TESOURO NACIONAL

REVISTA

CADERNOS DE FINANÇAS PÚBLICAS

2022

EDIÇÃO ESPECIAL

An economic evaluation of ProUni contrasting the salary mass of graduates with the program's tax expenditure

Vinícius Augusto Lima de Almeida
Economista da Confederação Nacional de Municípios
Francis Carlo Petterini Lourenço
Professor da UFSC

Abstract

ProUni is a federal public policy of tax exemptions for private Higher Education Institutions that grant scholarships to low-income students. The program has cost more than R$13 billion, and has provided degrees to roughly 500 thousand students. In an unprecedented way, this research tracked thousands of alumni using RAIS (Annual Social Information Survey), finding that a graduate from ProUni receives, on average, 30% more than a non-graduate equivalent. Converting this impact into salary mass, in order to contrast it with the tax expense in a cash flow, we conclude that, in a conservative scenario, the policy's Net Present Value would have reached over R$ 38 billion. Based on analyzes disaggregated by courses, the recommendation is that scholarships be focused on graduations that improve the collective benefit.

Keywords: ProUni, impact evaluation, economic return, tax exemption.

Classificação JEL: I28, J21, J24.

SUMÁRIO

1. Introduction

In recent decades there has been a significant increase in enrollments in the Brazilian higher education system, although with different configurations among the varied types of IES (Higher Education Institutions) (Durham 2005; Corbucci 2007, 2014; Tachibana et al. 2015). In public IES, there was a sudden expansion between the late 2000s and the middle of the 2010s, following the REUNI program (Support Program for Restructuring and Expansion Plans of Federal Universities), and the remodeling of the professional and technological education network (Faveri et al. 2018; Barbosa et al. 2019). In private IES, there has been a regular growth since the end of the 1990s due to several related facts. For example: many organizations realized the profitability of the sector and began to create new universities and courses throughout the country; distance learning (ODL) was popularized; student credit was largely offered with the student financing fund (Fies); and the number of scholarships increased with the University for All program (PROUNI) (TCU, 2009; Corbucci 2014; CGU, 2015; Corbucci et al. 2016).

Briefly, ProUni is a system of granting tax exemptions to private IES that offer scholarships to low-income students. In overall numbers, since 2005, when it began, we estimate that the policy has cost more than R$13 billion, in current value, and that it has promoted the graduation of about 500 thousand students.

Despite the impressive magnitude, there have been few attempts to assess ProUni's impact. From the existing ones, Lépene (2018) and Becker and Mendonça (2019) stand out for comparing the proficiency of beneficiaries with other students similar to the target audience, using microdata from the National Student Performance Exam (Enade). Both surveys indicate that the program would have a positive impact on grades of scholarship students, suggesting that they tend to be able to devote more time to their education when compared to control counterparts. Apparently, because they don't have to work to pay for college. Soon, this could improve national human capital, generating a future collective benefit that would justify tax expenditures.

Still, these studies fail to contrast the impact with the tax exemption resulting from ProUni, offering closure to the economic analysis. Perhaps because it is difficult to convert the results of Enade into monetary units in order to compare them with the tax expenditure in a cash flow, to then calculate, for example, a Net Present Value (NPV) of the investment that society makes when giving up revenue.

Additionally, in a counterpoint to proficiency, the salaries would be an impact measure more adherent to the program objective, which is: to encourage access to higher education and graduation of low-income people, in order to reduce social inequalities (Corbucci 2007, 2014). Because, from the perspective of economic theory and empirical evidence, it is the impact on wages that is best related to increase in productivity, and this, ultimately, is what externalizes the increase in human capital to the collective, culminating in the reduction of social inequalities (Mincer 1996; Barbosa-Filho and Pessoa 2010; Katovich and Maia 2018). Moreover, when assessing the impact through wages, there is no need to convert units of measurement to compare with tax expenditure, which favors an economic analysis.

With these points in mind, the general objective of this monograph is to advance the work of Lépene (2018) and Becker and Mendonça (2019), promoting an economic evaluation of ProUni considering its impact on the wages of graduates. Specifically, the idea is to infer quantities of monetary units that would not exist in the national economy without the scholarship generating mechanism, in order to apply financial mathematics tools to directly compare the estimated impact with the observed tax exemption.

Furthermore, one of the criticisms of ProUni is that scholarships are concentrated in few courses – in fact, almost half of them are for Administration, Law, Pedagogy and Accounting Sciences. When estimating the impact by course, a secondary objective of the monograph is to therefore promote suggestions for improvement in the design of the policy, indicating whether some graduations could be prioritized to the benefit of society.

With these objectives in mind, three microdata bases were crossed to operationalize the analysis: the list of scholarship students between 2005 and 2011, obtained in the Prouni system of the Ministry of Education (SISPROUNI, MEC); the records of the Census of Higher Education (CES), to identify characteristics of students, courses, universities, etc., and set up comparison groups; and the Annual Social Information Report (RAIS, identified), between 2010 and 2018, to track former alumni (and other students of the target audience) formally employed, and find out their salaries. In the end, this procedure resulted in the observation of several types of information from 346,810 former ProUni alumni.

