Office of Evaluation and Oversight

EVALUATING A PROGRAM OF PUBLIC FUNDING OF SCIENTIFIC ACTIVITY. A CASE STUDY OF FONCYT IN ARGENTINA

*Daniel Chudnovsky, Andrés López, Martín Rossi, and Diego Ubfal **

Working Paper: OVE / WP-12 / 06
November, 2006

Electronic version:
http://ove/oveIntranet/DefaultNoCache.aspx?Action=WUCPublications@ImpactEvaluations

Inter-American Development Bank
Washington, D.C.

Office of Evaluation and Oversight, OVE

Evaluating a Program of Public Funding of Scientific Activity. A Case Study of FONCYT in Argentina

Daniel Chudnovsky, Andrés López, Martín Rossi, and Diego Ubfal *

* This study was prepared by Daniel Chudnovsky, Andrés López, Martín Rossi, and Diego Ubfal from Centro de Investigaciones para la Información (CENIT), Buenos Aires, Argentina.

This working paper is one of several background reports to OVE's Science and Technology Thematic Evaluation carried out during the 2005-2006 Ex-Post Evaluation Cycle.

The team of this Thematic Study was comprised of Alessandro Maffioli, Bertha Briceño, and coordinated by Inder J. Ruprah. The information needed to elaborate this study was provided by CAICYT (Centro Argentino de Información Científica y Tecnológica). The findings and interpretations of the authors do not necessarily represent the views of the Inter-American Development Bank. The usual disclaimer applies. Correspondence to: Alessandro Maffioli, e-mail: alessandrom@iadb.org, Office of Evaluation and Oversight, Inter-American Development Bank, Stop B-719, 1300 New York Avenue, NW, Washington, D.C. 20577.

TABLE OF CONTENTS

INTRODUCTION

This report contains an evaluation of the Scientific and Technological Research Projects (PICT) funded by the Scientific and Technological Research Fund (FONCYT). FONCYT is managed by the National Agency of Scientific and Technological Promotion (ANPCYT), an organization that is part of the Argentine federal government. The main function of the FONCYT is to support, through the granting of subsidies, research projects aimed at generating scientific or technological knowledge.

The value of scientific research is widely recognized. Furthermore, since the seminal articles by Nelson (1959) and Arrow (1962) there has been a broad consensus on the need for public funding to compensate for the uncertainty and difficulty in the appropriation of scientific research outputs. However, the analysis of science funding mechanisms is sparse and empirical evidence on the effectiveness of public funding to increase research outputs is scarce and inconclusive.

In order to provide evidence to answer what the impact of public funding on scientific productivity and research quality is, this study applies the program evaluation literature, which has been widely utilized in labor economics and has spread throughout several economic fields, to the economics of science. This task has already been initiated by Arora and Gambardella (1998), Arora et al. (1998) and Goldfarb (2001), but none of them evaluates the impact of a program of grants for scientific research projects in a developing country. Furthermore, this work combines a more copious series of evaluation techniques, which increases the robustness of its results.

In particular, the focus is on whether the Scientific and Technological Research Fund (FONCYT) subsidies have improved the academic performance of supported researchers in Argentina. The study compares the academic performance of researchers with projects that were approved in terms of quality and financed, with that of researchers whose projects were also approved but not funded due to limited availability of resources.

Available data on scientific outputs in terms of quantity (number of publications in international peer-reviewed journals) and quality (impact factor of the journal in which articles were published), for both funded and non-funded researchers, is not necessarily appropriate to gauge the impact on scientific output. Research projects could have possibly had other kind of relevant impacts - for example, in terms of the generation of knowledge that could be patented, the establishment of linkages with other agents or the development of human resources-. However, FONCYT funds projects that aim at generating publicly disclosed results, and not appropriable knowledge or knowledge of practical utility. Hence, the use of scientific publications as output measure is appropriate in this case.

In order to carry out the evaluation, we have taken advantage of a series of econometric techniques that allow us to identify the impact of having received a subsidy from the FONCYT on the academic performance, distinguishing this impact from other factors that could also be affecting the performance of the researchers that are part of both analyzed groups (funded and not funded). In other words, we attempt to evaluate the impact of the subsidies on the academic performance and, at the same time, to establish a causal relationship between reception of subsidies and performance, and not a simple correlation between both phenomena. Both objectives can solely be reached by means of econometric techniques, since descriptive statistics or qualitative evaluations are not appropriate for testing causal relationships.

The structure of this work is organized as follows. Section 1 reviews the discussion on the rationale of a policy of public funding of scientific research and examines the debate on the forms that such policy should adopt as well as the mechanisms to evaluate its results. In section 2, the evolution of FONCYT and the main funding mechanisms that it offers are briefly outlined. Section 3 describes available data; section 4 includes a description of the methodology considered to make the evaluation, as well as the main results obtained through the econometric analysis. Finally, some concluding remarks are made in section 5.

I THEORETICAL BACKGROUND. PUBLIC PROMOTION OF SCIENTIFIC RESEARCH.

Impact evaluation studies do not question themselves the theoretical rationale for the program under study. They usually simply aim at measuring the effect of some policy through reduced form equations and ascertaining if that effect was manifested in the expected direction. This evaluation applies the same non-experimental techniques and therefore is not exempt from that fault. Nevertheless, this section tries to partially compensate for that failure by introducing the theoretical background that could lead to a better understanding of the logic behind the type of the programs like the one here evaluated. In fact, its purpose could be appreciated either independently as an up to date survey of the literature on the Economics of Science, or as the required input for a future structural evaluation based on a theoretical model.

To begin with, the arguments that economic literature has outlined in order to justify the public promotion of scientific research are briefly explored. Secondly, the particular mechanisms that have been proposed to promote the production, diffusion and use of scientific knowledge are analyzed. Finally, the main elements that have been suggested for the evaluation of these designs are indicated, with special emphasis on the determinants and results of scientific production.

A. The Rationale for Public Funding of Scientific Research

The seminal articles by Nelson (1959) and Arrow (1962) are the compelling references to begin any survey of the literature that justifies the public support of scientific research. Their focus lies on the public good character of knowledge, or the part of it that can be treated as information. In particular, scientific knowledge, which is considered as the product par excellence of scientific research, is conceived as a durable public good: the exclusion of its use is expensive[1], it is non-rival[2] and cumulative.

[1] The strict definition of a "pure public good" would imply, in fact, that knowledge could not be excluded, not even partially. In other words, because its benefits could not be internalized by an agent, it would not be possible to delimit property rights and, therefore, nor to prevent new knowledge from being used by charging a price. Nevertheless, the classic work by Coase (1974) makes it clear that no good is public in essence. In this sense, the use of knowledge can be limited, for example, through the establishment of patents and trade secrets.

[2] It has been argued that the "non-rival" term, popularized by Romer (1990) in relation to the characterization of knowledge as an economic good, would have to be replaced by the broader term of "infinite expansibility" proposed by David (1992). The first one describes a restriction in terms

At the same time, it is recognized that it is the factor of non-perfect exclusion, which goes hand in hand with the impossibility to appropriate all the benefits that arise in the form of a multiplicity of applications and combinations from the same knowledge - in other words, the impossibility to prevent externalities from arising out of the generation of new knowledge- , the one that generates the difference between the private marginal return and the social one, with the consequent underinvestment in a frame of competitive markets. On the other hand, the non-rival and cumulative character of knowledge intensifies the difficulty to compensate for the non appropriable profits, causing a greater sub-optimality in the allocation of resources, which is increased, as well, due to the high component of uncertainty and to the indivisibilities that entail the investments in knowledge.

This argument, which rests upon the traditional theory of "market failures", is the basic economic rationale for the position in favor of public funding of scientific research as stated mainly in the '50 and '60's. One of its first critiques is found in Demsetz (1969) who posits that the public good character of scientific knowledge is not inherent to this last one, but it is given by the institutional regime (that includes the property rights structure). Moreover, he argues that the need for public funding in the production of scientific knowledge could not be established by claiming that the market mechanism is not optimal because it would not work in an ideal theoretical way. Instead, Demsetz stresses that alternative real, and not ideal, institutional arrangements should be contrasted.

In this sense, program's impact evaluation, which rest upon a counterfactual notion of causality, could be a good point of departure to answer Demsetz's claims. It could confront the challenge to demonstrate that a particular government intervention generates a better result than the one that would be obtained without it or under other institutional regimes. This work only makes an initial step in that direction by estimating what would have happened with the research output of Argentinean researchers had they not been funded by one specific research grants program, which could help ascertain the convenience of the existence of such a program.

However, the main part of Demsetz's critique reassumed by the theoretical literature is concentrated on the institutional frame issue. The papers of Dasgupta and David (1987, 1994) are focused on noticing that the studies of what they designate as the "old economics of science" did not distinguish between the

of marginal utility (the marginal utility of an agent is not affected by the consumption of the same good by other people), whereas the second does it upon the socially available amount in a period of time at a cero marginal cost (the good is infinitely reproducible and therefore non-rival). See, for example, Quah (2003) for the implications of the use of term or the other.

particular character of the production of knowledge and the one of other public goods, precisely because they lacked an explanation of the distinctive institutional mechanism that governs the production of knowledge. The works by Dasgupta and David aimed at creating a "new economics of science" whose primary target would be to explain the existence of peculiar institutions -formal rules, informal norms and their enforcement- that distinguish the production of scientific and technological knowledge, institutions that are respectively labeled as "the Republic of Science" and " the realm of Technology".

Dasgupta and David (1994) [hereinafter DD] adopt an approach that conjugates the economic analysis, on the basis of the development of the seminal works by Nelson and Arrow, with elements of games theory and industrial organization (games with asymmetric and incomplete information, mechanism design, optimal contract, and theories of agency) and the vision of the sociology of science (with focus on the contributions of Polanyi and Merton). The reason to introduce these theoretical instruments is based on the necessity to explain the main characteristics that distinguish the production, diffusion and use of scientific knowledge.

With this purpose in view, the "new economics of science" reintroduces the difference between codified and tacit knowledge[3]. The codification of knowledge can be schematically understood as the process by which knowledge is reduced and turned into an easily transmissible instruction set -that is to say, it is transformed into information-. DD use the notion of tacit knowledge, developed originally by Polanyi (1967), to designate non-codified knowledge, unavailable in blueprints, with an expensive transfer process.

The key for DD analysis resides in considering codified and tacit knowledge as substitutable inputs, at least at the margin. It also consists in accepting that the quality of being non tacit is not inherent to knowledge, but that researchers can take the decision to codify knowledge and to make it available as information instead of maintaining it in its tacit form. This decision is described as a function of the incentive structure within which researchers act, as well as of the codification costs. Following the idea of Polanyi, what stands out is that the fraction of knowledge that is codified has to be determined endogenously on the basis of pecuniary and non-pecuniary rewards.

[3] Nelson and Winter (1982) had introduced this discussion in the economic literature. For a detailed analysis of the difference between both concepts, Cowan et al. (2000) can be consulted, the later distinguish between articulated knowledge (codified) and not articulated one (that can be tacit - non-codified -, or codified, but not available for its use). For a different perspective on tacit knowledge see Bartholomaei (2005).

Indeed, the focus of DD is centered on explaining the difference between the "Republic of Science" and the “realm of Technology" by means of their institutional structure and reward system. DD leave behind the traditional division between the studying of nature and the solving of practical problems that was used to distinguish science from technology. They claim that the aims, the norms of behavior with respect to the diffusion of knowledge and the structure of incentives are those that help explain the different operation of both systems. Under a functional definition, the institutions in the scope of science are associated with the objective of the maximization of the stock of scientific knowledge, whereas those in the realm of technology are related to the maximization of the flow of rents that is derived from the property rights upon knowledge.

In this way, research done with the objective of being sold under the protection of trade secret would belong to the sphere of technology; whereas the one that is carried out with the purpose of open disclosure would belong to the world of science. As tacit knowledge is easier to be kept in secret, it is considered as a partially excludible good, whose production generates appropriable benefits, and therefore it is thought to predominate in the realm of technology. On the other hand, the output of basic science continues being conceived as codified knowledge, freely available. For this reason, the argument in favor of public funding remains valid, but the same one is no longer based only on correcting the inefficiency of private production of scientific knowledge. It is now focused on creating adequate incentives in order to strike a balance in the allocation of researchers to tasks related to open science versus technological development.[4]

More recent works develop different reasons for public funding of scientific research. In this case, the questioning to the traditional vision of the first works of the Nelson[5]-Arrow’s style is even deeper since it is affirmed that the public good dilemma (key of the market failure’s argument) is valid solely for the fictitious world of codified knowledge with low costs of learning and transmission. Nevertheless, when these costs are high, knowledge could no longer be non-rival in essence and learning costs diminish its cumulative character as well.

[4] For a formal model of the effect of incentives in scientific research, see Lazear (1996).

[5] A reader who does not know Richard Nelson’s intellectual trajectory could be confused because of his being at the same time the father of the tradition that justifies the public subsidies to basic science on the basis of the market failure argument and the instigator of good part of the reflection upon tacit knowledge. In fact, it is a long route that took Nelson to incorporate progressively new concepts tied to the analysis of the generation and diffusion of knowledge processes and to inaugurate the modern "evolutionist" approach in economic theory, along with Sydney Winter, in the middle of the seventies (see Nelson and Winter, 1982).

