system in the form of a five-year contract that mimics a tenure-track position.

The Program was also intended to provide the PRCs with incentives to align their strategic priorities with their human resources practices. This was implemented by establishing a financial co-responsibility scheme between the PRCs and the Government. This scheme discouraged the PRCs from training and increasing the stock of researchers without a well-defined career path who were seeking the opportunity to hold positions within the system (Cruz-Castro and Sanz-Menéndez, 2005). Another concern, held both by legislators and in the public arena, was the need to eliminate the favoritism that prevailed in Spanish PRCs (see Bosch, 2001b). This concern was crucial in determining the selection procedures accomplished in the Program.

When the Ramon y Cajal Program was launched, there was a pervasive perception, spread by the mass media, of a significant brain drain on the Spanish S&T system and a belief that many Spanish PhD graduates working abroad could be enticed to return if they were provided with improved career opportunities.

The first call for applications by the Ramon y Cajal Program, in 2001, attracted approximately $2,800$ applicants and offered 800 contracts, with a total annual expenditure of 35 million euros. The recipients would receive a five-year contract with an annual wage of nearly $29,000$ euros, similar to the wage of a newly tenured professor ("profesor titular") in Spanish PRCs. In the second and third calls, in 2002 and 2003, 500 and 700 contracts, were offered, respectively, attracting more than $2,500$ applications in each call.[3] The number of contracts offered decreased substantially after 2003.

To ensure a transparent selection process and prevent the possibility of favoritism, the selection procedure was centralized in an evaluation agency, the "Agencia Nacional de Evaluación y Prospectiva" (ANEP). This procedure, centralized and external to the PRCs, was a novel feature of the public policy's design. Its success relied on the involvement of the PRCs, which agreed to be excluded from the selection process while co-financing the hiring of the selected researchers. Despite its transparency, candidate eligibility was subject to the endorsement of the PRCs in the first two calls: applicants were required to obtain an endorsement letter from at least one PRC, which committed the PRC to hire her if selected. This feature attracted international attention (see Bosch, 2001b), and Anna Birulés, then the Minister of Science and Technology, reported that some PRCs were jeopardizing the Program by only endorsing local candidates. This endorsement requirement, the effects of which were analyzed by Alonso-Borrego et al. (2013), became optional in the 2003 call,

[3] See Alonso-Borrego et al. (2013). It must be noted that applicants could apply to several areas, and hence the number of applications was even higher. However, most applicants, especially in the first call, only submitted a single application.

and was completely eliminated in the 2004 call.

The ANEP appraised all eligible researchers through a peer-review process, primarily based on the candidates' scientific records but also on their potential[4]. For this purpose, the evaluation agency formed 24 evaluation committees comprising national and international experts, one for each research field. The allocation of contracts among research areas was determined on technical grounds while accounting for the priorities established by the National R&D and Innovation Plan, as well as the demands of the different PRCs and the relative quality of researchers according to international standards (see Cruz-Castro and Sanz-Menéndez, 2005).

Finally, four years after each call, the performance of each program recipient during the benefit period was evaluated, on the basis of her scientific contributions generated until then. A positive evaluation implied the possibility of receiving a new contract that facilitated her access to a tenured contract at the PRC.

3 Preliminary evidence on the Program

The main data set, provided by the Dirección General de Investigación of the Spanish Ministry of Education, records all applications during the first seven calls of the Program, from 2001 to 2007. We excluded observations with missing values for individual characteristics, which represent less than one percent of all observations. Information on each applicant includes her research area, the institution and year in which she earned her PhD, her country of residence and nationality, and the score received in the assessment process and whether she was granted a contract. Given the particular characteristics of the first call, in 2001, which might reduce the comparability of its applications with those of subsequent calls (Alonso-Borrego et al., 2013), we concentrate on the second and the third calls, in 2002 and 2003, for which we are able to gather curricular information several years after the call. The elegibility conditions, which were similar in 2002 and 2003, required applicants to have a PhD and a minimum of an 18-months research stay, at a research center other

[4] In the first three calls (2001-2003), the curricular merits of the candidate accounted for 70 percent of the assessment, while the candidate's submitted scientific proposal and other merits (such as research stays in outstanding centers and letters of reference) each accounted for 15% percent.

than that from which the applicant's college degree was obtained.[5]

In Table 1, we provide the distribution of applications and contracts for these 2 years, broken down by gender, PhD tenure, and research area. We have aggregated the 24 areas designated by the ANEP into 10 broader areas.[6] The first six areas correspond to experimental disciplines (Physics, Earth Sciences, Chemistry, Agriculture, Biomedical and Engineering), followed by Mathematics, Economics, Social Sciences & Law, and Arts & Humanities.

Experimental sciences account for more than 80 percent of applications and nearly 90 percent of contracts. The reason for this allocation of contracts across research areas is the R&D priorities that policy-makers established in the National R&D and Innovation Plan.

We observe that applications are dominated by men. The gender distribution across research areas (not reported here) is highly unequal. Physics and Engineering are strongly dominated by men, amounting to 80 percent of applicants. In Chemistry, men represent approximately 60 percent of applicants. However, Social Sciences and Biomedical Sciences are balanced in terms of gender. With respect to the time elapsed since receiving their PhD, the majority of applicants earned their Ph.D. within 3 to 6 years before the call. Additionally, the success rate is higher for men than for women.

The curricular information has been collected from a complementary data source, the free online resource Publish or Perish (Harzing, 2007). Publish or Perish retrieves academic contributions by author using the Google Scholar database, which provides the title, source, year and authors of the contribution. Google Scholar is generally praised for its speed (Bosman et al., 2006) and high correlation with alternative bibliometric sources (see Harzing, 2012, and Harzing and van der Wal, 2011, for a comparison of citation analyses using different data sources). Whenever the contribution was published in a scientific journal, the journal information is also reported. For each applicant, we measure her number of distinct contributions and, among these, the number of published papers.

[5] Such conditions were changed in 2004, requiring applicants to having earned their PhD in the last 10 years, and a minimum 2-years postdoctoral stay in a research center other than that from which the Ph.D. was obtained.

[6] The correspondence with the areas designated by the ANEP is reported in Table A1 in the Data Appendix.

To assess the quality of each contribution, we use the Journal of Citation Reports (JCR), which provides the impact factors of the international journals listed in its database. A journal's impact factor is calculated on the basis of the average number of citations received by the contributions published in that journal. We use the JCR impact factors in 2006 to measure both the quality of each candidate, and the quality of the center from which each candidate earned her PhD, defined as the average number of citations to all the contributions published in JCR journals by all researchers affiliated with the center. We consider the journal impact factors from a single year to guarantee the comparability between contributions published in different years. The curricular information is updated until 2007.

We use three measures of the scientific quality of each applicant: her number of contributions listed in the JCR database, the average impact factor of her JCR publications, and the maximum impact factor among the JCR journals in which she has published. The two impact factor measures are based on the corresponding impact of the journal in which she published each contribution.

In Table 2 we summarize the curricular information of applicants by contract status. Furthermore, we break down the sample by applicants' characteristics: gender, research area, and time elapsed since PhD receipt.[7] For all categories considered, we observe that, at the time of the call, contract recipients have, on average, more published contributions and a greater scientific impact (either average or maximum impact) than non recipients. Nevertheless, given the high standard deviations, most differences are not significant. We also find that the three measures of scientific quality differ substantially by area, reflecting differences in the typical number of papers and citations across research fields, and therefore the impact indices are not comparable across areas.

