The Impact of Incentives for Researchers on the Gender Scientific Productivity Gap[1]

Diego Aboal (UA Macro, aboal@cinve.org.uy) **and Maren Vairo** (UA Micro, mvairo@cinve.org.uy)

August 2016

Abstract

This paper evaluates the impact of the Paraguayan incentives program for researchers (PRONII) on the gender scientific productivity gap, using data from electronic CVs provided by all applicants to the program and from bibliographic electronic databases. We first quantify the size of the gender scientific productivity gap previous to the program. Then, we estimate whether PRONII's selection process is gender-biased. Finally, we evaluate the gender differential impact of the program. The results show a pre-existent gender productivity gap among PRONII researchers. However, we find no evidence of discrimination against female researchers at the selection stage of the program. Finally, the results show that the impact of the program is not homogeneous across genders.

Keywords: Economics of Science, Economics of Gender, Scientific Subsidies, Policy Impact Evaluation.
JEL classification: O30, O38, J16, H43, C21.

1We thank Federico Veneri for the empirical work that led to the estimations reported in table 2 of the paper. We also thank comments and suggestions by Jose Miguel Benavente, Nestor Gandelman, Mateo Grazzi, Jacques Mairesse, Jocelyn Olivari, Janet Stotsky and participants at the Gender Summit 8 North & Latin America 2016. We greatly appreciate the funding provided by the Inter-American Development Bank for this research. All remaining errors and limitations are our responsibility.

1. Introduction

There is extensive evidence pointing to gender gap in labor market. Women receive lower wages, are underrepresented in several occupations and work fewer hours than men, while having less access to productive inputs (Cuberes and Teignier, 2016).[2] This, in turn, has negative consequences in terms of economic variables such as aggregate income, productivity and economic growth due to the underutilization of female human capital (Klasen and Lamanna, 2009; Hausmann et al., 2010; Cuberes and Teignier, 2016).

Science, technology and innovation (STI) activities are no different in displaying the same gender gaps. Entrance into knowledge production activities has historically been conditioned on gender (Halbert, 2006). In fact, in Latin America and the Caribbean (LAC), only 6.5% of all patents filed in 2006-2011 were registered solely by women, while this figure amounts to 69.6% in the case of men. Jointly filed patents, on the other hand, rise to 23.9% (Morales and Sifontes, 2014). This gap in patenting activities is most likely related to the low female participation in science, technology, engineering and mathematics (STEM) disciplines (Castillo et al., 2014).

In Latin America, despite the recent narrowing of educational gaps reflecting greater female access to higher education in the region, there still remain obstacles that impair women from making full use of their STI potential, limiting any positive externalities that might emerge from greater knowledge production and preventing society away from reaching its welfare optimum. This argument gives ground for policy action aiming at reducing STI gaps between men and women. However, few national STI policies reflect this objective (UNCTAD, 2011).

Why there is a gender gap in STI careers is a complex and multidimensional issue, since it involves a wide range of cultural factors (see Castillo et al., 2014, for a broader examination of this issue). Explanations associated with the labor market suggest that women do not face equal opportunities when applying for certain academic positions (Steinpreis et al., 1999) and the wages they are receive (Petersen et al., 2000, Moss-Racusin et al., 2012).

Gender-biased recruitment and hiring procedures remain a problem for women. Evidence points to a lower propensity by faculty members to write positive recommendations for women, which hampers the progression of their academic careers (Trix and Pzenka, 2003). Furthermore, women are frequently excluded from networking practices (such as social gatherings or collaborations for research and teaching) harming their access to available information on funding opportunities (UNESCO, 2007).

Explanations of gender gaps also focus on factors that inhibit female labor supply in STI occupations. First of all, the prevailing male-dominated culture in STI careers might generate an unpleasant environment for women to work, thus restraining their entrance (Fox, 2005). This is also reinforced by discrimination and cultural stereotypes

2Also see Klasen and Lamanna (2009), Olivetti and Petrongolo (2008, 2014) and Blau and Kahn (2007, 2013).

in higher education, which dissuade women from choosing a career in science in the first place (Blickenstaff, 2005).

Additionally, the demands of motherhood resulting in women's absence at work during maternity leave and their facing larger burden of childcare activities creates a conflict with professional development by limiting the time women can devote to their professional careers. In fact, as Muller et al. (2011) point out, the crucial stages in academic careers such as the PhD and Postdoc levels usually overlap with women's period of greatest fertility. Goldin (2014) suggests that STI occupations are relatively 'motherhood-friendly' because they have certain characteristics (such as greater time flexibility and independence) which produces a less direct relationship between hours worked and wages, leading to more gender equality in wages. This contrasts with business occupations where high pay is usually over-proportionally tied to working long hours.

All in all, the obstacles faced by women when building a career in STI disciplines inhibit their academic and technical output, which in turn results in underutilization of any contributions that these highly educated women might make to society. Therefore, it is relevant to analyze whether policies contribute to correcting these losses in human productivity.

The focus of this research is on the Paraguayan Incentives Program for Researchers (PRONII). PRONII was initiated in 2011 by the National Council for Science and Technology (CONACYT) to stimulate careers in research, by means of providing researchers with a fixed monthly subsidy according to their scientific productivity.[3] CONACYT is the institution responsible for the design and implementation of STI policies in Paraguay.

PRONII evaluates researchers using academic criteria. Gender is not considered in selecting beneficiaries, so that, in theory, this is a gender-neutral program.[4] However, as the gender budgeting literature asserts (see Stotsky, 2006a, 2006b), proper design of gender equality-sensitive policies involves examining the gender effects of all policies (not only those specifically aimed at reducing gender gaps). With this in mind, we intend to provide a quantitative evaluation of this program on the gender gaps in research achievements, using various measures of researchers' productivity (i.e. research production, technical production, own education attainment, and training of other researchers).[5]

As a result of the different opportunities faced by male and female researchers, one might expect a program such as PRONII to have differential gender impacts. Even though recent *gender mainstreaming*[6] still lacks a *full articulation of a theory of change* (Daly, 2005), one might hypothesize that a program providing equal opportunities can

---

3 Similar programs exist in other countries in the region, for example in Argentina, Mexico, and Uruguay. Therefore the lessons to be learnt in Paraguay could be useful for other countries in Latin America.

4 Part of our empirical approach involves analyzing whether this principle holds in practice.

5 We will define more precisely what we mean by research and technical production in the next sections.

help reverse the prevailing negative discrimination against women by reducing the obstacles that inhibit their academic careers. This could lead to larger scientific productivity at the individual level; while, in the long run, greater female presence in the area might favor the change of the current cultural standards that lead to gender segregation in STI.[7]

In the Latin American region, where policies to close gender gaps are scarce, a good starting point in terms of conceiving a gender mainstreaming approach to STI programs would be to evaluate the differential impact of existing policy incentives on men and women, and thus whether they are gender-neutral or help close gaps. Internationally, even though there are available evaluations of STI programs specifically directed at women, there is no evidence regarding how *ex-ante* gender-neutral programs supporting researchers affect men and women differently. Such assessments might provide insight into the on the existing gaps, while signaling where extra efforts should be made with the purpose of narrowing the STI gender gap.

The questions to be addressed in this paper are: Was there a pre-existing gender gap in academic productivity?, Does the program implicitly (at the selection stage) discriminate against women?, What is the impact of the program on academic productivity (publication, technical outputs, number of theses directed, education of researchers, etc.)?, Is there evidence of differential impacts on men and women?

In what follows, in Section 2 we characterize research and development (R&D) policy in Paraguay and describe PRONII. In section 3, data and methods are discussed. Section 4 presents the main results obtained in terms of quantifying the gender gap in STI, assessing the gender-neutrality of the program at the selection stage, and evaluating gender specific impacts of PRONII. Finally, in Section 5 the main findings are discussed and some conclusions and policy implications are presented.

2. R&D policy in Paraguay and the PRONII program

The Paraguayan government has undertaken considerable efforts to support R&D activities in recent years, by tripling its R&D investment from 6.5 million dollars in 2005 to $21.7 million in 2012 (CONACYT, 2012). However, given that growth has recently been strong in Paraguay, this increase in spending resulted in only a very minor increase in its share in GDP during this period, from 0.080% to 0.085%. Additionally, there was a considerable growth in the number of researchers from 543 in 2005 to 1,521 in 2012 (CONACYT, 2012). Related to this, the production of local knowledge increased significantly, so that publications indexed in the Science Citation Index (SCI) and in Scopus went from 41 and 45 respectively in 2005 to 101 and 135 in 2012 (CONACYT, 2012).

All of this happened in the context of an effort to strengthen the CONACYT. In particular, in 2011, CONACYT introduced PRONII, which was inspired by the National

6 Gender mainstreaming is the process of assessing the different implications for women and men of any policy action (United Nations, 2002).

7 However, available evaluations of programs looking to favor women participation in STI show that they have not resulted in structural changes at the institutional level (Muller et al, 2011)

Research Systems of Mexico and Uruguay with the objective of strengthening and expanding the scientific community of Paraguay. With that purpose, the program aims to promote careers in research and researchers' productivity by categorizing researchers according to their scientific and technological production, and providing them with a monetary subsidy according to their categorization in the system. In order to evaluate researchers and define their categorization, the program asks candidates to fill out a standardized CV through an electronic platform called CVPY, which is publicly available at CONACYT's website.