Indeed, the work by Pavitt (2005) is focused on the observation that the output of scientific research is not exclusively easily transmissible information. As the new developments of Nelson and in general of evolutionist authors emphasize, Pavitt highlights that the way from scientific research to its application is not linear. In this direction, Callon (1994) argues that the success of scientific research requires institutions, capabilities, equipment and networks. For this reason, he avers that its public character has been overestimated.

This observation, together with the increasing importance granted to tacit knowledge as output of scientific research, has led some authors to outline arguments that justify setting limits to public support of science. However, the line of reasoning that assert that tacit knowledge does not have the properties of a public good and, therefore, that there would not be a justification to subsidize scientific research, ignores the endogeneity of the decision to carry out the codification process when it is feasible, already raised by DD. Furthermore, the critique overlooks the more important social benefit of science: the development of researcher's capacities. This is considered the key of the learning process that tacit knowledge demands, as explained by Rosenberg (1990).

Consequently, in the new visions, the importance of embodied knowledge stands out. The relevance of fostering the production and diffusion of this kind of knowledge is stressed against the possibility of free riding on the one developed in other countries, since benefits are smaller if it comes from foreign sources. On the one hand because the marginal costs of transmission are higher and, on the other hand, because the development of endogenous capacities is not fostered, this is the basis of the explanation of why the most successful countries are not free-riders on the scientific results originated in other countries[6].

Actually, the new justifications for public support of scientific research come from the benefits that its domestic production triggers. Pavitt (2005) and Salter and Martin (2001) emphasize the training of scientists, the development of new methods (Rosenberg, 1992), the creation of networks (Lundvall, 1992) and the increase in the capacity to solve problems (Patel and Pavitt, 2000). In general, recent studies agree on the fact that the justification of investment in scientific research must be based on the quality of its being an investment in learning capabilities, in addition to the free disclosure of scientific knowledge that was emphasized by the previous literature[7]. In this direction, several authors argue

[6] Hicks et al. (1996) offer evidence that the bonds between scientific research and its application depend negatively on the distance and on the foreign origin of scientific developments.

[7] The empirical section of this work focus exclusively on bibliometric output due to data unavailability and, most importantly, since these additional objectives of public support to scientific research are not the primary ones for the program under analysis.

that it is important to consider scientific policies within the framework of the National System of Innovation approach, which considers innovation as an interactive process that implies the existence of networks of cooperation between the different agents participating in the system, especially firms, universities and organizations of science and technology [8].

B. Incentives Mechanisms to Scientific Research

If it is accepted that, to the present, the consensus in the literature on the subject is that, by one or more of the reasons mentioned in the previous section, it is not possible to trust only in the market to guarantee a volume of efficient scientific research from the social point of view, the issue is how this support must be carried out. Naturally, from the point of view of the models that identify scientific production with information, "the Schumpeterian" solution is inappropriate; since it implies an incentive in the form of some degree of monopolistic power that allows, at least temporarily, the appropriation of quasi-rents. In that way, the free use of knowledge, which is considered optimal from the social point of view due to its non-rival good character, would be restricted.

To confront this problem of the appropriation of benefits, which has been widely treated in the literature on the allocation of public goods, and to face the sub-production of scientific knowledge, three possible strategies have been outlined. In the first place, the possibility is contemplated that the government engages itself directly in the production and provision of knowledge, funding it through taxes. This is the case of government owned research and development laboratories where scientists are public employees[9]. The second option is the abovementioned granting of property rights to private producers to offer them some degree of monopolistic power through patents; which reintroduces market mechanisms. A third possible scheme is based on public subsidies, funded with taxes, to encourage the production of knowledge.[10,11]

The point of this last alternative that DD emphasize is that, under the same one, the exclusive property right upon the results of research is limited and open

---

[8] For an analysis of the Argentine case see Chudnovsky (2001) and López (2002).

[9] This is the model on which the creation of the CONICET in Argentina was based, following the French model of the CNRS, in which researchers are remunerated as public employees.

[10] We could find a similitude between the scheme of "prizes" to scientific activity and subsidies, with the difference that the mechanisms of evaluation in the former case are *ex-post*, whereas in the later they are *ex-ante*.

[11] Although the focus here is on the justification of public support of science, naturally, there is a vast array of private mechanisms of funding that promote the generation of freely available knowledge/information, either through organizations with permanent personnel or funds that grant subsidies and/or prizes to researchers.

disclosure is demanded, quality that this design shares with the first strategy (public laboratories)[12]-this it is not the case of the second alternative (patents), which guarantees a certain degree of disclosure of results, but only in counterpart to the granting of exclusive property rights-.

Additionally, a segment of the literature, based on ideas from the sociology of science, in particular on Merton (1973), attributes a fundamental role to the competition for priority in the organization of the scientific community. The same one would act as an impulse for both the production and diffusion of knowledge and therefore, it would be an incentive analogous in a certain way to the Schumpeterian monopolistic power[13].

The problem with this scheme resides in the fact that runners-up are not compensated, since a race unties in which the winner takes the greater prize. This would cause that all the risks fell on the shoulders of scientists, reason why in the case of risk averse individuals it would be necessary to pay a certain fixed amount so that they enter the scientific system -actually, this problem has been solved, in general, associating the scientific activity with teaching, which generates the fixed remuneration-.

Together with the competition for priority, there exists another mechanism that characterizes the scientific system: collective *ethos*[14] or cooperative culture. This predominant culture in the scientific system, along with the method of revision by peers, is the one that would allow to certify the quality of knowledge and manage to direct each researcher towards the problem which he/she is more enabled to solve -that is to say, it would favor the efficient allocation of resources in an autonomous way, once the existence of the fixed monetary incentives already mentioned are in place-.

---

[12] To be precise, in almost all countries a fraction of scientific production originated in public laboratories has secret character, in particular when it is tied to military aims. Nevertheless, this usually represents a small portion of the knowledge generated by this type of organizations.

[13] It must be clarified that the remuneration on the base of priority includes diverse variants: through the achievement of better results the scientist obtains greater recognition on the part of his/her peers, it is able to see his/her wage increased, to receive prizes, scholarships and to even send signals to the market through the acquired prestige.

[14] This notion of collective ethos was raised by Merton (1973) who describes it as a moral code that governs the behavior of researchers. Ziman (1994) summarizes it under the CUDOS term whose abbreviations represent: community, universality, disinterestedness, originality and skepticism. Dasgupta (2000) associates it to a community with a value system that has developed a "taste" for non-monetary rewards in science. In the same sense, Aghion et al. (2005) develop a model by which they estimate the necessary payments in order to prevent scientists from dedicating to other type of activities, and considering the presence of non-monetary own rewards from scientific research is a key quality of their work.

Taking these two mechanisms into account, several authors recommend adopting the system of payments that contemplates a fixed monetary salary to stimulate the selection of the scientific career, together with a variable compensation on the basis of performance. Although the fixed salary would curtail incentives to do research, the abovementioned reasons explain its necessity to induce researchers' incorporation into the scientific career (see DD). This intuitive affirmation is formally demonstrated in Lazear (1996) and Carrillo and Papagni (2004).

Carrillo and Papagni (2004) reference to the work by Shell (1966) and develop a tripartite remuneration scheme: a monetary prize to the "winner", a fixed salary and one non-monetary part based on the social prestige attributed to researchers. Thus, they demonstrate that the equilibrium level of researcher's effort is smaller with more public expenditure in research (mainly by the part of fixed salary and due also to tax increases), but at the same time they show that the flat salary is an imperative to guarantee the existence of a group of full time researchers. In practice, great part of scientific research works precisely on the basis of a combination of fixed salaries with an incentive scheme based on subsidies[15], so the next question is: how this last scheme would have to be designed?

The theoretical model of Lazear (1996) confronts this topic. For example, he shows that increasing the number of subsidized projects while decreasing the amount of the subsidy per project so as to leave expected amount received constant, would reduce researcher's effort. Furthermore, he demonstrates that if runners up in the contest for being granted a subsidy are constrained to receive no money, there is no scheme that induces efficiency and at the same time clears the market.

The more "traditional" option to grant subsidies is based, in practice, on excellence or quality criteria. Gambardella (2001) offers evidence that the funding of researchers who already were efficient (through a good publication profile) generates more and better publications. This goes against the implications of Lazear's model, which indicate that, in that case, older researchers would have fewer incentives to finish their projects and therefore would reduce their effort, so an age-contingent award would be needed. Furthermore, Molas-Gallart and Salter (2004) criticize the allocation based on excellence arguing that it reduces variety, which is considered essential to obtain a greater probability of valuable results, and generates a concentration of results in a limited number of topics and researchers. In this sense, Scherer and Harhoff

[15] Stephan (1996) observes that many European countries indirectly finance scientists when subsidizing the research institutes where they work, whereas in the United States the ordinary practice is based on each scientist obtaining his/her own funds through the submittal of projects to funding agencies.

(2000) propose an allocation based on a portfolio system to promote risk diversification, Lazear (1996) also advises moving into a higher risk funding direction to increase expected payoff.

While models based on "excellence" aim at reducing project failure to a minimum by choosing those proposals of greater quality and submitted by researchers with the best previous outputs, those based on portfolio selection accept certain rate of failure as a consequence of risk diversification[16]. Actually, several systems have been developed to guarantee variety and excellence[17]. In these cases, the aim is to give place also to projects submitted by researchers that do not have the best antecedents, but who make proposals that can be interesting and with a certain potential of very positive results, although they entail a greater implicit risk (Molas-Gallart and Salter, 2004).

At the same time, the need to take into account other criteria, such as "pertinence", has increasingly been considered in developed countries and in the developing ones as well. The possibility that the results of scientific activity should have practical applications or be "demand" driven is more frequently signaled (see Peterson and Sharp, 1998). In general, the preoccupation in these cases consists in directing scientific effort towards aims that would allow the increase in innovation opportunities -therefore, also firm's competitiveness and, more in general, the country one as a whole-. [18]

These tendencies have been criticized by diverse authors (see, for example, DD; Stern, 2004; David, 1998; Nelson, 2003). Among critiques, it is mentioned that the insistence on achieving results that could be patented[19]and/or with fast transferability to private sector could diminish, and not increase, scientific

[16] In this case it could be advisable to attempt a random allocation of part of the budget with the intention of making controlled experiments. In this direction, the work by Duflo and Kremer (2003) emphasize the convenience of making random experiments with the intention to give firmer foundations for public programs; Jaffe (2002) recommends it for the particular case of scientific programs evaluation. Naturally, randomization is neither always feasible nor desirable, and in addition we cannot forget the fact that it can imply a trade-off between confidence in the evaluation and a better design of the program to reach its specific goals.

[17] Molas-Gallart and Salter (2004) emphasize the cases of the Basque Government (Spain), the Experimental Program to Stimulate Competitive Research (EPSCoR) with the support of the National Science Foundation in the United States and the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom.

[18] For example, some works study the effect of science on growth through the scientific antecedents of innovations (Mansfield 1991, 1995, Sorenson and Fleming 2004) or the existence of spillovers between firms and universities (Jaffe, 1989).

[19] A key point in this connection was the sanctioning of the Bayh-Dole Act in 1980 in the U.S. that allowed universities to patent and to commercialize their discoveries. For a recent description of the situation in the United States and in the European Union see Mowery et al. (2001) and Geuna and Nesta (2005) respectively.

production, as well as limit the fast diffusion of new knowledge. For empiric results on the negative impact of commercially directed funds on publications see Goldfarb (2001)[20]. Nevertheless, the decision to redirect the funding of scientific activities towards applied fields with concrete economic benefits can be appropriate in emergent economies. As Pavitt (2005) indicates, scientific needs depend on the degree of development and the productive structure of a country. In emerging economies resources are limited and, because of being far away from the technological frontier, the uncertainty of the results can be smaller.

In conclusion, the consensus in favor of the public support of scientific research has not led to an agreement with respect to the suitable form to carry it out. Hence an empirical evaluation of the results of the different alternatives appears to be fruitful.

C. Scientific Research Outputs

Momentarily leaving aside the discussion about the modality that the public support of science must assume, if it is accepted that this last one is desirable, the question on how to evaluate results naturally arises. One of the more used variables in that direction is bibliometric output, since in general it is accepted that the number of publications can be a good measure of the production of codified knowledge and the possibility of accessing to this knowledge[21].

As Stephan (1996) notices, one usual way to measure the importance of a scientist's contribution is through the number of his/her publications with some weighting to consider the quality of articles[22]. The typical strategy to control for quality is to use the impact factor of the journals released by the Institute of Scientific Information (ISI). This is not more than a measurement of the frequency with which the "average" article of a journal was mentioned in a certain year[23]. Nevertheless, Amin and Mabe (2000) notice that this measurement of quality should not be adopted without circumspection, because it depends on

[20] It should be mentioned that some papers generate results that seem to contradict those fears (for example, Azoulay et al., 2005 find a positive relation between patenting and the number of publications -and what is more, they find that patenting does not have a negative effect on the quality of the respective publications -).