To provide a measure of scientific impact that permits comparisons across areas, we constructed the relative rank or position of each researcher with respect to the empirical distribution of impacts of all JCR journals in her research area. We computed the quan-

[7] The estimation results in Alonso-Borrego et al. (2013) indicate that applicant's curriculum, measured by the average impact factor of her contributions, the scientific quality of the center from which she earned her PhD, and her PhD tenure, increase the probability of receiving a contract.

tile (between 1 and 99) of the corresponding empirical distribution of the journal impact achieved by each researcher given her average and her maximum impact factor, which we have denoted "Rank average IF" and "Rank maximum IF", respectively. For instance, a researcher who published a paper in the journal with the highest impact in her area would have a Rank maximum IF of 99.

In Table 3, we report the sample median, 75th and 90th quantiles of the curricular information and the two aforementioned ranks, by contract status, and broken down by research areas. Regarding our measures of the applicants' scientific quality (number of papers, average impact factor and maximum impact factor), the empirical distributions differ substantially depending on the contract status. Specifically, those receiving a contract tend to have a larger number of papers and a greater impact than those without a contract. This evidence agrees with the criteria established by the committees in the selection of applicants, which emphasize their scientific quality.

However, we observe that for contract recipients, the median number of papers and the impact at the time of application is zero in most areas. We believe that selection committees were taking other quality features into consideration at the moment of application that are unobserved in our dataset. Specifically, we have measured each applicant's scientific merits using her contributions published in JCR journals up to the year of application. However, unlike the selection committee, we cannot observe papers under revision, forthcoming papers (not yet published in the year of the call) and, to a lesser extent, the quality of the candidate's research agenda, among others. For most areas, we observe strong differences by contract status for the highest quartile of the distribution. In terms of both the average and maximum impact, the highest quartile of researchers with a contract achieve high positions within the impact distribution in their corresponding area, approaching ranks above 85 in all experimental disciplines except Engineering. These disciplines amount to 75 percent of all contracts. In Economics, the highest quartile of researchers rank above 70. The pattern is substantially different in Mathematics, Social Sciences & Law, and, particularly, Arts & Humanities. In this latter discipline, the impact of the highest decile is relatively low.

For researchers without a contract, we observe that, in experimental disciplines except Engineering, the highest decile of researchers without a contract rank at least 85 in experimental disciplines. This evidence suggests that we can identify a sufficient number of researchers without contracts that are comparable, in terms of our curricular measures, with researchers with contracts. In the case of Economics, the highest decile of researchers without a contract are at least ranked approximately 70. For Engineering and the remaining non-experimental fields (Mathematics, Social Sciences & Law, Arts & Humanities), the highest decile of researchers without a contract rank much lower. This suggest that, in terms of curricular quality, the number of available comparable researchers to contract recipients might be much smaller in these areas. This finding might be evidence that most researchers in certain non-experimental disciplines, such as Arts & Humanities, have yet to adapt to international standards.[8]

The results suggest the quality of the researchers selected and the existence of comparable candidates in most of the research areas considered. However, as reported in Table 3, the areas of Social Sciences & Law and, particularly, Arts & Humanities and Engineering, exhibit a shortage of high-quality candidates, as reflected by the smaller number of papers and the low impact factor in the highest quartile of applicants, regardless of the outcome of the application process.

4 Empirical approach

We consider the scientific output of applicants in the four years after the call to assess the impact of contract status on the ex-post performance of researchers. Given the data constraints, we consider the time horizon selected to be sufficient to test the potential influence of the contract. Moreover, it is consistent with the usual time span the PRCs take to make tenure decisions, and with the maximum length of time required for scientific

[8] Jimenez-Contreras et al (2003) analyze the impact of the national evaluation of researchers' activity in Spain on their research output. They find that, "preference will be given to those articles which are published in journals of recognized prestige that is to say, those journals which occupy a notable position in the lists, organized by scientific field, which appear in the Subject Category Listing of the Journal Citation Reports of the Science Citation Index (Institute of Scientific Information, Philadelphia, PA, USA)". These criteria are applied to all the research areas except Arts & Humanities and Law, "which use alternative criteria".

contributions to undergo a peer-reviewed publication process.

Our relevant policy variable is a binary variable D_i indicating whether the individual was granted a Ramón y Cajal contract, with value 1 if the researcher i has been awarded a contract and 0 otherwise. Our concern is whether the contract status affects the researcher's productivity outcome Y_i in the four-year period after the call. We perform the analysis using three alternative outcome variables that measure researchers' scientific performance. These variables are the number of contributions published in journals listed in the JCR, the average impact of such contributions, and the maximum impact factor among the JCR journals in which a researcher has published.

The ideal evaluation problem, for researcher i, consists in comparing two potential outcomes depending on whether she received a contract or did not, denoted as Y_{1i} and Y_{0i}, respectively. If both counterfactual outcomes were observed, the treatment effect, i.e., the impact of the contract for researcher i, would simply be $(Y_{1i} - Y_{0i})$. Using $E(.)$ to denote the mean operator, we could then calculate the average impact of the contract on the population, or the average treatment effect (ATE), as $E(Y_{1i} - Y_{0i})$ (see Rosenbaum and Rubin, 1983). It may also be interesting to analyze the average impact of the contract for the subpopulation of contract recipients, $E(Y_{1i} - Y_{0i} | D_i = 1)$, termed the average treatment effect on the treated (ATT), and for non-recipients, $E(Y_{1i} - Y_{0i} | D_i = 0)$, termed the average treatment effect on the controls (ATC).

As receiving and not receiving a contract are mutually exclusive, for each researcher we only observe either $D_i = 1$ or $D_i = 0$, and therefore we only observe her outcome under one of the two situations, i.e.,

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) D_i. \tag{1}$$

If the contract status were purely random, and thus independent of the potential outcomes, then the three evaluation measures would be equivalent,

$$E(Y_{1i} - Y_{0i}) = E(Y_{1i} - Y_{0i} | D_i = 1) = E(Y_{1i} - Y_{0i} | D_i = 0) \tag{2}$$

and hence, the average effect of the contract would simply be $E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 0)$. In this case, a naive mean-difference estimator based on the sample means of the observed outcomes for recipients and non-recipients would consistently estimate the causal effect of the contract. This result also holds under the weaker assumption of mean-independence between the contract and the potential outcomes.

However, we know that contract status depends on researchers' characteristics, and therefore researchers' potential outcomes Y_{1i}, Y_{0i} are not independent of the contract status D_i. To see this, notice that the observed difference in outcomes between recipients and non-recipients can be written as

$$E\left(Y_{1i}|\, D_i = 1\right) - E\left(Y_{0i}|\, D_i = 0\right) = \text{ATT} \; + \left[E\left(Y_{0i}|\, D_i = 1\right) - E\left(Y_{0i}|\, D_i = 0\right)\right] \qquad (3)$$

where the second term on the right-hand side of the equation measures the potential selection bias arising because recipients and non-recipients could perform differently even in the absence of the contract. Under the (weaker) assumption that Y_{0i} is mean-independent of D_i (without imposing restrictions regarding Y_{1i} and D_i), the selection bias would be zero. Hence, a naive mean-differences estimator would yield a consistent estimate of the ATT. However, it is unlikely that the mean-independence assumption will be supported if contract status depends on the scientific quality of researchers, as observed at the time of application. Presumably, a naive mean-difference estimator is expected to exacerbate the positive impact of the contract, as contract recipients would likely be more productive than non recipients even in the absence of the contract.