With that information, the following criteria are used to evaluate applicants to the program:

1. Production of basic research, applied research and technological outputs of proven quality.
2. Level of education.
3. Participation in the development of other researchers' capabilities (mainly through the direction of undergraduate and graduate theses).
4. Participation in the creation and strengthening of institutional capacities for research and experimental development.
5. The quality of research, which is judged taking into account: publication in refereed journals, where indexed international journals are considered of greater value, followed by regional and national journals respectively; patents and original technological outputs; leadership in the field.

As a result of this assessment, researchers are categorized into the system where there are four possible levels: Candidate, Level I, Level II and Level III. Level I to Level III researchers received a monthly subsidy of approximately 700, 1,400 and 2,100 dollars respectively in the year 2012, while Candidate researchers are not subsidized by the program. Researchers in the Candidate level are not given any other material incentive, apart from the prestige to belong to the system and the hope to progress in it.[8] As a reference, the annual GDP per capita of Paraguay in the year 2012 was 3860 dollars (322 dollars per month). Therefore, the subsidies were very important for the context of Paraguay. The duration of the subsidy is 2, 3 and 5 years for Levels I-III respectively. After this period researchers are evaluated again and if they have a poor performance during these years, they are excluded from the system. Therefore, the incentive to improve performance in the items or dimensions listed in the previous paragraph comes from the threat to be excluded from the system and therefore to lose the subsidy.

3. Data, methods and descriptive statistics

8 Since the second call of the program in the year 2013, PRONII established a subsidy for Candidates as well.

3.1 Data

We will exploit data coming from the electronic CVs of PRONII's applicants. This information has been previously employed by Aboal et al. (2016) with the objective of assessing the impact of the program, but gender issues have not been previously addressed.

Some examples of the variables available in the CVPY database are presented in the following diagram (Figure 1).

When carrying out the impact evaluation, we mainly focus in four dimensions of researchers' performance: research production, technical production, level of education (i.e. the completion of a Master's or a PhD degree) and contribution to the formation of new researchers (through the direction of theses). Total research production includes working papers, conference papers, both published and accepted for publication papers, and books and book chapters. We also carry out a separate analysis for publication in scientific journals, publications in Scopus journals and the quality of the journals where the researcher is publishing (through the mean SCImago Journal Rank (SJR) indicator of the journals).[9] Technical production includes the following items: technical work (advisory activities, consulting, development of regulations and ordinances, etc.); technological products (such as new varieties of plants, prototypes, software, etc.); and the introduction of new processes or techniques (e.g. management processes or analytical techniques).

Figure 1. Examples of variables available in the database

3.2 Methodology

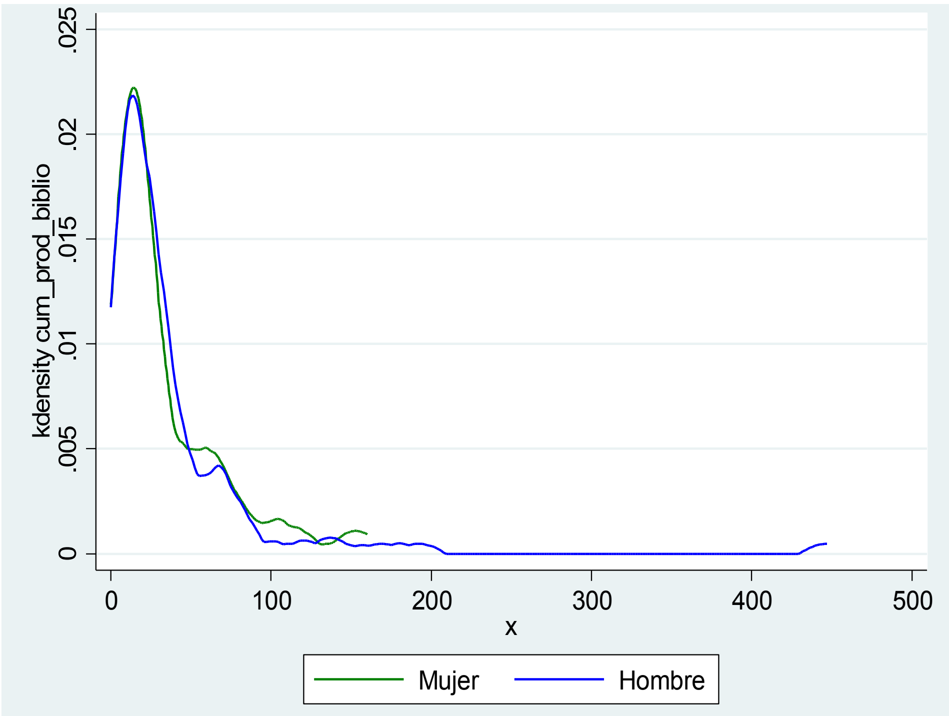
Firstly, we propose to assess the existence of a gender gap in scientific production prior to program implementation. This implies modeling a scientific production function, where relevant outputs are the result of certain inputs. Given the characteristics of the output variables (i.e. written research and technical production,

9 See Guerrero-Bote and Moya-Anegón (2012) for details about how the SJR indicator is computed for each journal. We computed for each researcher the mean of the SJR indicator of the journals where they published. This was done for every year where researchers have publications in journals ranked by the SJR (note that this indicator uses the Scopus database to compute the SJR indicator, therefore journals have to be indexed in Scopus in order to appear in the SJR ranking).

papers in scientific and in Scopus journals, and theses directed), the econometric estimation of that production function will be carried out through a negative binomial count model. To that extent, since the dependent variables are a count of the scientific outputs generated by the researcher, it is appropriate to choose a model from the count-data family in order to achieve consistent and efficient estimates.

Furthermore, as Figure 2 depicts, we find over-dispersion in the dependent variables in that, while most researchers have an accumulated number of publications below 50, there are some "rare" observations where the number of total publications is over 100.[10] This is typical of scientific production. Therefore, we choose to specify a negative binomial regression model (NBRM) to account for the over dispersion of the response variable[11] (Hausman et al., 1984, Long and Freese, 2006). Some variants of this model have been previously used in the literature to assess the production of researchers and the number of patents by universities. For example, Gonzalez-Brambila and Veloso (2007) use a negative binomial fixed effect model to address determinants of research output of Mexican researchers with a similar database.

Figure 2. Density of written research production until 2011, by gender



More formally the negative binomial model can be written as:

10Here we display the distribution of bibliographic production only. The graphs for the other response variables considered lead to the same conclusions, and can be found in the Annex.

11This contrasts with the alternative of using a Poisson count model which assumes a constant variance.

where is the expected number of outputs (where the output is) for individual *i* (it can number of publications, number of theses directed, etc.), is a vector of explanatory variables, is a vector of parameters and , with the assumption that $E(e^{\alpha i})=1$, is a fixed effect (unobserved heterogeneity among individuals) that in practice allows for overdispersion in the data. The exponential function in this count data model is key to avoid the prediction of negative number of outputs.

The distribution of observations given the value of and , , follows a Poisson. Under the assumption that is a draw from a Gamma distribution, follows a negative binomial distribution, what gives the name to the model.

We apply this model to the five selected researchers' output variables before program implementation. To that extent, we use the total accumulated production since the researcher attained his or her undergraduate degree until 2011 (the pre-program year). As for the selection of the relevant factors that explain scientific production, we follow Gonzalez-Brambila and Veloso (2007) (although some of the variables suggested by their work are not included here due to data availability). As a result, the explanatory variables used in the model are: gender, age, education (i.e. if the researcher has a Masters or/and a PhD degree) and the field of research.

We also include the number of years since the researcher attained his or her undergraduate degree in order to control for time exposure. Because we are counting outcomes, it is important to "normalize" the count by the number of potential years of production, we do not expect the same number of outcomes in young and old researchers (the later had potentially more production just because they had more years to produce). This is done in the following way (as it is usual in the literature),

where is the number of year after the researcher attained the undergraduate degree (is also known as the exposure time in the literature).

The main purpose of this exercise is to find out whether gender is a factor determining scientific output, which would provide evidence supporting the existence of a gender gap prior to the implementation of PRONII, thus justifying the relevance of studying the impact of the program on this gap.

Second, to evaluate the impact of PRONII on researchers' productivity and the possibility of heterogeneous treatment effects by gender we propose the use of propensity score matching (PSM) (Rosenbaum and Rubin 1983; Abadie and Imbens, 2006) with difference in difference (DiD) (Angrist and Krueger, 1999). To that extent, the control group for each level of PRONII researchers will be those who were categorized in the previous level in the 2011 call, and for the Candidates level controls will be the applicants that were rejected in the 2013 call and that did not enter the system in the 2011 call.[12] As an identification strategy, we will exploit the fact that the subsidy is increasing with the level in which researchers are categorized.

12We chose not to use rejected applicants from the 2011 call as a control group for Candidates because this is a very small group.

It is worth noting that, given the definition of the control group, impacts of PRONII will be estimated on the margin: i.e. it is the impact of belonging to a certain category compared to the possibility of being part of the category below. For instance, we will see how the greater subsidy received by Level III researchers leads to greater productivity when compared to their colleagues in Level II. Thus, our counterfactual scenario does not represent the situation in which the program did not take place (except for Candidates), so that the failure to identify any significant impacts does not mean that PRONII did not have any effects.

This strategy is used because individuals in the previous level of the program produce a better control group than the pool of individuals who did not enter the program. Note that we also use matching techniques, more specifically, propensity score matching, in order to find similar individuals in the control group (balancing test will be computed in order to see how good is the match).