[21] Diamond (1986) provides empirical evidence for the relevance of the number of publications to determine salaries increases and promotions at universities. Debackere and Glänzel (2003) analyze the results of an experiment that consisted in distributing funds to Flemish universities on the basis of bibliometric output.

[22] Hicks et al. (2000) show that the greater the quality of publications, the bigger is their impact upon innovation.

[23] In particular, the impact factor of a journal is calculated by dividing the number of citations to articles published in the two previous years by the number of publications in those years in each journal.

the field, the type and size of the journal and on the window of measurement used. Furthermore, they remark that its value fluctuates from year to year and that its use to measure the quality of a scientist's publications from the quality of those of the journal is not always appropriate. For this reason, it would be advisable to contrast it with a direct measurement of citations to the articles produced by each scientist, which is not usually available[24].

Crespi and Geuna (2004) offer evidence that indicates the importance of incorporating lags in the estimations of the results of scientific research, in order to catch the delay between the reception of the funds and actual publication[25]. In an econometric study with panel data they show that to gauge the total effect on publications and citations they must incorporate lags of 6 and 7 years respectively, although they reach their maximum level in the 5th and 6th years[26]. Moreover, Crespi and Geuna (2005) emphasize that distinct fields are characterized by a different propensity to publish in recognized journals, as well as by different times in reaching a publication.

D. Determinants of Scientific Productivity

A considerable number of empirical works studies the variables that affect the quantity of publications by scientists. They differ in methodology, sample size and range of years, for this reason their findings are not perfectly comparable. Nevertheless, they turn out to be a useful guide for this work as far as the selection of the variables that could influence scientific production is concerned.

With the aim of evaluating the effect of certain type of funding or simply understanding the determinants of scientific production, several characteristic of the scientist are included in regressions, being the most studied: gender, age, the type of institution in which he/she works and the level of education. Additionally, some studies incorporate other characteristics, like the score given in the evaluation - that partly determines the granting of the subsidy (Arora and Gambardella, 1998)- or elements of the background of the scientist; for example, the size and quality of the research laboratory (Turner and Mairesse, 2005).

With respect to personal characteristics, the effect of the age on the number of publications has been widely discussed, in particular in the studies on the labor

---

[24]Critiques have also been formulated towards the ISI because of its bias in favor of Anglo-Saxon publications and its underestimation of social sciences.

[25] Their work is based on the one by Adams and Griliches (1998), which highlights the fact that the transformation of an idea into codified knowledge and its publication takes time.

[26] In this direction, Arora and Gambardella (1998) measure the impact of public funding to economic researchers in the United States considering publications weighted by citations in a window of 5 years after the decision to grant the subsidy.

market of scientists[27]. Stephan (1996) presents the findings of these works, which have adapted the frame of human capital theory to develop life-cycle models. One of their implications is that the publication rate of a scientist grows initially, but soon begins to decline around half of its career. The empirical evidence, confirmed by works that use panel data and more sophisticated econometric techniques, usually shows that the turndown in the quadratic relation takes place between 45 and 52 years and that the same one can differ by area of knowledge[28].

The difference in productivity by gender has been a widely debated issue. From the finding of numerous works that the rate of women's publications is inferior to the men's one, which Cole and Zuckerman (1984) denominated "the productivity puzzle", several studies have tried to test this hypothesis. The works by Xie and Shauman (1998) and Turner and Mairesse (2005) would seem to verify it, but Long (2001) attributes this difference to the fact that similar positions were not being compared. Once the probability of acceding to a certain position in the academy is taken into account, gender differences apparently have no effect on scientific productivity.

Other variables mentioned to have a positive impact on the number of publications are the level of education, measured by the possession of a PhD degree and the prestige of the institution in which it was obtained[29], the availability of other sources of incomes (Stephan, 1996) and the number of previous publications (Arora and Gambardella, 1998).

E. Measuring the Impact of Public Funding on Publications

At the moment, there are few works that study the causal effect of a program of public funding of scientific production on the number of publications. Among the studies aimed at analyzing this issue are those of Arora and Gambardella (1998), Arora, David, and Gambardella (1998), and Goldfarb (2001). Arora, David and Gambardella (1998) test the effect of an Italian government program funding academic biotechnology research, finding a low average elasticity of research output with respect to funding. Arora and Gambardella (1998), in turn, focus on the impact of NSF funding on basic research in economics in the US and estimate a significant positive impact only for young economists. Goldfarb (2001) measures the impact of a NASA aerospace engineering program and find a significant positive average impact on research output, but only for scientists

[27] The article on the shortages of scientists in the labor market by Arrow and Capron (1959) had inspired the developments of the old economics of science.
[28] See for example: Bernier et al. (1975), Cole (1979), Levin and Stephan (1991), Turner and Mairesse (2003) and Gonzalez-Brambila (2003).
[29] See Buchmueller et al. (1999), Turner and Mairesse (2005) and Gonzalez-Brambila (2003).

with moderate academic reputation (measured through a low number of citations) and at the cost of a lower quality of publications.

Other studies, without measuring the impact of a specific program, estimate scientific knowledge production functions, based on a polynomial distributed lag model by which the research and development expenditure affects publications (Crespi and Geuna, 2005) or just try to see the effect of the determinants described in the previous section[30]. In this direction, Turner and Mairesse (2005) and Gonzalez-Brambila (2003) stand out. They take into account the integer and positive character of publications by using Poisson or Negative Binomial models[31]. In addition, with panel data they can consider fixed effects models that allow them to obtain estimations not affected by the presence of a possible time-invariant individual heterogeneity, thus correcting potential bias in the works that used cross section or times series data.

The literature that more vastly focuses on estimating causal effects and introduce program evaluation advance is the one that estimates the impact of scientific patenting activity on the number and quality of publications. In this literature, the patenting activity of a scientist represents the "treatment", whereas the scientists that do not patent constitute the control group. In order to consider the average treatment effect of the treatment on the treated they use methods that range from instrumental variables (Stephan et al., 2004) to strategies of difference in differences in panels of matched individuals (Markiewicz and DiMinin, 2004) on the basis of observable characteristics or the propensity score, or with estimations weighted by the inverse of the probability of receiving the treatment (Azoulay et al., 2005) [32].

It is relevant to combine the methods developed in the abovementioned papers with the advances in the analysis of the determinants of scientific productivity in order to estimate the causal effect of a public subsidy to scientific research. Despite the fact that the assumptions and strategies to use would have to be modified, many of the previous contributions can be of great value. That would

[30] Another branch of this literature analyzes scientific productivity using non-parametric methods, such as m-order frontiers (Bonnacorsi and Daraio, 2003).

[31] Basically, the conditional mean of publications is modeled as an exponential function of its determinants, instead of using a linear function as the ordinary least squares method assumes, that is to say: $E(y_i / x_i) = \exp(\beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki})$. In this way, it is possible to avoid the problem of negative predictions of the number of publications that would contradict the countable attribute of data.

[32] All these strategies, which are described in the methodological section, attempt to face the problem of the endogeneity of the treatment that bias OLS estimations and are based on particular assumptions in order to provide a causal interpretation of estimated parameters.

be the task of this work in section three where a series of evaluation techniques to estimate the impact of FONCYT grants are applied, but before that, in the next section the particular functioning of this program is described.

II FONCYT PROGRAM: EVOLUTION AND MAIN CHARACTERISTICS

The Fund for the Scientific and Technological Research (FONCYT) is one of the two funds[33] of the National Agency of Scientific and Technological Promotion. The Agency is a federal organism, created in 1996, that belongs to the Ministry of Education, Science and Technology. It was formed with the main objective of separating the promotion and funding functions from those directly related to the execution of scientific and technological activities34and to policy setting.

FONCYT is dedicated to grant non-reimbursable subsidies to scientific and technological projects that are considered of high-priority for national development. Researchers from public or private non-profit organizations located in Argentina must elaborate these projects. The activities of FONCYT began in 1997, within the framework of an ample reform to the scientific and technological system that included the objective to increase and to improve the mechanisms of public promotion of scientific and technological activities.

In particular, one of the objectives of the creation of FONCYT was to develop an instance of public funding of science on the basis of competitive mechanisms, based on the quality evaluation through peer review and pertinence criteria[35]. At the time the Agency was created, the practice of elaborating National Multi Annual Plans of Science and Technology began. Among other things, priorities and basic policy outlines are established in those Plans, which are taken into account to determine areas and activities towards which the funds of the Agency are preferentially oriented (see Chudnovsky, 2001).

In this study, the results of the PICT projects funded by FONCYT in 1998 and 1999 are considered. The PICT line of funding represents almost 75% of all the funding granted by FONCYT in the frame of the Second Modernization Technological Program from 1998 to 2004. The amount adjudicated per year was in the range from 30 to 37 million Argentinean pesos until 2001; afterwards the budget jumped up to a total of 72 to 81 million pesos per year in the period 2002-

[33] The other fund of the Agency is the Argentine Technological Fund (FONTAR), which gives credits and subsidies to technological projects.

[34] In the case of scientific activity, until the creation of the Agency, the National Council of Scientific and Technical Research (CONICET) was the responsible for funding scientific research in the country.

[35] It is important to bear in mind that CONICET, as has previously been mentioned, was organized, since its foundation in 1958, on the basis of the formula of a “career researcher” by which scientists are permanent staff of the Federal Government. For an analysis of the evolution of CONICET and a critical examination of the institutional mechanisms in which its working is based, see López (2002).

2004[36]. This change is explained, partly, by an increase in the number of funded projects from an average of 396 for the calls of the first period, to 485 for those of the second period. However, the main cause of this difference lies on the average amount granted by project: it switched from 85 thousand pesos in the first period to 158 thousands in the later one. At the same time, the number of submitted projects showed a decreasing trend, which implies that the percentage of funded projects constantly increased. Remember that this fact is consistent with an increase in researcher's effort according to Lazear's model, therefore; there is a reason to believe that the results of this evaluation could still be improved if one measured the impact for more recent calls.

Public institutions were the main recipients of the funds and Buenos Aires was the province that more funds received (34%), followed by the City of Buenos Aires (31.4%), Córdoba (10.2%) and Santa Fe (9.1%). With respect to the distribution of the value of subsidies by areas of knowledge, Biological Sciences took the lead with 19% and was followed by Medical Sciences (18%) and Agricultural, Forestry and Fishing Technologies (13%).

At the moment when the calls under analysis were made, the contest for awarding the subsidies was divided in two principal categories: Projects of Scientific and Technological Research (PICT) and Projects of Research and Development (PID) -these last ones were oriented to applications of economic or social interest of companies or public good entities prone to co-finance the respective projects together with FONCYT -. In the case of the PICT there were four categories of projects. The first one was the amplest in terms of disciplines including research projects opened to all areas of scientific and/or technological knowledge. The second category was for research projects on sectors and specific topics of the "National Multi Annual Science and Technology Plan" [37], whereas the third one was devoted to research projects on subjects related to the regional priorities of the Plan. Finally, there was a fourth category for projects co-financed within the framework of agreements with public and private organizations or companies, which is not considered for the analysis.

The projects were divided according to the experience of the group of researchers, distinguishing consolidated groups, made up of researchers of

[36] Since there was a devaluation of the peso in 2002, the amount received in dollars in fact was smaller; however, as long as the alternative funding of scientists is in pesos, this fact should not affect the claims in the text.

[37] In the case of the 1998-2000 Plan, prioritized areas were: Agro industrial Production, Mining, Education, Health, Natural Resources and Environment, Biotechnology, Argentinean Sea, Manufacturing Industry, Energy, Defence, Clean Technologies, Ozone and Climate Change, Biodiversity, Microelectronic Applications, National System of Innovation Studies, Violence and Urban Security and Gender Studies.

recognized trajectory, from the groups of recent formation - after 1994- and from young researchers (with less than 36 years) - these last projects were not included in the evaluation -. The maximum amount of the part of the subsidy in charge of the Agency for the two first types of projects was of 50,000 pesos per year, whereas for those of the last type it was of 5,000 per year. Besides, the implementation time could not surpass the 3 years horizon after the reception of the subsidy.

It is important to notice that, in contrast to typical subsidy programs in developed countries, FONCYT's funds cannot be used to pay researcher's salaries. It is assumed that these researchers count with a permanent source of incomes from the institution in which they work or from other source.[38] Funds are aimed at financing inputs, participations in scientific conferences, specialized technical services and equipment. This fact is extremely relevant for the evaluation, since, as almost all studies, data on researcher's other sources of income is not available. However, all the evaluated proposals –funded and not funded- comply with the requisite of having counterpart funds, so the bias introduced in the estimated impact because of varying external financial sources is mitigated in this particular case.

The project selection process can be summarized in three steps. The first one involves admissibility. In this stage it is verified that projects fulfill the minimum requirements of general nature that constitute admission criteria[39]. Once the project is admitted, the following step is the quality evaluation. For this phase, peers in the corresponding thematic area, whose identity remains anonymous, are consulted. The criteria to evaluate each project are, among others, the degree in which it suggests and explores creative and original concepts, the contribution of the proposed activity to the knowledge and understanding of the field, the possible impact of the research project on the socio-economic development of the country, the qualification of the research team (or of the individual researcher, according to the case) to develop the project, the antecedents in the proposed subject, the organization and methodology to be used and the correspondence between available and requested funds to carry out the research.