By establishing graduates and those who left the program as treatment and control groups, respectively, econometric techniques of Labor Economics literatures and Public Policy Evaluation were applied to data analysis, and the general result is that: in the first years after graduation, on average, a graduate earns 30% more than their non-graduate counterfactual would earn. This number is higher in courses such as Pharmacy or Medicine; on the other hand, in many other courses the effect is statistically zero.

Converting these impacts into mass income, we contrast it with the tax waiver accounting records, and then, in a conservative scenario, we estimate that: the internal rate of return (IRR) of the program would reach 16% per year; the social NPV would already be greater than R$ 38 billion; and each generation of ProUni graduates would equate the tax expenditure with their cohort in about 10 years. Although courses such as Administration and Accounting Sciences have a smaller individual NPV than Pharmacy and Medicine, for example, we notice that the social NPV of the former is substantially higher due to the number of beneficiaries being very large in these courses.

We conclude that it is possible to improve the focus of scholarships in order to enhance the collective benefit of the program. In particular, and in a different way from what is currently done, we propose the creation of a rule that conditions the amount of tax exemption of IES to the offer of scholarships in courses with evidence of impact on the employability of graduates. In particular, this rule could be modified annually, on the basis of RAIS data of immediately preceding years.

In addition to this introduction, this monograph is structured as follows: in Section 2 the ProUni program is explained; in Section 3 the economic evaluation strategy is discussed; in Section 4 the construction of the database and descriptive statistics are presented; in Section 5 the econometric results are presented; in Section 6 the financial mathematics of the program is discussed; and, finally, Section 7 details the conclusion.

2. ProUni

ProUni was established in 2005 by Law No. 11,096 (amending Law No. 10,891, of July 9, 2004), granting discounts on Corporate Income Tax (IRPJ), the Social Contribution on Net Profit (CSLL), the Social Contribution for Financing Social Security (Cofins) and the Social Integration Program (PIS), for private IES, with or without profit, that offer scholarships (full or partial - of 25% or 50%) in higher education courses (whether in undergraduate or graduate levels, face-to-face or ODL). Despite the legal nuances, the basic rule is to give full exemption when the offer reaches the minimum proportion of one full scholarship (per course) for every 11 paying students (in each IES course).

Table 1: Number of IES participating and eligible for ProUni (2005-2017)

IES	Participant [a]	Eligible [b]	% a/b
2005	928	1.777	52,2
2006	1.096	2.018	54,3
2007	1.203	2.095	57,4
2008	1.243	2.201	56,5
2009	1.287	2.270	56,7
2010	1.290	2.284	56,5
2011	1.283	2.352	54,5
2012	1.309	2.383	54,9
2013	1.130	2.398	47,1
2014	1.157	2.358	49,1
2015	1.141	2.373	48,1
2016	1.106	2.494	44,3
2017	1.226	2.695	45,5

Source: MEC (Ministry of Education). Elaborated by the authors.

Table 1 shows the number of institutions participating and eligible for ProUni in the period between 2005 and 2017, illustrating that about half of them show interest in the program. Probably, this fraction is not higher because most IES had already obtained full or partial exemption from taxes before 2005, through other instruments (TCU 2009; Haas and Pardo, 2017).

Candidates for a scholarship must: take the National High School Exam (Enem), not get a zero the writing section, and register in SISPROUNI choosing institutions, courses and schedules of preference. The system then makes a pre-selection based on this information and the available scholarships. If the student is pre-selected, they must prove that: they have not yet completed a higher education; they have a monthly family income of up to one and a half minimum wages (MWs) per capita for the full scholarship, or up to three MWs for the partial scholarship; and that they have attended public school or was recipient of scholarships in private institutions. There is also a percentage of scholarships destined towards students with disabilities, and teachers of the public basic education network; as well as a cost aid for full-time students, in certain circumstances. Once in the program, in order not to be cut off, the student must meet frequency and minimum grades in the subjects.

Table 2: Tax waiver and ProUni scholarships per year (2005-2017)

Year	Waiver (R$ million*) [a]	Scholarships (thousand) [b]	1.000×a/b
2005	598,3	95,6	6.259,3
2006	604,7	109,0	5.548,3
2007	578,1	105,5	5.478,1
2008	688,1	124,6	5.524,3
2009	1.002,5	161,3	6.217,2
2010	1.195,7	152,6	7.834,8
2011	922,3	170,7	5.404,7
2012	1.244,0	176,7	7.041,8
2013	1.202,7	177,2	6.787,0
2014	909,0	223,4	4.069,1
2015	1.379,1	252,7	5.464,5
2016	1.643,0	238,9	6.875,6
2017	1.603,1	240,6	6.661,6

*2020 values adjusted by IPCA (HICP). Source: RFB and MEC. Elaborated by the authors.