The impact of a program from the DD model is,

where is the variable of interest for the post program period and the value of the variable for the baseline period, and D is the treatment indicator.

To get an unbiased estimation using a DiD approach, assignment to the program must be exogenous (i.e. uncorrelated to any unobservable characteristics that might affect both treatment condition and results). This might not be the case in PRONII since researchers are selected into each level due to their previous performance, so that treated and control units are plausibly different. In this case, an alternative strategy that allows us to control for observable differences between both groups is to use DiD combined with PSM.

In order to implement a PSM strategy, the probability of being treated is estimated given the characteristics of the individual measured before treatment (age, gender, level of education, field of research, scientific production) with a probit model. Using this estimated probabilities (or *propensity scores*), we match each treated individual to its closest match (in terms of the propensity score) in the control group. This allows us to obtain a more precise identification of the average treatment effect, by controlling and matching for observable characteristics. As Tables A.3-6 in the Annex show, this matching procedure allows us to properly control for observable differences between treated and control researchers, achieving a balanced sample.

The average treatment effect can be expressed as follows,

.

The DiD with propensity score matching estimator is the difference in the relevant variable before and after the treatment, among the treated and the control group compared on the common support (P(X), using PSM).

Additionally, gender-specific effects are obtained by estimating the previous equations separately for male and female researchers. Any gender-differentiated impacts of the program would mean that it has some effect in terms of narrowing (or expanding) the gender gap in scientific production.

In summary, the estimation of a count model can be used to identify the existence of a gender gap among researchers previous to the program and the use of PSM with DiD can be used to test the existence of heterogeneous treatment effects by gender and thus the program's impact on gender gaps. Moreover, in first instance, by estimating the PSM including a gender variable, we will be able to identify if gender considerations were used at the selection stage of the program.

4. Results

4.1 Gender and researchers' outputs

In order to approach the issue of gender-based differences in scientific production, Table 1 shows the aggregate production of researchers belonging to PRONII before entering the program. The simple comparison of the mean by gender shows that women in our sample produce lower levels of output than men, for all the different types of outputs. Of course, the simple comparison of means overlooks the fact that other determinants of researchers' output are not held constant between women and men. In the regression analysis that follows, we will test more formally the differences in outputs across genders.

Table 1. Accumulated scientific production before PRONII, by gender

Gender	Stats	Written research production	Technical production	Papers in Scientific Journals	Papers Scopus	Thesis directed
Female	mean	33.5	3.6	13.4	2.0	7.8
	sd	34.5	7.9	19.1	3.7	13.9
	min	0	0	0	0	0
	max	160	56	130	23	83
Male	mean	36.6	4.2	13.9	2.6	8.8
	sd	54.7	8.1	30.0	5.1	12.0
	min	0	0	0	0	0
	max	447	49	235	34	53
Total	mean	34.7	3.8	13.6	2.2	8.2
	sd	43.8	8.0	24.1	4.4	13.2
	min	0	0	0	0	0
	max	447	56	235	34	83

Notes: sum of items in the period previous to the program and after obtaining the bachelor degree.

Additionally, in this section we estimate negative binomial models[13] in order to investigate the existence of a gender bias in researchers' aggregate output in the period since they got the bachelor degree until the year 2011 (pre-program year). Apart from the gender dummy (male=1) that is the main variable of interest in this

13 An alternative estimation strategy would be to follow the panel approach of Rivera, Mairesse, Cowan (2016).

section, we include the following control variables in the regressions: age dummies, dummies for level of education, dummies for research field and years of experience (number of years after obtaining the bachelor degree).

The results in Table 2 show that women produce fewer written research outputs and papers published in scientific journals. Women publish 0.42 and 0.58 per year fewer written research outputs and papers in scientific journals, respectively. This is evidence of existence of a gender productivity gap in scientific publications. We did not find a gender gap in the number of technical outputs, papers published in Scopus journals and theses directed. Also, we find that age is an explanatory factor when it comes to accumulated direction of theses (controlling for the number of years since the researcher obtained his or her degree), and that the field of science appears to be relevant when explaining almost all of our output variables. To this extent, Agricultural and Natural Sciences (Field 1) and Medical Sciences (Field 3) appear to be the most productive disciplines. The researcher's level of education, on the other hand, surprisingly does not appear as a relevant input in this production function.

In conclusion, this first analysis points to the existence of a gender gap in scientific production prior to the program's implementation. In what follows, we will investigate whether PRONII had any impact on this pre-existing gap.

Table 2. Estimation of scientific production function, marginal effects

Variables	Written research outputs	Technical outputs	Papers in scientific journals	Scopus Papers	Theses directed
Gender (Male=1)	0.418***	0.0670	0.582***	0.358	0.233
	(0.121)	(0.294)	(0.164)	(0.285)	(0.237)
Age 31-40	-0.119	0.497	0.205	0.281	0.368
	(0.221)	(0.550)	(0.311)	(0.554)	(0.441)
Age 41-50	-0.208	0.713	0.245	-0.0934	1.029**
	(0.214)	(0.530)	(0.303)	(0.541)	(0.431)
Age 51-60	0.0714	0.856	0.492	-0.215	1.169**
	(0.227)	(0.560)	(0.318)	(0.567)	(0.458)
Age >=61	-0.413	0.842	-0.341	-0.982	1.039**
	(0.272)	(0.673)	(0.376)	(0.676)	(0.507)
Master	-0.189	0.384	-0.301	-0.365	-0.274
	(0.143)	(0.338)	(0.189)	(0.347)	(0.277)
PhD	0.162	0.132	0.257	0.250	-0.102
	(0.144)	(0.334)	(0.185)	(0.315)	(0.273)
Field 2 (Engineering, etc.)	-0.118	0.0108	-1.058***	0.575	-0.145
	(0.199)	(0.457)	(0.273)	(0.435)	(0.354)
Field 3 (Health Sciences, etc.)	0.510***	-0.507	0.654***	0.550*	-0.743***
	(0.138)	(0.325)	(0.180)	(0.311)	(0.280)
Field 4 (Social Sciences, etc.)	-0.297**	-0.531	-0.830***	-1.398***	-0.674**
	(0.146)	(0.331)	(0.194)	(0.380)	(0.269)
Constant	0.308	-2.348***	-1.048***	-2.417***	-1.416***
	(0.225)	(0.588)	(0.312)	(0.584)	(0.483)

lnalpha	-0.569***	1.011***	-0.110	0.870***	0.566***
	(0.0955)	(0.125)	(0.104)	(0.155)	(0.110)
ln(years after bachelor degree)	YES	YES	YES	YES	YES
Observations	230	230	230	230	230

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Note that the model imposes the restriction that the coefficient of ln(years after bachelor degree) must be equal to 1. In the regression the excluded age is 21-30 and the excluded field is Agricultural and Natural Sciences, therefore, the coefficients of age and fields variables must be interpreted as differences with respect to these excluded categories.

4.2. PRONII's categorization and researchers' output

Table 3 shows the number of researchers by gender, field of science and the categorization awarded to them by PRONII during its 2011 call (with the exception of the *rejected* category in which we consider categorization from the 2013 call). Overall, 236 researchers were granted categorization during this call: 109 of them were admitted as Candidates, while 89, 26, and 12 researches were admitted under Level 1, 2 and 3 categories respectively. As for the gender composition of the different categories, we find that there is similar participation of male and female researchers in the rejected group, while the share of women inside the Candidates and Level 1 and 2 categories is larger than that of men. However, we find a considerable gender difference in favor of men among Level 3 researchers, where there was only one woman in this category, compared to 11 men.

The outcome for researchers who applied for the 2011 call are used in order to define the treatment and control groups for the impact evaluation exercise, so that rejected researchers serve as control units to evaluate the impact of being categorized as a Candidate. Candidates then form the control group when assessing the effect of belonging to Level 1, and so on. Table 3 provides information on the sample size in each of these exercises. To that extent, we see that, while we are quite comfortable with the number of observations for evaluating impacts at the lower PRONII categories, sample size becomes a concern when it comes to Levels 2 and 3. This concern worsens when attempting to divide the sample by gender. As a result, we omit the exercise of evaluating impacts at the Level 3 rank, where we would be working with only 38 control and treated units. Also, while separate results for male and female Level 2 researchers are presented, we warn the reader to read those results cautiously since sample size might be too low to properly identify impacts.[14]

Table 3. Number of researchers by field and category

14 In particular, the lower the sample size, the lower the power to identify small impacts (i.e. the lower the minimum detectable effect), so this becomes a concern whenever expected impacts are small.

Category\ Field	1	2	3	4	Total
Rejected	25	12	52	22	111
Female	10	7	31	4	52
% Female	*40%*	*58%*	*60%*	*18%*	*47%*
Candidate	32	46	24	7	109
Female	18	39	14	1	72
% Female	*56%*	*85%*	*58%*	*14%*	*66%*
Level 1	36	23	21	9	89
Female	15	21	12	3	51
% Female	*42%*	*91%*	*57%*	*33%*	*57%*
Level 2	7	11	5	3	26
Female	5	8	1	1	15
% Female	*71%*	*73%*	*20%*	*33%*	*58%*
Level 3	3	4	2	3	12
Female	0	1	0	0	1
% Female	*0%*	*25%*	*0%*	*0%*	*8%*
Total	103	96	104	44	347
Female	48	76	58	9	191
% Female	*47%*	*79%*	*56%*	*20%*	*55%*

Source: own elaboration based on CVPY

Notes: Fields of Science: (1) Agricultural and Natural Sciences (2). Health Sciences (3). Social Sciences and Humanities and (4) Engineering and Technology. The category "rejected" is composed by those who were rejected in the 2013 call and did not enter the system in 2011 (because they were rejected or because they did not apply in that year)

Furthermore, in order to carry out the DiD strategy, we need to gather information on researchers' performance (and other relevant characteristics) before and after program implementation. With that purpose, we choose to work with the period corresponding to the two years before the program (2010-11) and the two years after (2012-13). The logic behind the DiD exercise is illustrated in Table 4. There, we show the relevant results attained by the different groups of researchers before and after the implementation of PRONII. The impact of the program that would result from the most straightforward form of the DiD estimation can be derived from the table as the difference between the mean results in the treatment group before and after PRONII, minus the analogous difference in the control group.