As potential outcomes are not independent of contract status, identification requires the availability of individual pre-contract information and assumptions regarding the relationship between contract status and potential outcomes, conditional on such additional information. The main idea is that contract status is purely random for individuals with similar pre-contract information. We consider two alternative approaches: parametric (regression) and non-parametric (matching). We first describe the problem in the regression framework to illustrate how we can circumvent selection bias by exploiting additional information.

For each researcher i, we can write her two potential outcomes as $Y_{ji} = \mu_j + v_{ji}$, where $E\left(Y_{ji}\right) = \mu_j$ and v_{ji} captures the unobserved individual deviation of the potential outcome

j with respect to the mean population value μ_j, with $E(v_{ji}) = 0$ for $j = 0, 1$. The ATE is given by $\mu_1 - \mu_0$. We can write the observed outcome for researcher i as a simple linear projection on the contract status,

$$Y_i = \alpha + \rho D_i + u_i \tag{4}$$

where the slope $\rho = \mu_1 - \mu_0$ is the ATE and the error term is $u_i = v_{0i} + (v_{1i} - v_{0i}) D_i$. The simple OLS estimator of ρ in (4) yields the aforementioned naive mean-difference estimator based on the sample means of the observed outcome for recipients and non-recipients. This estimator would consistently estimate the ATE provided that the error term is mean-independent of contract status, i.e., $E(u_i | D_i) = 0$. However, we should expect $E(u_i | D_i = 1) \neq E(u_i | D_i = 0)$, as recipients are expected to be more productive than non-recipients. There is a selection bias, and therefore $E(u_i | D_i) \neq 0$.

Nevertheless, if we observed additional variables that contain pre-contract information, we could exploit them to circumvent selection bias. Let $\mathbf{X}_i$ be a vector of additional covariates, including the researcher's curricular information and other relevant variables at the time of application. Consider the assumption that, conditional on the covariates included in $\mathbf{X}_i$, the potential outcomes are mean-independent of D_i, i.e.,

$$E(Y_{ji} | D_i, \mathbf{X}_i) = E(Y_{ji} | \mathbf{X}_i) \qquad j = 0, 1. \tag{5}$$

This conditional mean-independence assumption is also called "selection on observables", as it states that $\mathbf{X}_i$ determines contract status. We also need to assume parametric specifications for $E(v_{1i} | \mathbf{X}_i)$ and $E(v_{0i} | \mathbf{X}_i)$, which are typically assumed to be linear in parameters. Under such assumptions, we can write the augmented model as

$$Y_i = \alpha + \rho D_i + \boldsymbol{\beta}' \mathbf{X}_i + \boldsymbol{\gamma}' D_i \mathbf{X}_i + u_i. \tag{6}$$

where now $E(u_i | D_i, \mathbf{X}_i) = 0$. It can be seen that the causal effect for researcher i is equal to $\rho + \boldsymbol{\gamma}' \mathbf{X}_i$, meaning that it varies with the values of the conditioning variables. To calculate the ATE, we must evaluate this expression at $E(\mathbf{X}_i)$, while to calculate the ATT and the ATC, it must be evaluated at $E(\mathbf{X}_i | D_i = 1)$ and $E(\mathbf{X}_i | D_i = 0)$, respectively (see Wooldridge, 2002). Under the above assumptions, OLS estimation of (6) will yield a consistent estimate of the impact of the contract.[9]

As an alternative to regression analysis, we can follow a non-parametric approach and produce matching estimators of the impact of the contract, using the individual pre-contract information mentioned above. If, for individuals with similar pre-contract information, contract status can be considered as purely random, we can estimate the corresponding counterfactual outcome for each contract recipient using the average outcome for non-recipients with similar pre-contract information. Matching estimators rely on a stronger version of the selection on observables assumption, by which, conditional on $\mathbf{X}_i$, treatment status is independent of potential outcomes (see Rosenbaum and Rubin, 1983). However, unlike regression analysis, matching estimators do not require parametric assumptions. The conditional independence assumption allows us to analyze our observational data as if they came from a randomized experiment.

Following Abadie and Imbens (2011), we implement bias-corrected matching estimators, which employ a regression adjustment to circumvent the finite-sample bias that arises when the matching is not exact. The matches are directly based on the same curricular covariates used in the regression. There are two reasons that we do not employ a propensity score approach to match treatment and comparison observations. First, given the small number of covariates, we do not have a serious dimensionality problem. Second, and more important, propensity score matching is based on first-step estimates of the unknown propensity score. It is difficult to derive the asymptotic variance of the matching estimator when estimated propensity scores, instead of (unknown) actual propensity scores, are used (see Abadie and Imbens, 2009). Moreover, the standard bootstrap variance employed in empirical work is not appropriate (Abadie and Imbens, 2008). Abadie and Imbens (2011)

[9] If we also assume that the average gain in productivity from receiving the contract has zero conditional mean, i.e., $E\left(v_{1i} - v_{0i} \middle| \mathbf{X}_i\right) = 0$, then $E\left(v_{1i} \middle| \mathbf{X}_i\right) = E\left(v_0 \middle| \mathbf{X}_i\right)$, and hence, by conditioning on the $\mathbf{X}_i$ we obtain the model

$$Y_i = \alpha + \rho D_i + \boldsymbol{\beta}' \mathbf{X}_i + u_i.$$

This specification establishes that, conditional on $\mathbf{X}_i$, the causal effect of the contract is the same for any applicant and equal to ρ. In particular, the causal effect of the contract for the entire population of applicants (ATE) coincides with the causal effect of the contract for recipients (ATT) and the causal effect for non-recipients (ATC).

derive the asymptotic variance for the bias-corrected matching estimator that we use, which is implemented in a Stata routine that is fully documented in Abadie, Drukker, Herr and Imbens (2003).

5 The performance of Ramón y Cajal researchers

We estimate the causal effects of the contract, conditional on researcher characteristics at the time of application, to overcome the selection bias due to the endogeneity of contract status. The validity of the conditional estimates of the causal effects relies on the absence of unobserved differences across researchers associated with contract status that affect their potential outcomes. The covariates we consider in the empirical analysis are related to the researcher's curricular information at the time of application, the time elapsed since PhD receipt and her research area. We consider alternative parametric (regression) and non-parametric (matching) procedures using this set of conditioning variables. Regarding the researcher's curricular information, we use the number of JCR papers and the average impact factor. In the case of regression, we use a second-order polynomial on these two variables, while in the case of matching we simply consider these two variables and the cross-product between them. In both cases, we compute three measures of the causal effects. First, the ATE, which measures the average effect of the contract irrespective of contract status. Second, we estimate the ATT, which measures the average effect of the contract for those researchers who actually had a contract. Finally, we consider the ATC, which measures the average effect of the contract for unsuccessful applicants. For the sake of comparison, we also calculate the naive unconditional estimates of the impact of the contract in (4), which presumably tend to overestimate the causal effect of the contract on scientific productivity.

The main outcomes that we consider to measure scientific productivity in the four-year period after the application are the number of JCR papers published by the researcher and the scientific impact measure of her JCR contributions during that period. As impact measures, we use the aforementioned average and maximum impact factors. For each outcome variable, we produce full sample estimates (for all researchers) and separate estimates by research area.