Table 2 presents an estimate of the tax expenditure with ProUni between the years 2005 and 2017, with values of 2020, using data from the Brazilian Federal Revenue Service (RFB). It also shows the number of scholarships granted per year – regularly, between 65% and 75% of them are full ride. The sums of the values of columns [a] and [b] reach the amounts of R$ 13,5 billion and 2.2 million scholarships.

Considering that a scholarship is taken by the same person for subsequent periods, and assuming that it takes an average of four years for a diploma (at least one non-sequential), 2,2 million are divided by four to conclude that the program had more than 500 thousand beneficiaries in the time span depicted. Given that, considering the figures presented by CGU (Comptroller General of Brazil, 2015), there is up to 30% of evasion or disconnection, and that this increases the turnover of the occupants of a scholarship, then we can deduce that the number of beneficiaries may have been greater than half a million. In this sense, even if no official statistics have been found on the number of graduates through ProUni, by this logic it is reasonable to state that about 500 thousand students have graduated through this program to date.

Looking at the ratio of the values of columns [a] and [b], we conclude that there is an approximate annual tax expenditure of R$ 6 thousand per scholarship, in 2020 values. Additionally, CGU (2015) makes a similar analysis of this value using a more accurate methodology, which controls a series of inconsistencies in the primary data, but still the estimate of annual expenditure per scholarship is close to this one. Moreover, considering the hypothesis that it takes an average of four years for a diploma, we assume that an average degree would cost the treasury something like R$ 24 thousand, or R$ 500 per month, in 2020 values (perhaps a little more due to evasion).

Another point of notice is the debate between supporters and critics of the program (Saraiva et al., 2011). Among the first, the main argument is that it is a policy well focused in terms of the target audience, because it promotes the access of low-income students to higher education, according to the performance (meritocracy) in Enem, which is a standardized and nationwide exam.

On the other hand, many critics argue that ProUni allows access to a higher education in a restricted subset of courses. Indeed, as shown in Table 3, Administration and Law courses alone accounted for 1/3 of all scholarships offered between 2005 and 2017; in addition to Pedagogy and Accounting Sciences, these four comprised almost half of the scholarships.

Table 3: Percentage of scholarships offered by course among the 30 degrees with the most beneficiaries in the period of 2005-2017

#	Course	%	% cumulative	#	Course	%	% cumulative
1	Administration	21,49	21,49	16	Civil Engineering	1,50	76,37
2	Law	12,21	33,70	17	Nutrition	1,41	77,78
3	Pedagogy	8,81	42,51	18	Eng. Production	1,35	79,13
4	Accounting Sciences	5,51	48,02	19	Architecture	1,22	80,35
5	Nursing	4,23	52,25	20	Medicine	1,06	81,41
6	Physical Education	3,26	55,51	21	Electrical Eng.	0,97	82,38
7	Psychology	3,22	58,73	22	History	0,93	83,31
8	Information Systems	2,52	61,25	23	Logistics	0,87	84,18
9	Physiotherapy	2,28	63,53	24	Odontology	0,86	85,04
10	Language and Literature	2,18	65,71	25	Marketing	0,86	85,90
11	Social Work	2,07	67,78	26	Mechanical Eng.	0,85	86,75
12	Pharmacy	2,04	69,82	27	Tourism	0,76	87,51
13	Social Communication	1,95	71,77	28	Economics	0,72	88,23
14	Biology	1,60	73,37	29	Biomedicine	0,72	88,95
15	Computer Science	1,50	74,87	30	Mathematics	0,66	89,61

Source: MEC (Ministry of Education). Elaborated by the authors.

However, in the face of the tax exemption rule, each IES is encouraged to offer scholarships for all the courses to which there is selection, in proportion to the enrolments. Therefore, such a concentration may simply be a reflection of the general supply concentrated in a few courses; it may not characterize any measure of bad faith or something of the sort; and, perhaps, it may simply be a failure in the policy design. The list in Table 3 follows with the 30 courses that accounted for almost 90% of the offer between 2005 and 2017, without differentiating between face-to-face and ODL courses, and the complete repertoire has hundreds of courses among undergraduate and graduate levels.

Eventually, some critics of ProUni also declare that the program benefits institutions of dubious quality. As for this argument, one can explore the general course Index (IGC) calculated by MEC in order to synthesize this point – its values range from one (worst) to five (best). In this sense, the average IGC in the participating institutions is 3,19, and in non-participating private institutions it is 3,11; the difference is not statistically different from zero. That is, by this indicator, there is no evidence that, on average, IES that participate in ProUni are better or worse because of it than non-participant ones.

In terms of the logical framework of public policy, as proposed by Gertler et al. (2018), we emphasize that the scholarships are the product of ProUni. Given this product, here are the following short-term results expected from the target audience: increase in entry and graduation rates in higher education; acceleration of training time; and dilation of time devoted to studies at the expense of work, with consequent improvement in proficiency.