When considering the entire sample, we find that the number of researchers whose maximum education level is a Master's degree increased over time in the rejected and candidate categories, but has remained stable in the higher levels, possibly due to the fact that having a PhD level is already a requirement for entering the latter. On the other hand, we find an increase in the number of researchers with a PhD in all levels, which could be associated with the incentives offered by PRONII, by providing monetary assistance, thus allowing researchers to devote themselves to their PhD studies (this would apply in the case of Level 1-3 researchers who are entitled to a subsidy). This incentive could work even in the case of non-paid researchers (rejected and candidates) through the expectation that they might rise to a higher level later on.

As for the other production variables, we find that outputs generally increase over time. This increase is observed in terms of the contribution to others' human capital through the direction of theses, and in technical and written research production. There are ambiguous results in terms of the quality of publications in Levels 1-3, given that in such categories the number of Scopus publications has moderately increased, while there is a decrease when measuring quality by the SJR indicator. Once again, it is no surprise that after the implementation of PRONII researchers' educational attainment and their scientific production increased, given that these variables are directly aimed at the program and are used as performance indicators in order to decide on the amount of subsidy received.

Furthermore, if we analyze researchers' production before PRONII (in 2010-11), we find that pretreatment production increases with the categorization received by the applicant in 2011. This means that PRONII was effective in directing its support, given that the program managed to select into each category the most deserving candidates according to the conditions of admission into each level. Here, it is worth pointing out the exception of Level 3 researchers, who show a very similar performance to their Level 2 colleagues. However, this is probably due to the fact that the requirements to reach Level 3 are less tied to scientific production as measured by the indicators in Table 4 and more related to alternative characteristics (such as whether the individual develops an academic network or helps to build institutional capacities).

When dividing the sample by gender, we find that the overall pattern in terms of pre-PRONII characteristics remains unchanged (with the exception that female researchers appear to be more active in terms of publishing in scientific journals in 2010-11 than males). Also, the behavior of both groups of researchers appears to have evolved similarly over time, in that both female and male researchers show a general improvement in their scientific performance. There is an exception here in the case of Candidates, given that male candidates show a larger proportional improvement in most production indicators than women.

Table 4. Researchers' performance before and after PRONII according to their category

	All sample	*Female researchers*							*Male research*
	Rejected	**Candidate**	**Level 1**	**Level 2**	**Level 3**	**Rejected**	**Candidate**	**Level 1**	**Level 2**

Master's (number of people who attained up to a Master's degree)									
2011	45	46	29	3	1	25	28	14	2
2013	59	53	31	3	1	30	36	16	2
PhD (number of people who attained up to a PhD)									
2011	16	20	27	20	11	9	10	12	10
2013	20	24	30	22	11	12	11	14	12
Theses directed-concluded (mean per year)									
2010-2011	0.653	0.771	2.219	2.385	1.792	1.038	0.757	2.108	2.500
2012-2013	0.824	0.972	2.865	3.538	1.667	0.904	0.868	2.814	3.867
Theses directed-in process (mean per year)									
2010-2011	0.054	0.170	0.258	0.365	0.708	0.106	0.132	0.216	0.367
2012-2013	0.509	0.408	1.652	0.942	0.750	0.365	0.306	1.461	0.933
Technical production (mean per year)									
2010-2011	0.297	0.335	0.438	0.596	1.667	0.250	0.382	0.382	0.700
2012-2013	0.347	0.454	0.798	0.654	3.125	0.433	0.424	0.804	0.667
Written research production (mean per year)									
2010-2011	0.914	2.413	4.612	5.462	6.167	0.731	2.569	4.941	5.767
2012-2013	1.536	2.193	5.567	7.442	7.792	1.269	1.951	5.039	8.500

Papers in scientific journals (mean per year)									
2010-2011	0.095	0.784	1.635	2.192	2.125	0.087	0.854	1.931	2.367
2012-2013	0.239	0.881	1.697	2.923	2.000	0.240	0.868	1.941	3.967
Scopus Papers (mean per year)									
2010-2011	0.041	0.101	0.438	0.865	0.917	0.038	0.104	0.441	1.033
2012-2013	0.032	0.161	0.483	1.212	0.875	0.029	0.132	0.392	1.733
Quality of papers-mean SJR (mean per year)									
2010-2011	0.022	0.078	0.217	0.482	0.446	0.020	0.098	0.208	0.625
2012-2013	0.029	0.082	0.128	0.373	0.147	0.029	0.048	0.124	0.509

Source: own elaboration based on CVPY.

Note: The category "rejected" is composed of those who were rejected in the 2013 call and did not enter the system in 2011 (

4.3. Probability of participation in the program

In this section we present the results from the first stage of the impact evaluation exercise, in which we estimate a propensity score that models the probability of being categorized into different PRONII levels, according to researchers' characteristics. This first stage allows us to identify whether there is gender-based discrimination at the program's selection stage, by analyzing whether gender is a significant factor explaining the propensity score (or the probability to enter the program).To that extent, Table 5 reports the results from the Probit estimation of the probability of entering the different categories defined by the program. This information is later used to match treatment units to their "nearest neighbors" in the control group.[15]

We do this exercise by separately taking into account the probability of being categorized as a Candidate, Level 1 and Level 2 researcher. Additionally, for Level 2 researchers, we carry out an evaluation after two and three years of the implementation of PRONII. The first exercise is similar to that of the remaining categories, since it involves comparing Level 2 researchers from the 2011 call with their Level 1 counterparts. However, in the second case, the control group is composed of those researchers who entered the Level 1 in the 2011 call and remained in that level after the 2013 call. We choose to perform this double exercise for this category because Level 2 researchers are requested to reapply to the program every three years in order to preserve their categorization, so that it is reasonable to assume that their electronic CVs were updated by 2014 (differently than the researchers belonging to the others categories who do not have this incentive to update their CVs after 2013). As was stated before, we do not report results on Level 3 researchers because there are too few observations.

The first 4 columns in Table 5 show the estimates that result from including every dimension that *a priori* could affect program participation: gender, age, education, different indicators of average yearly academic production and area of science. It is important for matching purposes that any variables included in this first step are measured before program implementation, so that information used in this equation is measured in 2011. This first specification allows us to assess which factors determine the researchers' categorization. One interesting result from this exercise is that gender does not significantly determine PRONII entrance in any level. This is to be expected since the program does not establish any gender-specific criteria for selection, thus confirming the program's neutrality in treatment allocation.[16]

Also, we find that the determinants of being admitted vary according to the different categories of PRONII. In this sense, probability of being categorized as a candidate is positively affected by age, having a PhD, previous written research production and publication in scientific journals, and is negatively related to being a researcher in the fields of Engineering and Technology or Social Sciences and Humanities (with respect to the omitted category which is Agricultural and Natural Sciences). This is consistent with the conditions for entering this level, such as demonstrating participation in research activities through publications and having completed or

---

15We also conducted the Probit estimations dividing the sample by gender. Results are qualitatively similar than those reported in Table 5.

16 This results are different from Bukstein and Gandelman's (2016) analysis for the analogous program in Uruguay (the National Research System), who find a 6.7% lower probability for female entrance into the system.

taking part in a graduate program. Bias against certain areas or selection based on researchers' age, however, is not a feature of the program's screening process.

Entrance to Level 1, on the other hand, is positively related to the number of concluded directed theses, written research production, and publications in Scopus journals, while being negatively related to being in the Health Sciences.[17] PRONII requires Level 1 researchers to have a Master's or PhD degree or equivalent scientific production, demonstrating the capability to carry out original research independently. However, it is scientific production (through publications), rather than the researcher's graduate education, that appears to influence most the entrance into Level 1. In a second specification in column 6 (which ends up being selected for estimating the propensity scores) we find that having a Master's degree negatively affects the possibility of belonging to Level 1. As Aboal and Tacsir (2016) point out, this result, which implies that having up to a Master's degree increases the probability of being categorized as a Candidate instead of belonging to Level 1, might mean that PRONII evaluators assessed that the applicant did not have sufficient education to enter Level 1.

Additionally, entrance into Level 2 is positively determined by having a Master's or a PhD, by the quality of publications and by belonging to the Health Sciences area. This is consistent with the fact that Level 2 researchers are required to have a PhD degree (or equivalent scientific production). Additional requirements for entering this level are to possess a "strong track record of work, particularly in the five years prior to each call of PRONII, having developed its own line of research with sustained production of original knowledge and activities aimed at capacity building for research" (Aboal and Tacsir, 2016). Because quality of publications might be an indicator of compliance with these requirements, it seems reasonable for it to enter the Probit estimation. However, contribution to *capacity building for research* through the direction of theses does not appear to affect entrance into this level. Lastly, we also find at this level, bias (in that it does not follow from the program's selection criteria) related to the researcher's area of science.

