Maastricht University

UNITED NATIONS UNIVERSITY
UNU-MERIT

Working Paper Series

#2018-031

A guide for the evaluation of programs of human capital training for science, technology and innovation

Diego Aboal, Marcelo Perera, Ezequiel Tacsir and Maren Vairo

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://www.maastrichtuniversity.nl/governance

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

Maastricht Graduate School of Governance
MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

A Guide for the Evaluation of Programs of Human Capital Training for Science, Technology and Innovation

Diego Aboal[1], Marcelo Perera[1], Ezequiel Tacsir[1,2] and Maren Vairo[1]

1 CINVE (Centro de Investigaciones Económicas), Montevideo, Uruguay

2 UNU-MERIT, Maastricht, the Netherlands

July 2018

Abstract

We provide a practical guide for impact evaluation of Training and Human Capital programs in Science Technology and Innovation (STI). This document addresses specific challenges that arise when evaluating this type of programs, discussing its logic, the advantages and drawbacks of the different sources of information, the strategies which may be appropriate for evaluation, and the suitability of applying the different experimental and quasi-experimental available methods. For each technique, the document highlights the characteristics and assumptions, the strengths and weaknesses, and the practical issues related to their application to programs of human capital training for STI. Also, some specific issues, as for example the time after which the effects and externalities are expected to materialize, are discussed. Discussion is based on specific examples of existing evaluations.

JEL Classification: H43, C01, O15, O22, O31, O32, O38.
Key words: Impact Evaluation, Human Capital, Training Programs, Science, Technology and Innovation, Effectiveness for Development.

1. Introduction[1]

Innovation is an essential determinant of productivity levels and economic diversification (Aboal and Garda, 2016; Crespi and Zuñiga, 2012; Crespi, Tacsir and Vargas, 2014). Available evidence for Latin America and Caribbean (LAC) stresses that the lack of highly qualified personnel particularly in the areas of science, technology, engineering and mathematics (STEM) is one of the obstacles deterring innovation (Crespi, Maffioli and Rasteletti,. 2014).

The demand for STEM skills is growing worldwide, and is even more evident in rapidly growing occupations. Becker and Park (2011) estimate that the 75% of the fastest growing professions require STEM knowledge and abilities. In the same line, employment in STEM occupations displays rates which double the rest of occupations (Craig et al, 2011), as a consequence of technological change and its close relation with STEM-based abilities and tasks (Autor, et al. 2003; CEA, 2009; Fast Future, 2010). It should be stressed that not only high-tech companies demand people with STEM qualifications, also firms in other sectors such as financing, insurance and public services look for graduates in areas such as mathematics, physics, and engineering capable of performing complex predictive analysis and risk modelling (Craig et al., 2011).

Rapid technological change should only intensify this shortage, highlighting the need for tailored public policy. In this context, the unfulfilled excess of demand results from a combination of factors, including market failures related with information asymmetries between employers and employees and incomplete and opaque information about the (future) labor market (Tacsir, 2014).[2]

[1] This report has benefited from the support of the Inter-American Development Bank though the project RG-K1397 “Highly Skilled Human Capital Strategies for Economic Growth”. We appreciate the comments received by Juan Carlos Navarro and Jocelyn Olivari.

[2] In recent years, several contributions emphasize the importance of experience and access to high-quality information, together with the new roles of vocational counselors as ways of improving the satisfaction with the chosen career. Cunha and Heckman (2005) indicate that 30% of individuals would have taken a different decision about their education if they had known their ex post income. On the other hand, Arcidiacono et al. (2012) conclude that a considerable portion of students would have chosen a different discipline if they had been able to project their future incomes in a better way. Jensen (2010) and Emerson and McGough (2011) list a number of reasons why young people in developing countries face a higher uncertainty when choosing their education due to lack of information.

Several LAC countries have recently revamped their efforts towards the design and implementation of policies aimed at talent formation (through the education of their own professionals and educators), as well as its acquisition by means of attraction of foreign human capital and incentives to promote diaspora repatriation and/or activation. These interventions are often justified by their expected benefits in terms of innovation and productivity improvements. However, policymakers face the challenge of producing evidence supporting that such benefits exist. So far, several obstacles inhibiting the generation of comparable and robust evidence of the impact of such programs remain; fundamentally, scarce information, restrictions in accessing administrative registers and limited evaluation capabilities.

In fact, the vast majority of LAC countries have not succeeded, so far, in taking advantage of the mechanisms that allow use and access of the information that is spread throughout several organisms and administrative registers (i.e., social security, national taxes institutions, university records, among others), or tapping from novel sources of information such as electronic CVs or publications and patents databases in order to perform impact evaluations. Moreover, evaluations available in the region are based in the use of different indicators and dissimilar periods. These have negatively affected the possibility of sharing information and knowledge among countries, and ultimately the implementation of good practices for more effective policies.

In this context, the present work seeks to develop a methodological guide for impact evaluation and to propose a consistent approach to be implemented in different contexts. The implementation of this guidelines should, in turn, contribute to understanding the effects of those programs aimed at supporting the generation and acquisition of highly qualified human capital both at the productive and the scientific and technologic sectors. In particular, productive impacts could arise in terms of labor productivity improvement at the firm level; while scientific and technological effects might involve higher academic and technical production of researchers. Even though both types of impacts are relevant in understanding the importance of this kind of programs, in this paper we will focus in the second category. This decision is based on the fact that evidence regarding the impact of training programs at the

firm level is very scarce, which may be due to lack of data availability to perform this type of evaluations.

The document is structured as follows. Section 2 discusses the rationale behind programs that support training of advanced human capital and the expected impacts for different time horizons. Section 3 summarizes the availability of this kind of programs in Latin America. In section 4 we discuss different impact evaluation techniques. Section 5 presents the sources of information usually available for impact evaluations. Finally, section 6 concludes.[3]

2. Rationale

Economic theory suggests that human capital training produces benefits for society which transcend individual benefit appropriation. In general, economic growth models (see Lucas, 1988) are based on the premise that an upgrade in the endowment of human capital leads to a productivity improvement (as it favors at the same time labor and physical capital productivity). Moreover, empirical evidence points to the existence of a relation between total factor productivity, economic growth, and economic development (Hall and Jones, 1999). In this context, aggregate levels of education appear as a key ingredient when defining the success of an economy. However, as we will argue below, knowledge as an asset holds some characteristics that distinguish it from other tradable goods, thus leading agents to invest below its optimum level (underinvestment) under perfect competition. This justifies public intervention aimed at obtaining the socially optimum level of investment in training and knowledge acquisition.

2.1 Rationale for the implementation of policies supporting human capital formation

According to the human capital theory, the acquisition of abilities through education leads workers to acquire knowledge that increases their productivity, and in that way obtain better salaries. In this context, individuals decide how much and in what way to invest in their education, by maximizing their expected income path. In that sense, the decision to acquire one extra year of education implies, not only direct costs (such as enrolment fees, study

[3] The annex illustrates the cost benefit analysis technique with an application for a program in Uruguay.

materials, etc.) but also an indirect cost that results from spending time out of the labor market and thus resigning their potential wage. This fact is particularly relevant in the case of post graduates studies as they typically involve working-age adults. Following Mincer (1974), individuals will opt for post graduate studies if the opportunity cost associated to not being able to work during the duration of the course is compensated by higher revenue streams in the future.

However, as was mentioned above, even though decisions about human capital acquisition are taken individually, they have effects over the rest of the economy (often called externalities). According to Nelson (1959), the amount of resources allocated to education will not be optimum if the marginal value it reports to society is greater than the marginal value it reports to the individual making the investment. This is due to the fact that the agent does not contemplate the positive externalities of her own education when deciding how much to invest in it. While this type of decisions are made by pursuing individual profit optimization and given the such variable does not incorporate social benefits, in absence of public support, competitive economies will tend to invest less in education than should be desirable. Thus, Nelson concludes that it is in society's best interest to collectively support knowledge production. In this context, governments have developed a kit of measures which seek to encourage basic and applied research, as well as the formation of highly qualified human resources.

Arrow (1962) contributes to this discussion by proposing the idea that knowledge is a non-rival and (to a large extent) non-excludable good. The first characteristic implies that, whenever an individual consumes or makes use of knowledge, it does not tire out or become less available for others' consumption. As a result, knowledge, when being produced in competitive market conditions, has a negligible marginal cost and therefore its price would tend to be zero. This fact discourages investing decisions. Moreover, such market failure is reinforced by the non-exclusion of knowledge, as its intangibility and easy transmission result in the impossibility (or extreme difficulty) of excluding from consumption those who do not pay for it. In fact, given its zero marginal cost, knowledge will be under-demanded whenever its price is above zero. In

synthesis, generation of knowledge in free market conditions and without public intervention is expected to be far from its social optimum.

Arrow also highlights that knowledge operates in markets where there is uncertainty, due to the fact that the consumer often does not know its value at the moment of acquiring it. Such uncertainty may constitute an inhibitor to demand. Another market failure which affects knowledge production is indivisibility: at the moment of using a certain piece of information, it is not possible to only pay for the piece we are going to consume. Therefore, price setting does not take into account the specific use that the consumer will give to knowledge. From the supply side, it is generally the case that the knowledge production function is unknown, meaning it is difficult to establish a clear relationship between inputs and output[4]. As a consequence, production takes place under more risky conditions. Given the scarcity of insurance mechanisms against this source of risk, knowledge production is less than socially desired, while at the same time there are smaller possibilities of finding credits to finance these kinds of activities.

In summary, the existing gap between the private and social benefits derived from knowledge leads, under free market conditions, to a sub-optimal amount of resources allocated to its production[5].This fact opens a window of opportunity for the public sector to support this kind of activities in order to reach optimal knowledge production, and therefore activate the mechanisms of economic growth. Moreover, in the context of the most recent theories of development (for example Sen, 1999) education is postulated as an end in itself, as it enlarges the capabilities that individuals possess in order to live in society. This has led the currently used development measures (as the human development index by UNDP) to incorporate education as a result variable. As a consequence, the support governments may give to

[4]This failure is more relevant in the case of basic sciences, in contrast with applied knowledge, given that it is less clear how the advances in basic sciences reflect in concrete applications in the productive sector. However, following Nelson (1959), since basic research is not restricted by the necessity of solving a certain practical problem, the potential social gains which can be derived from discoveries in this area are higher.

[5] This situation is even more serious if the information the individuals use to determine their individual optimum investment is based on information which is not of easy access or processing, or that contains mistakes. Several contributions have focused on analyzing the effects of the instruments of delivery of information to future and current students. Some examples include: Wiswalland and Zafar (2014), Zafar (2011 & 2013), Arcidiacono, Hotz, and Kang (2012), Stinebrickner and Stinebrickner (2012a & 2012b) and Jensen (2010).

education may seek broader objectives than just the mere pursuit of higher productivity levels and economic growth.

2.2. Expected results: theory of change

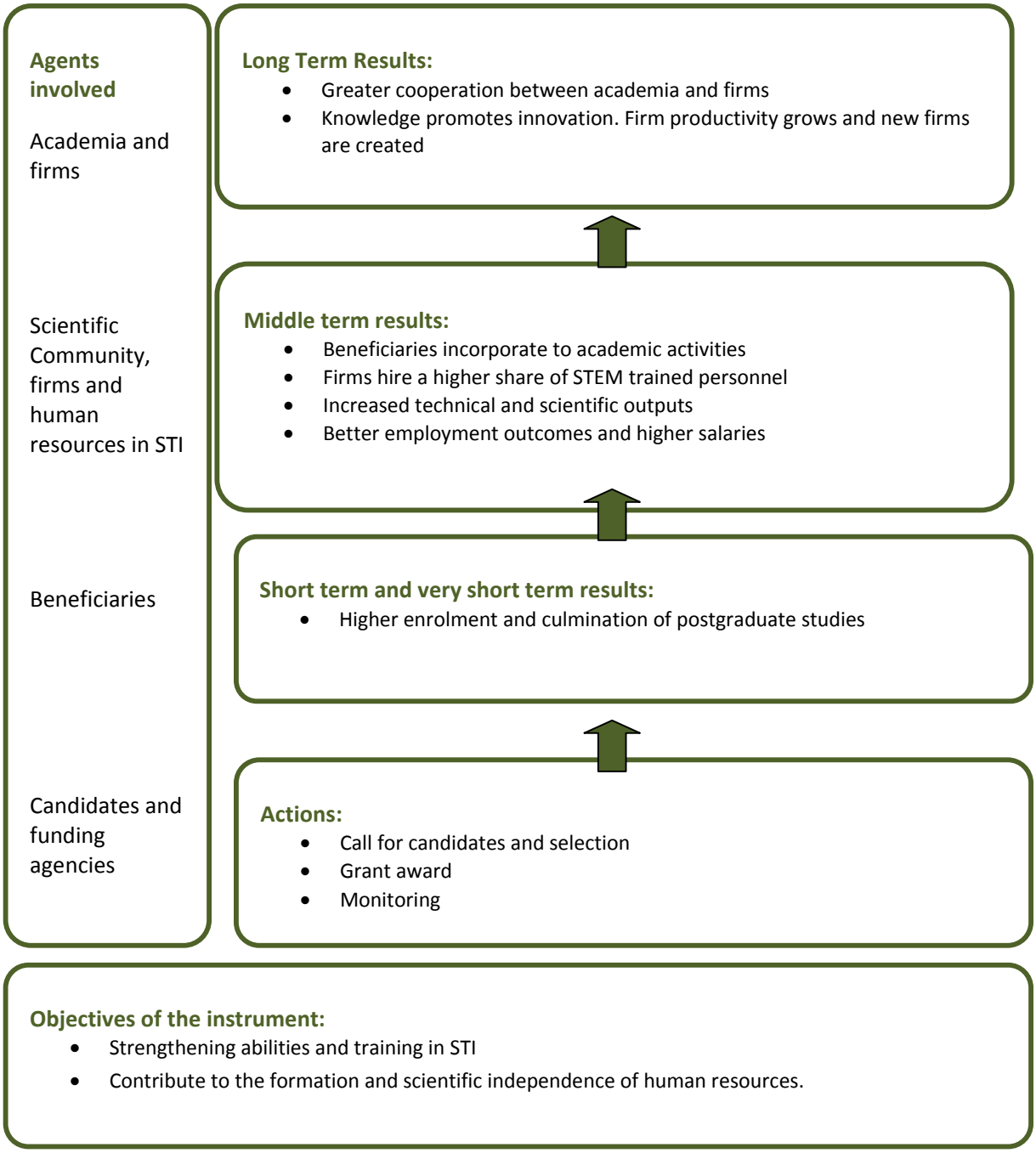
Figure 1 summarizes the main mechanisms associated to a generic support program to postgraduate studies. In general, programs supporting higher levels of education aim to encourage the scientific community and to enable the development of innovations through knowledge creation that attends local demands. For that, the selection guidelines for these programs seek to capture the candidates with the best academic level, and to monitor the beneficiaries in order to verify compliance with the scholarship requirements.