In fact, Lépene (2018) and Becker and Mendonça (2019) verify an improvement in proficiency in beneficiaries, especially for those who receive full scholarships; and, Andriola and Barrozo-Filho (2020) point out that training time has been accelerating. Furthermore, although nothing specific has been found about ProUni, Dynarski (2000, 2003), Stinebrickner and Stinebrickner (2003) and Darolia (2014) provide evidence that these short-term results occur in similar programs in the US.

In longer terms, the program would aim to reduce social inequalities by encouraging access to higher education and graduation for people with low family income (Corbucci 2007, 2014). Then, ProUni's impact indicators would involve increasing employability and wages of the beneficiaries, and this, in turn, would reflect on more well-being for the collective.

3. Economic assessment

The economic assessment of a public policy involves estimating its impact on the target audience, and contrasting it with the costs. Following the praxis of the literature, this monograph defines potential results w_{1i} e w_{0i} as factual and counterfactual salaries, respectively, of the graduate i from ProUni. Put another way, w_{1i} is an observed salary and w_{0i} is a salary that would be observed if i had not completed a higher education degree. That way, w_{1i}-w_{0i} represents an impact of the program – i.e., an effect of "being more qualified", on the individual perspective of a beneficiary's salary.

As a result, $B=\Sigma^i (w_{1i}-w_{0i})=n\times(\bar{w}_{1i}-\bar{w}_{0i})$ is a mass salary caused by the program, considering n students and the arithmetic means of w_{1i} e w_{0i} represented by $\bar{w}_{1i}$ e $\bar{w}_{0i}$, respectively. So, $\bar{w}_{1i}-\bar{w}_{0i}$ represents an "average treatment effect".

This B amount lingers for post-graduation periods of a cohort, and represents the impact of the program on the perspective of society, because it is money that will be invested, consumed, taxed, etc. This mass can be contrasted with the costs of the program in order to provide closure to an economic analysis. However, it is impossible to compute B directly because the counterfactual salary is not observed w_{0i} – this is the fundamental problem of causal inference of a policy (Gertler et al., 2018). It is, first and foremost, important to find a strategy to identify $\bar{w}_{1i}-\bar{w}_{0i}$.

3.1. Adaptation of the Mincer equation

To get around the fundamental problem of causal inference, the strategy adopted in this monograph starts with applying the logic of Mincer (1970), establishing: d=0 and d=1 as generic schooling plans; w_d for the salary with the respective plan; and s_d and n_d as periods of study and dedication to work from these plans, so that $s_1>s_0$ e $s_d<n_d$ (which implies that only the period of work after the completion of the schooling plan is considered in the modeling).

Thus, in continuous time, the present value of the plan d in the labor market is $V_d=\int_{sd}^{nd} w_d \exp(-\gamma t)dt$, where $\gamma>0$ is an intertemporal discount rate. Seeing as d=1 involves more schooling than d=0, the difference V_1-V_0 is the NPV of personal investment in d=1 vis-à-vis d=0. Therefore, the equality V_1-V_0=0 implicates two results. First, considering $\gamma=\ln(1+r)$, the r term represents an IRR of personal effort in obtaining more schooling. Second, with some algebra, Mincer (1970) demonstrates that the following equation is true: $w_1 \exp(-\gamma s_1)(1-\exp(-\gamma n_1))=w_0 \exp(-\gamma s_0)(1-\exp(-\gamma n_0))$.

With this last result, if $n_1=n_0$, ou se $n_1 \neq n_0$ but both are large enough (i.e., people work for many years), it occurs that $w_1/w_0 \cong \exp(\gamma(s_1-s_0))$. Without loss in generality, one can write $s_1=1$ e $s_0=0$, meaning, in particular, having or not having graduated, respectively. Conveniently, this implies that: d=1 and d=0 come to mean the same as $s_1=1$ and $s_0=0$; and, the ratio $w_1/w_0 \cong \exp(\gamma)$ being true.

Soon, defining a tautology $w \equiv w_0(w_1/w_0)^d$ for any w salary observed (whether factual or counterfactual), and associating it with the last result, $w \cong w_0 (\exp(\gamma))^d$. is found. This, in turn, implicates that $\ln w = \ln w_0 + \gamma d + u$, with u being an error term added to transform the approximation sign into an equal sign. That way, with this modeling we arrive at a version of the Mincer equation adapted to the context of this monograph.

3.2. Econometric Strategy

Starting from the Mincer equation, $\ln w_0 = cte + \Sigma^L_{l=1} \beta_l x_l$, rewritten, in which "cte" is a constant and each β_l and x_l represent L salary parameters and covariates, respectively – e.g., experience and its square (Taber, 2001; Heckman et al ., 2006; Barbosa-Filho E Pessoa 2010). Thus, given a sample $\{w_i, d_i, x_{1i}, \ldots, x_{Li}\}^N_{i=1}$, in which i indexes individuals, but n<N of them graduated via ProUni, the basic equation of the econometric exercise is:

$$\ln w_i = cte + \gamma d_i + \Sigma^L_{l=1} \beta_l x_{li} + u_i \qquad (1)$$

That way, since d_i appropriately indicates belonging to treatment and control groups, γ can also be estimated and interpreted as "the average treatment effect on the treated" (by degree), or the English acronym ATT, the details of which are didactically presented in Wooldridge (2010, p.605). Notably, in this particular case, γ it should be interpreted in terms of semi-elasticity of salary, because the dependent variable is in logarithm and d_i is a dummy.