[38] What is more, these salaries can be taken as counterpart funds to those coming from FONCYT since funding requires a monetary counterpart by the research group, by itself or through third funds, equivalent as a minimum to the contribution of FONCYT. In the 1998 call, the funds taken as counterpart were, as a minimum, 33% of the contribution of the FONCYT.

[39] Requirements vary by area, but it is assumed that the researchers of the group have a labor relationship with an Argentine institution of science and technology, dedicate as a minimum 50% of their time to the execution of the project and have research antecedents (which take into account the graduate studies as well as the publications in academic journals).

The last step to determine which projects are funded is the evaluation of their pertinence. The collegiate bodies that have the responsibility to establish the merit of the projects -that turns out to be a combination of the quality evaluation and pertinence- are *ad-hoc* commissions, which must be composed by not more than 8 recognized members of the scientific and technological community, national and/or foreigners experts, with the necessary experience to globally analyze the projects. Members of the *ad-hoc* commissions are named by the Directory of the Agency.

Criteria taken into account to analyze pertinence differ according to the category of the projects[40], but in general they involve topics related to the intrinsic relevance of the proposal, its possible impacts on the socioeconomic development of the country or the region and on the development of human resources -in addition, there were projects funded in the categories corresponding to priorities by sector, theme and/or region, within which, pertinence criteria were also applied-. Each *ad-hoc* commission classified the respective projects into three levels of pertinence -high, medium and low-, constituting three equal groups. Moreover, when there was a tie on the merit criterion, the group projects and those submitted by new groups were given preference over the individual ones and over those of consolidated groups respectively.

In the case of the studied calls, only those projects that had been evaluated as good, very good or excellent (those excluded were from regular or non-acceptable quality) were declared as fundable. The distribution of resources between categories I, II and III was determined in proportions 1/2/1, respectively[41]. On the other hand, the distribution within categories I and II was done so that resources adjudicated to each thematic area (or high-priority area) were proportional to the amount asked for by projects approved in quality terms. In the case of categories II and III neither the good projects in terms of quality from medium levels of pertinence nor the very good ones from the low level of pertinence were considered with sufficient merit to be funded -in other words, it was accepted to fund projects of smaller quality whenever they were considered with high or medium levels of pertinence-. The order of merit for the projects in condition of being funded was the following one: excellent, very good and good projects of high pertinence; excellent and very good projects of medium pertinence and excellent projects of low pertinence. Although the formal lines of the calls and the posterior information do not clarify it explicitly, the pertinence criterion was also taken into account in category I to discriminate between

[40] As shown below, in our data base quality evaluation data are available, but we were not provided with data on the degree of pertinence attributed to each project.

[41] Given the absence of enough projects to be funded after the evaluation, the funds that were not used in Categories II and III, were destined to fund projects from Category I.

selected and not selected projects. Nevertheless, it has not been possible to distinguish if rules as precise as the ones mentioned for categories II and III were actually applied.[42]

[42] Taking this description into account, it must be emphasized that it is not possible for us to identify a function that determines the selection into the program. There is no threshold that separates funded from non-funded projects implied by the quality measure, which is the only index of evaluation on which we count. If pertinence evaluation data were available it would be possible to construct a step function, although it is not even clear how and if evaluators in fact have based their decisions on such kind of function.

III DATA

The sample used in this study includes data of 323 projects that applied for subsidies from FONCYT under the modality PICT. They correspond to the calls of the years 1998 and 1999 and include 218 funded projects and 105 non-funded projects[43] -the 323 projects obtained a qualification of good in the respective evaluation, which means that they were declared eligible, though some of them did not obtain funding because available resources were not enough.[44]

Descriptive statistics of the variables used later in this work are presented in Table 1. The set of variables includes Gender (a dummy variable that takes the value of one if the researcher is a man), Age (age of the researcher in 2005), Doctorate (a dummy variable that takes the value of one if the researcher has a doctorate degree), Evaluation (the average peer review score received by the proposals), New Group (a dummy variable that takes the value of one if the group was constituted after 1994), Region (a set of dummy variables for the regions of Cuyo, Buenos Aires, Center, Patagonia, the Northeast and the Northwest), Institution (a dummy variable that takes the value of one if the institution where the main researcher works is private), Category (a set of dummy variables for the three categories described in section 2), and Discipline (main field of the principal researcher).[45]

Data on publications and impact factors were collected from Science Citation Index (SCI), which is developed by the Institute for Scientific Information (ISI) and covers approximately 3200 journals.[46] The variable Publications_1 measures

[43] The sample has been chosen with the condition of avoiding member repetition among funded and not funded projects. Moreover, we only considered, among funded projects, those projects that were completed at the time of the evaluation.

[44] Our sample contains only non-funded projects whose participants never received funding in the PICT's line.

[45] There are 11 fields grouped in 3 broadly defined areas: Biomedical Sciences (Biological Sciences and Medical Sciences), Exact Sciences (Physical and Mathematical Sciences, Chemical Sciences, and Earth and Hydro-atmospheric Sciences) and Technologies (Food Technology, Mining, and Energy Technology, Information, Electronic and Communication Technology, Mechanic and Material Technology, Environmental Technology and Chemical Technology). In this study Human Sciences has not been considered because the database from SCI does not provide appropriate coverage of publications in that field; this implied discarding 34 projects from our database of PICT 98 and 99 calls.

[46] Publication data from 1994 to 2004 was obtained for the 323 projects; there are no reasons to doubt on the quality of these data provided by the CAICYT (Argentine Center of Scientific and Technological Information), which depends on the CONICET. However, in relation to impact factors, available data do not include values for around a 10% of the magazines in which publications of the researchers included in the sample were found. Impact factors before 1998 were also missing; therefore the values of the nearest year were attributed in this last case, whereas a null

the number of publications after the granting of the subsidy (more specifically, the number of publications in the window of four years that starts one year after receiving the subsidy)[47], whereas Publications_0 includes the number of publications in the period previous to the one corresponding to Publications_1. This means that for those researchers who received the subsidy in 1998 Publications_1 is the number of publications corresponding to the period 2000 to 2003, and for those that received the subsidy in 1999 Publications_1 is the number of publications corresponding to the period 2001 to 2004. Analogously, Publications_0 is the total number of publications from 1996 to 1999 for projects submitted in 1998 and from 1997 to 2000 for those submitted in 1999. The idea behind using a four-year window after receiving the subsidy is to take into consideration the lag between receiving the subsidy and the publication of the findings of the project (see Crespi and Geuna, 2004, 2005).

In order to account for the quality of publications, the sum of the impact factors of publications in a four-year window (Impact_i is the sum of the impact factors associated to Publications_i) was also considered as an alternative measure for academic output.

Table 1. Summary Statistics

	Full sample 323 projects		**FONCYT = 0 105 observations**		**FONCYT = 1 218 observations**	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Publications_0	6.08	61.3	4.28	49.55	6.96	64.9
Publications_1	6.56	53.2	4.07	19	7.77	65.5
Impact_0	10.4	272.1	5.97	140.5	12.54	322.4
Impact_1	13.3	323.1	7.17	101.1	16.18	20.11
New group	0.44	0.25	0.50	0.25	0.41	0.24
Cuyo	0.04	0.04	0.07	0.07	0.02	0.02
Buenos Aires	0.59	0.24	0.60	0.24	0.59	0.24
Center	0.21	0.17	0.15	0.13	0.24	0.18
Patagonia	0.08	0.07	0.04	0.19	0.09	0.08
Northwest	0.06	0.06	0.08	0.07	0.05	0.05
Northeast	0.02	0.02	0.06	0.05	0.01	0.01

value of impact has been assigned to the other cases. In addition, other inconsistencies on impact factor data have been detected.

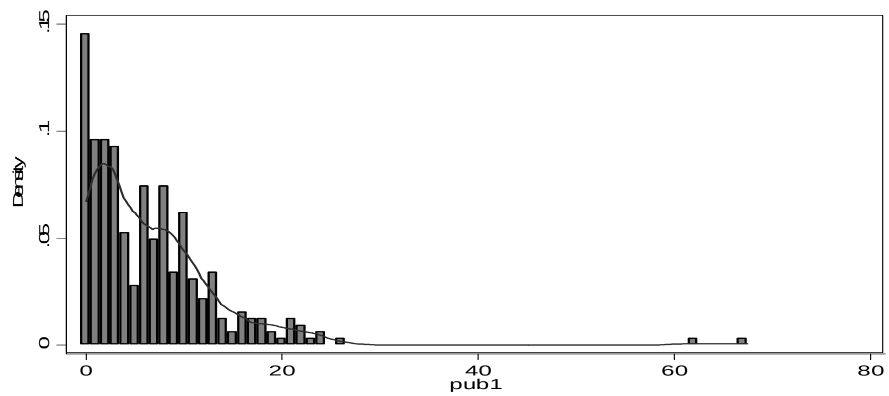
[47] Although projects can be submitted by a group of researchers, each project must have a responsible researcher. In order to develop our evaluation we have taken as the performance variable the number of publications of the researcher who is in charge of the project. The other possible strategy –to take the publications of all the members of the group- would have been highly troublesome, considering in particular that a considerable proportion of the publications that we are analyzing are made in co-authorship, being foreseeable that those co-authorships take shape to a great extent between the members of the group.

	Full sample 323 projects		FONCYT = 0 105 observations		FONCYT = 1 218 observations	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Year 1998	0.47	0.50	0.35	0.23	0.52	0.25
Gender	0.65	0.23	0.63	0.24	0.66	0.23
Age	55.6	69.9	56.7	74.8	55	66.8
Doctorate	0.82	0.15	0.77	0.17	0.84	0.13
Institution	0.02	0.02	0.03	0.03	0.02	0.02
Evaluation	7.81	1.30	6.83	0.63	8.28	0.94
Category 1	0.53	0.25	0.76	0.18	0.42	0.49
Category 2	0.35	0.23	0.21	0.17	0.41	0.24
Category 3	0.12	0.11	0.03	0.03	0.17	0.14

The measure of publications and impact factors do not distinguish those that are a direct product of the subsidized project. Nevertheless, there are reasons to believe that this is not a problem for the analysis, since scientific activity is rarely developed in the form of closed compartments, being on the contrary the norm that there is "crossed fertilization" between results derived from different projects -which generates that it is not always possible to distinguish results derived from a specific project. Moreover, the subsidy from FONCYT allows researchers to pay inputs, i.e., hire research assistants, buy equipment, etc., that can be useful for other projects.

Figure 1 shows that the dependent variable in the econometric model (Publications$_1$) takes positive discrete values. Its distribution also presents an important accumulation of observations in low values (especially at zero), thus motivating the use of an appropriate model for count data explained in the following sections.

Figure 1. Publications in the Period After Treatment



IV IMPACT EVALUATION OF FONCYT PROGRAM: ECONOMETRIC METHODOLOGY AND RESULTS.

The aim of program evaluation methods is to answer counterfactual questions, i.e., what would have taken place without the policy under study. Whenever possible, random assignment of the sample from the population and to treated and control groups is advised[48]. The first stage of randomization -a random sample from the population- would make possible to generalize results found for a particular project (external validity), the second one -the random assignment of the sample to treated and control groups- would avoid all systematic differences between those receiving and those not receiving the treatment, with the exception of the project grant that is received only by the treated group. Any subsequent differences in average outcomes between the two groups could be reliably attributed to the effects of the project (internal validity).

None of the two stages of randomization was undertaken, however, in order to allocate FONCYT funding, neither researchers receiving support, nor the non-funded ones, can be considered random draws. Constructing a valid control group in this setting is quite challenging and it is necessary to draw upon the discussion of recent advances in econometric methods for evaluation studies based on non-experimental data in order to deal with potential biases. In our case, to evaluate the impact of a scientific research subsidy program, the outcome of non-supported researchers is used to estimate the counterfactual scenario, i.e., what the funded researchers would have experienced had they not been funded. The difference in performance between supported and non-supported researchers is the estimated gross impact of the funding scheme.

The performance of non-funded researchers may, however, differ from what funded researchers would have experienced in the absence of the support scheme. For example, in table 1 we can appreciate that the mean value of publications and impact factors is quite smaller for non-funded researchers. In this case, besides the effect of the program there could be systematic differences between supported and non-supported researchers, this is the selection bias problem widely discussed in the evaluation literature.

Formally, if Y_i^T is the research outcome (i.e., number of publications, impact index) of a funded researcher i and Y_i^C is the outcome of the same researcher in case he had not been funded, the effect of the grant program is equal to Y_i^T- Y_i^C.

[48] For the advantages of randomization in development economics program evaluation studies see Duflo and Kremer (2003), for the particular case of evaluation in R&D and scientific programs see Jaffe (2002).

However, the two outcomes cannot be observed at the same time, thus it is not possible to get the causal effect for a single researcher; this is the fundamental problem of causal inference. Nevertheless, the average causal effect is estimable, in particular, for the purpose of this work, $E\left(Y_{\mathrm{i}}^{\mathrm{T}} - Y_{\mathrm{i}}^{C}/D=1\right)$ the average effect of the grant program for those researchers receiving grants, is the object of interest[49], where D is a dummy equal to one if a researcher received a subsidy from FONCYT.