It is to be expected for a program of this kind to affect several groups of agents, generating results in different economic variables and in different periods of time. The transmission mechanisms through which the program creates impacts may be sketched through the theory of changes illustrated in Figure 1. A first expected result, in the very short term, is that the program leads its beneficiaries to enter a post graduate program. This result is not that evident as one may think, given the potential *crowding out* between public and private funds. This means that the beneficiary may have taken up post graduate education even when she had not been selected to participate in the program, by means of her own financial resources or through funding from another program. In the case of *crowding out*, the impact of the program in the decision of acquiring post-graduate education will be less than expected or non-existent. The same happens for the conclusion of postgraduate studies: the program will have impacts in this variable as long as it affects the decisions of entering a program of this type and of continuing in it while complying with the scholarship requirements.

On the other hand, one would expect to observe middle-run impacts on the type of labor participation of the scholarship holder once her studies are finished. Thus, an expected result would be a higher propensity to devote (totally or partially) to academic activities, given that the academia tends to concentrate highly qualified human resources. Consequently, this might also lead to higher bibliographic production from the beneficiary, given her decision to take up academic activities. At the same time, the greater level of education may lead the beneficiary to

have bigger chances of being selected for tenders or calls for non-academic projects (as for example professional consultancies or development of technological products), resulting in a higher level of technical production. As a consequence of their productivity growth, beneficiaries could also access higher salaries or better working conditions.

Figure 1. Theory of change programs aimed at advanced human capital formation



Source: Own elaboration.

In the long term, and as it was outlined in the previous section, it is expected that the the increase in individual productivity and the increase in knowledge generation should create impacts over other actors (what was previously referred to as externalities)[6]. In the first place, the conformation of a solid academic community, equipped with first class human resources, would contribute to the consolidation of a knowledge-based society, in which knowledge generation would penetrate entrepreneurial activity. This would lead to firms investing more in Research and Development (R&D), through the hiring of external services or even through the conformation of R&D teams within the firm. Following the model by Crepon, Duguet and Mairesse (CDM)[7] (1998), an increase in firms' R&D expenditure may result in the developing of innovations. In turn, according to the CDM model, such innovations will plausibly affect firms' productivity.

There are other mechanisms through which a bigger number of workers with post graduate formation could affect the economy's aggregate productivity. In particular, a worker with a higher education level might also generate positive productivity spillovers to her coworkers (Moretti, 2004; Winters, 2015). In the first place, knowledge transmission from the beneficiary to its colleagues may lead to a productivity upgrade involving her entire division. Moreover, synergies may emerge from the interaction between highly qualified workers which lead to new ideas and more efficient processes. Lastly, as Borgas (2002) states, substitution between qualified and unqualified workers may occur. As workers move towards more qualified jobs and the demand for unqualified workers remains unchanged, non-qualified workers may see their productivity, and hence their salary, augmented.

[6] It is important to mention that in order for post graduate studies to have an impact in the economy it is important that the beneficiary resides in her country after culmination (or at least to maintain strong bonds with the national academic community). This is not a minor consideration, as it is usual that the programs finance studies abroad, and due to this, it is possible that the beneficiary decides not to return to her country after completion. In this sense, it is common for the scholarships programs to include a special clause in which they demand beneficiaries to return to their country once the subsidy is finished.

[7] The Crépon-Duguet-Mairesse (1998) – CDM – model is a recursive system of three blocks of equations to assess the impact of innovation on productivity at the firm level. Specifically, a first block explains the determinants of the probability to do R&D and of the intensity of R&D. A second block explains the determinants of the probability to be innovative and the extent of product and/or process innovation. Finally, the productivity equation depends on innovation output.

Although it is always important to think of the theory of change behind the program at the time of evaluating it, we shall notice that a diagram as the one presented above is always a simplification of a rather much complex reality. In particular, the delimitation between different time horizons might not always be clear. For instance, it is likely that beneficiaries increase their academic production even before culminating their postgraduate studies. On the other hand, that the impacts of a program materialize at different periods, it is important to be careful when selecting the outcome variables for the analysis. In that regard, it may be that program implementation is too recent in order to identify long term impacts.

Next, we list some indicators that might be useful when measuring the above mentioned outcomes.

Table 1. Outcome measurement

Term	Direct impacts (on the beneficiary)	Indicator
Very short-term	• Entry to a postgraduate program	• Entry to a postgraduate program
Short term	• Culmination of postgraduate studies	• Culminated postgraduate studies • Years of education
Middle term	• Participation in academic production • Increase in academic and technical production • Occupational choice	• Works as a researcher • Belongs to the National System of Researchers or receives funds from research programs • Bibliographical publications • Citations • Technical outputs • Salary • Duration of unemployment
	Indirect impacts **(on other agents)**	

Long-term	• Greater cooperation between academic and productive sectors. • Innovations • Higher productivity	• % of firms which perform R&D • Entrepreneurial investment in R&D • Employment in R&D activities • Publications in cooperation with entrepreneurs • Patents • % of innovative sales in total sales • Labor Productivity • Total Factor Productivity

In the first place, the expected impacts for the very short, short, and medium terms would affect mainly the scholarship beneficiary, so that we would need information at the beneficiary level in order to assess those impacts. Information about entrance to post graduate programs and its culmination may be collected directly, and the same goes for the incorporation of individuals to academic activities after program completion. On the other hand, bibliometric indicators, frequently used in the literature, would allow to quantify some indicators of researchers´ academic production, such as the number of publications weighted by the number of citations or by the journal's ranking. Technical production may be quantified also by counting the number of technical products created by the beneficiary. Finally, labor conditions may be measured mostly through its salary dimension, given that it is rather difficult to observe other working conditions.

Further, measuring the externalities caused by these programs is even more cumbersome. Generally, the existence of spillovers makes it more difficult to identify policy impacts, given the resulting need to design an identification strategy that copes with the transmission effects from beneficiaries to non-beneficiaries. For instance, an identification strategy for impacts over labor productivity that relies in the comparison between those who were scholarship holders against those who were not would be under-estimating the scholarships impact whenever the program generates spillovers which positively affects the "control group's" productivity.

In the Table, we propose a few variables that approximate outcomes that might be affected as a consequence of the program's spillover effects. The effects in terms of cooperation between

the academy and the productive sector may be reflected in changes in some ratio of adoption of R&D by firms, or by the number of published articles as a result of collaborations. On the other hand, new innovations are frequently quantified through patent achievement or through the share of innovative sales over total sales. Lastly, the identification of the program's impact on firm productivity could make use of measures of labor productivity (a possible measurement arises from the number of workers to total sales ratio) or of total factor productivity.

3. Advanced Human Capital Formation Programs in LAC

In the context of the adoption of a systemic approach towards STI that took place in LAC during the beginning of the 2000's, policy actions have seeked to complement the efforts aimed at fostering innovation at the firm level with measures which tend to consolidate a structural framework of human resources and a research system which favors innovative development (Crespi et al., 2013). This has led to an important diversification of the existing STI policies. Particularly, and in a context of emerging new agencies focused on designing and implementing this type of policies, different instruments aimed at favoring the formation of highly qualified human resources have started to proliferate.

Table 2 shows a general outlook of the existing measures that support training of highly qualified human capital in LAC. First of all, we confirm that almost all countries considered use traditional support measures through post-graduate scholarships programs in the country or abroad. Generally, these programs select the most qualified candidates on the basis of their own merits and the relevance of the study programs, and provide a subsidy which covers totally or partially the costs associated with attending the program. Also, in order to ensure that the financed studies will have an effective impact in the country of origin, the vast majority of scholarships for post-graduate programs abroad establish clauses where the beneficiary commits to return to the country once her studies are over.

Additionally, the data shows that there are other policies, not that widespread, which also form part of the group of STI policies in the region. To this respect, some countries have identified strategic areas of priority in terms of knowledge production and thus, they have been implementing specific programs to give support to the training of professionals in those areas. These prioritized areas may be chosen due to productive purposes (as for example the case of the program of "Fortalecimiento de capacidades y desarrollo del conocimiento científico, tecnológico y de innovación del sector hidrocarburífero" from Colombia or the scholarships for post-graduates and PhD programs in extractive industries in Uruguay), while at the same time others respond to broader social objectives. For instance, the program CAPES/SPM in Brazil aims to finance human capital training in topics of sex and violence against women; while in Argentina there is a program which is specifically aimed at training human resources in policy and management in STI. It is important to highlight that the traditional scholarship programs aforementioned (not assigned to a specific sector), in several cases propose priority areas that are favored when allocating funding, even though belonging to those areas is not an excluding requirement.

Recently, another area in which there have been developments is international cooperation, where postgraduate studies are financed bilaterally by both parts involved. In some cases, cooperation results from agreements between the governments of both countries (as for example the case of Fulbright scholarships to study in USA which are operative in countries such as Brazil and Uruguay), and allow the beneficiary to choose between different post-graduate programs in the country; while in other cases the innovation agency establishes bonds with some university or specific post-graduate program (see Uruguay's ANII agreement for Phd in Health Engineering in the University of Delft, in the Netherlands, for example). This type of agreements is desirable as they favor the establishment of scientific bonds and knowledge exchanges with other countries, while at the same time allowing both interested parts to share out the economic burden of the subsidy.

Another type of scholarships is that which results from coordination between the public sector authorities and the productive sector. For instance, in Argentina the novel program *"Becas de*

formación de postgrado y posdoctorales cofinanciadas po rempresas del CONICET" has companies associate with the public agency in order to finance postgraduate studies of their members. Given that the productive sector also has interests in the appropriation of the benefits derived from workers' productivity increase, is seems sensible that firms will devote their resources to this kind of investment, thus alleviating the burden of the public sector. Another positive fact of this funding scheme is that subsidies are given with the objective to address certain productive needs, affording higher guarantees that the generated knowledge will be in fact of applicable in the productive sector,

In a similar way, researchers' support programs seek to consolidate the link between advanced human capital and firms. An example is the case of the program *"Recursos humanos altamente calificados en la Empresa de Uruguay"*, where the Government funds the companies up to an 85% of the costs associated to hiring a highly qualified worker during two years. Such hiring shall be associated to the participation in a concrete project presented by the firm, and selection favors those which have the objective of introducing an R&D unit in the firm. Although these programs are not conceived with the purpose of supporting postgraduate studies, but are aimed to help firms' development, they may be relevant at the time of identifying how the endowment of highly qualified human resources impacts over firm performance. As was mentioned before, the impacts at a firm level are more difficult to identify, since this implies evaluating how a particular program in which the beneficiary is the worker, affects an agent external to her, as is the firm. In this case, as the firms are the beneficiaries, it becomes more feasible to carry out this type of evaluations.

Table 2. Public policies supporting the formation of highly qualified human capital in selected countries in LAC.

	AR	BR	CH	CO	CR	EC	MX	PY	PE	UY
Scholarships of national postgraduate studies										
Scholarships of postgraduate studies abroad										
Scholarships for studies in specific areas										

Scholarships of postgraduate studies abroad provided by International Cooperation Agreements.										
Scholarships coordinated with the Productive Sector										
Programs of support to the incorporation of researchers and interns in firms.										
Visits of foreign professors										
Stay in research centres										
Scholarships of initiation in research										

Source: www.politicascti.net (accessed December 2015).
Note: Green cells stand for the country having at least one policy instrument in the corresponding category.

On the other hand, international mobility programs, in which foreign professors visits are funded contribute to knowledge generation, since they enable contact with foreign skills. For instance, the Uruguayan program "Visita de profesores del exterior" favors scientific diffusion and high level formation by funding scholars' visits to impart postgraduate courses, participate in seminars and workshops, and collaborate in research. In a similar way, the funding programs for internships at research centers favor knowledge exchange, while at the same time enabling training and research experience. Moreover, this type of opportunities for gaining knowledge through experience may be conducted in different stages of an individual's education. In this sense, "introduction to research" scholarship programs allow students to acquire training in research during early stages in their educational path, while also encouraging students to choose an academic career later on and to undertake postgraduate studies in the future.

Even though this revision is not intended to be exhaustive, given that there might be spread programs not gathered here, it gives an insight with respect to the kind of strategies which have been adopted in the region. In general, it appears that programs of support to the formation of advanced human capital are wide-range in LAC, as all the countries considered have at least one type of support in this sense. Moreover, there are some cases which stand out (as Uruguay and Brazil) in which a broad package of support measures has been consolidated, which allows

to tackle the matter from different stand points and hence attempt to achieve an adequate endowment of researchers and highly qualified workers.

4. The tool box

4.1 The evaluation problem

The impact evaluation of programs sets to identify a causal relation between policy implementation and some outcome variable of interest. For example, in the case of scholarships programs for postgraduate studies, the objective could be to establish if having received a scholarship has an effect over the number of publications in reviewed journals by the beneficiary as well as to quantify the magnitude of the effect. In an ideal situation, we would obtain such effect by comparing the results obtained by the grant holder after receiving the subsidy against the potential results that same individual would have obtained in case of not having received the subsidy. The result which would have been observed in the absence of the program is denominated *counterfactual*.