In order to avoid over-specification of the Probit model, we chose to include in the estimation of the propensity score only those variables that were significant at the 10% level. To that extent, results for the final specification of the estimated propensity scores are presented in columns 5-8.[18]

Table 5. Probit estimates for probability of participation, by PRONII category

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Candidates	Level 1	Level 2- 2 years	Level 2- 3 years	Candidates	Level 1

17 In a second specification (column 6 of Table 3), where other non significant variables are removed, age appears as another factor that positively affects the probability of being categorized into Level 1, while researchers whose maximum education level is a Master's degree appear to be less likely to enter this level.

18 In order to achieve a balanced matching procedure, we removed PhD as a control in the Candidates estimation (this variable turned out non-significant when removing other non-significant variables included in column 1), and added age and Master's degree into the Level 1 estimation (both variables are non-significant in column 2, but turn out to be significant in the new specification in column 6).

Female	0.162	-0.0426	0.0216	0.00879		
	(0.100)	(0.104)	(0.0749)	(0.0929)		
Age	0.116***	0.00376	0.0466	0.0521	0.127***	0.0222***
	(0.0451)	(0.0430)	(0.0343)	(0.0406)	(0.0391)	(0.00503)
Age^2	-0.00139**	0.000266	-0.000372	-0.000440	-0.0015***	
	(0.000541)	(0.000488)	(0.000336)	(0.000395)	(0.000474)	
Master’s	0.175	-0.160	0.378*	0.490**		-0.170*
	(0.104)	(0.119)	(0.206)	(0.231)		(0.0924)
PhD	0.295**	-0.0148	0.603***	0.699***		
	(0.114)	(0.138)	(0.123)	(0.131)		
Theses directed_concluded	-0.0282	0.0740***	0.00401	0.0141		0.0722***
	(0.0294)	(0.0275)	(0.0126)	(0.0152)		(0.0263)
Theses directed_ongoing	0.0866	0.00213	0.0343	-0.0411		
	(0.101)	(0.107)	(0.0657)	(0.0800)		
Technical production	0.0268	-0.0532	0.0168	0.00734		
	(0.0571)	(0.0505)	(0.0427)	(0.0476)		
Written research production	0.0511**	0.0717***	-0.00693	0.0105	0.0446**	0.0645***
	(0.0236)	(0.0263)	(0.0110)	(0.0175)	(0.0211)	(0.0201)
Papers in scientific journals	0.648***	-0.0138	0.00230	-0.0289	0.619***	
	(0.129)	(0.0585)	(0.0220)	(0.0294)	(0.116)	
Papers in Scopus	0.239	0.281*	-0.00552	0.116		0.419***
	(0.291)	(0.162)	(0.0522)	(0.0753)		(0.125)
Mean SJR	-0.0800	0.231	0.223**	0.214**		
	(0.388)	(0.196)	(0.0932)	(0.104)		
Field 2	-0.289**	0.0810	0.0910	0.0780	-0.315***	
	(0.138)	(0.177)	(0.168)	(0.181)	(0.115)	
Field 3	0.0551	-0.476***	0.262	0.345*		-0.417***
	(0.139)	(0.110)	(0.166)	(0.201)		(0.0938)
Field 4	-0.375***	-0.182	-0.0162	0.0976	-0.350***	
	(0.116)	(0.116)	(0.101)	(0.167)	(0.0892)	
Observations	205	186	113	100	210	186

Source: own elaboration based on CVPY.

Note: The 0 category is rejected applicants in 2013 in columns 1 and 5; Candidates in 2011 in columns 2 and 6; Level 1 researchers in 2011 in columns 3 and 7; Level 1 researchers in 2011 and 2013 in columns 4 and 8. The omitted field of science is Agricultural and Natural Sciences. Mean of variables in 2010-2011, except in the case of Masters and PhD dummies where we use degree attainment by 2011.

4.4. The gender specific impact of PRONII

In this section we review the results from the impact evaluation of PRONII. In Table 6, we report the average treatment effects of PRONII on the different researchers' performance indicators, distinguishing them by categorization and by gender. As was explained above, we use a 5-neighbor PSM matching approach combined with DiD techniques.[19] This implies adopting a traditional matching approach, while defining the dependent variable as the change in the result of interest before and after program implementation.

At the Candidates level, we only find a positive and significant impact of the program in the case of publications in scientific journals, in that being categorized as a Candidate in PRONII derives in publishing 0.45 more papers yearly compared to those researchers who were rejected by the program. Additionally, this effect appears to be explained by the behavior of male researchers, in that the program has no impact on women at this level. Thus, in terms of the program's contribution to the gender gap, we can conclude that at the Candidates level (where no monetary incentives are at stake), PRONII had no impact on the gap, except in the case of publications in scientific journals where it appears to contribute to broadening the existing gender gap.

The results are a little different when we look at Level 1 researchers. Here, we find a positive impact of entering this category (as compared to entering the system as a Candidate) in terms of the ongoing direction of theses (0.79 more theses directed), and technical and written research production (with 0.75 additional technical outputs and 1.10 additional written research outputs per year). This result is reasonable since all of these indicators are used to assess researchers' permanence in Level 1 or their promotion to Level 2. Also, since Level 1 researchers do receive a monetary subsidy, one might argue that they have bigger incentives to enhance their production and less financial needs that might distract them from their academic activities compared to Candidates. We also find that the impacts on technical and written research outputs are due to a positive effect on female written research output, so that subsidies provided at this level appear to be closing the gender gap in this dimension. Lastly, we find a negative impact of Level 1 categorization on male researchers' quality of publications. Here the concern might be that the program is over-encouraging the number of publications, and that researchers might be disregarding their quality in turn. However, no positive impacts on the number of publications by male researchers are found at this stage.

---

19 Results from using 1 neighbor matching are shown in Table A.2 in the Appendix. In the 5-neighbor PSM each treated individual is compared with a weighted average (of the relevant variable) of the 5 closest neighbors (the 5 most similar individuals in the control group).

Table 6. Impact of PRONII on researchers' performance: overall and by gender.
Difference in Difference estimates with 5-neighbors PSM

	(1)	(2)	(3)	(4)	(5)
VARIABLES	**Master**	**PhD**	**Theses directed (concluded)**	**Theses directed (ongoing)**	**Technical production**
Candidates					
ATT_All	-0.003	0.009	0.326	0.006	-0.245
Obs_All	174	174	174	174	174
ATT_Women	0.089	-0.036	0.909259	0.025	-0.410
Obs_Women	77	77	77	77	77
ATT_Men	-0.169	0.069	-0.004	-0.165	0.250
Obs_Men	82	82	82	82	82
Level I					
ATT_All	-0.070	0.038	0.752	0.786**	0.752**
Obs_All	162	162	162	162	162
ATT_Women	-0.057	0.057	0.683	0.380	0.971**
Obs_Women	101	101	101	101	101
ATT_Men	-0.024	0.012	0.106	0.176	0.365
Obs_Men	50	50	50	50	50
Level II-2 years					
ATT_All	0.001	0.068	0.705	-0.563	-0.186
Obs_All	114	114	114	114	114
ATT_Women	-0.119*	0.133	0.056	-0.196	-0.152
Obs_Women	65	65	65	65	65
ATT_Men	0.008	-0.008	-0.436	-2.056**	0.689
Obs_Men	45	45	45	45	45
Level II-3 years					
ATT_All	-0.070*	0.104	-0.206	0.211	-0.242
Obs_All	93	93	93	93	93
ATT_Women	-0.129**	0.143	0.105	0.993**	-0.043
Obs_Women	58	58	58	58	58
ATT_Men	0.010	-0.010	-0.115	-0.899	0.848*
Obs_Men	39	39	39	39	39

Source: own elaboration based on CVPY

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Dependent variable in columns 1 and 2 is the change in the accomplishment of a Master's or PhD degree from 2011 to 2013. For columns 3-9, the dependent variable is the difference between the mean production in 2010-11 and the mean production in 2012-13. The only exception is in the case of the evaluation of PRONII in 3 years (bottom panel) where we use change in Master's and PhD from 2011 to 2014, and changes in mean production from 2010-11 to 2012-14. Control variables used in each panel are the ones that result from columns 5-8 in Table 5. ATT is the average treatment effect on treated. Obs is the number of observations.

As was stated before, results for Level 2 researchers ought to be interpreted cautiously due to the small size of the sample (especially when it comes to gender-specific impacts). In any case, results at this stage of the system show a negative effect in female Level 2 researchers (with respect to their Level 1 female colleagues) in terms of contributing to Master's degrees in students. This negative effect is observed both after two and after three years of program implementation. Also, we find a positive impact on male Level 2 researchers' technical production (0.85 more technical outputs per year) that takes place after three years of PRONII, which would contribute to broadening the gender gap in this dimension. On the contrary, the negative effect found on the ongoing direction of theses in men after two years (2 fewer theses per year) and the positive effect found for women after three years (1 more yearly) would both contribute to closing the gap. Similar conclusions might be drawn from the positive impact on women belonging to Level 2 in the quality of their publications.

All in all, the results suggest that the program has little impact at the Candidates level, both in terms of aggregate scientific production and in terms of affecting the gender gap associated with such output. However, we do find more consistent evidence of impacts of the program at Level 1, in particular in terms of narrowing the gender gap. This suggests that monetary incentives are important if we want to alter the gender distribution of scientific production. Finally, results at Level 2 are more ambiguous, possibly due to robustness issues that result from the low sample size.