Now, the problem lies in estimating $E\left(Y_{\mathrm{i}}^{C}/D=1\right)$, the average counterfactual outcome for supported researchers, which is done using $E\left(Y_{\mathrm{i}}^{C}/D=0\right)$, the observable outcome of non-funded researchers. It is easy to see that:

$$E\left(Y_{\mathrm{i}}^{\mathrm{T}}/D{=}1\right)\text{-}E\left(Y_{\mathrm{i}}^{C}/D=0\right)=[E(Y_i^T - Y_i^C / D=1)]+[E(Y_i^C / D=1)-E(Y_i^C / D=0)]. \quad (1)$$

Therefore, equation (1) shows that the estimated impact will be equal to the desired effect of treatment on the treated (first term in brackets) plus the abovementioned selection bias (second term in brackets), which captures systematic differences between funded and non-funded researchers. Random experiments do not eliminate this bias, but balance it in the treatment and control groups.[50] Otherwise, non-experimental methods need to rely on assumptions in order to construct the more similar control group as possible in order for it to act as an appropriate counterfactual.

In this section, various strategies to mitigate the selection bias problem are explored. These are: (i) regression with controls; (ii) matching methods; (iii) difference-in-differences and (iv) difference-in-differences matching estimator[51].

[49] Note at this point that the impact of the program is associated with a dichotomous grant (D=1) or no grant (D=0) condition. Thus, it is assumed that the magnitude of the research grant does not matter.

[50] Not even randomization is a panacea, it has its own assumptions and problems as Heckman (1992) and Heckman et al. (2000), for instance, emphasize.

[51] Two widely used non-experimental techniques, instrumental variable methodology and regression discontinuity design, were not applicable in our case. The former was not used because we found no variable that affect the probability of selection but have no impact on performance (publications) beyond its effect on the probability of selection. Neither gender nor geographical location of researchers, which were two potential instruments, was relevant selection criteria. Furthermore, selection process involved variables that are likely to be correlated with the number of publications and, therefore, they are not suitable instruments. Regression discontinuity was not applicable since, as it has already been explained (see footnote 44), selection into FONCYT program was not based on an observable variable nor on a threshold process that imply a discontinuity in the probability of receiving treatment at some point.

In what follows each of these alternative strategies is described together with their results for our sample.

A. Regression with Controls

The most common approach to mitigate selection bias is to use regression in order to control for variables that affect the outcome and since they are different for treated and non-treated units, they may "confound" the effect of the treatment. In the case of FONCYT the typical regression would be:

$$Y_i = \beta D_i + X_i^{'} \gamma + \delta \varpi_i + \varepsilon_i \qquad (2)$$

Where Y_i is the research output (the number of publications in refereed journals) of applicant *i*, D_i is a dummy variable that takes the value of 1 if applicant *i* receives the grant, X_i is a vector of observable determinants of output (i.e., age of researcher), and the usual error term, which is assumed to be uncorrelated with the *X*'s and *D*, is represented by ε_i. The only non-standard entry in Equation (2) is ϖ_i which represents applicant-specific variation in research productivity that is in general unobservable by the econometrician, but observable by the granting agency –for example, through the evaluation of the quality and potentiality of the project. In our case, the average peer review score received by the projects is available and used as a relevant additional control.

Since this method compares conditional mean outcomes of funded and non-funded researchers, it requires a key identifying assumption in order to estimate the average effect of FONCYT on funded researchers. This is a conditional mean independence assumption, which would hold if, in the case of not being funded, supported researchers would have on average the same number of publications as non-supported researchers, once we control for differences in the other observable determinants of publications. That is: $E(Y_i^C / D=1, X_i) = E(Y_i^C / D=0, X_i) = E(Y_i^C / X_i)$[52], under this assumption, $\beta = E(Y/D=1, X_i) - E(Y/D=0, X_i) = E(Y_i^T / D=1, X_i) - E(Y_i^C / D=0, X_i) = E(Y_i^T - Y_i^C / D=1, X_i)$, since the selection bias term in equation (1) cancels out. This gives us the conditional average treatment effect on the treated, and integrating out the X's, it leads to the marginal effect on the treated.[53]

[52] For simplicity we include *w in X.*

[53] As marked by Angrist (1998) in the case of heterogeneous treatment effects regression with controls gives a weighted average with an unclear meaning.

In conclusion, it is necessary to find a set of observable controls as detailed as possible to be confident in the mean independence assumption, which is also known as "selection on observables" since its holding requires that selection into the process be related to the no-treatment outcome only through observable variables. The rule is to choose among controls those variables that affect the output and the treatment, but are not affected by this last one.[54] For this, this work draws upon the literature on the determinants of scientific production surveyed in section 2; included variables are those explained in section 3, with summary statistics presented in table 1.

Results are presented in Table 2. Given the abovementioned characteristics of the distribution of Publications_1, appreciated in Figure 1, it is necessary to use one of the several models that have been proposed in order to deal with a dependent variable that takes non-negative discrete values, such as the Poisson model and the Negative Binomial (Negbin) model. This last one is the more appropriate in certain cases of over-dispersion in the data. One intuitive way of checking that we are in the presence of over-dispersion consists in comparing the sample mean and the sample variance. As a rule of thumb, if the sample variance is greater than the double of the sample mean then it is likely that the Maximum Likelihood regression model would tend to conserve the difference in its conditional values (Cameron and Trivedi, 1998 pp.77). In our case this condition is fulfilled since the variance of publications is around eight times greater than its mean (see Table 1).

In order to test the hypothesis of equal dispersion against the alternative of over-dispersion in a more formal way we use the parameters estimated using the Negbin model. This model assumes over-dispersion and contemplates the Poisson model, which assumes equal dispersion, as a particular case. The test of the coefficients that measure over-dispersion are single tailed (since the Negbin model allows over-dispersion but not under-dispersion) and can be performed by means of a Likelihood Ratio (LR) test. The LR test is based on the difference between the log-likelihoods of the Poisson and Negbin models.[55] In our case we find strong evidence against the null hypothesis of equal dispersion (LR equal to 233).

[54] For a detailed analysis on which kind of controls should be included see Lee (2005).
[55] To test with a 95% of confidence we must use the 90% statistic from the chi-square distribution with one degree of freedom (See Cameron and Trivedi, 1998 pp.78 and Verbeek, 2000 pp.214).

Table 2: Regression with controls

	Dependent variable: Publications$_1$		Dependent variable: Impact$_1$	
	(1)	(2)	(3)	(4)
FONCYT	1.648***	1.128**	6.712***	4.065*
	(3.52)	(1.92)	(3.22)	(1.64)
Publications$_0$	0.306***	0.299***		
	(4.65)	(4.58)		
Impact$_0$			0.676***	0.660***
			(5.94)	(5.74)
New group	-0.242	-0.216	-2.492	-2.161
	(-0.51)	(0.46)	(-1.46)	(-1.22)
Evaluation		0.388*		1.793*
		(1.89)		(1.71)
Age	-0.083***	-0.082***	-0.272***	-0.279***
	(-3.29)	(-3.23)	(-2.59)	(-2.67)
Doctorate	2.050***	2.090***	5.341***	5.630***
	(3.35)	(3.41)	(2.72)	(2.81)
Gender	0.238	0.149	0.530	0.246
	(0.60)	(0.37)	(0.36)	(0.16)
Observations	323	323	323	323
Pseudos R2	0.15	0.15	0.12	0.12
Type of estimation	Negbin	Negbin	Tobit	Tobit
Alpha	0.272***	0.266***		
	[0.041]	[0.041]		

Notes: all models include, as additional regressors, the type of institution (private or public), the year in which the subsidy was granted, and a set of dummy variables for the region, category and discipline of the project.
Robust standard errors are shown in parentheses. Coefficients correspond to marginal effects.
**Coefficient significant at the 10% level;*
***Coefficient significant at the 5% level;*
****Coefficient significant at the 1% level.*

It is important to notice that the value of the coefficients in the Negbin model cannot be read in the usual way. In order to simplify the interpretation, results are presented in terms of marginal effects evaluated at the mean values of the regressors.

Column (1) in Table 2 presents the baseline model. In this model, the coefficient associated to FONCYT is positive and significant at the 1% level. The value and sign of the marginal effect on FONCYT suggests that the "representative" funded researcher has managed to publish approximately 1.6 more articles than those non-funded researchers, always evaluating at the mean value of the regressors.

As expected, the coefficient associated to Publications$_0$ is positive and significant at the 1% level.[56] Other variables with significant explanatory power are Age and Doctorate. The coefficient on Doctorate is positive and significant at the 1% level, indicating that researchers with a doctorate publish more, *ceteris paribus*, than researchers without a doctorate. Age's coefficient is negative and significant at the 1% level, suggesting a decrease in researcher's productivity, in terms of the number of publications, with the passage of years.[57] The gender variable is not significant, implying that there is no "productivity puzzle", coinciding with the findings of Long (2001).

A common problem in regression analysis is the one of relevant omitted variables. In our case, the average peer review score is a measure of the quality of the project and other characteristics of the researcher exposed in the proposal that are only observed by evaluators. This variable is positively correlated with the granting of the subsidy and could also positively affect the outcome; therefore, omitting it from the regression generates a positive bias, which means that the coefficient on FONCYT estimated in column 1 is biased upwards. In fact, since we count with this variable, column (2) incorporates it and shows that it has a significant coefficient. In this specification, FONCYT continues showing a positive and significant (at the 5% level) coefficient, but its marginal effect is lower than the previous one, indicating that an omitted variable bias was actually present in our estimation. Once this bias is considered, results do not change in a significant way.

1. Robustness Checks

a) Model specification

An important aspect in this type of analysis consists in testing the specification of the model (i.e., the use of a Negbin model). The particular Negbin model used in this study, the Negbin II in Cameron and Trivedi (1998), assumes that output dispersion is a function of the expected average of the publications for each researcher. In order to justify the use of this specification instead of the other alternative Negbin one, which assumes constant over-dispersion for all observations, we can say that the former was the one with higher log-likelihood value in all the cases considered in this paper. Furthermore, the maximum

---

[56] Conclusions remain unchanged when Publications$_0$ is included in logs. In particular, the coefficient on FONCYT remains positive and significant at the 1% level. The coefficient on the log of Publications$_0$, which is interpreted as elasticity, is positive (its value is equal to 3.92) and significant at the 1% level.

[57] We tried a non-linear specification for Age, but the square term was not significant, thus the quadratic turndown usually found in the literature seems not to be present in our case.

likelihood estimation of the Negbin II model is robust to a misspecification of the form of over-dispersion posed for the distribution of the dependent variable.

Nevertheless, as the standard errors of the Negbin model are not consistently estimated if the exact specification is incorrect, regressions using the Poisson model were estimated by quasi-maximum likelihood with robust -sandwich- standard errors. In this case, results remain unchanged. The same happens when we consider the model in logs using Ordinary Least Squares with Huber-White standard errors.

b) Indirect test of the mean independence assumption

Although the mean independence assumption is not directly testable, this section applies one of its indirect tests analyzed in Imbens (2003). In particular, the model is re-estimated, but using $Publications_0$, a variable that should not be affected by FONCYT subsidies because it contains the number of publications prior to the granting of the subsidy.

In Table 3 results are presented. It is important to notice that, as can be appreciated in column 1, the coefficient on FONCYT is significant; signaling a bias in the estimation, and therefore, it is probable that a bias be also present when using $publications_1$ as dependent variable. Nevertheless, when the Evaluation variable is included in column 2, the effect of FONCYT vanishes, which is consistent with the previous finding on the importance of including this variable. It could be the case that the peer review score reflected some inherent qualities of researchers, such as ability, thus its significance for explaining the number of pre-FONCYT publications. Once this is taken into account, there is some evidence to support the selection on observables assumption.

Table 3: Indirect test of mean independence assumption

	Dependent variable: $Publications_0$	
	(1)	**(2)**
FONCYT	2.311***	1.115
	(0.544)	(0.725)
New group	-0.323	-0.254
	(-0.709)	(-0.668)
Evaluation		0.938**
		(0.311)
Age	0.055	0.043
	(0.043)	(0.039)
Doctorate	2.305***	2.342***

	Dependent variable: $Publications_0$	
	(1)	**(2)**
	(0.627)	(0.583)
Gender	0.472	0.279
	(0.552)	(0.549)
Observations	323	323
Type of estimation	Negbin	Negbin

*Notes: all models include, as additional regressors, the type of institution (private or public), the year in which the subsidy was granted, and a set of dummy variables for the region, category and discipline of the project. Robust standard errors are shown in parentheses. Coefficients correspond to marginal effects. **Coefficient significant at the 5% level; ***Coefficient significant at the 1% level.*

c) New control group

Data on a group of 88 scientists that are members of the CONICET[58], and have never applied to FONCYT funding are available. As Heckman et al. (1997) do, it is possible to test indirectly the assumption of mean independence by estimating a treatment effect using only researchers that applied and were not funded and those that never applied, where the treatment variable is being a member of one of the two control groups.

A new dummy variable was created that takes the value of 1 if a researcher is member of the CONICET, and considering both non-funded groups the model was re-estimated. The result indicates that this new variable was negative and significantly different from cero, indicating that a "representative" researcher of the CONICET that never applied to FONCYT publishes approximately 1.3 papers less than those who did apply and were not funded, at the mean value of regressors.[59] Besides the logic of this result, i.e., the possibility that those not applying do not do it because of their antecedents; this non-zero treatment effect indicates that at least one of the controls groups is not a valid comparison.