The evaluation problem may be formalized through the Roy-Rubin model (Roy, 1951 y Rubin, 1974). If $Y_i(1)$ and $Y_i(0)$ denote the observed results for individual i in presence and absence of the program respectively, the effect of the program for such individual would be given by:

$$\tau_i = Y_i(1) - Y_i(0) \qquad (1)$$

The observed result would be $Y_i(1)$ in the case of receiving treatment, while we would observe $Y_i(0)$ otherwise. If we define a variable D which indicates the individual's treatment status, where D takes value 1 if the individual was treated and 0 if she was not, we can write the observed result as:

$$Y_i = Y_i(0) * (1 - D_i) + Y_i(1) * D_i \qquad (2)$$

In the case the interest is in identifying the impact of a program over the whole population and not over a particular individual, then the effect of interest is the Average Treatment Effect (ATE) which is formally given by the mean of τ_i:

$$ATE = E(\tau_i) = E[Y_i(1) - Y_i(0)] \quad (3)$$

Another option would be to obtain the impact of the program for the treated group, which results from the mean of τ_i only for the population who received treatment (in this case, the scholarship holders). Such effect is known as ATT (*Average Treatment Effect on the Treated*) and is obtained as:

$$ATT = E(\tau_i|D_i = 1) = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] \quad (4)$$

Likewise, we can define the impact over the non-treated (ATU, *Average Treatment Effect on the Untreated*) as:

$$ATU = E(\tau_i|D_i = 0) = E[Y_i(1)|D_i = 0] - E[Y_i(0)|D_i = 0] \quad (5)$$

While in the case of ATT it is not possible to observe the term $E[Y_i(0)|D_i = 1]$ that represents the expected value of the outcome variable for those who were treated in absence of treatment, in the case of ATU the counterfactual (or unobserved) result is given by $E[Y_i(1)|D_i = 0]$, as computing this requires knowing the result of the non-treated in the case of receiving treatment.

A possibility to obtain the treatment effect would be given by comparing the results obtained by those who were affected by the intervention (*treatment group*) with the results of those who were not treated (*control group*) after program implementation. This approach would be valid as long as the results of the control group are a good approximation to the *counterfactual* scenario. However, whenever individuals in the treatment group are systematically different in their characteristics from the control group, we would be mistakenly attributing an effect to the program which in fact would be (at least partially) due to these differences and not because of the intervention. In general, it is argued that the decision of participating in a program depends on individual characteristics, so that it is very likely that treatment and control groups are not

comparable without making further adjustments. As a result of this identification problem, known as *selection bias*, the impact evaluation literature has proposed a set of techniques which allow causal identification of treatment effects.

4.2 Randomized experiments

A possible strategy to evaluate programs while at the same time addressing the selection bias problem is the experimental design. In an experimental framework, the evaluator divides the target population in two, using a purely random criterion, and assigns treatment to one of the groups (treatment group) while the second group (control group) is not treated, serving as a reference to compare to the results of the treatment group. For instance, in cases in which there is an excess of demand to participate in a program, the policy-makers may randomly draw between all the enrolled, in order to select who are going to be beneficiaries of the program and who are not. As a result of this, treatment intake is not a decision variable anymore, and thus, the selection bias would no longer be a constraint. The randomized design is considered as a "gold standard" for impact evaluations, since it guarantees that the treatment and control groups are, in average, similar in terms of their members' observed and unobserved characteristics. In such context, it is possible to causally infer the impact of the program from the comparison between both groups.

In formal terms, the experimental methodology ensures that the mean results for the control group ($E[Y_i(0)|D_i = 0]$) constitute a good approximation to the desired counterfactual for the treated ($E[Y_i(0)|D_i = 1]$). Therefore, the basic model consists in estimating the following linear regression:

$$Y_i = \beta_0 + \beta_1 D_i + \gamma X_i + u_i \quad (6)$$

Where β_0, $\beta_1$and the vector γ represent the parameters to estimate. In particular, β_1 (the coefficient associated to the treatment variable) is the parameter of interest for the evaluation as it indicates the impact in the outcome variable Y_i that results from having participated in the program. X_i is a vector of individual characteristics. Even though it is not strictly necessary, it is

common to include this type of information in the estimation of equation (6) in order to improve estimators' efficiency. Lastly, u_i represents the error term of the equation.

Therefore, obtaining the treatment effect from an experimental sample is very simple since it boils down to comparing the result between the treatment and the control group. Nonetheless, the experimental design requires some key stages which are not trivial[8].In particular, the design must include a statistical power analysis. This will allow us to determine the required size to conduct the experiment or, in other words, the size of the effect that we will be able to estimate with a reasonable probability. The objective of the impact assessment is to contrast the null hypothesis that the treatment has no impact on the average value of the outcome variable. When determining the significance level, we ensure a limit of type 1 error, that is to say, to the probability of rejecting the null hypothesis when the treatment does not have an impact. However, when setting the significance level we are not bounding the probability of a type II error, i.e. we do not assign a value to the probability of accepting the null hypothesis when it is false, (this mean concluding that the program has no effect when in fact it does). The power of a test is measured precisely as the probability of not committing a type II error. In general, it is considered reasonable for the statistical power of a test to be at least 80% (Bloom, 2006; Duflo et al., 2006). This way, in the design of the evaluations, the power calculation is used to define the sample size of the experiment which enables to identify a certain impact with a power of at least 80%.

An advantage of the experimental framework is its capacity of identifying the impact of the program without resorting to sophisticated econometric techniques. This implies results that are clearer and more easily transmitted to policy authorities and the public in general. However, the first constraint of this method is its difficult application in social sciences given that it is not that frequent to find policies which have been implemented in an experimental setting. Moreover, despite the method being very strong in terms of internal validity (this means it provides reliable results), its external validity (the ability of extrapolating the results to the entire of the population) is more questionable. This is due to the fact that many times the

[8] A very good guide about experimental evaluations is Glennerster and Takavarasha (2013).

experiment is carried out under particular conditions that might not be representative of actual program implementation. The same can be said in the cases in which the experiment has a much smaller scale than the program itself. Lastly, there are ethical criticisms, as in some cases it may seem reprehensible not exposing the entire targeted population to the expected benefits of the intervention.

In the case of programs supporting postgraduate studies, we have not found any evidence that uses the experimental framework for impact evaluation. This may be explained by the fact that these are not massive programs, and there might not always be excess of demand for the program so that random allocation between those who are interested in the program is not possible. Even if this was the case, candidate selection is strongly linked to academic merits and qualifications so that treatment allocation is far from being random. In the next sections, we will discuss the empirical strategies to identify the impact of this type of programs when the implementation is not carried out in a purely experimental framework.

It is important to keep in mind that our knowledge about the impact of a program can be highly improved if program design includes an impact evaluation strategy for the future. It is good to bear in mind that even when the experimental design is the optimal way to evaluate the effect of a policy in terms of internal validity, this criterion comes in conflict with other policy and ethical issues. It is important to consider that the experimental design would determine the scholarship to be randomly denied to some candidates after it was recommended by an evaluating committee using pre-established selection criteria. It also implies awarding the benefit to candidates whose evaluation may be unfavorable.

Anyway, it can be useful during the design stage of the program to look for allocation criteria that, without being 100% random, introduce some exogenous variable in the assignment of the scholarship in at least one sub sample of candidates. For instance, the applications may be grouped after evaluation by the selection committee as 1)”Imperative”, 2)”marginal”, 3)”rejectable”.

Ideally, treatment inside each group could be randomized but with different assignment probabilities. Meaning, we can make a lottery in which the probability that the scholarship is

denied for the first group is small and is high for the third group. The limit of such probabilities is imposed by the number of cases (treatment and control) needed to obtain precise estimations of the program's impact (i.e. the statistical power of the experiment must be considered). In the limit, where ethical and political restrictions are priorities, we can set to zero the probability of not giving the scholarship to the first group and to one the probability of not giving it to the third group, letting allocation to the second group be random. This way, agencies in charge of the programs might be more open to randomization, at least in one subsample of candidates, even though these consensus solutions drives us away from the ideal experiment and reduces internal validity.

It is important to highlight that in the case of randomizing in the "marginal" group we will be in conditions of inferring a local effect of the program (Local Average Treatment Effect- LATE) and not an ATT. Under the reasonable assumption that the overall evaluation given to the applications by the evaluating committee is positively related to the expected impact of the scholarship (i.e. scholarship is recommended to those people who are expected to take a bigger advantage of it), then it is possible for the local effect to be an inferior limit of the ATT. This shall be informed to the program's authority so as to point out the "costs" that result from deviating from the "golden standard" of randomization.

Sometimes, even when it's not intended, the allocation rules for some programs is very close to random. This happens when the evaluation of the applicants is done through a score and the beneficiaries are those who exceed certain threshold of such score or those who are among the N higher scores (where N is the number of scholarships to be allocated). If the threshold (or the number of quotas) is reasonably exogenous it may be argued that applicants whose score is in a neighborhood of such threshold are reasonably similar and that program assignment (being above or under the threshold) is arbitrary. This particular case results in the Discontinuous Regression Method which we discussed later on.

In other situations, it may occur that the assignment rules impose restrictions such as to guarantee certain quotas for some minorities or population with certain attributes (affirmative actions). This situation may be exploited whenever it is reasonable to assume that such

attributes are not correlated with the expected result (i.e. exogenous variables) and at the same time that they are good predictors of the probability of receiving treatment. The idea is to exploit the exogenous variation given by those variables, called instruments, in order to identify the impact of the program. This is another method that we will present below, known as Instrumental Variables.

4.3 The matching method: Propensity Score Matching

A possible strategy that can be used for impact evaluation under non-experimental conditions consists of matching participant individuals with non-participants according to observable characteristics, so as to "control" for the selection or observable differences between treated and controls. *Propensity Score Matching* (PSM), the most common of these methods, proposes to compare each individual of the treatment group with its most similar counterpart in the control group in terms of their probability of receiving treatment, conditional on observed characteristics. This probability, which is a function of individual characteristics, has to be estimated. The PSM has then the particularity of using only one dimension to perform the matching: the conditional probability of receiving treatment. In this context, the impact of the program would be given by the average of the differences obtained after comparing every member of the treatment group with its matched counterpart in the control group.

The first step in the implementation consists of estimating the probability of program participation conditional on a set of observed characteristics. This allows synthesizing the multiple dimensions which affect program participation in a single measurement called Propensity Score (PS). Secondly, for each unit of the treatment group, we select the nearest control in terms of the mentioned measurement. The PS is estimated with a Probit or Logit model for the binary variable indicator of the treatment[9]:

$$p(X_i) = P(D_i = 1|X_i) \qquad (7)$$

[9]Estimation methods which are typically used when the dependent variable in the estimation is binary. Following Bernal and Peña (2011), both models generate similar results, and therefore, the decision on which one to use is not critical.

Where $\mathrm{p}(\mathrm{X_i})$ is the probability of participation or PS and $\mathrm{X_i}$ is the vector of observable characteristics not affected by program participation (predetermined). The important assumption on which this method relies is that vector X includes all the variables which cause the selection bias, meaning, the variables which are correlated with program participation and, at the same time, have influence in the outcome variable.

The validity of this method hinges on the assumption that, conditional on the PS, the potential result is independent of treatment assignment. The identification assumption is known as the Conditional Independence Assumption (CIA). In this context, in order to eliminate the selection bias it is enough to introduce all relevant variables as controls in our estimation. It is important to notice, however, that the assumption of CIA is rather strong, as it fails whenever participation is determined by unobservable variables[10]. The assumption is not formally contrastable although it is possible to discuss its validity through falsification tests (Imbens and Rubin,2010). It is important to consider that the X vector cannot include variables which have been affected by program participation. In this context, it would be desirable to include variables which remain unchanged in time or that were measured before the program was implemented.

Another requirement for the construction of the counterfactual is that all the members of the control group are sufficiently similar to their counterpart in the treatment group. This is called the overlap or common support condition (CS) of PSM which imposes that all individuals who are used for the comparison have a strictly positive probability of being treated as well as of not being treated. Formally, it is assumed that for every individual *i*:

$$0 < P(D_i = 1|X_i) < 1 \qquad (8)$$

Therefore, the next step in the evaluation would be to select the group of individuals which form the CS and to exclude from the estimation those who do not form part of it. Typically, this implies excluding from the control group those individuals which have a very small probability of receiving treatment, and eventually excluding from the treatment group those individuals

[10] Innate cognitive ability is an example of an unobserved variable which is expected to affect participation in educational programs.

whose PS is too high. However, there is not a unique way to choose the common support. Among the simplest, is the graphical analysis of the control and treatment groups' PS density, which defines as common support the region in which both distributions overlap. Another option consists of eliminating all the observations whose PS is above the maximum or under the minimum of the other group.

Once the common support is determined, we have constructed a potential control group that is comparable with the observations of the treatment group. The next step is the matching, in other words, selecting the controls for each treated unit. There are several matching algorithms to choose from. Although the discussion about each of them exceeds the scope of this work, some of the most frequently used options are the following[11]:

1. **Nearest neighbor estimator.** Every individual in the treatment group is matched with the individual in the control group which has the nearest PS. This option may be extended to consider several nearest neighbors. In such case, we compare the result of each treated individual with the average of the results of the n nearest neighbors, where n is the number of neighbors to be used and it is a parameter selected by the evaluator. Moreover, when computing such average, we may assign the same weight to the n neighbors or weight them differently according to the distance between their PS and the PS of the treatment unit we are trying to match.
2. **Maximum distance Matching.** In order to avoid matching one observation of the treatment group with a neighbor of the control group which is not sufficiently alike, we may impose a maximum level of tolerance for the distance between the probabilities of participation. Therefore, the individual of the treatment group is compared with all the individuals of the control group within the tolerated distance. In this case, the same criterion as in the case of the nearest neighbor is followed in order to compare the result of the treated unit with the average of the matched units in the control group.
3. **PSM estimator by stratification.** The probabilities of participation are divided into strata or intervals. Then, we compute the impact inside each stratum and the ATT is obtained as the weighted average of the impacts in each stratum, being the weights assigned to

[11] See for example Caliendo and Kopeinig (2005) for a discussion about matching algorithms.

each stratum equal to the proportion of individuals in the common support that belong to the that stratum.

4. **Kernel estimator and local linear regression.** Each unit of the treatment group is compared with a weighted average of all or a subsample of the individuals in the control group, in which the weight assigned to each observation depends on the distance between the respective PSs.

In general, the choice of an algorithm involves a *trade-off* between quality and quantity of matching, or between bias and estimation precision. For instance, the bigger the number of observations which form the control group for each treatment (the more the neighbors), the smaller will be the variance of the estimator (the more the precision); but the quality of the matching will be smaller (the more the bias). The opposite occurs when reducing the number of neighbors. However, in big samples, the different matching methods tend to give similar results. Therefore, it would be desirable to check that results do not change significantly when changing the matching method.

Once the algorithm is chosen, we must check the quality of the matching, by verifying that the treatment and the control group are similar in terms of the chosen characteristics (X). The purpose is to determine whether after conditioning for the probabilities or participation there are still differences between both groups. Two easy procedures which are generally used at this point are to test the differences in the mean of covariates X, and to graphically analyze the distributions of the propensity score between the treatment and control group. In both cases, the desirable outcome is to find evidence of similarity between both groups.