The estimation of the parameters of equation (1) should consider the endogeneity of d_i, because the salary is influenced not only by observable characteristics x_{li}, but also by unobservable variables embedded in u_i, and which can be correlated with d_i. Generically, the literature calls this unobserved component "skills".

Wooldridge (2010, P. 61) provides a didactic presentation of this matter, such that one can rewrite the error as $u_i=\delta o_i+v_i$, where δ is a parameter that associates the omitted variable o_i with d_i in (1), and v_i is the error term by not omitting o_i. With some algebra, we can see that the convergence in probability of the Ordinary Least Squares (OLS) estimation of γ, denoted by $\hat{\gamma}$, is such that "plim" $\hat{\gamma}$ = $\gamma + \delta \times$ "Cov" (d_i, o_i)/"Var" (d_i), where "Cov" and "Var" indicate covariance and variance, respectively.

Therefore, if $\delta>0$ (higher salaries being associated with higher skills), and if there are more skilled people among graduates ("Cov" $(d_i, o_i) > 0$), it occurs that $\hat{\gamma}$ overestimates γ. On the other hand, if o_i represents an unobserved experience, δ>0 is also expected (because higher salaries should be associated with more experienced people), but if there are fewer experienced people among graduates, we have "Cov" (d_i, o_i) < 0 and then $\hat{\gamma}$ underestimates γ.

Therefore, determining the signal and magnitude of such bias is often complicated, and to mitigate this problem it is common to apply instrumental variables, usually using Least Squares in two stages (2SLS). In the literature of labor economics there is a fruitful discussion about the potential instruments associated with schooling (in this case, d_i), but not with skill. Whereas in the works of analysis of Higher Education, the common thread seems to be the use of the parents' level of education; unfortunately, this was not possible in the tabulation of the data (discussed in the next section).

On the other hand, as mentioned earlier, MEC provides the IGC to assess the quality of the courses, and it is also possible to identify for how long the IES have existed. So, in a similar way to the works of Dale and Krueger (2002) and Borgen (2014), we suppose that higher quality and longer history when it comes to the IES can motivate students to graduate, either because the course would be more interesting, or because the expectation of employability would be higher in view of the completion of courses better evaluated and in better known institutions. We also consider that such variables have nothing to do with the intrinsic ability of a student.

So, considering that the chance of graduation would be higher in courses of better quality and older history within the IES, and that such variables would not be related to the student's ability, then these would work as instruments. As discussed below, we do not reject the hypothesis that the chances of graduation are greater in better evaluated and/or older institutions, and the Durbin-Wu-Hausman test corroborates with the perspective of this instrumentalization.

Another source of bias in estimating the parameters of an equation such as (1) is that wages are only observed among employed people, and these are potentially more skilled than unemployed people. Therefore, disregarding observations with missing salaries, there is potentially similar bias to an omission of important explanatory variable. In this sense, the literature often applies the so-called "Heckman correction" to mitigate the problem, in which Equation (1) is estimated conditionally to the status of (formally) employed or (formally) unemployed – and may or may not use instrumentalization.

To put it briefly, one way to operationalize this fix, commonly called a Heckit, is as follows: dummy "being (formally) employed", and we estimate the probability of observing someone (formally) employed in the face of covariates; with these estimates, computing the so-called "inverse Mills ratio", or IMR by the acronym in English; finally, we estimate me the regression (2) adding to IMR as covariate [by OLS or 2sls]. Such a procedure (and its nuances) is presented in detail in Mullahy (1998) or Wooldridge (2010, p. 563-564), and has the advantage of mitigating sample selection biases in the estimation of γ.

Lastly, and no less relevant, there is the potential bias of self-selection when it comes to defining the control group, which is widely discussed in the literature of evaluation of public policies (Gertler et al., 2018). In short, the problem is that a considerable part of the most skilled people in the target audience can self-select to participate in the program, so γ it would not reflect the ATT of the program, but the effect of high skill intrinsic to the treatment group. Therefore, the control group should be defined in order to further mitigate this source of bias.

We can then notice that the expected impact of the program occurs through the graduation of the beneficiary, which in turn would improve their condition in the labor market. As the CGU (2015) informs, there is a significant number of students who drop out or are disconnected from the program prior to graduation, and then have no enrollment found in any IES for subsequent years. Therefore, a control group composed of these people would belong to the same target audience, whose only apparent difference from the treatment group would be graduation, and that would mitigate the bias of self-selection.