However, one should take the previous result with caution. When the regression in the previous paragraph is repeated using $Publications_0$ as dependent variable, there continues to be a significant negative treatment effect indicating that a bias is present. Moreover, in this case, the evaluation received by the projects from the granting agency is not an available control (of course, those researchers that did not apply do not have an evaluation score), so it is not possible to avoid the bias as before by including this variable. This implies that the control group of

[58] See footnotes 8 and 35.
[59] All results mentioned but not showed are available upon request.

scientists that applied but were not funded is the more appropriate one to gauge the treatment effect of FONCYT subsidies.[60]

d) Quality of research

Since our dependent variable is the number of publications, a potential bias could also arise if the quality of publications were correlated with FONCYT. For example, a possibility would be that funded researchers had incentives to publish more papers in order to show an increase in productivity. If the increase in the number of publications occurs at the expense of a loss in its academic quality, then the coefficient on FONCYT would be overestimating the impact of the program. This is the case in Goldfarb (2001), who finds that the NASA program, being an oriented source of funding, increases the publications of low-cited researchers, impinging on the quality of their publications.

In order to address this potential problem, in Column (3) of Table 2 is presented a model having as dependent variable the sum of the impact index of publications in the period after the subsidy ($Impact_1$)[61]. In this model, FONCYT continues being positive and significant at the 1% level, indicating that it has not fostered publications at the cost of their quality. When controlling for the quality of the project (peer evaluation), in Column (4), it is possible to see that the coefficient on FONCYT continues showing a positive and significant coefficient at the 10% level. However, its value is reduced, which adds evidence to the omitted variable bias previously mentioned, together with the fact that the coefficient associated to Evaluation is also positive and significant at the 10% level.

B. Matching Methods

Matching estimators are another way to deal with selection on observables. They rely upon the same conditional mean independence assumption as regression with controls. Traditional matching estimators pair each program participant with an observably similar non-participant in terms of some selected variables and interpret the difference in their outcomes as the effect of the program.

The key difference between matching and the regression with controls approach is that regression makes the additional assumption that simply conditioning linearly on X suffices to eliminate selection bias. Matching is non-parametric

[60] Nevertheless, regressions were re-estimated using only researchers that did not apply as the control group, and the full sample contemplating both control groups. In all cases FONCYT remained positive and significant.

[61] A Tobit model was used in order to take into account the “left-censored” character of the dependent variable.

and, as such, avoids the functional forms restrictions implicit in running a linear regression (Smith, 2000). Avoiding functional forms restriction can be important to reducing bias (Smith and Todd, 2003)[62]. Moreover, matching highlights the so-called common support problem, in the sense that it makes it easy to see when there is no non-participant to match with for some participants.

The main purpose of matching is to re-establish the conditions of an experiment when no randomized control group is available. The matching method aims at constructing the correct sample counterpart for the missing information on the treated outcomes had they not been treated by pairing each participant with members of the non-treated group. Under the matching assumption, the only remaining difference between the two groups is program participation. All that is required in order to estimate the average treatment effect on the treated is that there be analogues for each treated researcher in the non-treated sample.

In the literature on the matching samples construction one can find several approaches to construct the control group. Supposing that X contains only one variable, it would be intuitive to look for an individual as control observation that has exactly the same value in X as the corresponding participant. If the number of matching criteria is large, however, it would hardly be possible to find any control observation. To solve this problem, Rosenbaum and Rubin (1983) developed the so-called 'propensity score matching'. The idea is to estimate the propensity score of participation, *P(X),* for the whole sample and find pairs of participants and non-participants that have similar probability value of participation. In this way, by using the propensity score, matching reduces the multidimensional problem of several matching criteria to one single measure of distance.[63]

When using the propensity score, the comparison group for each treated individual is established by a pre-defined measure of proximity. Having defined the common support, the next issue is that of choosing the appropriate weights to associate the selected set of non-treated observations for each participant one. Several possibilities are commonly used; this work presents matching estimates for two different schemes of weighting[64] : kernel matching and radius matching. In kernel matching each treated observation is matched with a weighted average

[62] Another important difference when treatment effects are heterogeneous consists in the different weighting schemes they apply.
[63] Rosenbaum and Rubin (1983) show that if mean independence assumption holds conditional on X, it also holds once one condition on the propensity score estimated by means of that vector X.
[64] Asymptotically, all different matching estimators produce the same estimate, because in an arbitrarily large sample, they all compare only exact matches. In finite samples, different matching estimators produce different estimates because of systematic differences between them in which observations they assign positive weight and how much weight they assign them.

of all controls units, and the weights are constructed on an inversely proportional way to the distance between treatment and controls' estimated propensity scores. In radius matching each treated observation is matched with the control units that have an estimated propensity score differing less than an established distance from the score of the corresponding treated unit.

Formally, weights for kernel matching are given by

$$w_{ij}^{K} = \frac{G\left(\frac{p_j - p_i}{b_n}\right)}{\sum_{m \in C} G\left(\frac{p_m - p_i}{b_n}\right)},$$

where p is the estimated propensity score, $G(.)$ is a kernel function and b_n is the selected bandwidth parameter. We used the Gaussian kernel and chose a bandwidth parameter of 0.16 on the basis of visual selection combined with sensitivity analysis.

In radius matching, the weights are constructed as

$$w_{ij}^{R} = \frac{1}{N_{C_i}} \; if\ j \in C_i ; w_{ij}^{R} = 0 \;\; if\ j \notin C_i ; \;\; C_i = \left\{ p_j : \left| p_i - p_j \right| < r \right\};$$

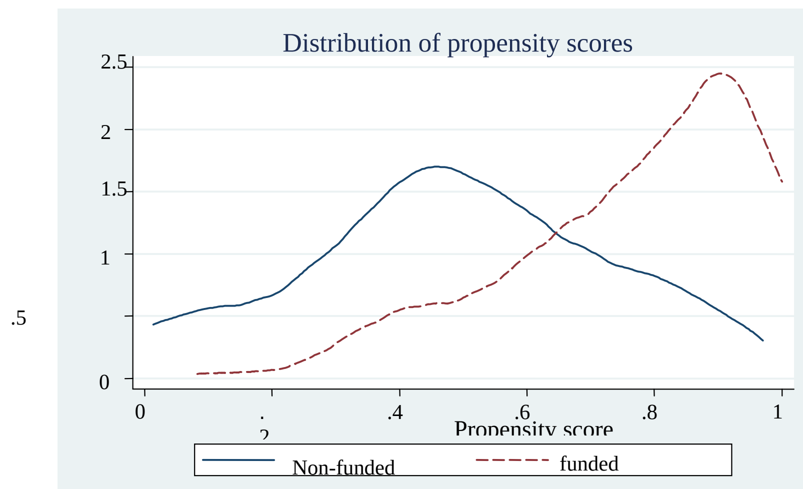
where, C_i is the set of control units matched to the treated i, and r is the radius that determines the maximum propensity score distance for the controls to be considered for matching.

It could be the case that for some funded researchers there are not comparable non-funded researchers in terms of the propensity score. Figure 2 presents a non-parametric kernel estimation of the propensity score densities for both groups; there it is possible to appreciate the lack of overlap, especially for low values of the propensity score. To deal with this problem, it is possible to restrict the estimation to the common support, which in fact comprises all observations whose propensities scores are in the intersection of the supports of both groups. In our case, the estimation is restricted to the sub-sample obtained by deleting all observations of non treated researchers with an estimated propensity score smaller than the minimum one estimated for the treated group.

This way of estimating the common support is consistent with the estimation of the effect of the treatment on the treatment. In contrast, Heckman et al. (1998) suggests a way that also excludes some treated observations, this could improve the quality of the matches, but it is no longer clear whether the estimate obtained in that way could be interpreted as an average treatment on the treated estimate[65].

The propensity score used to obtain the common support sample was obtained by means of a Probit model, where the dependent variable is FONCYT and the independent variables are Publications_0, Impact_0, Age, Doctorate, Gender, New Group, and a set of dummy variables for the region, category, year in which the subsidy was granted, and discipline of the project. That is, the same variables used as controls in the regression with control approach[66].

Figure 2. Non-parametric kernel estimation of propensity score's densities.



The general form of the matching estimator is given by:

$$\beta^M = \frac{1}{N_T} \sum_{i \in T} \left\{ Y_i^T - \sum_{j \in C_i} w_{ij} Y_j^C \right\}$$, where C_i and w_{ij} are defined as before, and T and C indicate, respectively, the treatment and the matched control groups.

[65] I thank Andrea Ichino for this observation. All matching regressions in this work were done using the STATA program he developed together with Sascha Becker, for references see Becker and Ichino (2002).

[66] This specification of the Probit model satisfies the tests for the "balancing property", that is the balancing of pre-treatment variables given the propensity score. This implies that treatment and controls matched on the propensity score should be observationally identical on average. When we introduce the "evaluation" variable, the specification must be slightly modified in order for it to accomplish the balancing property, but results practically do not change.

Results are reported in Table 4. In the two cases standard errors were obtained by bootstrapping with 500 replications. We re-estimated the propensity score at each replication of the bootstrap in order to account for the error that comes from both Probit estimation and determination of the common support.[67] All matching estimates are significant at the 1% level and their values and higher than the ones obtained through parametric regressions.[68]

Table 4. Matching estimates

	Observations: 218 treated and 97 controls	
	Dependent variable: Publications	**Dependent variable: Impact Index**
	(1)	**(2)**
Kernel matching[a]	3.315***	7.508***
	(0.857)	(2.221)
Radius matching[b]	3.370***	7.608***
	(0.818)	(2.085)

Notes: bootstrapped standard errors (500 reps) are shown in parentheses.
[a] The bandwidth used is 0.16. [b] The radius used is 0.05
**Coefficient significant at the 10% level, **significant at the 5% level.*

C. Difference-in-Differences

Since we count with data for two different periods (pre and post FONCYT), it is advisable to apply a method that takes profit of the panel character of our data. Precisely, the advantage of the difference-in-differences approach is that of relaxing the selection-on-observables assumption, essential both for matching and regression with controls approaches. Now, selection on unobservable variables is allowed, as long as these variables are candidate specific and time invariant. This means that one important limitation of this approach is that it only controls for time-invariant candidate's unobservable variables.

The required assumption is that the mean change in the no-treatment outcome is the same for funded and non-funded researchers, formally:

[67] When using matching methods, it is important to get the correct standard errors. As pointed out by Smith (2000), in practice most researchers report bootstrapped standard errors. However, Abadie and Imbens (2006) show that bootstrapping is not valid for nearest neighbor matching due to lack of smoothness, so it could be not valid either for radius matching, nevertheless they pose that it works for kernel matching. For radius matching analytical standard errors were also computed and results did not result altered.

[68] Stratification and nearest neighbor matching were also applied. The former method gives very similar results to the ones presented in Table 4, the latter gives quite higher values; in all cases they remain positive and significantly different from cero.

$E(Y_{it}^{C} - Y_{it'}^{C} / D = 1) = E(Y_{it}^{C} - Y_{it'}^{C} / D = 0)$, where t is the time period after treatment and t' is the time period before treatment.

Under this assumption, the difference-in-difference estimator, which consists of the difference between the before-after difference in output for participants and the before-after difference in output for non-participants, consistently estimates the treatment effect on the treated. The validity or not of the difference-in-differences identification assumptions depends on the particular case under analysis. If one believes that part of the self-selection mechanism work through the observed covariates and that, given these covariates, what determines whether or not a researcher is granted a subsidy are researcher characteristics that stay more or less constant across time, then the difference-in-difference estimator is an acceptable estimation procedure. However, one could imagine sources of unobserved performance differences that vary across individuals and time. For example, applicants may decide to enter the grant competition when they have been enjoying unusually good or bad recent performance[69]. Any unobserved variation of this kind, if not experienced by non-funded researchers as well, biases the estimator.

In a regression framework, the difference-in-differences model is equivalent to:

$$Y_{it} = \beta_i D_{it} + \lambda X_{it} + \alpha_i + \mu_t + \varepsilon_{it} \qquad (3)$$

where Y_{it} is the research output of applicant *i* in time *t*, D_{it} is a dummy variable that takes the value of 1 if applicant *i* receives the grant in time *t*, β_i is the impact for candidate *i* from receiving a grant[70], and X_{it} is a vector of observable determinants of output. The unobservable determinants of research output are reflected by the last three terms. There is a time-invariant 'applicant effect' that represents permanent differences in performance among candidates (α_i) and a time-period effect common to all applicants (μ_t). The usual error term, which is assumed to be uncorrelated with the *X*'s and *D*, is represented by ε_{it} and accounts for temporary fluctuations in performance around the specific mean of each applicant.

[69] This last situation is known in the literature as the Ashenfelter's dip.