After analyzing the balancing of the sample, we can estimate the program effect. Otherwise, we should go back to previous steps and reconsider the specification of our PSM model.

The estimation of the impact of the program will be given by:

$$\tau_{ATT} = E_{P(X_i)|D_i=1}\{E[Y_i(1)|D_i = 1, P(X_i)] - E[Y_i(0)|D_i = 0, P(X_i)]\} \qquad (9)$$

This way, the program impact consists of the average of the differences in the outcome variable between the treatment group and its respective controls.

Lastly, we need to compute the standard errors from the ATT in order to assess the statistical significance of the estimated impact, or to obtain a confidence interval for the estimate. This step is not trivial; the variance of the treatment effect should contemplate not only the sampling variance but also the uncertainty given by the estimation of the PS and the imputation of the CS. A common option is to use bootstrapping (see for example Caliendo and Koepeing, 2005).

The PSM procedure for program evaluation can be summarized:

1. Estimate the probability of program participation conditional on observable characteristics and compute the predicted probabilities for each individual in the treatment and control group.
2. Restrict the sample to the common support
3. Select a matching algorithm
4. Check the balance between the treatment and control groups using the weights obtained from the algorithm selected in step 3.
5. Compute the effect of the program using the average of the differences in the outcome variable between the treated units and their respective controls
6. Obtain the standard errors of the estimation to assess the significance of the effect

The PSM methodology is frequently used in the impact evaluation literature due to its small information requirements. In fact, this strategy can be implemented with a single cross sectional dataset that collects the required information (outcome variable after treatment, treatment status, observable characteristics) for a sample of treated and control individuals. This fact distinguishes PSM from the rest of the methods we are going to see below, as they require information through different periods of time, thus requiring longitudinal data. Hence, baseline (before program implementation) and follow-up (post-treatment) information is required. However, the reliability of the results casted by PSM depends to a high extent on compliance with the CIA assumption. In cases in which it is reasonable to think that program participation is in some way explained by factors that are not observable or unavailable in our

dataset, or that the observed differences between treated and controls are too significant, the evaluator will have to resort to alternative identification methods.

Some elements to keep in mind when using PSM to evaluate postgraduate scholarships programs.

As was mentioned above, the main problem we face when evaluating the impact of a program is selection bias in treatment assignment. This means that the difference in the outcome variable between program beneficiaries and non-beneficiaries cannot be exclusively attributed to the effect of the program. This is due to the fact that both groups are different in several characteristics (observable and unobservable) which may also cause differences in the results.

The selection process has at least two instances; the first one is the candidates' self-selection. It is reasonable to expect that those who apply for a scholarship program are a target population, not only because of observable variables (that makes them eligible) but also because of unobservables. For instance, one might expect differences in proactivity, risk aversion, natural abilities, etc. The self-selection tends to be bigger when there is a higher cost of application (or when the process of applying is more complex). Therefore, to control for this self-selection bias, it is desirable for the control group to be formed by individuals who applied to the scholarship.

The second stage of selection occurs by program implementation and the decision of who will receive the scholarship. Usually, this decision constrained by the assessment of an evaluation committee, which may be binding depending on the program's design, available resources, etc. In any case, the population of beneficiaries is always non-randomly selected. The implementation of an experiment would imply replacing this selection stage with a random draw. If we discard that possibility, the matching technique arises as an alternative to balance the treatment and control groups in a set of observable variables, assuming that this are the only variables affecting selection.

The validity of this identification assumption will depend on the mechanism behind scholarship assignment and on the information available to make the matching. If the selection criteria consists on a clear rule based for example in a score assigned to each candidate, and such score is related to the expected impact of the scholarship, it is most likely that PSM is not a good

strategy. In other words, when estimating the PS with the score as the main determinant of treatment, the common support condition would rarely hold. In this context, the regression discontinuity design might be a better strategy.

When selection criteria are vaguer and clear scholarship assignment rules are not distinguishable, PSM becomes more feasible to implement. In that case, we would need variables from different nature such as characteristics of the candidates (sex, age, education, academic records) and information from the evaluating committee.

A type of variable whose inclusion in the PS must be a priority is some measurement or proxy of the outcome variable in the pre-treatment period. For instance, if our interest is focused in the effect of the scholarship on academic production, it is a good option to include in the PS the candidates' academic production at the moment of applying for the scholarship. Another option, in case we do have a pre-treatment measure of the outcome variable, is to redefine the variable over which the impact is estimated. That is to say, instead of evaluating the impact in the post treatment value of the outcome variable, evaluate it over the growth of such variable (i.e. the difference between the post-treatment and pre-treatment values). This way, the PSM will be accounting for baseline differences in the outcome variable between treated and controls. In some way, this would suppose a relaxation of the assumptions for identification as we are controlling also by unobservable differences that persist over time. This will be discussed in the next section when the difference in difference method is presented. Anyway, in presence of baseline information about the value of the outcome variable, when implementing PSM, a decision has to be made between the two positive alternatives: i) treat such information as another X when estimating the PS or ii) redefine the outcome variable and estimate the impact over the growth of Y. A criterion which may guide this decision is the nature of the outcome variable. If it is a highly persistent variable, as could be income, estimating the impact over the growth of the variable is an adequate option. If, on the contrary, the outcome variable is less persistent, as is the case of some measure of scientific production, it is advisable to estimate the impact over the post-treatment value of such variable and use

the pre-treatment value as a regressor in the PS estimation. Anyway, it is always advisable to confirm the sensibility of the results to the aforementioned alternatives.

An example of a PSM application in a Scholarship program can be found in the report by the Unidad de Evaluaciòn y Monitoreo de la Agencia Nacional de Investigación e Innovación"(ANII), for Uruguay and in Nunez et al.(2014) for the case of Colombia[12].

In the case of Uruguay, the impact of a scholarship program: "Becas de Posgrado de la ANII" is estimated. The information used to estimate the PSM is provided by the application forms and essentially from the system CVuy, a web application administered by the ANII which gathers the curriculums of researchers who are registered in the National Research System. CVUy is a public database which allows measuring some outcome variables such as academic production and researchers' education. For instance, it is possible to measure the number of publications in scientific journals, chapters in books, presentation of work in congresses, technical outputs and access to scholarships and other instruments of support to research activities such as the subsidy for researchers by the National Research System. Further, information about scientific production, education and access to support instruments (scholarships, subsidies, etc.) are referenced over time, which allows to collect the value of certain indicators (e.g. number of publications) in different moments of time, for instance before and after the program is evaluated.

Therefore, this type of data base, combined with some information about the process of evaluation of the applications by the evaluating committee, is a very useful source for the evaluation of scholarships programs, as long as the candidates have their curriculums updated.

Another possibility in order to broaden the group of outcome variables is to match the dataset of curriculums with social security information or firms' administrative records. This for example is done in Nuñez et al. (2014) for Colombia, where the information about Doctorate Scholarships candidates and young researchers of "Colciencias" with was matched with monthly information about social security contributions from the Health Ministry in order to

[12] Aboal and Tacsir (2017) use a similar methodology in order to evaluate the impact of a similar program of research support in Paraguay.

have information about wages. At the same time, this dataset included identification of the firms where workers from the treatment and control groups worked. This was matched to information from the survey of "Superintendencia de Sociedades" which allows to measure sales, exports, investment and firm productivity with PhD's (treatment) and without PhD's (control). This way, the possibility of matching the information from the scholarship applicants (using their identity number) with the administrative records from social security and the firms' recrods is potentially a very good source of information for impact evaluation. In the first place, it broadens the group of outcome variables over which we can identify impacts over different horizons. In second place, given that the data is longitudinal, it is possible to analyze impacts over longer horizons. In third place, this type of data will allow to explore impacts beyond the average treatment effect on the treated. The candidates' information matched with information about curriculums, social security and firms would allow inquiring into spillovers effects of this kind of programs. A possibility, following a logic similar to the one by Castillo et al. (2014), is to define as indirect beneficiaries of the program those researchers who, without being part of the program, work in the same firm as the beneficiaries. Under the hypothesis that these may be indirectly affected by the program, the existence of externalities involves identifying a differential behavior, in some outcome variable, between this group and the control group.

4.4 The Difference in Differences method

Even though the experimental setting discussed above is difficult to come across, there are certain conditions under which we are able to identify the effect of the program in a setting similar to that of a controlled experiment. This is the case in which the program is assigned to the different population segments following a criterion that is not associated with the results of interest. For instance, this occurs when policies are implemented in a differentiated manner by geographic zones. However, as far as these criteria partially explain program participation, the differences between the control and treatment group remain. A way of dealing with self-selection under these circumstances is given by the Difference in Differences strategy (DiD).

Such method allows to control for both observable and unobservable differences between the treatment and control groups.

The DiD method consists of comparing the outcome of the treatment and the control group before and after treatment. Intuitively, the impact of the program identified with this method consists of the expected change in the outcome variable Y for the treatment group minus the analogous expected change for the control group. Formally, if we denote the baseline outcome as $\mathrm{Y_i}(1)$ and $\mathrm{Y_i}(2)$ as the post-treatment outcome, the impact of the program would be provided by the sample analogue of:

$$\tau_{DiD} = [E(Y_i(2)|D_i = 1) - E(Y_i(1)|D_i = 1)] - [E(Y_i(2)|D_i = 0) - E(Y_i(1)|D_i = 0)] \quad (10)$$

Although DiD allows for systematic differences between the treatment and control group, it does require that such differences remain constant over time (in the absence of the program), or in other words, that the effect of the unobservable factors is persistent over time. This assumption is known as the parallel trends and it requires that the natural evolution (in absence of the program) of the outcome variable would have been the same for both groups. Once more, we find ourselves facing the problem of ignoring the counterfactual outcome. However, there are some simple procedures under which we can assess compliance with the parallel trends assumption. This is basically done by comparing the evolution of the outcome variable between the treatment and the control group in a period before the treatment (see Khandker et al., 2010).

Formally, the DiD estimator can be obtained from the following regression model:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 t_i + \beta_3 (D_i * t_i) + \gamma X_i + u_i \qquad (11)$$

Where $\mathrm{D_i}$is the variable which indicates treatment exposure and $\mathrm{t_i}$ is a temporal variable which takes value 1 after the intervention and value 0 in the baseline. $\mathrm{X_i}$ is a vector containing observable characteristics which may be included in order to improve efficiency and $\mathrm{u_i}$ is a residual term. Here, the ATT is represented by the parameter β_3 in equation (11).

Regarding data requirements, DiD may be used in cases in which there is longitudinal data so that the treatment and control groups are followed through time (starting at the baseline and until after program implementation). Moreover, even when panel data is not available, it is possible to use DiD from repeated cross-sectional data, where individuals in the control and treatment group are observed more than once in time, even though the individuals we observe in the different waves may not be the same. For instance, this would be the case if we worked with a series of household surveys. The requirement in this case is to have representative samples of both groups, before and after the treatment.

In sum, the DiD methodology holds advantages against the matching technique as it imposes less restrictive assumptions, allowing treatment and control group to be different in observable and unobservable characteristics (as long as the latter are persistent in time). Moreover, similarly to the experimental method, the results are of easy interpretation for the general public. However, this method has bigger data requirements, since it calls for pre and post-program information.

Some elements to consider when applying DiD to evaluate Postgraduate Scholarship Programs

A particular feature regarding the information we may access when evaluating scholarships programs is that, in general, we will have data before and after the treatment for at least some outcome variable. For example, in a case in which we only have information from program's application forms and some subsequent survey of the outcome variable (for the treatment and the control group), it is possible to measure some result in the baseline. This should come from the application forms where we probably have the academic records and candidate´s education.

The possibility of matching candidate´s information with curriculums and administrative records broadens the possibility of applying the difference in differences method. In this case, we will be able to reconstruct the academic production of candidates before the program. With

social security administrative records we will be able to have the labor and income paths for a large period. The advantages of having this type of information available were previously outlined when discussing the PSM application. In this case it is important to highlight that the possibility of reconstructing the path for the outcome variable in a period previous to the program offers the possibility of analyzing the parallel trends assumption. It is important in this case to report graphs on the outcome variable path for both groups as well as to carry out statistical contrasts.

One possibility is to estimate equation (11) using data previous to the program and redefining the variable t as a temporal trend (t=1,2,3,….). In that case the parallel trends contrast is the same as the significance contrast of the parameter β_3 . Another possibility is to maintain variable t as a dummy but which takes value 1 since the moment t-k (k=1,2,….) and, for different values of k, estimate the model with the data from the period previous to the program. The latter may be interpreted as a contrast of a placebo treatment as we are estimating the DiD equation but supposing that the program started in the period t-k. If the coefficient β_3 is significant, there is evidence against the parallel trends assumption (we are finding an impact in a program which did not exist!). The equations to estimate in order to contrast the parallel trends may or may not include controls, to account for the fact that the parallel trends assumption may be met conditional on the X´s. This would not invalidate the method, although it would be necessary to include such variables in the DiD equation when estimating the program impact. Although it is utterly important to include this type of analysis when applying the DiD method, it is important to highlight that such analysis is not a formal demonstration of compliance with the parallel trends assumption, since the assumption requires trends to be parallel after the application of the program and not before.

Finally, it is important to mention that in some situations non-compliance with the parallel trends assumption indicates something more than just invalidity of the DiD estimator. Suppose we found coefficient $\beta_3$to be significant when we estimate the model (11) with data previous to the program and where the variable t takes value 1 at time t-1. In this case, we could be finding a significant impact of a false treatment in period t-1. Although this is against the parallel trends

assumption, by the same token it deserves to be analyzed carefully as it could be evidence of an anticipated impact of the program. For instance, imagine that the expectation of "winning" the scholarship is really high within the treatment group with respect to the control group and that this fact determines important decisions which affect an outcome variable during the year previous to the grant. For example, people with high expectations of winning the scholarship may postpone their entrance to study programs or discard certain employments, foreseeing their future situation. It is very important to reflect upon this type of behaviors when evaluating the impact of a scholarship program. In this case we would be in presence of what is known as "Ashenfleter dip" (due to Ashenfleter, 1978), which refers to this particular type of behavior of the outcome variable before program implementation. What is recommended in this cases is to discard the period immediately before the starting of the program as a baseline for calculating the DiD estimator, and to consider for instance, the value of the outcome variable 2 years before. Anyway, this should be analyzed for every outcome variable separately, meaning that the "Ashenfelter" effect may be present in some variables but not in others.