Given the differences in the impact levels that affect the comparability of impact measures across areas, we produced normalized outcome variables of the average impact and the maximum impact, the units of measurement of which are comparable across areas. These outcome variables measure the rank or relative position that each researcher would achieve within her research area with respect to the empirical distribution of impact of the journals in that area. The rank provides the percentile that the researcher would reach within this empirical impact distribution. The ranks have been calculated for our two impact measures, average and maximum impact.

In Table 4, we report the estimates of the impact of the contract on each outcome measure. The naive estimates are both positive and significant for every outcome considered. These estimates may suffer from a positive selection bias, which is confirmed by the conditional estimates (both regression and matching) of the causal effects, which employ the researchers' pre-contract characteristics as covariates.

The conditional estimates are also positive, but their magnitudes are much lower. Further, the magnitudes of the regression and matching estimates are very similar. For all of the quantity (the number of JCR papers) and the quality (impact) measures, the ATE is positive and significant. These results indicate that the contract would have, on average, a positive effect on both the quantity and quality of the scientific production of a randomly chosen applicant. However, when we distinguish by actual contract status, we observe remarkable differences.

In the case of contract recipients, the causal effect of the contract is positive, but not significant, on the number of papers, but positive and significant on any impact measure (average impact, maximum impact, and the two corresponding rank impact measures). These results are similar for the regression and matching estimations. Consequently, contract status does not imply a significant increase in the number of published contributions, but affects the scientific influence of the recipients. Using either of the two rank measures, we find that, on average, the receipt of a contract causes recipients to shift 3 percentiles upwards within the impact distribution. In Table 5, we have calculated the causal effects (in percentages) of the contract on each outcome variable, using the sample average of the corresponding outcome at the time of application. Using these reference values, the causal effect on quality for contract recipients is approximately 17 and 10 percent if we consider the average impact and the maximum impact, respectively.

Our finding of significant positive effects of the contract on the impact of the researcher's contribution, suggests that the Ramon y Cajal contract has a persistent effect on the academic career of the selected researchers. Under the hypothesis of cumulative advantage, or state-dependence, the distribution of productivity among scientists (measured by publications and citations) will exhibit increasing inequality as a cohort of scientists passes through its career. Allison and Stewart (1974), using counts of publications and citations, and Weiss and Lillard (1982), using counts of publications only, find empirical support of this hypothesis. The success associated with the productivity of scientists in their early years, measured by publications and impact engenders early recognition from peers, leveraging opportunities in their subsequent scientific career, in different instances. First, for a given intrinsic merit, further research might receive differential recognition (and, therefore, different publication and citation prospects) if the researchers are unequal in prestige (Merton, 1968). Second, higher reputation in the early years will ease access to research funding, thus increasing the chances to undertake independent research.

In the case of non-recipients, the evidence is mixed. Although the estimated ATC is positive for every outcome, the significance of these estimates depends on the estimation method. The regression estimates exhibit positive and significant effects for quantity, measured by the number of papers, but only for one quality measure: the rank based on the average impact factor. However, in the case of the matching estimates, the effect on quantity is not significant while the effects on the quality measures are positive and significant. Moreover, the ATC estimates are generally smaller than the ATT estimates, but because the levels of the curricular variables at the time of application are also smaller, the relative differences between ATT and ATC are low. If we consider the regression estimates that are significant, the causal effect of the contract on quantity for non-recipients would be approximately 16 percent, and having a contract would shift non-recipients 2 percentiles upward in the average impact distribution.

In Tables 6 to 8 we have reported the estimates of the causal effects, by research area, on the number of papers and on the average and the maximum impact. When we disaggregate by areas, differences appear. For the experimental disciplines, the results generally resemble the full sample estimates in Table 4, but the precision of the estimates decreases. We must recall that experimental sciences account for 80 percent of the contracts, and hence the full sample results are primarily driven by these disciplines. Further, the number of observations in non-experimental disciplines is much lower than in experimental disciplines, as shown by Table 1. Specifically, in Mathematics, Economics and Social Sciences & Law, the number of contract recipients is 31 or lower in each area. This may affect the precision and the sensitivity of our results to the estimation method or the existence of outliers or influential observations.

Regarding the causal effect on quantity, Table 6 reports naive estimates that are generally positive and significant, except for Arts & Humanities, which exhibits a negative and significant coefficient. When we consider the conditional estimates, the causal effects are generally positive, but much smaller than the naive estimates, and mostly non significant, again with exception of the negative effect for Arts & Humanities. We also observe substantial differences between the regression and matching estimates with respect to the magnitudes of the estimated coefficients. The ATT is significant and positive for Chemistry and Mathematics, and the ATC is significant and positive for Earth Sciences, Chemistry and Engineering, but only for one of the estimation methods (either regression or matching).

Regarding the causal effect of the contract on the researchers' average scholarly impact, reported in Table 7, we observe that the naive estimates are significantly positive in most areas. The conditional estimates are generally positive, but smaller in magnitude and less precise than the naive estimates. The regression and matching estimators yield very similar results. In the case of experimental disciplines, the ATT estimates are positive and significant for three areas: Physics, Biomedical and Engineering. The causal effect on researchers with a contract exceeds 40 percent in Engineering and Physics. In this latter case, the effect seems remarkable, as the average impact factor for researchers at the time of application was approximately 2.1. Chemistry is the only area in which the ATC is significant. This area received a substantial number of contracts (more than 200), and the result suggests that we could have expected gains in scientific productivity if additional non-recipients had benefitted from a contract. With some differences in magnitude and significance, the estimates of the causal effects on the maximum impact, presented in Table 8, align with the results for the average impact. We do not find significant effects in Biomedical but obtain significantly positive effects in Earth Sciences, for both recipients and non-recipients.

The results in Arts & Humanities are the opposite of the full sample results and the findings for most of the areas. We find that the causal effect of a contract is significantly negative on both quantity and quality, which, as mentioned above, might be related to the fact that researchers in certain non-experimental areas have yet to adopt international standards.

We have also considered alternative estimates, using alternative conditioning sets and a longer time span after the application, to evaluate the sensitivity of the results. Specifically, we have considered additional conditioning variables, such as the gender and the score that the assessment committee assigned to the researcher at the time of application, as well as the inclusion of the maximum IF among the curricular measures and different degrees of the polynomial in the curricular measures. The results (not reported here but available upon request) can be summarized as follows. First, we find that both the gender and the score were not significant and did not increase the explanatory power of the regression estimates. The estimated causal effects remain similar, yet the precision of the estimates is lower than in our reported results, when either of these two variables were included. The finding that the score is not significant is particularly interesting, as it suggests that there are no other substantial differences between researchers, not captured by our observed curricular measures, which might affect the comparison between them. Moreover, the selection of either the curricular measures or the functional form does not alter the main empirical results. When we considered the longer five-year time span after the application, the qualitative results, particularly those concerning the ATT, are unchanged. However, as we were only able to gather curricular information until 2007, we can only use applicant data from 2002, and hence the precision of the estimates is reduced. Overall, our results appear to be robust to the choice of conditioning variables, functional form, and time span after application.

6 Conclusions

The Ramon y Cajal Program constituted a novel policy measure and has become a relevant S&T instrument in Spain. The Program was designed to improve the working conditions and long-term employment prospects of a sizeable stock of postdoctoral researchers within the S&T system and attract numerous Spanish PhD graduates with high-quality scientific records who were working abroad.