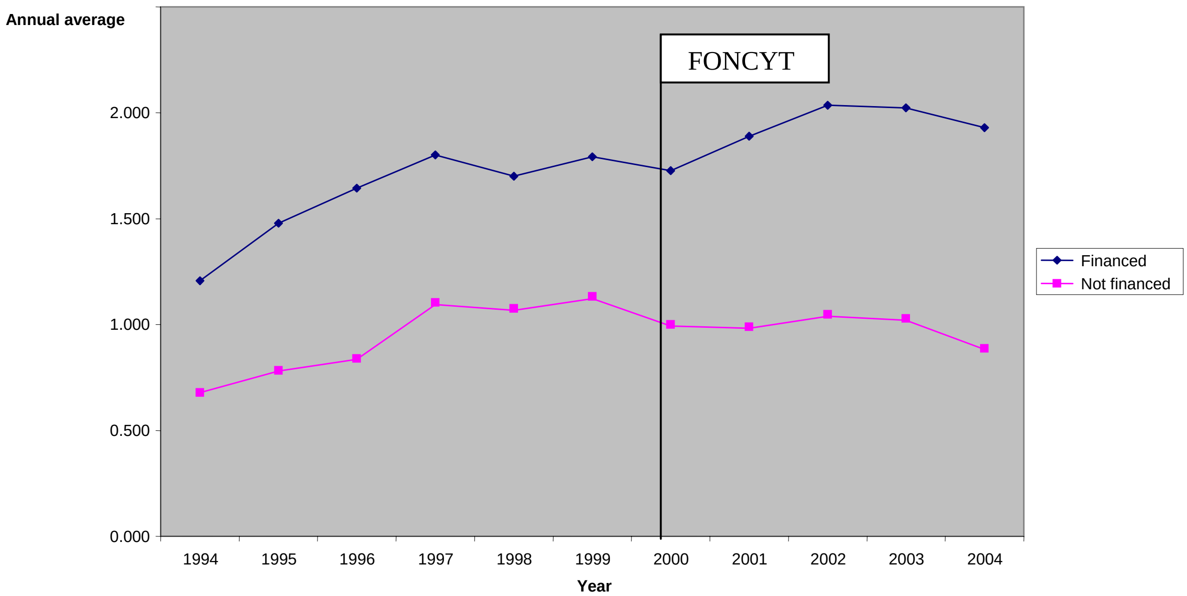
[70] Note that in this case, in contrast with regression with controls, the average treatment effect on the treated can be consistently estimated even under heterogeneous treatment effects.

The second crucial assumption of the method is that the time-period macro effect does not have a differential impact across the two groups. The same one can be informally checked by appreciating if there was a 'parallel trend' in the outcomes for the treated and control groups before the treatment, indicating that those parallel trends would have persisted if there program had not been introduced.

As shown in Figure 3, the trends in our main measure of academic output, the number of publications of funded and non-funded researchers were the same previous to the year 2000, when subsidies should have started to show their impact. This behavior in the evolution of the number of publications in treatment and control groups in the pre-treatment period validates our difference-in-differences identification strategy. It is worth noting that after the year 2000, publications of funded researchers continue increasing, while those of non-funded researchers remained at similar levels as in previous periods, suggesting a causal relationship between FONCYT funding and academic performance.

The selection of the control group is a crucial decision for any evaluation. In this case, data on scientists that never applied for the program are also available; they represent an alternative control group to the one of non-funded researchers who did apply for the subsidies. In the strategies based on the "selection on observables" assumption the first group was discarded since data on the evaluation score, which was considered a relevant observable control, were not available for researchers that never applied and, therefore, results using that control group would have been biased. In the same sense, this new potential control group does not provide appropriate data for difference-in-differences strategies, since it does not pass the informal test of parallel trends in the years before the treatment –i.e., if figure 3 is replicated using researchers that never applied instead of researchers that did apply and were not funded, non-parallel trends before the year 2000 are observed-.

Figure 3. Trends in the number of publications



Under the previously mentioned assumptions, the average effect of the grant program for those researchers receiving grants: $E\left(\beta_i / D = 1\right)$ can be consistently estimated.[71] Results corresponding to the difference-in-differences model are reported in Table 5.

Table 5. Difference-in-differences estimates

	Dependent variable: Publications				Dependent variable: Impact Index			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Foncyt	1.01	9.71***	1.09	9.79***	2.45*	13.13**	2.41*	13.09**
	(0.65)	(2.74)	(0.69)	(2.75)	(1.41)	(5.92)	(1.47)	(5.95)
Foncyt*Age		-0.17***		-0.17***		-0.22**		-0.22**
		(0.05)		(0.05)		(0.10)		(0.10)
Foncyt* Doctorate		0.47		0.47		2.73*		2.73*
		(0.70)		(0.70)		(1.55)		(1.55)
Foncyt* Gender		0.25		0.25		-1.41		-1.42
		(0.70)		(0.71)		(1.74)		(1.74)
Researcher	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[71] A typical concern when using difference-in-differences is the potential problem of serial correlation, which results in biased standard errors and generates over-rejection. As pointed out by Bertrand et al. (2004), however, the problem of over-rejection is mitigated when the panel is of length two (where observations have been aggregated into pre and post-treatment periods) and the number of treated units is large, as in our case.

	Dependent variable: Publications				Dependent variable: Impact Index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fixed-effect Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	646	646	630	630	646	646	630	630
R-squared	0.88	0.88	0.88	0.88	0.86	0.86	0.86	0.86
Type of estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

*Notes: Huber-White robust standard errors are shown in parentheses. Results in Columns (3), (4), (7), and (8) use the sample restricted to common support. *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.*

In Column (1) results corresponding to the baseline model are presented, without interaction effects. The coefficient associated to FONCYT in this specification is only significant at the 12% level, and its value indicates that, comparing the pre- and post- subsidy periods, the increase in the number of publications of financed researchers was greater than the increase in the number of publications of non-financed researchers in about one article.

Given the evidence that age, gender, and education may have an impact on researcher's productivity, it is worth exploring if FONCYT funds have a different effect according to these characteristics. It is possible to address this issue by estimating a model that includes the interaction terms between FONCYT and Age, Gender, and Doctorate. As shown in Column (2) of Table 2 the coefficient of the interaction term between FONCYT and Age is statistically significant, and its negative value suggests that the impact of FONCYT is more important for young researchers. Differential effects by gender or education were not found.[72] The coefficient on FONCYT remains positive, is now significant at the 1% level and its impact on publications, evaluated at the mean value of the regressors, is equal to 0.92, a figure similar to the one obtained in the model in Column (1).[73]

As pointed out by Heckman et al. (1998), one important source of bias in the difference-in-differences approach could arise if for some funded researchers there are not comparable non-funded researchers and vice versa. One way to deal with this potential source of bias is by applying the difference-in-differences

[72] The interaction between FONCYT and the average peer review score received by the proposals was also allowed, but the coefficient was highly not significant in all specifications. Similar results were obtained in the cases of the interaction effect with New Group and with the scientific areas defined in Section II.

[73] We reproduced the estimations in columns (1) and (2) using the Poisson and the Negative Binomial fixed effects models, used in Haussman et al. (1984) and also explained in Cameron and Trivedi (1998); and the principal results remain unchanged.

approach to the common support, which is defined and calculated as was done for matching estimates.

The common-support sample has 632 observations, 14 observations less than the full sample. As shown in Columns (3) and (4) in Table 5, results corresponding to the difference-in-differences in the common support approach are consistent with previous ones, suggesting that funded researchers increased the amount of publications in the four-year period after the subsidy in about one article *vis á vis* those researchers that were not funded, and that the impact is larger for young researchers.

Columns (5) to (8) in Table 5 report results using the sum of the impact index of publications as the measure of academic output. Results using impact indexes are similar to the ones using publications as the measure of academic output, thus providing additional evidence in favor of the hypothesis that FONCYT subsidy had a positive impact on the academic performance of funded researchers.

Again, the interaction term between FONCYT and Age is negative and significant, indicating a higher impact of the subsidy on young researchers. Interestingly, while the interaction term between doctorate and FONCYT was not significant when Publications was used as the measure of academic output, it becomes now positive and significant at the 10% level. This result suggests that although the impact of the subsidy is not significantly different for researchers with a doctorate degree and researchers without it, the impact of the subsidy on the quality of research is higher for researchers with a doctorate degree. In this specification the impact of FONCYT on the impact index, evaluated at the mean value of the regressors, is equal to 2.39, a figure similar to the one obtained in the model in Column (5).

D. Difference-in-Differences Matching Estimator

Another source of bias could arise in difference-in-differences estimations when the distribution of the variables on which we condition differs between treated and non-treated researchers, even within the common support. To avoid this bias, control group observations must be re-weighted. The difference-in-differences matching estimator accomplishes this task by combining both matching and difference-in-differences estimators (see Heckman et al., 1998 and Blundell and Costa-Dias, 2002). The estimator can be specified as:

$$\hat{\beta}^{M-DID} = \frac{1}{N_{T_S}} \sum_{i \in T_S} \left(\left[Y_{it} - Y_{it'} \right] - \sum_{j \in C_S} w_{ij} \left[Y_{jt} - Y_{jt'} \right] \right), \quad (2)$$

where t is the time period after treatment and t' is the time period before treatment, T_s and C_s indicate the treatment group and the matched control group in the common support, and w_{ij} represent the weights corresponding to researcher j matched to a treated researcher i.

Difference-in-differences matching estimates are presented for the two same schemes of weighting used in the case of simple matching: kernel matching and radius matching. For both options standard errors were again obtained by bootstrapping with 500 replications and the propensity score was re-estimated at each replication of the bootstrap.

Results are reported in Table 6. It is possible to appreciate that when matching is added to the difference-in-differences procedure, estimates are significant, and their values are higher than the ones reported in Table 5, but smaller that the ones for simple matching presented in Table 4.

Table 6. Difference-in-differences matching estimates

	Observations: 218 treated and 97 controls	
	Dependent variable: Publications	**Dependent variable: Impact Index**
	(1)	**(2)**
Kernel matching[a]	2.368*	4.008
	(1.413)	(2.721)
Radius matching[b]	2.599**	4.451*
	(1.330)	(2.730)

Notes: bootstrapped standard errors (500 reps) are shown in parentheses.
[a] The bandwidth used is 0.16. [b] The radius used is 0.05
**Coefficient significant at the 10% level, **significant at the 5% level.*

Column 2 of Table 6 shows results for the quality of publications. As before, difference-in-differences matching estimates are higher than those corresponding to the simple difference-in-differences estimator, but smaller than simple matching estimates.

V CONCLUSION

In this study the impact of FONCYT subsidies on academic performance of Argentine researchers has been evaluated, introducing a series of non-experimental program evaluation techniques into the Economics of Science. The performance of researchers with supported projects has been compared *vis á vis* a control group constructed using researchers that submitted projects that were accepted in terms of merit but not supported because of unavailability of funds.

Four alternative methods have been considered -with different robustness checks of their assumptions-: regression with controls, matching, differences-in-differences and differences-in-differences matching, they are based on different assumptions and thus it is no surprise that they generate dissimilar results. Nevertheless, in almost all cases it has been verified that the effect of having received the FONCYT subsidy is positive and statistically significant. Moreover, it has been shown that this effect is not obtained at the cost of lower quality publications. Thus, the empirical evidence indicates that FONCYT funding improves the academic performance of supported researchers, and according to some estimates, the effect of the subsidy is even stronger for young researchers.

Two caveats should be considered. First, this evaluation of FONCYT program focused on the bibliometric output of supported researchers since the funds were created to stimulate scientific research that could generate knowledge to be diffused within the public domain. Of course, as it has been analyzed in section 1, there could be other objectives in funding scientific activities and thus other relevant outputs to consider. Second, this study assumes, as most evaluation studies, that there are no spillover effects to the non-supported researchers. The question that merits further investigation is whether the performance of non-funded researchers can be considered independent of the support given to funded researchers, which would require the estimation of general equilibrium effects. In this direction, the developing of a structural model that describes the functioning of FONCYT would be a valuable input for future evaluations; the material presented in the first sections of this work can be relevant for that purpose.

Taking these caveats into account, results suggest that funding scientific research could be an effective way of promoting scientific activity in developing countries. This work could be the first step in the evaluation of Scientific Research Programs, especially in developing countries, which requires further evidence in order to ascertain the real value of fostering scientific activity for national development.

REFERENCES

Abadie, A. and Imbens, G. (2006) "On the failure of the Bootstrap for Matching Estimators", working paper available at http://elsa.berkeley.edu/~imbens/wp.shtml.

Adams, J. and Griliches, Z. (1998) "Research Productivity in a System of Universities", NBER Working Paper 5833.

Aghion, P., Dewatripont, M. and Stein, J. (2005) "Academia, the Private Sector and the Process of Innovation", mimeo.

Amin, M. and Mabe, M. (2000) "Impact Factors: Use and Abuse", *Perspectives in Publishing* No 1.

Angrist, J. (1998) "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants", *Econometrica*, 66(2), pp. 249-288.

Arora, A. and Gambardella, A. (1998) "The Impact of NSF Support for Basic Research in Economics", Available at SSRN: http://ssrn.com/abstract=163309.

Arora, A., David, A. and Gambardella, A. (1998) "Reputation and Competence in Publicly Funded Science". *Les Annales d'Economie et de Statistiques*, 49/50, pp. 163-198.

Arrow, K. (1962) "Economic Welfare and the Allocation of Resources for Invention", in R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton UP.