4.5 Differences in Differences Estimator with Matching

In case we have observations for different moments in time and provided that the data base has a wide group of observable characteristics which explain program participation, it is possible to combine the DiD and PSM strategies in order to achieve higher robustness in the results.

An example in which the combination of both methods may be useful is when the parallel trends assumption is not satisfied, invalidating the DiD estimator. Suppose we found different paths between the outcome variable in the treatment and control group. In that case, we could look for a subsample within the control group which does satisfy parallel trend with respect to the treatment group. For that purpose, we can apply a PSM and select the matched sample as the control group in order to estimate the DiD model. When estimating PS, it is recommended to include the pre-treatment values of the outcome variable or even the growth of such variable. This way, we would be obtaining balance between treatment and control group in the

outcome variable before the treatment and thus will be able to afterwards apply the DiD estimator.

Another less data demanding possibility is to use the PS in order to restrict the estimation of equation (11) to the common support. Finally, another way of combining both methods which was previously mentioned in the section 4.4, is to estimate the PS with the group of variables X and the variable Y in the baseline, and obtain the matching estimator for the first difference of the outcome variable ($\Delta Y_i = Y_i(2)\text{-}Y_i(1)$).

In sum, when there is baseline information of the outcome variable available, we may improve our estimation of the impact of the program, since such information can be used to obtain the DiD estimator, which in turn relies on weaker assumptions than PSM. Moreover, we can use it to obtain the PSM estimator considering baseline information to estimate the PS (i.e. balancing the treatment and control group in the outcome variable before treatment). Finally, we can combine both methods, for instance, by obtaining the DiD estimator in the matched sample yielded by the PSM.

4.6 Instrumental Variables

The instrumental variables (IV) strategy allows dealing with selection bias, by instrumenting program participation through some exogenous source of variation. In general terms, the strategy consists in finding an observable variable (an instrument) which explains the participation but has no relation with the rest of the variables which influence the outcome variable. Once we identify an instrument satisfying such characteristics, the IV method may reproduce the randomized conditions that are ideal for evaluation. Card (1993) is a typical example, where the purpose is to identify the impact of attending University over future incomes. Given that it is reasonable to think that there is self-selection in university attendance, as there are unobservable factors which affect the decision of studying and at the same time may affect the labor income (as ability, for example), the author instruments such variable using the distance between the individual's residence and the nearest university. It is argued

that the distance variable is a good instrument as it explains attendance, while at the same time it is reasonable to think that it has no direct relation with the outcome variable. The central idea of the strategy is then to find an exogenous source of variation which explains program participation.

Formally, a good instrument must satisfy two conditions:

1. Relevance. The instrument explains the decision to participate in the program. Formally, if Z_i is the instrumental variable, it should satisfy:

$$Cov(D_i, Z_i) \neq 0 \qquad (12)$$

2. Exogeneity. The instrument only affects the outcome variable through its effect on program participation:

$$Cov(u_i, Z_i) = 0 \qquad (13)$$

Where$u_i$is the error term from regressing the outcome variable against the treatment variable and a group of observed characteristics:

$$Y_i = \alpha + \beta D_i + \theta X_i + u_i \qquad (14)$$

Although condition 1 is may be contrasted using statistical criteria, the impossibility of observing the real error term $\mathrm{u_i}$, impedes empirically testing the second one. As a result, the evaluation by IV in general requires efforts to argue the exogeneity of the instrument from an economic point of view. For that purpose, the evaluator will have to make use of its knowledge about the program implementation, the institutional setting and the transmission mechanisms of the program.

Once an appropriate instrument is found and its relevance and validity is justified, the impact estimation is done by 2 Stages Least Squares (2SLS). Specifically, the procedure consists of estimating a first equation in which program participation ($\mathrm{D_i}$) is explained using the instrumental variable ($\mathrm{Z_i}$) and a group of exogenous characteristics ($\mathrm{X_i}$).In this first stage, the predicted values of the probability of participation are obtained ($\widehat{\mathrm{D}_\mathrm{i}}$). Moreover, we may use the first stage to test the relevance of the instrument, based on the significance test of the

coefficient associated to Z_i and the joint significance of the model[13]. Then, the second stage is estimated by OLS in an equation similar to (14), but using the predicted participation ($\widehat{D_i}$) obtained in the previous stage instead of the participation itself (D_i) as the treatment variable. This is done under the assumption that $\widehat{D_i}$ gathers the exogenous variation of the treatment variable. The coefficient associated to $\widehat{D_i}$represents the measure of the impact identified by IV.

It is worth being precise over this procedure. In the first place, the estimation by 2SLS requires making corrections to the standard errors obtained in the OLS estimation, in order to be able to conduct proper inference regarding the significance of the estimated parameters[14]. Secondly, it is important that the variables which compose the vector X_i are the same in the first and in the second stage. This condition, called exclusion restriction, requires the estimation of the first equation to include the same regressors as the second one plus the instrument Z_i. Additionally, when it is possible to find more than one exogenous source to explain program participation, it is desirable to make the estimation using more than one instrument. This enhances the efficiency of the estimator, while also allowing us to test exogeneity of the instruments through contrasts which can only be applied whenever there is more than one instrument for the participation variable (i.e. whenever the model is overidentified).

Even though the IV method is powerful as it has strong internal validity, it presents some limitations in terms of the external validity of the results. More specifically, the impact identified through IV is of local nature. This means that the identified effect can only be ascribed to the portion of individuals who effectively decide whether or not to participate in the program depending on the instrument (the compliers group). Moreover, another difficulty arises since we are not able to identify which individuals are compliers. This is not a problem whenever the effects are homogeneous for all the population. However, we might expect that many programs have heterogeneous effects for different groups of people. In any case, it is important to have in mind the scope of the results at the time of interpreting them.

[13]For example, according to Staiger and Stock (1997), requiring an F statistical not less than 5.

[14]These corrections are already included and are automatically made by the IV estimation commands in the vast majority of statistical programs (for example using the command ivreg in Stata).

The availability of valid instruments is not something frequent unless we are in presence of a program with certain characteristics in its allocation rules. A particular case is the setting of the Regression Discontinuity Design which we discuss next. In that case, we will see that the program's allocation rule, for a subsample of the candidates, can be assumed exogenous and therefore be used as an instrument for treatment intake.

Sometimes, the program may impose selection criteria which give certain exogeneity to the assignment of treatment. For instance, when there is a quota for some minorities of the population with certain characteristics (affirmative actions), advantage may be taken of this situation in case it is reasonable to assume that such characteristics are not correlated with the expected result (i.e. that they are exogenous variables), and at the same time, are good predictors of the probability of receiving treatment.

4.7 Regression Discontinuity Design

In some cases, the program implementation offers a source of variation in treatment intake which allows replicating the experimental framework using relatively weak assumptions. In particular, the Regression Discontinuity Design (RDD) is widely used in cases in which assignment to treatment is strongly related to the value of a certain continuous variable Z_i. Such is the case in which the policy itself is designed in a way in which only those individuals with a value Z_i, above or under certain arbitrary threshold ($\bar{Z}$) receive treatment.

For example, consider the case of a scholarship program in which a score which summarizes the different merits of the candidates can be constructed (education, labor experience, academic production, etc.) and it is decided to provide the grant to those whose score is above a certain threshold. In this case, the assignment depends on the score (running variable), and due to this, the treatment probability presents a discontinuous jump in Z around the threshold. The RDD exploits the feature that, in a neighborhood of the threshold, the individuals at both sides are very similar, with the only difference that, as a consequence of the assignment rule, some

received treatment and others did not. This allows to construct comparable treatment and control groups, yielding a valid estimator of the program impact.

The first thing we have to consider when conceiving an RDD consists in determining the relationship between the running variable and program participation. When there is a deterministic relationship, meaning that program participation is completely determined by Z_i, program evaluation is conducted under a Sharp Regression Discontinuity design (SRD). These cases are easily identified through a graphical analysis of the running variable against program participation. In cases of SRD, we see that the probability of participation is 0 and 1 at one side and the other of the threshold, respectively. This implies that all and only the individuals who fulfill the assignment condition receive treatment. Different is the case of the Fuzzy Regression Discontinuity design (FRD), in which the probability of participation is largely but not totally determined by Z_i. For instance, if we think in a treatment which is only assigned to those individuals who fulfill $Z_i > \bar{Z}$, a case of FRD would be that in which we observe some individuals above the threshold who do not receive treatment or/and some below it who do receive it. This situation would be identifiable in a graphical analysis as a discontinuous leap in the probability of participation around the threshold, so that the probability is near 1 to the right of $\bar{Z}$ and near 0 to its left.

Moreover, in order to ensure that RDD implementation is appropriate, it is necessary that two assumptions are fulfilled. First, the local continuity assumption requires balance in terms of the characteristics which influence the outcome variable (the vector X_i), between individuals at both sides of the threshold. Graphically, it is hoped that, when analyzing the relation between some variable of X_i (for example, the educative level), we do not find great discontinuities around $\bar{Z}$. Second, it is hoped that the individuals as well as the program administrators are not able to manipulate the value of Z_i or the threshold $\bar{Z}$ in order to influence selection of beneficiaries. This can be checked by analyzing the density of Z_i, so that we shouldn't find any discontinuities around $\bar{Z}$.

At the time of implementing the RDD, it is possible to make the estimation through parametric or non- parametric strategies. The first option implies making assumptions about the

underlying relationship between the outcome variable and Z_i. In its most simple version, the parametric method assumes a linear relationship between Y_i and Z_i. In case the program has an impact, a discontinuity in such relationship around $\bar{Z}$ would be observed. For instance, in the case of the scholarship program aforementioned, such discontinuity would be given by a difference in the outcome variable between those who had a score above and under the threshold. It is understood that such difference represents the impact of the program as it would be explained exclusively by the fact that the individuals who are above the threshold receive the scholarship while those who are below it do not.

We will see formally the case of the sharp regression discontinuity design. Suppose the assignment rule is such that those who are above the threshold receive treatment. Then, the treatment variable D_i will be defined by:

$$D_i = \begin{cases} 1 \; if \; Z_i \geq \bar{Z} \\ 0 \; if \; Z_i < \bar{Z} \end{cases} \qquad (15)$$

In that case the effect of the program is obtained as the β from estimating:

$$Y_i = \alpha + \beta D_i + \gamma Z_i + \varepsilon_i \qquad (16)$$

It is also feasible to assume a non-linear relationship between the outcome variable and the running variable, in which case we would estimate:

$$Y_i = \alpha + \beta D_i + f(Z_i) + \varepsilon_i \qquad (17)$$

Where$f(Z_i)$ assumes a polynomic form in Z_i. Other specifications which are usually used consist of including interactions between the treatment variable and the running variable (by adding terms of the form $D_i * {Z_i}^n$) and thus allowing trend differences in one side and another of the threshold. In such cases, it is suggested to use the difference between the running variable and the threshold ($\widetilde{Z_\iota} = Z_i - \bar{Z}$) instead of Z_i in order to get the treatment impact around the threshold.In this context, a flexible specification would have the form:

$$Y_i = \alpha + \beta D_i + \gamma_1 \widetilde{Z_\iota} + \gamma_2 {\widetilde{Z_\iota}}^2 + \cdots + \gamma_n {\widetilde{Z_\iota}}^n + \delta_1 \widetilde{Z_\iota} * D_i + \delta_2 {\widetilde{Z_\iota}}^2 * D_i + \cdots + \delta_n {\widetilde{Z_\iota}}^n * D_i + \varepsilon_i$$

(18)

In general, the procedure consists of starting off with a simple specification of $f(.)$ and afterwards proceed to add terms of higher order while evaluating the robustness of the results throughout the process. Moreover, one can implement statistical tests in order to compare the goodness of fit of the different specifications (Lee and Lemieux, 2010).

On the other hand, the estimation through non-parametric methods holds the advantage that it does not rely on functional form assumptions, reducing the potential bias due to errors in the specification of $f(.)$. Broadly, the estimation consists of estimating a non-parametric regression of Y_i over Z_i and another one of D_i over Z_i on both sides of the threshold. Then, we can use the average of the predicted values for both variables on both sides of $\bar{Z}$ to obtain the program impact:

$$\tau_{RDD} = \overline{\hat{Y}(\bar{Z}^+)} - \overline{\hat{Y}(\bar{Z}^-)} \Big/ \overline{\widehat{Pr}(D=1|\bar{Z}^+)} - \overline{\widehat{Pr}(D=1|\bar{Z}^-)} \quad (19)$$

There are several techniques to estimate this kind of local regressions at both sides of the threshold (kernel estimator, local linear regression, local polynomial regression, and estimator of linear partial model, among others). In general, they fit some function (which assumes a different form depending on the method) to the observations which are inside an arbitrarily small neighborhood at one side and the other of $\bar{Z}$. In this context, an important decision which has to be taken at the time of implementing the non-parametric method is the width of the neighborhood at the sides of the threshold considered when making the estimation. Generally, there is a trade-off, as a bigger bandwidth provides gives more efficient estimates as it takes advantage of a bigger number of observations, but at the same time it may induce a higher bias as the comparison between individuals far from $\bar{Z}$ may turn dubious[15].

In general, the parametric and non-parametric strategies are not seen as competing methods, in fact it is suggested to use both methods as a check for robustness and credibility of the results (Lee and Lemieux, 2010).

[15] See Imbens and Lemieux (2008) for a criterion which allows to formally establish the optimal bandwidth

In presence of a fuzzy regression discontinuity design, the estimation procedure changes slightly. We will see the parametric alternative. In this case, the assignment rule does not fully determine treatment assignment, given that some individuals with values of Z above the threshold do not receive treatment and others with values below the threshold do receive it. What does happen is that the probability of receiving treatment presents a clear discontinuity at the threshold. As a result, what we have in this case is a relevant instrument (being above or below $\bar{Z}$) for the treatment variable (D). At the same time, being above or below the threshold in a neighborhood of it can be seen as an accidental event and thus, the instrument is exogenous.

Therefore, the estimation in this case uses the indicator variable of being above $\bar{Z}$ as an instrument of D. Meaning, we estimate for example the equation (16) through the instrumental variables method.

To sum up, the RDD has the advantage of working under plausible assumptions, providing the strategy with good internal validity. However, there are some limitations in what regards the external validity of the method. The same as in instrumental variables, the RDD estimator is local, as the identified impact can only be attributed to the individuals who are around the threshold $\bar{Z}$. On the other hand, although RDD involves techniques of certain econometric sophistication, its implementation may be carried out following a rather standardized method. We suggest, for example, the work of Lee and Lemieux (2010) in which a tool box and steps to follow when implementing RDD are proposed.