Finally, we performed some tests to validate the identification strategy. First of all, we carried out balancing tests on the means of the control variables chosen for the estimation of the propensity scores between treated and controls. Balance of the sample is a key assumption of the PSM strategy in terms of guaranteeing comparability between the treatment and control groups (in the common support). The results of these tests were satisfactory and are reported in Tables A.3-A.6 in the Appendix.

Also, Figures A.5-A.7 allow us to analyze compliance with the parallel trends assumption imposed by the DiD model by illustrating the separate evolution of the dependent variables in the treatment and control groups before 2011. As the charts show, some variables are not showing a parallel pre-PRONII evolution. This might question the fulfillment of the time-invariant evolution of unobserved heterogeneity assumption on which the DiD strategy relies. However, the matching strategy proposed here aims to address any issues that might result from this by controlling for observed characteristics and matching each treated individual to a comparable complement in the control group.

4. Conclusions

This paper investigates three important issues related to the existence of gender gaps in researchers' productivity in Paraguay. First, we analyze the existence of a gender gap in researchers' productivity in the pre-program (pre-PRONII) period. Second, we investigate if the program implicitly discriminated, at the selection stage, against female researchers. Finally, we evaluate the differential impacts of the program on researchers' productivity across genders.

The findings show a pre-existent gender gap among PRONII researchers. Other things equal, female researchers have a smaller number of written research outputs and papers published in scientific journals. This implies that there is room for policy action that might help narrow this gap in academic achievement between male and female researchers with similar merits, thus also calling for the analysis of the impact of existing policy actions.

We find no evidence of discrimination against female researchers at the selection stage for participation in the program. This confirms the notion that the program is gender-neutral in that it guarantees equal evaluation of female and male applicants. This result is not obvious since there is abundant evidence of gender discrimination against women when being considered for academic positions. However, the program does exert a certain degree of non-gender based discrimination, according to age and field of science which is not intended by the selection criteria established for the program.

Finally, the results show that the impact of the program is not homogeneous across genders or levels. In particular, we found that PRONII contributed to closing the gender gap, by improving female researchers' production relative to their male colleagues, in terms of technical and written research production and quality of publications in Level 1 of PRONII, and in terms of direction of theses and quality of publications for Level 2 researchers. However, we do find that the program didn't have any impacts or even contributed to widening the gap at other stages.

In sum, results found at this stage in terms of the impact of PRONII on the gender gap are mixed, in that certain levels of the categorization or/and certain scientific production variables appear to be associated with a narrowing of the gap that results from program implementation, while we find that the gap increases or remains the same for other levels/outputs. This ambiguity in results might be caused by idiosyncratic elements that lead to specific patterns in each case, which are not considered in PRONII's design, thus not being affected by program implementation. For example, previous evidence about the impact of policies specially aimed at favoring women's STI careers shows that these initiatives were not effective in terms of dismantling institutional and cultural factors that lead to gender discrimination in STI activities (Muller et al., 2011). It is likely that the same happened with PRONII (even more being a relatively new program and one that has no gender specific objectives), so that it seems reasonable that no impacts were found in some cases (such cases might represent those areas of performance or stages in academic career in which male dominant behaviors are more rooted).

It is worth noting that, given that PRONII is a recently implemented program, the impacts we identified are those associated with the very short run. Long run impacts require further program maturity in order to be empirically identifiable. Additionally, we ought to bear in mind that there are issues with the sample size that hamper impact identification, especially at the higher levels of the program.[20] To that extent, a future implementation of this exercise spanning a more extensive time period might lead to more robust results, while also accounting for longer term impacts of the policy.

However, some policy implications might be derived from the results obtained so far. First of all, even though the program appears to have some positive effects on gender gaps, it is also

20 More precisely, we have small power to identify relatively small impacts.

evident that additional efforts ought to be made in order to properly tackle the issue. A good starting point might be to incorporate gender-specific considerations into PRONII's objectives. In addition, we see that at Level 2 results are not as conclusive in terms of PRONII's impact on the gender gap. It may be the case that, at more advanced stages in the academic career, women face greater barriers due to male-dominant practices,[21] so that additional efforts ought to be made to balance the scale at this stage. Such efforts might be made inside the orbit of PRONII by adapting its evaluation criteria at Levels 2 and 3 in order to take these factors into account, thus favoring female entrance into this stage of the system. Finally, it is probably the case that additional gender-specific measures (outside the PRONII orbit) aiming at improving female scientific productivity should be designed in order to properly address the issue of gender inequality in STI activities.

The research findings discussed previously have also shed light into the magnitude of gender gaps in researchers' productivity in Paraguay. And should help initiate policies to close them. In addition, the evidence of a gender-differentiated impact of PRONII on scientific productivity shows that gender-neutral programs can have non-neutral impacts. This evidence calls for additional research to better understand the mechanisms through which these supposedly gender neutral incentives have non-neutral impacts. This is key to implement ex-post (on results) neutral incentives. Also a cost-benefit analysis in which the macroeconomic costs of the gender gap are quantified would provide very valuable information when assessing the optimal amount of resources to be invested into these types of programs.

21 This is asserted by the body of literature on the existence of a *pipeline* that determines that female academic development becomes harder the more advanced the career stage (see Castillo et al., 2014; Blickenstaff, 2005).

References

Angrist, J. D. and Krueger, A. B. (1999). Empirical Strategies in Labor Economics. *Handbook of Labor Economics*, 3: 1277-1366

Abadie, A. and Imbens, G. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica* 74 (1): 235–267.

Aboal, D. and Tacsir, E. (2016). The Impact of Ex-ante Subsidies to Researchers on Researcher's Productivity: Evidence from a Developing Country. UNU-Merit Working Paper #2016-019.

Blickenstaff, J. (2005). Women and Science Careers: Leaky Pipeline or Gender Filter?. *Gender and Education*, 17(4): 369-386.

Blau, F.D., and Kahn, L.M. (2007). The Gender Pay Gap: Have Women Gone as Far as They Can? *Academy of Management Perspectives* 21(1): 7.

Blau, F.D., and Kahn, L.M. (2013). Female Labor Supply: Why Is the US Falling Behind? NBER Working Paper 18702.

Castillo, R., Grazzi, M., and Tacsir, E. (2014). Women in Science and Technology: What Does the Literature Say? Inter-American Development Bank Technical Note No. IDB-TN-637.

CONACYT (2012). "Estadísticas e Indicadores de Ciencia y Tecnología de Paraguay – 2012". Available online at www.conacyt.gov.py

Cuberes, D., and Teignier, M. (2016). Aggregate Effects of Gender Gaps in the Labor Market: A Quantitative Estimate. Journal of Human Capital 10(1): 1-32.

Daly, M. (2005). Gender Mainstreaming in Theory and Practice. *Social Politics: International Studies in Gender, State & Society*, 12(3), 433-450.

Fox, F. M. (2005). Gender, Family Characteristics, and Publication Productivity among Scientists. *Social Studies of Science*, 35(1): 131−50.

Guerrero-Bote, VP, and Moya-Anegón, F. (2012). A Further Step Forward in Measuring Journals' Scientific Prestige: The SJR2 Indicator. *Journal of Informetrics* 6(4): 674–688.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *The American Economic Review*, 104(4), 1091-1119.

Gonzalez-Brambila, C. and Veloso F.M (2007). The Determinants of Research Output and Impact: A study of Mexican Researchers. *Research Policy*, 36(7): 1035-1051.

Halbert, D. (2006). Feminist Interpretations of Intellectual Property. *American University Journal of Gender, Social Policy & the Law*, 14(3): 431-460.

Hausman, J., Hall, B. and Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, 52(4): 909-938.

Klasen, S., and Lamanna, F. (2009). The Impact of Gender Inequality in Education and Employment on Economic Growth: New Evidence for a Panel of Countries. *Feminist Economics*, 15(3): 91-132.

Long, J. S., and Freese, J. (2006). Regression Models for Categorical Dependent Variables using Stata. Stata Press.

Morales, R., and Sifontes, D. (2014). Desigualdad de Género en Ciencia y Tecnología: Un Estudio para América Latina. *Observatorio Laboral Revista Venezolana*, 7(13): 95-110.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science Faculty's Subtle Gender Biases Favor Male Students. *Proceedings of the National Academy of Sciences*, 109(41): 16474-16479.

Müller, J, Castano, C., Gonzalez, A. and R. Palmén. (2011). Policy towards Gender Equality in Science and Research. *Brussels Economic Review*, 54 (2/3): 295−316.

Olivetti, C., and Petrongolo, B. (2008). Unequal Pay or Unequal Employment? A Cross-country Analysis of Gender Gaps. *Journal of Labor Economics* 26(4): 621-654.

Olivetti, C., and Petrongolo, B. (2014). Gender Gaps across Countries and Skills: Demand, Supply and the Industry Structure. *Review of Economic Dynamics* 17(4): 842-859.

Petersen, T., I. Saporta, and Seidel, M-D. L. (2000). Offering a Job: Meritocracy and Social Networks. *American Journal of Sociology*, 106: 763−816.

Rivera, L., J. Mairesse, and Cowan, R. (2016). An Econometric Investigation of the Productivity Gender Gap in Mexican Research, and a Simulation Study of the Effects on Scientific Performance of Policy Scenarios to Promote Gender Equality. Unpublished manuscript.

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

Steinpreis, R., Anders, K.A., and Ritzke, D. (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: a National Empirical Study. S*ex Roles*, 41 (7/8): 509−28.