Our paper studies whether a Ramon y Cajal contract affects the subsequent research output of the researchers receiving it. We analyzed the effect of the Program on the productivity of the selected researchers and compared them with scholars with similar curricular characteristics that were not awarded a Ramón y Cajal contract. We employed two alternative approaches to estimate the causal effect of the contract: conditional regression and matching procedures. Overall, the results provided by the two methods are similar. They indicate that the Ramon y Cajal researchers were able to maintain a quality level above that of comparable applicants not selected by the Program.

In particular, our results demonstrate the success of the Program in increasing the scientific impact of young researchers in the Spanish S&T system, in several research areas. This is an important result, which supports policies designed to increase the stock of human resources in scientific research to raise the international impact of the Spanish R&D system. An early higher impact of a researcher contributes to boosting her scientific reputation, but also her future recognition through cumulative advantage. The evidence that early recognition in science disproportionately favors success in a scientific career (Allison and

Stewart, 19674; Weiss and Lillard, 1982) emphasizes the importance of our finding that contract status increases the scientific impact of the researchers.

The Program also had favorable effects on the Spanish S&T system. We have summarized the contributions of previous researchers who established that the Program has earned a solid reputation abroad and increased the chances of PRCs with solid reputations to attract high-quality young researchers. It has also provided short-term improvements in the employment opportunities, working conditions and academic career prospects of PhDs.

However, in recent years, the S&T system has failed to provide employment opportunities for all of the researchers who, in the spirit of a tenure-track position, were evaluated positively at the end of the Ramon y Cajal contract. The first researchers selected finished their contracts in 2007, coinciding with the beginning of the Spanish economic crisis. This has limited many of these researchers' prospects within the Spanish S&T system, jeopardizing the Program's achievements in previous years.[10]

[10] This has generated the perception that Spain has shifted from aiming at "premier league" status (Schiermeier, 2004) to "scientific suicide" (Moro-Martin, 2012).

References

[1] Abadie, A., D. Drukker, J.L. Herr, and G. Imbens (2004), "Implementing Matching Estimators for Average Treatment Effects in Stata". The Stata Journal 4(3), 290-311.

[2] Abadie, A., and G.W. Imbens (2008), "On the Failure of the Bootstrap for Matching Estimators". Econometrica 76(6), 1537-1557. DOI: 10.3982/ECTA6474.

[3] Abadie, A., and G.W. Imbens (2009), "Matching on the Estimated Propensity Score". National Bureau of Economic Research, Working Paper 15301, August 2009.

[4] Abadie, A. and G. Imbens (2011), "Bias-Corrected Matching Estimators for Average Treatment Effects". Journal of Business & Economic Statistics 29(1), 1-11. DOI: 10.1198/jbes.2009.07333.

[5] Allison, P.D. and J.A. Stewart (1974), "Productivity Differences among Scientists: Evidence for Accumulative Advantage". American Sociological Review 39, 596-606.

[6] Alonso-Borrego, C., Romero-Medina, A. Sánchez-Mangas, and Triossi, M. (2013), "Boosting Scientific Research: Evidence from a Public Program". Revista de Economía Aplicada 21(2), 115-128, Autumn.

[7] Arora, A., P. David and A. Gambardella (1998), "Reputation and competence in publicly funded science: Estimating the effects of research group productivity". Annales d'Économie et de Statistique, 49/50, 163-198.

[8] Arora, A. and A. Gambardella (2005), "The impact of NSF support for basic research in Economics". Annales d'Économie et de Statistique, 79/80, 91-117.

[9] Azoulay, P., J.G. Graff Zivin and G. Manso (2011), "Incentives and creativity: Evidence from the academic life science", The Rand Journal of Economics, 42, 527-554. DOI: 10.1111/j.1756-2171.2011.00140.x.

[10] Benavente, J.M., G. Crespi, L. Figal Garone and A. Maffioli (2012), "The impact of national research funds: a regression discontinuity approach to the Chilean FONDECYT". Research Policy, 41(8), 1461-1475. DOI: 10.1016/j.respol.2012.04.007.

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

[12] Bosch, X. (2001a), "Tenure-track plan aims to end university inbreeding". Nature 410, 1014 (26 April 2001). DOI: 10.1038/35074267.

[13] Bosch, X. (2001b), "Spain's postdoc plan takes off". Nature 413, 556 (11 October 2001). DOI: 10.1038/35098219.

[14] Bosman, J, Mourik, I. van, Rasch, M., Sieverts, E. and Verhoeff, H. (2006), "Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar". Utrecht: Utrecht University Library, http://igitur-archive.library.uu.nl/DARLIN/2006-1220-200432/Scopus doorgelicht & vergeleken - translated.pdf.

[15] Carayol, N. and M. Lanoe (2013), "The impact of blue skye project based funding of academic research". Université de Bordeaux, mimeo.

[16] Chudnovsky, D., A. López, M.A. Rossi and D. Ubfal (2008), "Money for Science? The impact of research grants on academic output". Fiscal Studies, 29(1), 75-87. DOI: 10.1111/j.1475-5890.2008.00069.x.

[17] Cruz-Castro, L., and L. Sanz-Menéndez (2005), "Bringing science and technology human resources back in: the Spanish Ramón y Cajal programme". Science and Public Policy 32(1), 39-53. DOI: 10.3152/147154305781779687.

[18] DGI - Direccion General de Investigacion y Gestión del Plan Nacional de I+D+i (2010), "10 años del Programa Ramón y Cajal" Secretaría de Estado de Investigación. Ministerio de Ciencia e Innovación.

http://www.idi.mineco.gob.es/stfls/MICINN/Investigacion/FICHEROS/FicherosNotasPrensa/Res
Visited July 10th, 2014.

[19] Fernández Esquinas, M. (2002), "La formación de los investigadores científicos en España", Madrid: Centro de Investigaciones Sociológicas-Siglo XXI Editores.

[20] Eisenberg, T. and M. T. Wells (2000), "Inbreeding in Law School Hiring: Assessing the Performance of Faculty Hired From Within". Journal of Legal Studies, 29(1), 369–388. DOI: 10.1086/468077.

[21] Harzing, A.W. (2013), "A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners". Scientometrics 93(3), 1057-1075. DOI: 10.1007/s11192-012-0777-7.

[22] Harzing, A.W. and van der Wal, R. (2011), "Google Scholar as a new source of citation analysis". Ethics in Science and Environmental Politics 8(1), 63-71. DOI:10.3354/esep00076.

[23] Jacob, B.A. and L. Lefgren (2011), "The impact of research grant funding on scientific productivity". Journal of Public Economics, 95, 1168-1177. DOI:10.1016/j.jpubeco.2011.05.005.

[24] Jiménez-Contreras, E., F. de Moya Anegón and E. Delgado López-Cózar (2003), "The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI)". Research Policy 32(1), 123–142. DOI: 10.1016/S0048-7333(02)00008-2.

[25] Merton, R.K. (1968), "The Matthew Effect in Science". Science 159,56-63. DOI: 10.1126/science.159.3810.56

[26] Moro-Martin, A. (2012), "Spanish changes are scientific suicide". Nature 482, 277 (16 February 2012) DOI:10.1038/482277a.