Oosterbeek and Webbink (2006) use RDD to evaluate a program of subsidy to young talents in the Netherlands to study abroad. The authors use a sample of graduates who requested a specific subsidy to study abroad between 1997 and 2004 and compare the results of those who received the grant against those who did not. At the same time, they exploit the information about the candidates' ranking made by the selection committee in order to obtain a more credible comparison group.

First, they intend to identify the impact of the subsidy in the probability of studying abroad. It is reasonable to evaluate the impact of the program in the accomplishment of its most immediate objective. Does the program effectively change the history of the beneficiaries? Or would the beneficiaries have studied abroad anyway?

In order to answer this question, the authors restrict the analysis to a subsample of candidates for which they have information after postulation to the program. Basically, they use two strategies in order to estimate the impact. On the one hand, they directly estimate the effect of the treatment variable (scholarship) in a regression in which the probability of studying abroad is explained and another one in which the length of such studies is explained. They use variants of this regressions in which they include different controls, among them the ranking obtained in the evaluation process of the postulations. Then, they restrict the sample to the individuals who were ranked in the surroundings of the cut off (they select 4 different subsamples of different range). In this last case, the authors are obtaining a variant of the regression discontinuity estimator. The results show that the scholarship has a fundamental effect in the decision of studying abroad (between 23 and 42%).

Secondly, the authors propose to estimate the effects (or returns) of studying abroad in several dimensions (residing abroad, taking up graduate studies, wages, etc.). Observe in this case the focus is not strictly in the direct impact of the program but in the indirect impact the program could have through augmenting the probability of young talents to study abroad. Therefore, the result of this analysis should not be interpreted directly as the impact of the program, even though they do contribute to understanding the mechanism by which the program may affect those dimensions[16]. In order to estimate the impact of studying abroad over this other variables, the authors use the assignment rule of the program as an instrument and estimate for the subsample of candidates in the surroundings of the selection threshold. That is to say, for the candidates in the surroundings of the cut off, the scholarship may be considered a random or accidental event which guarantees that the instrument is exogenous. At the same time, it was previously shown that the scholarship has a significant effect over the probability of

[16] It is true that once we know the impact of the program over the probability of studying abroad, and the impact of the latter in a Y variable, it is possible to infer a value for the impact of the program in the Y variable.

studying abroad, so that it is a relevant instrument. The estimations show that studies abroad increase the probability that the students will live abroad after finishing their studies. Also, it was found that studying abroad is associated to higher wages although the estimated effect was not robust and thus, cannot be interpreted as an impact of studying abroad.

Jacob and Lefgren (2011) is another application of a regression discontinuity estimator, in this case used to identify the impact of a subsidy for research by the National Institute of Health (NIH) of the United States. The analyzed sample contains the candidates to such subsidy between the years 1980 and 2000 and the impact on their scientific production (publications and citations) is evaluated. Given the characteristics of the program, the authors use a fuzzy regression discontinuity design, or what is equivalent, they use the threshold of selection as an instrument for the treatment. The estimation shows that on average receiving the subsidy leads to one additional publication after a period of 5 years. The authors point out that this low impact is consistent with a model of competitive credit markets for research which allows candidates rejected by the NIH grant to find other sources of financial support.

4.8 Effects at different time horizons and externalities

The magnitude of a program impact over an outcome variable is, in general, dependent of the time elapsed since the program was implemented. As figure 1 shows, postgraduate scholarship programs may have impacts in a short, medium or long term. It is necessary to lay out the hypothesis about the term in which the effect will materialize over each outcome variable and to ensure that the measurement of the variable corresponds to a horizon consistent with such hypothesis. In case the available information allows to measure the result of a variable over which we expect impacts in the mid or long term only in the short term, this must be considered when interpreting the obtained results. For example, if the expected impact is positive in the long term, the short term estimator will underestimate the impact of the program in the long term and thus may be interpreted as a low bound for the latter.

In other cases, the available information will allow us to evaluate the temporal profile of the effect. Meaning, in the case we have longitudinal information which allows us to observe the outcome variable at different temporal horizons, it is the data itself which can inform us about the timing of the effect. Suppose we have longitudinal information for individuals in the treatment group as well as for those in the control group, then we can specify the following equation:

$$Y_{it} = \sum_{k=0}^{K} \beta_k D_{it}^k + \gamma_t + u_i + \varepsilon_{it} \qquad (20)$$

Where γ_t is a fixed temporal effect, u_i is the fixed individual effect (or individual unobserved heterogeneity) and ε_{it} is an idiosyncratic error[17]. In this case the variables D_{it}^k form a set of dummies which take value 1 in the k-esim period after the beginning of the program and 0 in the remaining periods. For example, D_{it}^0 is a variable which indicates the year in which the beneficiaries receive the scholarship for the first time. Therefore, every coefficient of β_k is the estimator of the impact after k periods since program implementation. In this case, it is a DiD estimator, as we are controlling for the existence of unobservables that are fixed in time. An application of this estimator may be found in Crespi et al. (2011) where the impact of programs supporting innovation in Colombia over some firms' results (employment, investment, productivity, new products) is analyzed.

Estimating equation (20), we will be able to identify heterogeneous impacts in time. Consider, for example, the impact of a scholarship program on academic production. Suppose we have already ensured that the program effectively increases beneficiaries' probability of enrollment and completion of post graduate studies. In this case, it may happen that when analyzing the effect over academic production, we find a negative impact in the short term (i.e. the beneficiaries produce less than their counterpart in the control group) but a positive one in the

[17] This specification is a variant of equation (11) for the case in which we have panel data, in other words, observations before and after the program for all the individuals which participate in the evaluation. In that case, not only will be able to control for the fixed effect that is common to the individuals in the treatment group (the term $\beta_1 D_i$ in equation 11), but for the individual fixed effect which includes the previous one (the term u_i in equation 20).

long term (i.e. after completing their studies, the scholarship beneficiaries show a higher production than those individuals in the control group).

The presence of externalities is another aspect to consider when evaluating a program of human capital formation. The externality refers to the indirect effect of the program over individuals or areas beyond those of its direct application. On this regard, it is advisable to distinguish the case in which the indirect impact is limited to the performance of a group of people who are not in the treatment group, from the situation in which the impact affects a broader scope such as the labor market, the innovation system or the economy as a whole. In this last case, when the scope of the program is wider, the externalities are the same as the general equilibrium effects of the program.

Let us consider the first case, when the program affects people who are not in the treatment group. The presence of this kind of effects may be common in the case of highly qualified human capital formation programs given that, for example, research activity generally entails interaction and collaboration between researchers. Therefore, the effect of the program may spillover on the rest of the researchers with whom they interact (creating a positive externality).

That said, at the time of evaluating the impact of the program the first thing we have to ask is whether it is possible that the presence of this type of indirect effects affect our control group. If the answer is yes, the externality becomes an identification problem. That is to say, our control group is not a good counterfactual, so that in the case of a positive externality our estimator would underestimate the ATT. If we suspect the existence of external effects over the control group, could at best identify a lower or upper bound (depending on the sign of the externality) to the real impact of the program over the treatment group.

On the other hand, identification of the externality can be by itself a target of a program evaluation. In order to do that, we should find a subgroup within the control group which is presumably indirectly affected by the program and another one which is not (pure control). We can think of the individuals affected by the externality as a second treatment group. The externality estimator results from the comparison of the outcome variable between this group

and the pure control, while the ATT is obtained from the comparison between the treatment group and the pure control. For instance, we could define as indirect beneficiaries of the postgraduate scholarship program those researchers who, without being beneficiaries, work in the same firm or research team as the grant beneficiaries. At the same time, the pure control group would be formed by researchers who did not receive the scholarship or collaborate with the beneficiaries in the work environment. The identification of the externality would involve comparing the behavior in some outcome variable between the indirect beneficiaries and the pure control group. Obviously, all the considerations made previously about the validity of the control group as a counterfactual for the indirect beneficiaries also apply to this case. An application following this logic, that aims to evaluate the effects of a productive cluster support program in Argentina, can be seen in Castillo et al. (2014)[18].

When the externality takes the form of a general equilibrium effect, the problem we face when evaluating the impact of the program is rather more complex. This is the case, for example, of a large scale program or when the treatment group has a relevant size in the target universe (i.e. in the researchers' national system). Ideally, the impact could be identified if the program is applied in some regions and if the general equilibrium effects appear in that scope only, and at the same time we have other regions where the program is not applied. In this case, the comparison of regions with and without the program would allow identifying the total impact of the program, namely, including the general equilibrium effects. Except for this particular situation, the techniques presented in this study are not the most appropriate to identify general equilibrium effects. The alternative is to use a general equilibrium model, meaning, an abstract representation of the economy which, based on a large number of assumptions about behavior and interaction between agents, allows to estimate the impact (ex-ante) of a policy or shock considering the general equilibrium effects.

[18] Another possibility is to identify natural experiments which allow estimating external effects. An example can be seen in Azoulay et al. (2007) where the magnitude of the spillovers generated by distinguished academics in their academic activity is measured. In order to do that they use information about the scientific production of the co-authors of 137 academics who died while they were active, providing this way, an exogenous source of variation in the collaborators coauthors networks' structure.

Even though in general that the ultimate aim of human capital formation programs is precisely to generate positive externalities for the economy as a whole (i.e. over innovation, aggregated productivity, etc.), this rarely can be identified with the techniques developed in this study. However, these techniques are appropriate to answer the first questions we should formulate, which refer to the program effectiveness in the attainment of its immediate objectives.

However, even in the case where we are evaluating a small scale program and so that we can assume no general equilibrium effects, we should ask whether general equilibrium effects would happen when scaling such program. This has to do with the external validity of our evaluation. If general equilibrium effects are expected when applying the program at large scale, then our impact estimation at small scale tells us half of the story. In these cases, it is worthwhile to discuss, with the support of theory, the mechanisms and the expected sign of the general equilibrium effects. This analysis is relevant even when the sign of the effects is ambiguous and thus we cannot assert whether our estimation is overestimating or underestimating the impact. Let us consider the effects of a program of highly qualified human capital formation over the private returns of training. What would happen if we scaled our postgraduate scholarship program? On the one hand, the scaling would produce a supply shock for skilled labor which would reduce wages for this type of work (in presence of an inelastic demand). On the other hand, the concentration of qualified workers could give rise to more innovations or attract foreign direct investment and increase the demand for this kind of jobs and thus its remuneration. This is an example in which we could not conclude whether our estimation of the impact of the program at a small scale over the private returns to training underestimates or overestimates the effect that would be observed if the program was implemented at a large scale.

5. Sources of Information

5.1 Electronic CV

One of the sources of information available in several countries of Latin America (i.e. Uruguay, Colombia, Paraguay, Brazil, Argentina, Mexico) are the electronic CV's which cover at least a

good part of the scientific community of those countries[19].These contain very useful information which can be used to evaluate scholarship programs, instruments of support to international mobility and of support to the development of researchers' careers, as well as longitudinal studies which study the progression in the professional career[20]. In general, these CV's contain the necessary information to at least evaluate the impact of these programs on the following dimensions:

1. Bibliographic Production
2. Technical Production
3. Human Resources Formation
4. Own education
5. Distribution of hours of work over institutions, their characteristics, etc.
6. Public funds granted for research.

The first step, in order to make the information available, is to use electronic CVs for the process of evaluation of applications. Apart from enabling and standardizing information in a centralized record this allows to generate information before the program. Further, if the institution providing the scholarship can get applicants to update their electronic CV in a regular way (whether or not the received the subsidy) during a certain period after program implementation, then we would have a database which will contain information of the beneficiaries as well as from a potential control group.

It is good to complement the information provided by the CV with information resulting from records of publications or patents, as for example, Scopus. This will give hard information which to contrast problems stemming from sub reports (and also over reports) which may be present in the CV. This information is obtained through searches in these bases using the name of the candidate and extracting information such as the number of citations to her work, the number of indexed publications she has, or the quality of the journals in which she publishes.

[19] They even contain common information, comparable between the different countries.

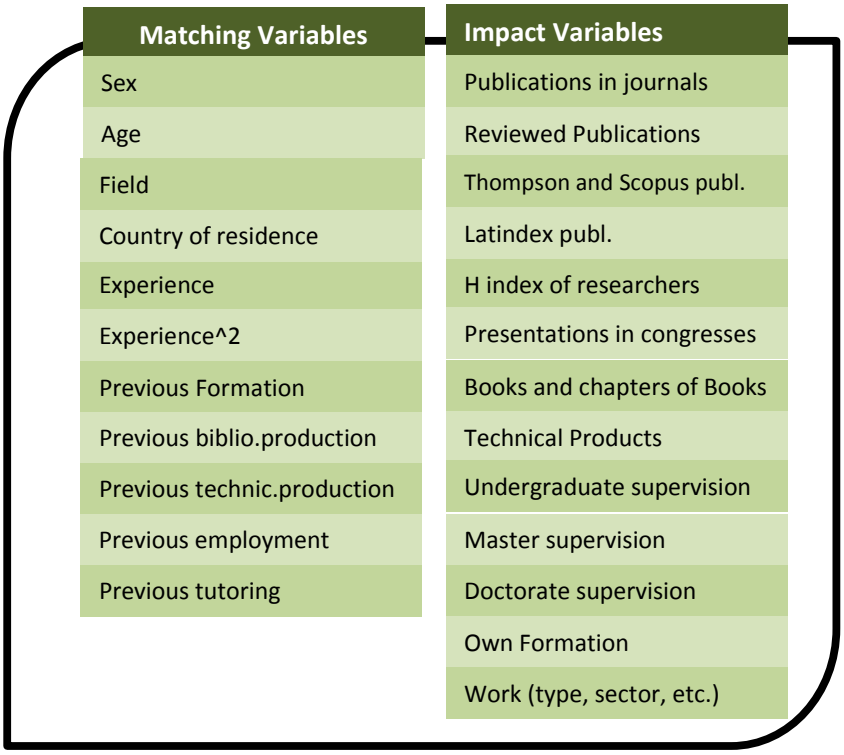
Usually, in order to estimate the impact of the scholarships with this data, it will be necessary to employ matching techniques together with DID. The idea is to compare individuals in the beneficiaries group with similar individuals in the group of rejected candidates or non-beneficiaries, as it can be seen in the next figure.