Arrow, K. and Capron, W. (1959) "Dynamic Shortages and Price Rises: The Engineer-Scientist Case", *The Quarterly Journal of Economics*, Vol. 73 (May), pp. 292-308.

Azoulay, P., Ding, W. and Stuart, T. (2005) "The Impact of Academic Patenting on the Rate, Quality and Direction of the (Public) Research Output", mimeo, September 2005.

Bartholomaei, M. (2005) "To Know is to Be: Three Perspectives on the Codification of Knowledge", Electronic Working Paper, No.131, SPRU: Science and Technology Policy Research.

Becker, S. and Ichino, A. (2002) “Estimation of Average Treatment Effects Based on Propensity Scores”, *forthcoming*, *The Stata Journal*, pp.1-19.

Bernier, C., Gill, W. and Hunt, R. (1975) “Measures of Excellence of Engineering and Science Departments: A Chemical Engineering Example”, *Chemical Engineering Education*, pp. 194-197.

Blundell R. and Costa-Dias, M. (2002) “Alternative Approaches to Evaluation in Empirical Microeconomics,” London: CEMMAP Working Paper CWP10/02.

Bonnacorsi, A. and Daraio, C. (2003) “A Robust Nonparametric Approach to the Analysis of Scientific Productivity”, *Research Evaluation*, 12, pp. 47-69.

Buchmueller, T., Dominitz, J. and Hansen, L. (1999) “Graduate Training and the Early Career Productivity of PhD Economists”, *Economics of Education Review*, 14, pp. 65-77.

Callon, M. (1994) "Is Science a Public Good?" *Science, Technology and Human Values*, 19, pp. 395-424.

Cameron, C. and Trivedi, P. (1998) *The Regression Analysis of Count Data*, Econometric Society Monograph No. 30, Cambridge University Press, Cambridge.

Carillo, M. and Papagni, E. (2004) “Incentive Structure in Basic Research and Economic Growth”, Working Paper No. 9.2004, Università degli Studi di Napoli “Parthenope”.

Chudnovsky, D. (2001) “Políticas de Ciencia y Tecnología y el Sistema Nacional de Innovación en la Argentina”, *Revista de la CEPAL* 67, Separata.

Coase, R. (1974) “The Lighthouse in Economics”, *Journal of Law & Economics*, University of Chicago Press, vol. 17(2), pp. 357-76.

Cole, S. (1979) “Age and Scientific Performance”, *The American Journal of Sociology*, 84, 4, pp. 958-977.

Cole, J. and Zuckerman, H. (1984) “The Productivity Puzzle: Persistence and Change in Patterns of Publications of Men and Women Scientists”, *Advances in Motivation and Achievement*, 2, pp. 217-258.

Cowan, R., David, P. and Foray, D. (2000) “The Explicit Economics of Knowledge Codification and Tacitness”, *Industrial and Corporate Change*, 9(2), pp. 211-254.

Crespi, G. and Geuna, A. (2004). "The Productivity of Science", Brighton: SPRU Report prepared for the Office of Science and Technology (OST), Department of Trade and Industry (DTI), UK. http://www.sussex.ac.uk/spru/publications/reports/CrespiOST.pdf

Crespi, G. and Geuna, A. (2005) "Modeling and Measuring Scientific Production: Results for a Panel of OECD Countries", SPRU Electronic Working Paper Series, No. 133, The Freeman Centre, University of Sussex.

Dasgupta, P. (2000) “Science as an Institution: Setting Priorities in a New Socio-Economic Context”, Text of a lecture delivered at the Plenary Session on Science in Society at the UNESCO/ICSU. World Conference on Science, held in Budapest, 26 June - 1 July, 1999.

Dasgupta P. and David P. (1987) “Information Disclosure and the Economics of Science and Technology”, in George R. Feiwel (ed.), Arrow and the Ascent of Modern Economic Theory, Chapter 16, Macmillan Press, London, pp. 519-42

Dasgupta P. and David P. (1994) “Toward a New Economics of Science”, *Research Policy*, vol. 23, n. 5, September, pp. 487-521.

David, P. (1992) “Knowledge, Property, and the System Dynamics of Technological Change”, *Proceedings of the World Bank Annual Conference on Development Economics* pp. 215–248.

David, P. (1998) "Common Agency Contracting and the Emergence of ‘Open Science’ Institutions", *American Economic Review*, 88 (Papers & Proceedings), pp. 15-21.

Debackere, K. and Glänzel, W. (2003) “Using a Bibliometric Approach to Support Research Policy Decisions: The Case of The Flemish BOF-key”, mimeo.

Demsetz, H. (1969) “Information and Efficiency: Another Viewpoint”, *Journal of Law and Economics* 12 (1), pp. 1–21.

Diamond, A. (1986) "The Life-Cycle Research Productivity of Mathematicians and Scientists", *Journal of Gerontology*, 41, pp. 520-525.

Duflo, E. and Kremer, M. (2003) "Use of Randomization in the Evaluation of Development Effectiveness", paper prepared for the World Bank Operations Evaluation Development (OED) Conference on Evaluation and Development Effectiveness in Washington, D.C.

Gambardella, A. (2001) "Economic Tools and Methodologies for the Design of Research Programmes in the Socio-economic Field", Expert's Evaluation prepared for the DG XII. Pisa, Sant'Anna School of Advanced Studies.

Geuna, A. and Nesta, L. (2005) "University Patenting and its Effects on Academic Research: the Emerging European Evidence", Forthcoming in *Research Policy*. Available in: http://www.sussex.ac.uk/Users/prff0/Publications/Geuna_Nesta%20University20patenting.pdf

Goldfarb, B. (2001) "The Effect of Government Contracting on Academic Research: An Empirical Analysis of Reputation in Research Procurement," Discussion Paper No 00-24, Stanford Institute for Economic Policy Research.

Gonzalez-Brambila, C. (2003) "The Determinants of Research Productivity: A Case Study of Mexico", mimeo.

Hausman, J., Hall, B. and Griliches, Z. (1984) "Econometric Models for Count Data and an Application to the Patents-R&D Relationship", *Econometrica*, 52, pp. 909-938.

Heckman, J. (1992) "Randomization and Social Policy Evaluation", in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs*, Harvard University Press.

Heckman, J. H., Ichimura, H., Smith, J., and Todd, P. (1998) "Characterizing Selection Bias Using Experimental Data", *Econometrica* 66, pp. 1017-1098.

Heckman, J., Ichimura, H., and Todd, P. (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", Review of Economic Studies 64 (4), pp. 605-654.

Heckman, J., LaLonde, R., and Smith, J. (2000) “The Economics and Econometrics of Active Labor Market Programs”, in *The Handbook of Labor Econ*omics, Volume 3A, O. Ashenfelter and D. Card (eds.), North-Holland.

Hicks, D., Breitzman, T. Hamilton, K. and Narin, F. (2000) "Research Excellence and Patented Innovation", *Science and Public Policy* 27 (5), pp. 310-320.

Hicks, D., Isard, P. and Martin, B. (1996) "A Morphology of Japanese and European Corporate Research Networks", *Research Policy*, 25, pp. 359-378.

Imbens, G. (2003) “Sensitivity to Exogeneity Assumptions in Program Evaluation”, *American Economic Review, Papers and Proceedings*, 93, pp. 126-132.

Jaffe, A. (1989) "Real Effects of Academic Research", *American Economic Review*, 79, pp. 957-970.

Jaffe, A. (2002) “Building Program Evaluation into the Design of Public Research Support Programs”, *Oxford Review of Economic Policy*, Spring (18), pp. 22-34.

Lazear, E. (1996) “Incentives in Basic Research”, *Journal of Labor Economics*, 15(1), pp. 167-197.

Levin, S. and Stephan, P. (1991) “Research Productivity Over the Life Cycle: Evidence for Academic Scientists”, *American Economic Review*, vol. 81, n. 1, March, pp. 114-32.

Lee, M.-J. (2005) Micro-Econometrics for Policy Program, and Treatment Effects. Oxford University Press.

Long, J.S. (editor) (2001) *From Scarcity to Visibility: Gender Differences in the Careers of Doctoral Scientists and Engineers*. Washington, D.C.: National Academy Press. Chapter 6: “The Academic Career”.

López, A. (2002) “Sistema nacional de innovación y desarrollo económico: una interpretación del caso argentino”, Tesis Doctoral, Buenos Aires.

Lundvall, B. (1992) *National systems of innovation*, Pinter Publishers, London.

Mansfield, E. (1991) “Academic Research and Industrial Innovation”, *Research Policy*, vol. 20, No 1, February, pp. 1-12.

Mansfield, E. (1995) “Academic Research Underlying Industrial Innovations: Sources, Characteristics and Financing", *Review of Economics and Statistics*, 77, pp. 55-65.

Markiewicz, K. and DiMinin, A. (2004) “Commercializing the Laboratory: The Relationship Between Faculty Patenting and Publishing”, Working Paper.

Merton, R. (1973) [1942] “The Normative Structure of Science”, in *The Sociology of Science: Theoretical and Empirical Investigations*. Robert K. Merton, ed. Chicago, IL: The University of Chicago Press.

Mollas-Gallart, J. and Salter, A. (2004) “Diversidad y Excelencia: Consideraciones sobre Política Científica”, Electronic Working Paper, SPRU: Science and Technology Policy Research.

Mowery, D., Nelson, R., Sampat, B. and Ziedonis, A. (2001) “The Growth of Patenting and Licensing by US Universities: an Assessment of the Effects of the Bayh-Dole Act of 1980”, *Research Policy*, 30:1, pp. 99-119.

Nelson, R. (1959) “The Simple Economics of Basic Scientific Research”, *Journal of Political Economy* 67(3), pp. 297-306.

Nelson, R. (2003) "The Market Economy, and the Scientific Commons," *LEM Papers Series* 2003/24, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Italy.

Nelson, R. and Winter, S. (1982) *An Evolutionary Theory of Economical Change*, Cambridge, MA and London: Harvard University Press.

Patel, P. and Pavitt, K. (2000) "National Systems of Innovation under Strain: The Internationalisation of Corporate R & D" in R. Barrell, G. Mason and M. O'Mahoney (eds.) *Productivity, Innovation and Economic Performance*. Cambridge UP.

Pavitt, K. (2005) “Public policies to support basic research: What can the rest of the world learn from US theory and practice? (And what they should not learn)”, Electronic Working Paper, SPRU: Science and Technology Policy Research.

Peterson, J. and Sharp, M. (1998) *Technology Policy in the European Union*, Macmillan, Basingstoke.

Polanyi, M. (1967) *The Tacit Dimension*, London: Routledge & Keagan Paul.

Quah, D. (2003) “Digital Goods and the New Economy”, Centre for Economic Performance London School of Economics and Political Science.

Romer, P. (1990) “Endogenous technological change”, *Journal of political economy*, vol. 98, n. 5, pp. 71-102.

Rosenbaum, P. and Rubin, D. (1983) “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 70 (1), pp. 41-55.

Rosenberg, N. (1990) “Why Do Firms Do Basic Research (with Their Own Money)?” *Research Policy*, 19, pp. 165-74.

Rosenberg, N. (1992) “Scientific Instrumentation and University Research”, *Research Policy* 21, pp. 381-390.

Salter, A. and Martin, B. (2001) “The Economic Benefit of Publicly Funded Basic Research: A Critical Review”, *Research Policy*, 30, pp. 509-532.

Scherer, F. and Harhoff, D. (2000) “Technology Policy for a World of Skew-distributed Outcomes”, *Research Policy*, 29, pp. 559-566.

Shell, K. (1966) "A Model of Inventive Activity and Capital Accumulation", in *Essays on the Theory of Economic Growth*, Cambridge: MIT Press (1967), 67-85.

Smith, J. (2000) “A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies”, UWO Department of Economics Working Papers 20006, University of Western Ontario, Department of Economics.

Smith, J. and Todd, P. (2003) “Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?” CIBC Working Paper Series No. 2003-5.

Sorenson, O. and Fleming, L. (2004) "Science and the Diffusion of Knowledge", *Research Policy* 33, No. 10, pp. 1615-1634.

Stephan, P. (1996) “The Economics of Science”, *Journal of Economic Literature*, vol. 34, n. 3, September, pp. 1199-1235.

Stephan, P., Gurmu, S., Sumell, A. and Black, G. (2004) “Who's Patenting in the University? Evidence from the Survey of Doctorate Recipients”, Working Paper.

Stern, S. (2004) “Do Scientists Pay To Be Scientists?” *Management Science*, 50:6, pp. 835-53.

Turner, L. and Mairesse, J. (2005) “Productivity differences across individuals in public research: an econometric study of French physicists' publications and citations (1980-1997)”. Forthcoming in *Annales d’Economie et de Statistiques*.

Verbeek, M. (2000) *A Guide to Modern Econometrics,* John Wiley and Sons.

Xie, Y. and Shauman, K. (1998) “Sex Differences in Research Productivity: New Evidence About an Old Puzzle”, *American Sociological Review*, 63, pp. 847-870.

Ziman, J. (1994) *Prometheus Bound: Science in a Dynamic Steady State*. New York: Cambridge University Press.

Inter-American Development Bank
Washington, D.C.