[27] Muñoz-Pinedo, C., M.A. del Pozo, J.A. Gálvez, E.M. Moreno, J. Buceta-Fernández, M. Garcia-Sanchez, J.M. Valdivielso, and E. Bascones (2003), "Spanish Scientists Working Abroad". Science Vol. 300 no. 5616, 51 (4 April 2003). DOI: 10.1126/science.300.5616.51b.

[28] OECD, Organisation for Economic Co-operation and Development (2007), Main Science and Technology Indicators, Volume 2007 Issue 1 OECD Publication Date : 11 June 2007 DOI: 10.1787/msti-v2007-1-en-fr.

[29] Rosenbaum, P. R. and D. B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects". Biometrika 70(1), 41–55. DOI: 10.1093/biomet/70.1.41.

[30] Sanz-Menéndez, L. (1997), "Estado, ciencia y tecnología en España 1939-1997," Madrid: Alianza Universidad.

[31] Sanz Menéndez, L. (2003), "Coping with researchers' labour market problems through public policy: The Spanish Ramón y Cajal Program", in: Avveduto, S. (Ed.), *Fostering the Development of Human Resources for Science and Technology*, 241-274. Roma: Biblink Editori.

[32] Sanz Menéndez, L., M. J. Jerez, A. Romero-Medina, I. Marqués and A. Martínez (2002), "Una nueva política de recursos humanos en I+D: el Programa Ramón y Cajal". Economía Industrial 343, 149–160.

[33] Schiermeier, Q. (2004), "Spain aims at preimier league". Nature 431, 488-489 (22 September 2004). DOI:10.1038/nj7007-488a.

[34] Soler, M. (2001), "How inbreeding affects productivity in Europe". Nature 411, 132 (10 May 2001). DOI: 10.1038/35075637.

[35] Stephan, P.E. (1996), "The Economics of Science". Journal of Economic Literature 34(3), 1199-1235. DOI: 10.1007/BF02696298.

[36] Weiss, Y. and L. Lillard (1982), "Output Variability, Academic Labor Contracts, and Waiting Times for Promotion". Research in Labor Economics 5, 157-188.

Table 1. Distribution of applications and contracts
Absolute and relative (%) frequencies in each category

	Applications		Contracts	
	4137		1154	
Gender				
Female	1744	43.9	409	36.9
Male	2227	56.1	699	63.1
Ph.D tenure				
Up to 2 years	910	23.0	187	17.1
3-6 years	1813	45.8	543	49.6
More than 6 years	1233	31.2	364	33.3
Research Area				
Physics	387	9.4	104	9.0
Earth Sciences & Ecology	502	12.1	126	10.9
Chemistry	633	15.3	213	18.5
Agriculture, Livestock and Fishery	543	13.1	148	12.8
Biomedical Sciences	986	23.8	289	25.0
Engineering & Computing Sciences	311	7.5	134	11.6
Mathematics	147	3.6	31	2.7
Economics	95	2.3	23	2.0
Social Sciences & Law	196	4.7	26	2.3
Arts & Humanities	337	8.2	60	5.2

Data from the 2nd (2002) and 3rd (2003) calls.
We have excluded observations with missing values in any of the variables.

Table 2. Curricular information at time of application by contract status (Yes/no)

	Number of papers		Average Impact factor		Maximum Impact factor	
	Yes	no	Yes	no	Yes	no
All						
	2.2	1.2	2.0	1.0	3.2	1.5
	(4.1)	(2.8)	(3.3)	(1.8)	(6.7)	(3.9)
Gender						
Female	2.4	1.4	2.1	1.1	3.9	1.8
	(4.5)	(3.1)	(3.1)	(2.0)	(7.4)	(4.2)
Male	2.0	1.1	1.7	0.9	2.9	1.4
	(3.9)	(2.5)	(3.4)	(1.9)	(6.1)	(3.7)
Ph.D. tenure						
Up to 2 years	0.9	0.6	1.2	0.6	1.8	0.7
	(2.2)	(1.8)	(3.2)	(1.7)	(5.0)	(2.2)
3-6 years	2.1	1.1	2.1	1.0	3.7	1.5
	(3.6)	(2.4)	(3.7)	(1.9)	(7.2)	(3.7)
More than 6 years	2.8	1.9	1.8	1.3	3.4	2.3
	(5.1)	(3.7)	(2.6)	(2.3)	(6.7)	(5.4)
Research Area						
Physics	1.7	0.8	2.1	0.8	2.6	1.1
	(2.4)	(1.7)	(3.1)	(1.6)	(4.4)	(2.2)
Earth Sciences & Ecology	2.3	1.4	1.6	0.9	2.6	1.6
	(3.6)	(2.8)	(3.4)	(1.4)	(4.8)	(3.8)
Chemistry	2.8	1.5	1.8	1.1	2.7	1.4
	(4.0)	(2.9)	(2.6)	(1.5)	(4.0)	(2.3)
Agriculture, Livestock and Fishery	2.1	1.5	1.6	1.2	2.8	1.8
	(3.1)	(3.0)	(2.4)	(2.0)	(5.5)	(4.0)
Biomedical Sciences	3.3	2.1	3.2	1.8	6.4	3.1
	(5.9)	(3.8)	(4.3)	(2.9)	(9.7)	(6.1)
Engineering & Computing Sciences	0.4	0.2	0.3	0.1	0.4	0.1
	(0.9)	(0.8)	(0.9)	(0.4)	(0.5)	(0.5)
Mathematics	1.2	0.5	0.3	0.2	0.4	0.2
	(2.0)	(1.2)	(0.5)	(0.3)	(0.5)	(0.4)
Economics	2.0	0.5	0.8	0.4	0.9	0.5
	(2.0)	(0.9)	(0.9)	(0.9)	(1.0)	(1.1)
Social Sciences & Law	0.5	0.2	0.5	0.1	0.7	0.1
	(1.2)	(0.7)	(1.2)	(0.4)	(1.5)	(0.4)
Arts & Humanities	0.3	0.1	0.8	0.2	1.9	0.2
	(0.7)	(0.6)	(3.4)	(0.7)	(8.2)	(1.1)

Mean of the outcome variables, and standard deviation in parentheses.

Table 3. Curricular information at time of application by contract status (Yes/no)
Quantiles of the empirical distributions.