Figure 2. Treatment and Control Group

The comparison of individuals between the beneficiary group and the rejected group must be done in such a way that these individuals are similar before the program. For this, we generally employ matching techniques (as for example PSM) as described above. Amongst the variables we can use for matching, we can consider those that appear on the right side in the next figure. On the left side, we consider some variables over wich we can measure impacts. Given that CV's also provide a temporal dimension, we can usually combine matching techniques with DiD, which, as was previously discussed, provides more robust results than simple matching.

Even when in the country where the scholarship program is implemented there is not a standardized electronic CV, some of the templates available in the region may be adopted and request that all applicants fill them in, and also update them periodically.

Figure 3. Possible impact and matching variables using information from electronic CV's



5.2 Administrative Information: Social Security and Fiscal Institutions

A source of information which can be really useful in order to estimate the impact of the scholarship programs over labor and income paths are the (usually public) agencies which keep administrative records referred to the work history, tax declaration and social security of individuals.

For the purposes of the evaluation, it is enough to have anonymous information. Usually, the procedure in order to access this information consists of sending the list of the program applicants and require the information about their work history, incomes, etc. in an anonymized way (deleting all the information which allows to identify the individual, namely,

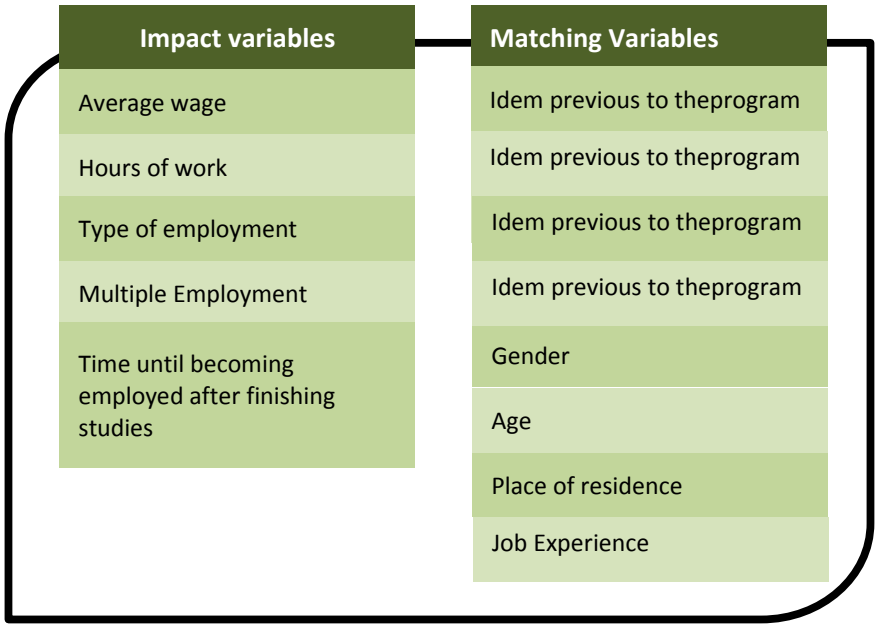
name, ID number, etc.), with an indicator which only states whether or not the individual received the scholarship.[21]

Once this information is obtained, it is possible to resort to matching and DID techniques in order to estimate the impacts. In figure 4, we show some of the variables over which we can usually measure impacts using these sources of information. The matching variables may be the same but measured at the baseline. Another characteristic of this type of data is that they allow assessing the impacts on labor productivity measured by wages, which broadens the scope of the evaluation as it allows assessing the impact on the corporate sector.

As an example, we can consider Chile's Mifuturo.cl -a website by the Ministry of Education- which provides plenty of information about the labor market for University graduates. The information about incomes and employment associated to careers comes from higher education institutes, which provide complete records of their alumni; and this data is matched to taxes declarations of taxpayers by the Subdirection of Studies of the SII. Finally, the processing and validation is carried out by the "Servicio de Informacion de Educacion Superior" (SIES), of the Ministry of Education. In this site, it is possible to find 1650 combinations of careers and institutions, allowing the comparison of the differences of incomes for a same career, depending on the institution where the individual studied. At the same time, it is possible to find generic information about 220 careers (without distinguishing by educational institution). It also has data about number of enrollments, composition by gender and number of graduated, rate of retention, real duration and range of tariffs of the different options.

[21] I must be said that social security and tax authorities tend to be very private about this information. Obtaining access to this information then requires efforts when designing the instrument that ensure that this information will be available when carrying out the evaluation.

Figure 4. Possible impact variables and matching from administrative information



5.3 Surveys

All the variables mentioned in the previous sections can be obtained from surveys that the institution which provide the scholarships may carry out at the beginning of the program and afterwards over time.[22]One of the main problems of this strategy is that in many cases it is difficult to elicit answers from non-beneficiaries of the program, especially if it has been a while since the application. Anyway, the institution granting the scholarship may include as a requirement at the time of the candidature that candidates must answer a determined number of surveys in future. A way of ensuring they are completed is by inhibiting access to future public support schemes to those persons who have not answered the required surveys. Another possible strategy in order to achieve greater success is to manage an agreement with the statistics institute of the country and include beneficiaries in periodic surveys (such as the household survey) with a special module which collects information of interest for evaluating

[22]Examples in this sense are the monitoring of graduates survey of the labor observatory in Colombia, the monitoring of graduates performed by CONARE and OPES in Costa Rica.

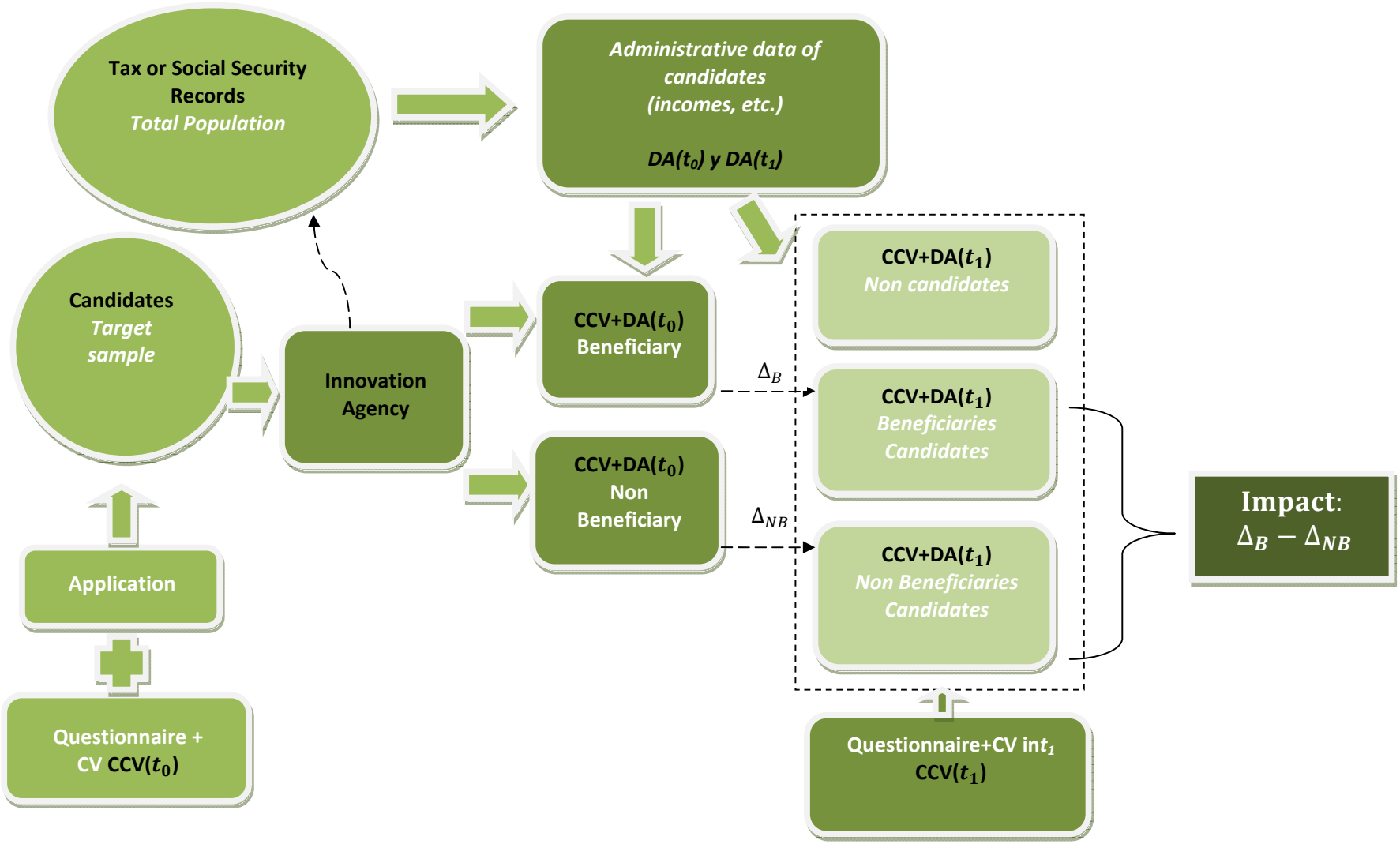
the program. However, it is necessary to have in mind that developing a system to collect information requires a certain specific budget for that purpose.

The next table and figure resume the discussion of this section

Table 3.Sources of information to evaluate programs of human capital formation

Dimension	Electronic CV	Social Security and/or Internal Revenue	Monitoring Survey of candidates
Coverage	Candidates (rejected and beneficiaries). Only a fraction of the total population	Working Force	Only a fraction of the total population
Way of collecting information	First round obtained with the application form. Successive rounds referred to new calls or promotion processes.	Automatic	Specific survey with additional costs of design and collection
Legal framework/agreements	Centralized in the own entity responsible for the program	Requires specific access and agreements with authorities	Without specific precautions. With agreements when implemented with other institutions or statistics direction
Available Information	Rather standardized in the region. Information about personal characteristics, production, etc.	Information oriented to objectives of tax or social security records ("Little" information to characterize the individual) but with higher accuracy over the relevant variables	To be designed. Greater degree of specificity

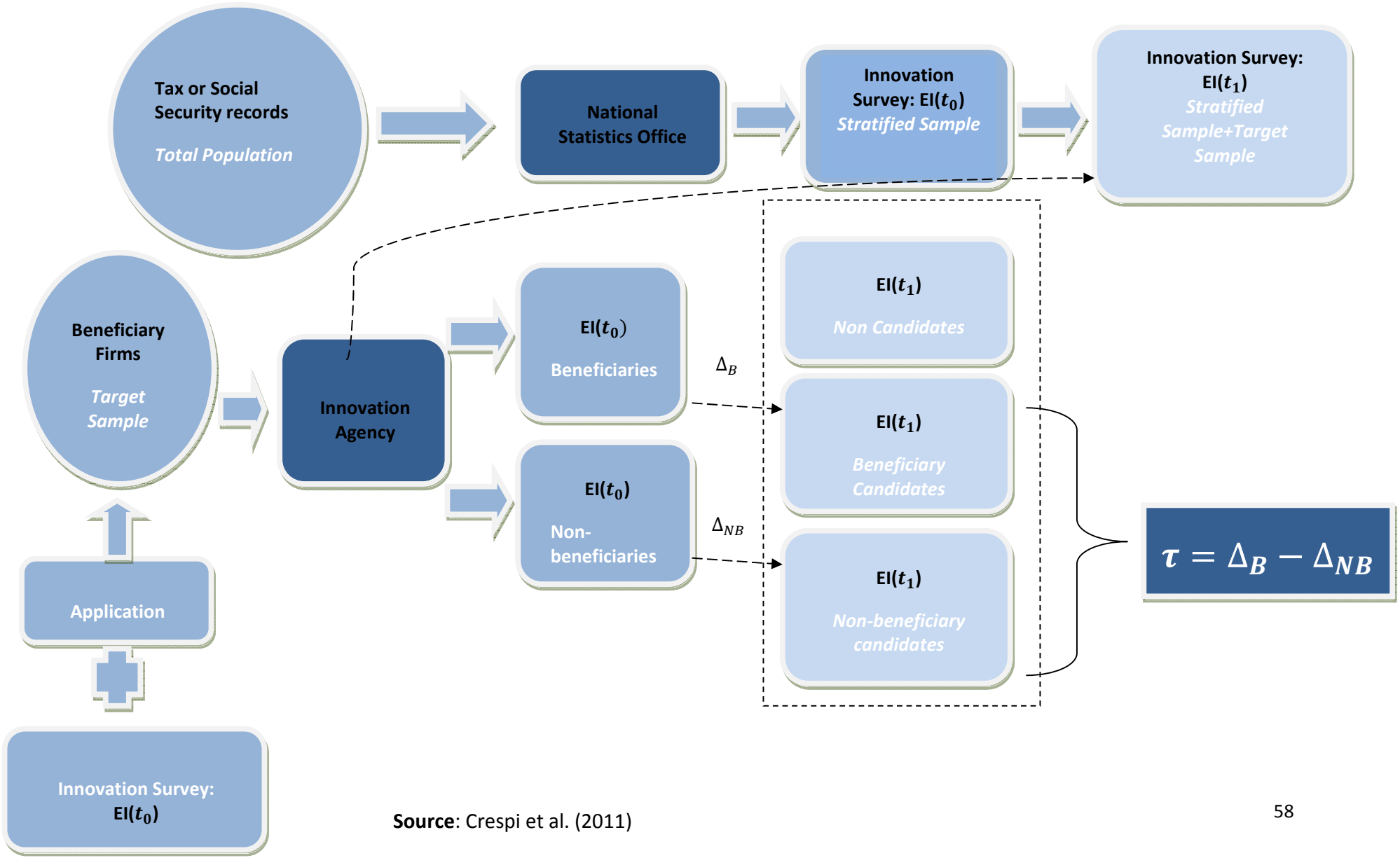
Figure5. Evaluating with data from CV's, surveys and administrative record



At the same time, we can use a logic similar to Figure 5 to carry out evaluations where the unit of study are the firms, with the aim of identifying impacts over firm productivity. In figure 6, we present the logic to perform evaluations of this kind, following the example of the "Agencia Nacional de Investigacion e Innovacion" (ANII) from Uruguay. In the Uruguayan model, the National Institute of Statistics regularly collects an innovation survey aimed at firms. The survey is representative of the firm sector and, at the same time, is compulsory for all those firms who apply to ANII programs. Also, after applying once, the firm remains as part of the sample for future surveys. This system provides base line and follow-up measurements referred to those who candidate to the agency's programs.