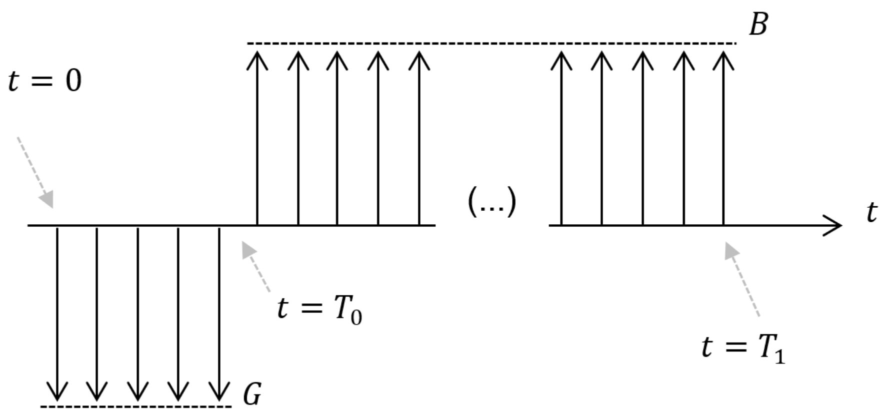
Still, one can imagine that an omitted variable would remain, given by the "resilience" of those who continue to study to the point of obtaining a diploma. In the terms discussed above, if o_i is this variable, "Cov" $(d_i, o_i) > 0$ occurs (because resilience is associated with higher chances of graduation); but if o_i is also not observed by the labor market, it is possible to imagine that $\delta=0$ and so this eventual bias would be negligible. Anyway, to deal with this issue, the idea is to promote a propensity score (PSM) pairing in order to mitigate this potential bias arising from resilience (Gertler et al. 2018).

In view of what is discussed here, the estimation exercises of Equation (1) will be repeated using: the estimators typically used in the literature (OLS, 2sls and Heckit with/without instrumentalization through the IGC and the age of the IES); the Durbin-Wu-Hausman test to evaluate the endogeneity of d_i; non-graduate former student will be used as a control group; and samples with and without PSM. Based on the estimation of γ with these procedures, the subsequent financial mathematics exercise will still consider a scenario in which the ATT may be overestimated.

3.3. Social Net Present Value

Having an estimate of γ, the problem of non-observed counterfactual salary is then circumvented with result $w_{0i} \cong w_{1i}/\exp(\gamma) = w_{1i}/(1+r)$, because $B \cong \Sigma_i (w_{1i}-w_{1i}/(1+r))=(r/(1+r))\Sigma_i w_{1i}$ occurs. That is $\Sigma_i w_{1i}$ is the observed salary mass of graduates from ProUni, whose fraction $r/(1+r)$ is computable.

Figure 1: Social Cash Flow Diagram of the Program



$$VPL^{social} = -\sum_{t=0}^{T_0} \frac{G}{(1+\tau)^t} + \sum_{t=T_0+1}^{T_1} \frac{B}{(1+\tau)^t} \quad (2)$$

To provide the closing balance of social cash flow, we notice that every year there is a tax expense G with each cohort, for a representative period of T_0 years until graduation. If the cohort maintains the benefit for T_1-T_0 years, then this flow would be represented by a diagram like the one in Figure 1. That way, the NPV of the program in the social perspective would be NVP^{social} according to Formula (2), for a reference rate τ; this way, we can estimate the IRR and the payback for the collective.

4. Data

The construction of the database started with the record of students by cohort of admission between 2005 and 2011, at SISPROUNI, containing: the social security number and the id-MEC; the type of scholarship (partial or full); the IES identification, the course, the shift and the modality (face-to-face or ODL); and the gender, date of birth and the markers of physical disabilities. With id-MEC, the cohorts were crossed using the CES (Superior Education Census) of the respective year, where we observed: gender, age and if they were a ProUni student; the status of enrollment (regular, in the completion phase, locked, discontinued, transferred and indications of graduation or death); and characteristics of the course and the IES. Then, a process of data deduplication and elimination of inconsistencies – e.g., when a person is called “female scholarship student” on one basis and “non-scholarship male student” on another. This procedure culminated in the observation of 346.810 beneficiaries.

Subsequently, these individuals were tracked throughout the editions of the CES until 2018, in order to classify the graduates and those who did not graduate, nor were deceased and were no longer found in other IES – these being defined as “dropouts”. Among the individuals observed, 84,17% and 15,83% had graduated (treatment group) and dropped out (control group), respectively, as discussed in the last section.