		Number of papers		Average Impact factor		Maximum Impact factor		Rank Avg. IF		Rank Max. IF	
		Yes	no	Yes	no	Yes	no	Yes	no	Yes	no
ALL	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	3.0	1.0	2.8	1.4	4.1	1.5	84	57	92	66
	q_{90}	6.0	4.0	5.4	3.5	7.8	4.9	95	89	98	94
Physics	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	2.0	1.0	3.1	1.0	5.1	1.2	88	45	90	49
	q_{90}	5.0	3.0	5.5	3.0	7.1	4.3	95	87	96	89
Earth Sciences	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	4.0	2.0	2.5	2.2	3.5	2.9	85	78	94	88
	q_{90}	8.0	5.0	3.8	3.5	6.7	4.8	95	93	98	97
Chemistry	q_{50}	1.0	0.0	1.1	0.0	1.1	0.0	38	1	38	1
	q_{75}	4.0	2.0	2.8	2.3	4.3	2.8	81	67	86	73
	q_{90}	8.0	5.0	3.8	3.5	7.7	4.5	93	85	96	93
Agriculture	q_{50}	1.0	0.0	1.1	0.0	1.2	0.0	68	1	71	1
	q_{75}	3.0	2.0	2.7	2.1	3.6	2.8	98	94	99	99
	q_{90}	6.0	5.0	4.0	3.6	5.8	5.6	99	99	99	99
Biomedical	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	5.0	3.0	5.3	3.2	8.1	4.2	90	75	95	87
	q_{90}	9.0	7.0	8.7	4.7	26.7	7.5	95	88	99	95
Engineering	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	0.0	0.0	0.0	0.0	0.0	0.0	2	1	1	1
	q_{90}	2.0	1.0	1.5	0.2	2.0	0.2	84	15	92	15
Mathematics	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	2.0	0.0	0.6	0.0	0.8	0.0	44	1	60	1
	q_{90}	4.0	2.0	1.2	0.7	1.2	0.8	80	55	83	58
Economics	q_{50}	1.0	0.0	0.6	0.0	0.0	0.0	38	1	53	1
	q_{75}	4.0	1.0	1.1	0.4	1.6	0.4	69	22	85	22
	q_{90}	5.0	2.0	1.8	1.2	2.4	1.4	88	73	94	79
Social Sciences	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	1.0	0.0	0.4	0.0	0.4	0.0	28	1	29	1
	q_{90}	2.0	1.0	3.0	0.9	3.8	0.9	94	54	97	59
Arts & Hum.	q_{50}	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	1
	q_{75}	0.0	0.0	0.0	0.0	0.0	0.0	4	3	1	1
	q_{90}	1.0	0.0	0.6	0.0	0.7	0.0	63	4	66	1

q_j is the jth percentile of the empirical distribution of the corresponding outcome variable.

Table 4. Full sample estimates of the causal effect of the contract

Outcome	Naive	Regression ATE	Regression ATT	Regression ATC	Matching ATE	Matching ATT	Matching ATC
Number of papers	1.00§	0.21*	0.23	0.20†	0.19*	0.24	0.17
	(0.18)	(0.11)	(0.15)	(0.10)	(0.11)	(0.15)	(0.11)
Average IF	0.71§	0.17†	0.31§	0.12	0.23§	0.33§	0.19†
	(0.11)	(0.07)	(0.09)	(0.07)	(0.07)	(0.09)	(0.08)
Maximum IF	1.20§	0.24*	0.33*	0.20	0.28†	0.32*	0.26†
	(0.19)	(0.14)	(0.18)	(0.12)	(0.13)	(0.17)	(0.13)
Rank Avg. IF	10.59§	2.37†	3.33§	1.96*	2.93†	2.96†	2.92†
	(1.57)	(1.16)	(1.21)	(1.18)	(1.26)	(1.30)	(1.37)
Rank Max. IF	11.26§	2.44†	3.42§	2.02	2.89†	2.82†	2.92†
	(1.68)	(1.21)	(1.27)	(1.23)	(1.31)	(1.36)	(1.42)

*,†,§Significant at 10%, 5% and 1% levels, respectively. Robust standard errors in parentheses.
Full sample estimates include controls for research areas.
ATE is the sample average treatment effect. ATT is average treatment effect on treated (with contract). ATC is the average treatment effect on controls (without contract).
Matching procedure is implemented with replacement, using 4 matches per observation, and the Mahalanobis metric, based on the sample covariance matrix of the covariates, is used to measure the distance among covariate values.

Table 5. Percentage causal effect of the contract
(relative to the outcome at the time of application)

Outcome	Naive	Regression ATE	Regression ATT	Regression ATC	Matching ATE	Matching ATT	Matching ATC
Number of papers	67.5	14.1	11.0	16.1	13.2	11.6	14.0
Average IF	57.1	13.9	16.7	11.4	18.8	17.8	19.0
Maximum IF	58.4	11.6	10.2	12.3	13.6	10.0	16.3
Rank Avg. IF	37.0	8.3	9.0	7.7	10.2	8.0	11.5
Rank Max. IF	36.8	8.0	8.8	7.4	9.5	7.2	10.7

See notes to Table 4.
The reference value is the average of the outcome variable at the time of the application for all the individuals (ATE), for the contract recipients (ATT) and for the non-recipients (ATC), correspondingly.

Table 6. Estimates of the causal effect of the contract by research areas
Outcome variable: Number of papers

		Regression			Matching		
	Naive	ATE	ATT	ATC	ATE	ATT	ATC
Physics							
	$0.64^{\dagger}$	0.09	0.22	0.04	0.10	0.40	−0.02
	(0.30)	(0.22)	(0.32)	(0.20)	(0.21)	(0.31)	(0.19)
Earth Sciences							
	$0.92^{\dagger}$	0.39	0.29	0.42	0.52^{*}	0.49	0.53^{*}
	(0.44)	(0.34)	(0.40)	(0.32)	(0.31)	(0.38)	(0.30)
Chemistry							
	$1.41^{\S}$	0.54^{*}	0.58	$0.52^{\dagger}$	0.42	0.63^{*}	0.32
	(0.39)	(0.28)	(0.38)	(0.24)	(0.27)	(0.37)	(0.24)
Agriculture							
	0.59	0.23	0.16	0.25	0.10	0.30	0.03
	(0.41)	(0.27)	(0.32)	(0.26)	(0.28)	(0.35)	(0.26)
Biomedical							
	$1.12^{\dagger}$	0.13	0.13	0.13	0.12	0.04	0.15
	(0.45)	(0.25)	(0.31)	(0.26)	(0.24)	(0.30)	(0.25)
Engineering							
	$0.58^{\S}$	0.14	0.04	0.22^{*}	0.21	0.27	0.17
	(0.20)	(0.25)	(0.48)	(0.12)	(0.14)	(0.18)	(0.12)
Mathematics							
	0.76	0.91^{*}	$2.82^{\dagger}$	0.40	$0.90^{\dagger}$	3.58	0.17
	(0.62)	(0.50)	(1.20)	(0.46)	(0.37)	(0.58)	(0.34)
Economics							
	$1.45^{\S}$	0.62	0.12	0.78	0.49	0.31	0.55
	(0.45)	(0.53)	(0.60)	(0.58)	(0.51)	(0.50)	(0.56)
Social Sciences							
	0.24	−0.13	−0.29	0.10	−0.08	−0.08	−0.08
	(0.36)	(0.22)	(0.45)	(0.20)	(0.13)	(0.28)	(0.12)
Arts & Hum.							
	-0.18^{*}	$-0.48^{\S}$	$-1.32^{\S}$	$-0.30^{\S}$	$-0.42^{\S}$	$-0.96^{\S}$	$-0.31^{\S}$
	(0.10)	(0.06)	(0.17)	(0.05)	(0.06)	(0.07)	(0.06)

*,†,§Significant at 10%, 5% and 1% levels, respectively. Robust standard errors in parentheses. See Notes to Table 4.