In this framework, one can perform evaluations of the programs mentioned in section 3 in which incorporation of highly qualified human resources in the firms is subsidized. In such cases, given that the firms themselves are the beneficiaries of the program, creation of a treatment and a control group requires data from firms who did not candidate to the program, but also from those who did apply and were rejected and from those who applied and were accepted. On the other hand, it would also be possible to use a scheme like this one in the case in which the beneficiary is the individual and not the firm, as it happens with the vast majority of scholarship programs for advanced training. This is made possible by identifying as "treated" firms those in which at least one worker is a beneficiary of the program, while the control group would be formed by those where there are no workers participating in the program. A strategy of this type would allow identifying the externalities of these programs in terms of variables such as firm productivity.

Figure 6. Evaluation with Innovation Surveys. The ANII model



Source: Crespi et al. (2011)

6. Conclusions

The purpose of this work is to provide a guide on how to measure the effectiveness or the impact of programs of human capital training for Science, Technology and Innovation (STI). The paper addresses the specific challenges which arise when evaluating this type of programs, discussing its logic, the pros and cons of the different data sources, the strategies which may enable the evaluation and the adequateness of the application of the different experimental and quasi-experimental methods available.

For each method, the document highlights the characteristics and assumptions, the practical issues related to the application, and the strengths and weaknesses related to applying them to evaluate programs of human capital training for STI. Also, we considered some more specific issues related to the evaluation of programs of human capital formation for STI, as for example the time in which the effects and externalities are expected to realize. Concrete examples of evaluations were used for the discussion of the several issues during the paper. To this respect, it is important to highlight that evaluations of this kind of programs are scarce and that this is therefore a fertile field for researchers.

This guide is an introduction to the evaluation of programs of advanced human capital training and, because of this, is incomplete. The bibliographic references are a very important source to complement the different aspects discussed here.

References

Aboal, D. and Tacsir, E. (2017), The impact of subsidies on researcher's productivity: Evidence from a developing country. *Research Evaluation*, 26(4): 269–283.

Aboal, D. and Garda, P. (2016). Technological and Non-Technological Innovation and Productivity in Services vis-à-vis Manufacturing Sectors. *Economics of Innovation and New Technology*, 25(5): 435-454.

ANII (2014). "Informe de evaluación ex post. Becas de posgrado nacionales 2008". http://www.anii.org.uy/upcms/files/listado-documentos/documentos/n8-evaluaci-n-de-impacto-de-becas-de-posgrado-nacionales-2008-1-.pdf

Arcidiacono, P., Hotz, J. and Kang, S. (2012). Modeling College Major Choice Using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics* 166(1), 3-16.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors* (pp. 609-626). Princeton University Press.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 47-57.

Autor, D., F. Levy, and Murnane, R.J. (2003). The Skill Content of Recent Technological Change: an Empirical Exploration. *Quarterly Journal of Economics*, 118(4):1279−1333.

Becker, K. y Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education: Innovations and Research* 12.5/6: 23.

Belli, P., Anderson, J., Barnum, H., Dixon, J. and Tan, J-P. (1998), "Handbook on Economic Analysis of Investment Operations", Operational Core Services Network Learning and Leadership Center, Banco Mundial, Washington, D.C.

Bernal, R., and Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Universidad de los Andes, Facultad de Economía, Centro de Estudios sobre Desarrollo Económico.

Bloom H. (2006). "The Core Analytics of Randomized Experiments of Social Research", MDRC Working Papers on Research Methodology.

Borjas, G. J. (2005). The labor market impact of high-skill immigration (No. w11217). National Bureau of Economic Research.

Caliendo, M., and Kopeinig, S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. IZA Discussion Paper No. 1588.The Institute for the Study of Labor. Bonn.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling (No. w4483). National Bureau of Economic Research.

Castillo, V., Maffioli, A., Rojo, S. and Stucchi, R. (2014).Knowledge Spillovers of Innovation Policy through Labor Mobility: An Impact Evaluation of the FONTAR Program in Argentina. IDB Working Paper IDB-WP-488. Inter-American Development Bank.

Council of Economic Advisers (CEA) (2009). Preparing the Workers of Today for the Jobs of Tomorrow. Washington, DC: Council of Economic Advisers.

Craig, E., Thomas, R., Hou, C., y Mathur, S. (2011). No Shortage of Talent: How the Global Market is Producing the STEM Skills Needed for Growth. Accenture Institute for High Performance.www. accenture. com/institute.

Crépon, B., E. Duguet, y J. Mairesse. (1998). "Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level." NBER Working Paper 1-43. Cambridge, MA: National Bureau of Economic Research.

Crespi, G., Maffioli, A. and Rasteletti, A. (2013). "Business Innovation Policies in LAC. Experiences and Pending Challenges." Banco Interamericano de Desarrollo (BID), *mimeo*.

Crespi, G., Maffioli, A., Mohnen, P., and Vázquez, G. (2011). Evaluating the impact of science, technology and innovation programs: a methodological toolkit. Inter-American Development Bank Technical Note 333

Crespi, G., Tacsir, E. and Vargas, F. (2014). Innovation and productivity in services: Empirical evidence from Latin America. Inter-American Development Bank Technical Note 690.

Crespi, G. and Zúniga, P. (2012). Innovation and Productivity: Evidence from Six Latin American Countries. *World Development* 4(2): 273-90.

Cunha, F. y Heckman, J., (2005). "The Technology of Skill Formation". *American Economic Review*, 97(2), 31-47, May.

Duflo E., Glennerster R. and M. Kremer (2006). "Using Randomization in Development Economics Research: A Toolkit." NBER Technical Working Papers 0333, National Bureau of Economic Research.

Emerson, P. and Mcgough, B. (2011). "Learning about Education." IZA Discussion Paper Series, IZA DP No 5756, June.

Fast Future (2010). The Shape of Jobs to Come: Possible New Careers Emerging from Advances in Science and Technology (2010–2030). Final Report.

Glennerster, R., and Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.

Hall, R. E., and Jones, C. I. (1999). "Why do some countries produce so much more output per worker than others?"., NBER Technical Working Papers No 6564, National Bureau of Economic Research.

National bureau of economic research.

Imbens, G. W., and Lemieux, T. (2008). "Regression discontinuity designs: A guide to practice." *Journal of econometrics*, 142(2), 615-635.

Imbens, G., and Rubin, D. (2010). Causal inference. *Unpublished manuscript*, Harvard University.

Jacob, B. A., and Lefgren, L. (2011). "The impact of research grant funding on scientific productivity." *Journal of Public Economics*, 95(9), 1168-1177.

Jensen, R. (2010). "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2), 515-548.

Khandker, S. R., Koolwal, G. B., and Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.

Lee, D. and Lemieux, T. (2010). "Regression Discontinuity Designs in Economics". *Journal of Economic Literature* 48(2): 281-355.

Loomis, J.B. (2011), "Incorporating distributional issues into Benefit Cost Analysis: why, how, and two empirical examples using non-market valuation", *Journal of Benefit-Cost Analysis*, 2 (1), article 5. (ttp://www.bepress.com/jbca/vol2/ iss1/5).

Lucas, R. E. (1988). "On the mechanics of economic development". *Journal of monetary economics*, 22(1), 3-42.

Mincer, J. (1974). "Schooling, Experience, and Earnings." *Human Behavior & Social Institutions* No. 2.

Moretti, E. (2004). "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *Journal of econometrics*, 121(1), 175-212.

Nelson, R. R. (1959). "The economics of invention: A survey of the literature." *The Journal of Business*, 32(2), 101-127.

Núñez, J., Castro, F., Estupiñán, F., Gordillo, A., Martínez, A., Medina, C., Castañeda, C., Hurtado-Caycedo, C. and Trujillo, L. (2014). Evaluación de impacto de los programas de formación de capital intelectual ofertados por Colciencias, Jóvenes Investigadores y Becas de Doctorados. *Mimeo*

Oosterbeek, H., and Webbink, D. (2006).*Assessing the returns to studying abroad* (No. 64).CPB Netherlands Bureau for Economic Policy Analysis.

Psacharopoulos, G., (1995), “The Profitability of Investment in Education: Concepts and Methods”, Human Capital Development and Operations Policy, Banco Mundial, Washington, D.C.

Roy, A. D. (1951). “Some thoughts on the distribution of earnings.” *Oxford economic papers*, 3(2), 135-146.

Rubin, D. B. (1974). “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of educational Psychology*, 66(5), 688.

Sen, A. (1999). Commodities and capabilities. *OUP Catalogue*.

Staiger, D., and Stock, J. H. (1997). Instrumental variables regression with weak instruments *Econometrica* 1997, 557-86.

Stinebrickner, T.R. and Stinebrickner, R. (2012a). Learning about Academic Ability and the College Dropout Decision.*Journal of Labor Economics* 30(4), 707-748.

Stinebrickner, R. and Stinebrickner, T.R. (2012b).A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout.*NBER Working Paper Series*, Working Paper 19165, Cambridge.

Tacsir, E. (2014). I'm sorry, Miss: You've Missed! Information on Wages and CareerChoice. Forthcoming as UNU-MERIT Working Paper. Paper presented at AnniversaryConference of UNU-MERIT, Maastricht, Netherlands, 26-28 November 2014.

Winters, J. V. (2015). Do higher levels of education and skills in an area benefit wider society?IZA World of Labor.

Wiswall, M. and Zafar, B. (2013), “How Do College Students Respond to Public Information about Earnings?” *Federal Reserve Bank of New York Staff Reports*, no. 516

Zafar,B., (2011), “How Do College Students Form Expectations?”, *Journal of Labor Economics,* 29(2), 301 - 348.

Zafar, B. (2013), “College Major Choice and the Gender Gap”, *Journal of Human Resources*, vol. 48(3): 545-595.

The UNU-MERIT WORKING Paper Series

2018-01 *The serendipity theorem for an endogenous open economy growth model* by Thomas Ziesemer
2018-02 *Instability constraints and development traps: An empirical analysis of growth cycles and economic volatility in Latin America* by Danilo Sartorello Spinola
2018-03 *Natural, effective and BOP-constrained rates of growth: Adjustment mechanisms and closure equations* by Gabriel Porcile and Danilo Sartorello Spinola
2018-04 *The relevance of local structures for economic multiplier effects of social pensions in Uganda* by Maria Klara Kuss, Franziska Gassmann and Firminus Mugumya
2018-05 *Regulating the digital economy: Are we moving towards a 'win-win' or a 'lose-lose'?* by Padmashree Gehl Sampath
2018-06 *The economic impacts of a social pension on recipient households with unequal access to markets in Uganda* by Maria Klara Kuss, Patrick Llewellin and Franziska Gassmann
2018-07 *The effect of weather index insurance on social capital: Experimental evidence from Ethiopia* by Halefom Y. Nigus, Eleonora Nillesen and Pierre Mohnen
2018-08 *Evaluating intergenerational persistence of economic preferences: A large scale experiment with families in Bangladesh* by Shyamal Chowdhury, Matthias Sutter and Klaus F. Zimmermann
2018-09 *Agricultural extension and input subsidies to reduce food insecurity. Evidence from a field experiment in the Congo* by Koen Leuveld, Eleonora Nillesen, Janneke Pieters, Martha Ross, Maarten Voors and Elise Wang Sonne
2018-10 *Market integration and pro-social behaviour in rural Liberia* by Stephan Dietrich Gonne Beekman, and Eleonora Nillesen
2018-11 *Sanctioning Regimes and Chief Quality: Evidence from Liberia* by Gonne Beekman, Eleonora Nillesen and Maarten Voors
2018-12 *Chinese development assistance and household welfare in sub-Saharan Africa* by Bruno Martorano, Laura Metzger, Marco Sanfilippo
2018-13 *Foreign direct investment in sub-Saharan Africa: Beyond its growth effect* by Hassen Abda Wako
2018-14 *Fluctuations in renewable electricity supply: Gains from international trade through infrastructure?* by Thomas Ziesemer
2018-15 *What is the potential of natural resource based industrialisation in Latin America? An Input-Output analysis of the extractive sectors* by Beatriz Calzada Olvera and Neil Foster-McGregor
2018-16 *Global Value Chains and Upgrading: What, When and How?* by Padmashree Gehl Sampath and Bertha Vallejo
2018-17 *Institutional factors and people's preferences in the implementation of social protection: the case of Ethiopia* by Vincenzo Vinci and Keetie Roelen
2018-18 *Towards a European full employment policy* by Jo Ritzen and Klaus F. Zimmermann
2018-19 *Drivers of growth in Tunisia: Young firms vs incumbents* by Hassan Arouri, Adel Ben Youssef, Francesco Quatraro and Marco Vivarelli
2018-20 *Africa Sector Database (ASD): Expansion and update* Emmanuel Buadi Mensah and Adam Szirmai
2018-21 *Combatting corruption in higher education in Ukraine* by Anna Vasylyeva and Ortrun Merkle

2018-22 *What more can we learn from R&D alliances? A review and research agenda* by Andrea Martinez-Noya and Rajneesh Narula

2018-23 *Perpetual growth, distribution, and robots* Önder Nomaler and Bart Verspagen

2018-24 *R&D, embodied technological change and employment: Evidence from Spain* Gabriele Pellegrino, Mariacristina Piva and Marco Vivarelli

2018-25 *Structural change, productivity growth and labour market turbulence in Africa* by Emmanuel Buadi Mensah, Solomon Owusu, Neil Foster-McGregor and Adam Szirmai

2018-26 *Domestic quality certification and growth of Vietnamese MSMEs* by Elisa Calza and Micheline Goedhuys

2018-27 *Knowledge convergence in European regions: Towards a cohesion?* By Semih Akçomak, Erkan Erdil and Umut Yılmaz Çetinkaya

2018-28 *Modern industrial policy and public-private councils at the subnational level: Mexico's experience in an international perspective* by Robert Devlin and Carlo Pietrobelli

2018-29 *The Eurasian customs union and the economic geography of Belarus: A panel convergence approach* Mehmet Güney Celbi Pui-Hang Wong and Tatjana Guznajeva

2018-30 *Informal sector innovation in Ghana: Data set and descriptive analysis* by Elvis Korku Avenyo

2018-31 *A guide for the evaluation of programs of human capital training for science, technology and innovation* by Diego Aboal, Marcelo Perera, Ezequiel Tacsir and Maren Vairo