Table 4: Observations in the treatment and control groups by cohort of admission to ProUni and presence in RAIS after the year expected for graduation

Coorte	Treatment (Graduates)		Control (Dropouts)	
	Not found in RAIS	Found in RAIS	Not found in RAIS	Found in RAIS
2005	1.685 15,3%	9.340 84,7%	182 19.8%	739 80.2%
2006	3.911 12,1%	28.444 87,9%	565 21,7%	2.041 78,3%
2007	3.607 14,1%	22.063 85,9%	1.141 31,3%	2.505 68,7%
2008	4.143 14,3%	24.896 85,7%	1.665 29,8%	3.926 70,2%
2009	7.122 13,6%	45.160 86,4%	2.439 21,5%	8.918 78,5%
2010	11.476 16,7%	57.245 83,3%	3.405 23,7%	10.975 76,3%
2011	15.040 20,7%	57.785 79,3%	4.352 26,5%	12.040 73,5%

Total	46.984 16,1%	244.933 83,9%	13.749 25,0%	41.144 75,0%

Source: MEC and RAIS. Elaborated by the authors.

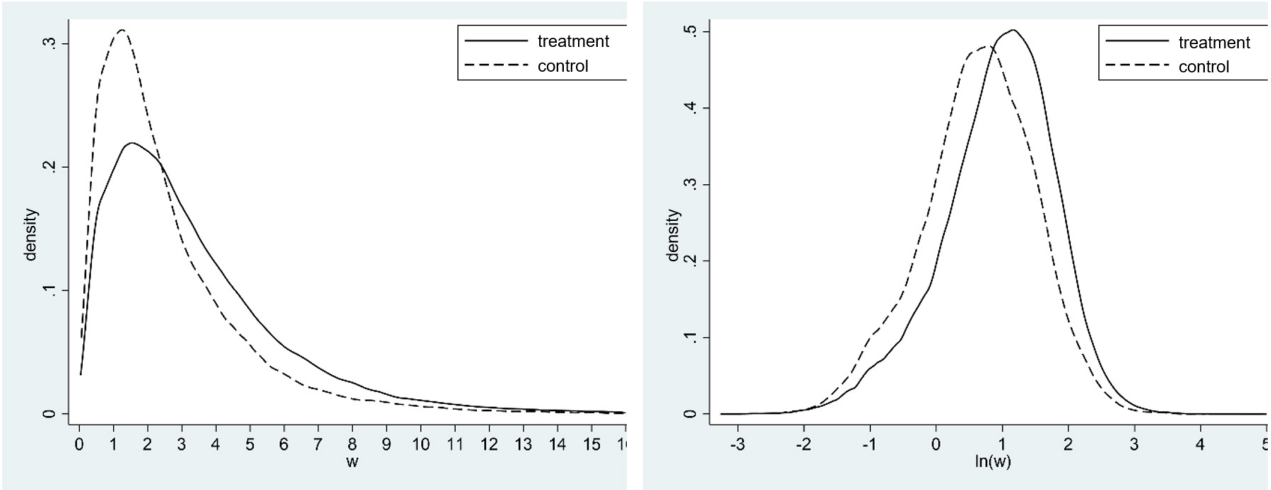
Then, the social security number was crossed using RAIS (identified), where information about salaries in the formal labor market can be found. For this point, we considered the representative time of graduation to be four years, with an extra year for people to fit into a job; therefore, the cohort who joined in 2011 would potentially be in the labor market with a higher education in 2016. Given that the most recent identified RAIS that was available at the time of tabulation was that of 2018, the 2011 cohort was tracked in 2016, 2017 and 2018; the 2010 cohort in 2015, 2016, 2017 and 2018;... ; and the cohort of 2005 in 2010, 2011,..., 2017 and 2018. Table 4 shows the number of observations in the treatment and control groups by cohort of admission to ProUni and presence in RAIS, between the year after the expected graduation and 2018.

In all cohorts, the percentage of those found in RAIS is higher among graduates (treatment) than among those who dropped out (control). In total, 83,9% of the treated presented formal employment records, compared with 75,0% in the case of controls – a difference of 8,9 percentage points (p.p.). Therefore, greater employability is noted among graduates.

From RAIS, we defined: w_{ia} as the average monthly earnings of the individual i in the year a, in minimum wages (MW), according to the microdata dictionary; and w_i as the average of w_{ia} in the years when the individual i is observed. w_i is an indicator of the average monthly salary of the individual i in the formal labor market in the period after graduation (in the case of treatment) or potential graduation (in the case of control). Thus, in the final step of tabulation, a cross-sectional arrangement of the data was made by cohort.

Figure 2 (a) shows a smoothed histogram (i.e., density kernel) of w_i, between treated and controls. The average observed between the treated is 3,50 MW, and between the controls is 2,65 MW; the difference in means, therefore, is 0,85 MW or R$ 888 in 2020 [with a 95% confidence interval between 0,82 and 0,88 MW]. Figure 2 (b) is analogous, considering ln w_i. In this sense, the average observed between the treated is 0,93, and between the controls is 0,63; the difference in averages, therefore, is 0,30 [with a 95% confidence interval between 0,29 and 0,31]. Consequently, we have a preliminary estimate of γ, without considering any covariate in Equation (1), representing an ATT close to 30% in terms of semi-elasticity of salary to treatment.

Figure 2: Smoothed histogram of the monthly salary indicator in MW (w) across treatment and control groups



(a) w (b) $\ln w$

Source: MEC and RAIS. Elaborated by the authors.