Stotsky, J. G. (2006a). Gender Budgeting. IMF Working Papers 06/232.

Stotsky, J. G. (2006b). Gender and its relevance to macroeconomic policy: A survey. IMF Working Papers 06/233.

Trix, F. and Psenka, C. (2003). Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse and Society*, 14(2): 191−220.

United Nations (2002). Gender Mainstreaming. An Overview. United Nations.

United Nations Conference on Trade (UNCTAD) (2011). Applying a Gender Lens to Science, Technology, and Innovation. Geneva: UNCTAD.

UNESCO. (2007). Science, Technology and Gender: An International Report. Science and Technology for Development Series. United Nations Educational, Scientific and Cultural Organization.

Appendix

Table A.1. Definition of indicators of researchers' performance used in the empirical exercises

Variable	Definition
1. Researchers' performance indicators	
Master's	Indicator variable, =1 if the researcher's maximum education level is a Master's degree
PhD	Indicator variable, =1 if the researcher's maximum education level is a Doctorate degree
Theses directed concluded	Number of concluded direction of undergraduate and graduate theses per year
Theses directed ongoing	Number of ongoing direction of undergraduate and graduate theses per year
Technical production	Number of yearly technical outputs (this includes technical work, technological products and new processes or techniques)
Written research production	Number of yearly written research publications (this includes papers in both scientific and non-scientific publications, works published in events, publication of books and book chapters, and working papers)
Papers in scientific journals	Number of yearly papers published or accepted for publication in scientific journals
Papers in Scopus	Number of yearly papers published in Scopus journals
Mean SJR	Mean SJR rank of the journals in which the researcher published that year
2. Area of science	
Field 1	Indicator variable, =1 if the researcher's specialization is in Agricultural and Natural Sciences
Field 2	Indicator variable, =1 if the researcher's specialization is in Engineering and Technology
Field 3	Indicator variable, =1 if the researcher's specialization is in Health Sciences
Field 4	Indicator variable, =1 if the researcher's specialization is in Social Sciences and Humanities

Note: when estimating the Probit models for the estimation of propensity scores we use the pre-treatment values of the performance indicators. In those cases we are using educational level attained by 2011 in the case of *Master's* and *PhD*, and the mean values for 2010 and 2011 for the remaining variables. On the other hand, the variables used for DiD impact evaluation are defined as the change in variables before and after PRONII. As a result, we use the change in Master's and PhD attainment between 2013 and 2011, and the change in mean production in 2012 and 2013 versus mean production in 2010 and 2011 for the remaining variables.

Figure A.1. Density of accumulated technical production until 2011, by gender

Figure A.2. Density of accumulated number of publications in scientific journals until 2011, by gender

Figure A.3. Density of accumulated theses directed until 2011, by gender

Figure A.4. Density of accumulated number of publications in Scopus journals until 2011, by gender

Table A.2. Impact of PRONII using 1 neighbor matching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Master's	PhD	Theses directed (concluded)	Theses directed (in progress)	Technical production	Written research production	Papers in scientific journals	Scopus papers	Quality of papers (mean SJR)
Candidates									
ATT_All	0.066	-0.022	0.243	0.147	-0.408	0.29	0.460**	0.074**	0.001
Obs_All	174	174	174	174	174	174	174	174	174
ATT_Women	0.093	-0.019	1.296**	0.111	-0.769*	0.102	0.102	0.000	-0.084
Obs_Women	77	77	77	77	77	77	77	77	77
ATT_Men	-0.192	0.038	0.019	-0.096	0.154	1.423	0.481	0.115*	0.080
Obs_Men	82	82	82	82	82	82	82	82	82
Level I									
ATT_All	-0.175**	0.032	0.810	0.762	0.556	1.167	0.452	0.183*	0.168*
Obs_All	162	162	162	162	162	162	162	162	162
ATT_Women	-0.086	0.057	0.271	0.657	0.771	0.757	0.629*	0.286*	0.179
Obs_Women	101	101	101	101	101	101	101	101	101
ATT_Men	0.000	0.000	-0.382	0.176	0.176	0.618	0.000	-0.059	-0.341*
Obs_Men	50	50	50	50	50	50	50	50	50
Level II-2 years									
ATT_All	0.010	0.066	0.741	-0.594	-0.006	-0.090	0.071	0.060	0.102
Obs_All	114	114	114	114	114	114	114	114	114
ATT_Women	0.013	0.130	0.364	-0.608	-0.002	3.572**	2.281***	0.913	0.157
Obs_Women	65	65	65	65	65	65	65	65	65
ATT_Men	0.011	-0.011	-0.233	-2.011**	0.438	-0.801	-0.844	-0.065	0.157
Obs_Men	45	45	45	45	45	45	45	45	45
Level II-3 years									
ATT_All	-0.046	0.102	-0.387	0.312	-0.132	1.515	0.392	0.078	0.451***
Obs_All	93	93	93	93	93	93	93	93	93
ATT_Women	0.000	0.143	-0.041	0.953**	0.152	-0.450	0.292	0.000	0.157
Obs_Women	58	58	58	58	58	58	58	58	58
ATT_Men	0.014	-0.014	-0.127	-0.604	0.627	0.591	-0.625	-0.190	0.005
Obs_Men	39	39	39	39	39	39	39	39	39

Source: own elaboration based on CVPY.

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Dependant variable in columns 1 and 2 is the change in the accomplishment of a Masters or PhD degree from 2011 to 2013. For columns 3-9, the dependant variable is the difference between the mean production in 2010-11 and the mean production in 2012-13. The only exception is in the case of the evaluation of PRONII in 3 years (bottom panel) where we use change in Masters and PhD from 2011 to 2014, and changes in mean production from 2010-11 to 2012-14. Control variables used in each panel are the ones that result from columns 5-8 in Table 5. ATT is the average treatment effect on treated. Obs is the number of observations.

Table A.3. Balancing tests for matching variables in 5 neighbors-PSM. Candidates

Variable	Unmatched/ Matched	Mean Treated	Control	% bias	% reduct bias	t-test t	p>t	V(T)/V(C)
	U	39.667	39.302	3.5		0.25	0.803	0.58*
Age	M	38.574	40.526	-18.8	-435.4	-1.42	0.159	1.69*
	U	1652.2	1679.9	-3.2		-0.22	0.823	0.53*
Age^2	M	1568.1	1689.7	-13.8	-338.8	-1.07	0.286	1.6
Written research	U	2.4949	0.90094	74.6		5.32	0	0.74
production	M	1.75	1.2517	23.3	68.7	1.8	0.074	0.8
Papers in scientific	U	0.82323	0.08962	132.9		9.65	0	10.85*
journals	M	0.41176	0.37328	7	94.8	0.55	0.586	1.02
	U	0.07071	0.20755	-40.1		-2.85	0.005	.
Field 2	M	0.10294	0.17598	-21.4	46.6	-1.23	0.222	.
	U	0.23232	0.46226	-49.5		-3.53	0.001	.
Field 4	M	0.26471	0.34118	-16.5	66.7	-0.97	0.336	.

Source: own elaboration based on CVPY.

Table A.4. Balancing tests for matching variables in 5 neighbors-PSM. Level 1 researchers

Variable	Unmatched/ Matched	Mean Treated	Control	% bias	% reduct bias	t-test t	p>t	V(T)/V(C)
	U	2.2701	0.73737	55.6		3.88	0	6.82*
Age	M	1.5317	1.5429	-0.4	99.3	-0.03	0.979	0.55*
	U	4.6897	2.4949	60.3		4.2	0	5.80*
Master’s	M	3.2222	2.8794	9.4	84.4	0.88	0.382	2.09*
Theses directed	U	0.44828	0.11111	51.8		3.62	0	10.42*
concluded	M	0.22222	0.14921	11.2	78.3	1.06	0.291	1.6
Written	U	0.25287	0.40404	-32.4		-2.2	0.029	.
res. production	M	0.22222	0.16825	11.6	64.3	0.76	0.449	.
	U	46.322	39.667	71.9		4.91	0	1.15
Papers in Scopus	M	45.476	44.273	13	81.9	0.77	0.441	1.19
	U	0.32184	0.43434	-23.2		-1.58	0.116	.
Field 3	M	0.39683	0.33968	11.8	49.2	0.66	0.51	.

Source: own elaboration based on CVPY.

Table A.5. Balancing tests for matching variables in 5 neighbors-PSM. Level 2 researchers-2 year evaluation

		Mean				t-test		
Variable	Unmatched/ Matched	Treated	Control	% bias	% reduct bias	t	p>t	V(T)/V(C)
	U	0.11538	0.32955	-52.8		-2.16	0.033	.
Master's	M	0.11538	0.23077	-28.4	46.1	-1.09	0.281	.
	U	0.76923	0.30682	103.4		4.54	0	.
PhD	M	0.76923	0.73077	8.6	91.7	0.31	0.755	.
	U	0.48217	0.21919	56.1		2.7	0.008	1.63
Mean SJR	M	0.48217	0.51485	-7	87.6	-0.21	0.833	0.78
	U	0.42308	0.25	36.8		1.72	0.089	.
Field 3	M	0.42308	0.26154	34.3	6.7	1.22	0.228	.