Table 7. Estimates of the causal effect of the contract by research areas
Outcome variable: Average IF

		Regression			Matching		
	Naive	ATE	ATT	ATC	ATE	ATT	ATC
Physics							
	$1.39^{\S}$	0.69^{*}	$1.28^{\S}$	0.47	$0.85^{\dagger}$	$1.26^{\S}$	0.70^{*}
	(0.49)	(0.38)	(0.43)	(0.38)	(0.41)	(0.41)	(0.42)
Earth Sciences							
	0.44	0.09	0.21	0.05	0.13	0.20	0.11
	(0.28)	(0.15)	(0.17)	(0.14)	(0.14)	(0.17)	(0.15)
Chemistry							
	$0.57^{\S}$	$0.27^{\dagger}$	0.19	$0.31^{\S}$	0.23^{*}	0.18	$0.26^{\dagger}$
	(0.15)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)
Agriculture							
	0.35^{*}	0.08	−0.16	0.17	−0.01	−0.09	0.02
	(0.20)	(0.17)	(0.19)	(0.17)	(0.15)	(0.17)	(0.14)
Biomedical							
	$0.91^{\S}$	0.13	0.41^{*}	0.01	0.28	$0.41^{\dagger}$	0.23
	(0.26)	(0.17)	(0.21)	(0.18)	(0.19)	(0.21)	(0.20)
Engineering							
	$0.20^{\dagger}$	0.13^{*}	0.18^{*}	0.09	0.13^{*}	$0.19^{\dagger}$	0.08
	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)	(0.08)	(0.08)
Mathematics							
	0.24^{*}	0.12	0.01	0.15	0.06	0.01	0.08
	(0.13)	(0.14)	(0.15)	(0.15)	(0.14)	(0.11)	(0.15)
Economics							
	0.27^{*}	0.10	−0.02	0.14	0.04	0.03	0.04
	(0.16)	(0.19)	(0.21)	(0.21)	(0.19)	(0.19)	(0.21)
Social Sciences							
	0.46	−0.05	0.23^{*}	−0.09	0.08	0.24	0.05
	(0.33)	(0.12)	(0.14)	(0.12)	(0.07)	(0.22)	(0.06)
Arts & Hum.							
	0.02	$-0.15^{\S}$	−0.10	$-0.16^{\S}$	$-0.16^{\S}$	$-0.15^{\dagger}$	$-0.16^{\S}$
	(0.14)	(0.04)	(0.08)	(0.04)	(0.04)	(0.06)	(0.04)

*, †, §Significant at 10%, 5% and 1% levels, respectively. Robust standard errors in parentheses. See Notes to Table 4.

Table 8. Estimates of the causal effect of the contract by research areas
Outcome variable: Maximum IF

		Regression			Matching		
	Naive	ATE	ATT	ATC	ATE	ATT	ATC
Physics							
	$1.74^{§}$	0.83^{*}	$1.56^{§}$	0.56	$1.30^{\dagger}$	$1.55^{§}$	1.20
	(0.60)	(0.46)	(0.55)	(0.45)	(0.66)	(0.54)	(0.73)
Earth Sciences							
	$1.51^{\dagger}$	$0.95^{\dagger}$	$1.17^{\dagger}$	$0.87^{\dagger}$	$0.81^{\dagger}$	$1.10^{\dagger}$	$0.71^{\dagger}$
	(0.60)	(0.43)	(0.49)	(0.41)	(0.34)	(0.48)	(0.32)
Chemistry							
	$1.06^{§}$	$0.52^{\dagger}$	0.39	$0.59^{\dagger}$	0.35	0.33	0.36
	(0.28)	(0.24)	(0.26)	(0.24)	(0.22)	(0.25)	(0.23)
Agriculture							
	0.60	0.19	−0.24	0.35	−0.12	−0.24	−0.08
	(0.44)	(0.36)	(0.38)	(0.36)	(0.30)	(0.38)	(0.29)
Biomedical							
	$1.48^{\dagger}$	0.18	0.03	0.24	0.13	−0.20	0.27
	(0.47)	(0.37)	(0.49)	(0.36)	(0.34)	(0.50)	(0.34)
Engineering							
	$0.28^{§}$	0.19^{*}	0.25^{*}	0.14	0.19^{*}	$0.27^{§}$	0.13
	(0.10)	(0.10)	(0.13)	(0.09)	(0.10)	(0.10)	(0.10)
Mathematics							
	0.27^{*}	0.12	−0.03	0.16	0.05	0.02	0.06
	(0.15)	(0.16)	(0.25)	(0.16)	(0.16)	(0.14)	(0.17)
Economics							
	0.33	0.10	−0.32	0.23	0.04	−0.08	0.08
	(0.24)	(0.31)	(0.30)	(0.35)	(0.28)	(0.28)	(0.30)
Social Sciences							
	0.33	−0.28	−0.29	−0.28	−0.15	−0.03	−0.17
	(0.43)	(0.23)	(0.47)	(0.20)	(0.13)	(0.29)	(0.12)
Arts & Hum.							
	0.08	$-0.26^{§}$	$-0.28^{\dagger}$	$-0.26^{§}$	$-0.27^{§}$	$-0.31^{§}$	$-0.26^{§}$
	(0.27)	(0.08)	(0.11)	(0.08)	(0.07)	(0.12)	(0.07)

*,†,§Significant at 10%, 5% and 1% levels, respectively. Robust standard errors in parentheses. See Notes to Table 4.

Data Appendix

The initial sample from 2002-2003 is composed of $6,247$ applications, corresponding to $4,023$ researchers. There are more applications than candidates, as researchers can apply multiple times, for two non-exclusive reasons: i) they do not receive a contract in 2002 and decide to apply again in 2003; and, ii) they apply in two or more different research areas in a given year, thus accounting for different applications in that call. The total number of contracts in 2002-2003 is $1,197$.

There are some factors leading to a loss of information on researchers, applications and contracts. First, a small percentage of selected applicants refused to offered contracts (59 researchers out of $4,023$). Second, applicants in 2002 or 2003 who did not receive a contract in these years but obtained one in one of the 4 subsequent years (within the evaluation period) were omitted from the analysis. This affected 327 reseachers. Third, the unit of analysis is the "applicant - research area" pair. The following criterion has been used for applicants submitting several applications. For each researcher, if one of her applications resulted in a contract, we retain that application and discard the rest of them. If none of the applications resulted in a contract and there are several records for the same "applicant - research area" pair, we retain the application that received the highest score. This selection criterion involves discarding applications, but not individuals or contracts. Finally, some researchers have very common surnames. We performed an exhaustive search to guarantee that the contributions retrieved from Google Scholar correspond to the specified researchers and not other researchers. We eliminated 37 individuals for whom we could not guarantee the authorship of all the contributions assigned to them.

These selection criteria led us to eliminate 423 researchers and 43 contracts. The final sample is composed of $3,600$ researchers (89.5 percent of the initial sample) and $1,154$ contracts (96.4 percent of the initial sample). The total number of applications in the final sample is $4,137$.

Table A1
Research areas and its correspondence with ANEP classification

Area	**ANEP classification**
Physics	Physics and Space Sciences (1)
Earth Sciences & Ecology	Earth Sciences (2) Plant and Animal Biology, Ecology (6)
Chemistry	Science and Technology of Materials (3) Chemistry (4) Chemical Technology (5)
Agriculture, Livestock & Fishery	Agriculture (7) Livestock and Fishery (8) Food Science and Technology (9)
Biomedical Sciences	Molecular and Cell Biology and Genetics (10) Physiology and Pharmacology (11) Medicine (12)
Engineering and Computing Science	Mechanical, Ship and Aeronautical Engineering (13) Electric and Electronic Engineering (14) Civil Engineering and Architecture (15) Computing Sciences and Computer Technology (17) Electronic and Communications Technology (18)
Mathematics	Mathematics (16)
Economics	Economics, Finance and Business (19)
Social Sciences and Law	Law (20) Social Sciences (21) Psychology and Education Sciences (22)
Arts & Humanities	Philology and Philosophy (23) History and Art (24)