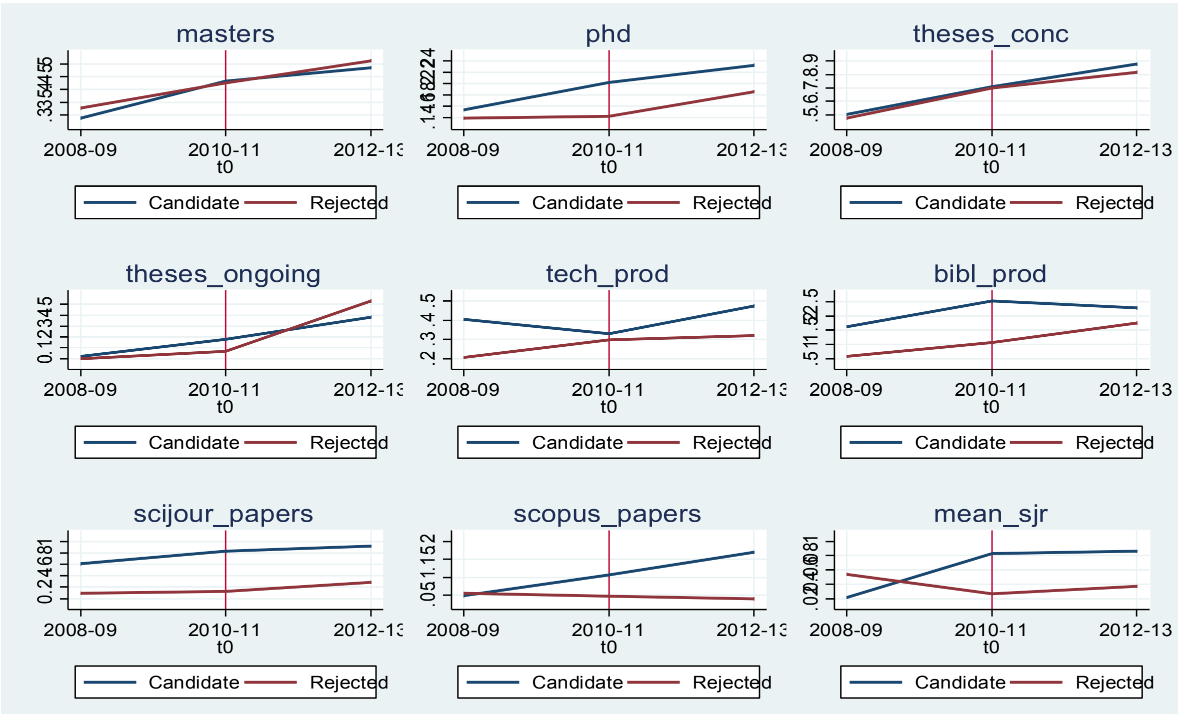
Source: own elaboration based on CVPY.

Table A.6. Balancing tests for control variables in 5 neighbors-PSM. Level 2 researchers-3 year evaluation

		Mean				t-test		
Variable	Unmatched/ Matched	Treated	Control	% bias	% reduct bias	t	p>t	V(T)/V(C)
	U	0.11538	0.33333	-53.5		-2.17	0.033	.
Master's	M	0.16667	0.33333	-40.9	23.5	-1.14	0.261	.
	U	0.76923	0.29333	107.1		4.63	0	.
PhD	M	0.66667	0.55556	25	76.7	0.67	0.508	.
	U	0.48217	0.21017	58.4		2.74	0.007	1.68
Mean SJR	M	0.38165	0.39342	-2.5	95.7	-0.06	0.952	0.77
	U	0.42308	0.25333	36		1.64	0.105	.
Field 3	M	0.22222	0.17778	9.4	73.8	0.32	0.748	.

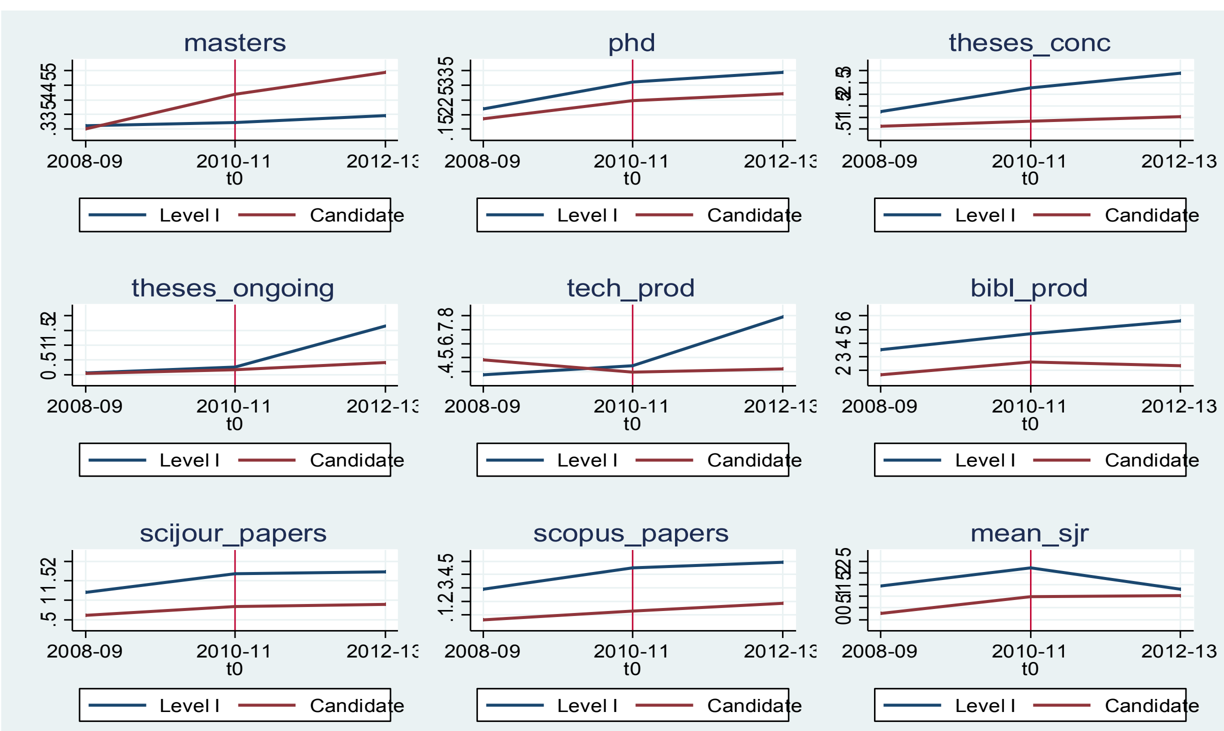
Source: own elaboration based on CVPY.

Figure A.5. Parallel trends analysis: Candidates and Rejected



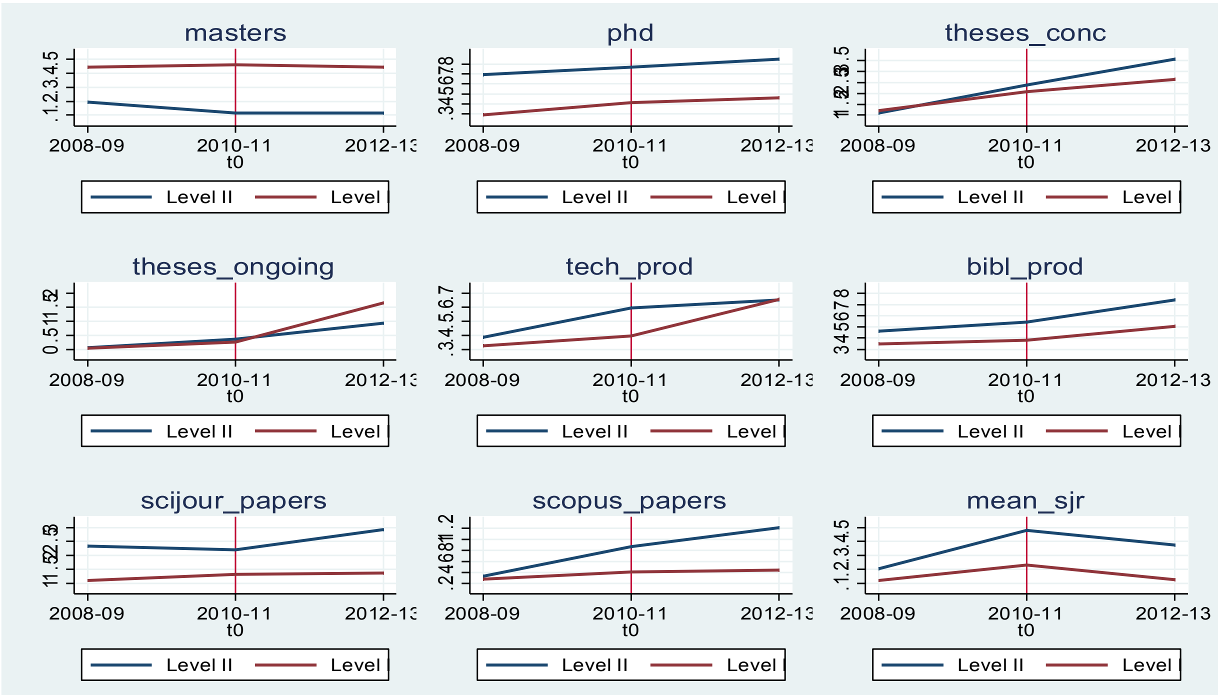
Source: own elaboration based on CVPY

Figure A.6. Parallel trends analysis: Level 1 researchers and Candidates



Source: own elaboration based on CVPY

Figure A.7. Parallel trends analysis: Level 1 and Level 2 researchers



Source: own elaboration based on CVPY