Table 5 shows the average salaries among selected courses, separating in columns (a) and (b) the comparison groups. Column (c) shows the difference in wages, (a)-(b), such that the courses in which this difference was not statistically different from zero were not listed – this is the case of pedagogy, for example. We can see that in Administration the averages of the treated and controls are 3,31 and 2,55 MW/month, respectively; and that the difference was 0,76 (or R$ 794 in 2020 values). In Medicine, the averages of the treatments and controls are 9,76 and 7,84 MW/month, respectively; and the difference was 1,92 (or R$ 2.006 in 2020 values).

There are at least two points to be registered in relation to the numbers above. First, this exercise indicates that the program has different impacts among courses. Second, the average salary of graduates in certain courses is higher than the average salary of graduates in other courses – an example being the cases of Medicine (7,84) vis-à-vis Administration (3,31). We imagine that this is the result of unobserved variables (e.g., abilities), in the sense that, on average, a Medical School graduate may have more of these characteristics than a student of Administration. But regardless of whether this is true or not, we can understand it is important when estimating the equation (1) that it be controlled by the course the individual attended.

Table 5: Average salary (w in monthly MWs) by selected courses, across treatment (d=1)) and control (d=0) groups, and preliminary estimate of the value potentially created by the program (B)

Course	average w		Difference	Graduates	Preview of B
	d=1 (a)	d=0 (b)	(c) = (a)-(b)	(d)	(c) × (d)
Administration	3,31	2,55	0,76	60.629	46.078
Law	3,62	3,09	0,53	37.000	19.610
Accounting Sciences	3,58	2,59	0,99	15.919	15.760
Nursing	4,13	2,94	1,19	13.197	15.704
Information Systems	4,49	2,83	1,66	6.739	11.187
Pharmacy	4,93	3,59	1,34	6.445	8.636

Medicine	9,76	7,84	1,92	3.564	6.843
Civil Engineering	4,34	3,09	1,25	4.579	5.724
Computer Science	4,35	3,09	1,26	4.048	5.100
Social Communication	3,11	2,34	0,77	5.623	4.330
Electrical Engineering	5,56	4,04	1,52	2.808	4.268
Production Engineering	4,53	3,46	1,07	3.934	4.209
Mechanical Engineering	5,22	3,54	1,68	2.477	4.161
Psychology	2,75	2,36	0,39	9.899	3.861
Nutrition	2,83	2,01	0,82	4.389	3.599
Economics	4,67	2,90	1,77	1.962	3.473
Architecture and Urbanism	2,78	2,05	0,73	3.756	2.742
Chemistry	3,96	2,47	1,49	1.747	2.603
Foreign Trade	4,17	2,83	1,34	1.243	1.666
Computer Engineering	4,81	3,18	1,63	980	1.597
Marketing	3,10	2,42	0,68	2.330	1.584
Physiotherapy	2,39	2,18	0,21	7.052	1.481
Biology	2,66	2,37	0,29	4.710	1.366
Agronomy	3,88	2,68	1,20	1.127	1.352
Automation Engineering	4,69	3,75	0,94	1.414	1.329
Mathematics	3,60	2,90	0,70	1.885	1.320
Environmental Engineering	3,43	2,73	0,70	1.659	1.161
Biomedicine	2,99	2,48	0,51	2.218	1.131
System Analysis	3,86	3,16	0,70	1.536	1.075
Veterinary Medicine	2,68	2,17	0,51	2.038	1.039

Source: MEC and RAIS. Elaborated by the authors.

Column (d) shows the number of graduates (treatment) observed per course in the database. Thus, remembering that the amount $B = \Sigma_i (w_{1i} - w_{0i}) = n \times (\bar{w}_{1i} - \bar{w}_{0i})$ represents a value potentially created by the program, where w_{1i} and w_{0i} represent factual and counterfactual salaries of graduates, column (e) presents a previous estimate of B through the product between columns (c) and (d). The premise of this exercise is that (c) = (a) - (b) would be an initial indicator for $\bar{w}_{1i} - \bar{w}_{0i}$, which is not necessarily true, because of the various sources of bias described in the previous section, but serves to give a prior notion of the impact of ProUni on the salary of graduates.

In the terms mentioned above, the Administration degree would generate greater social benefit, with an amount close to 46 thousand MW per month, by the parameters of the illustration – something like R$ 577 million per year, in 2020 values. Additionally, the exercise indicates that courses such as Medicine generate relatively larger individual benefits (1,92 MW/month), but from the collective perspective generate relatively smaller benefits (6,8 thousand MW/month), because Medicine has a much smaller number of scholarships than, for example, Administration.

In this line of reasoning, Table 6 repeats the previous exercise by replacing courses with a set of Federal Units – not all, only the ones where the difference (a)-(b) was statistically different from zero. That way, we can see that although the individual benefit of graduation is potentially higher in the Federal District (1,25 MW), the collective result would be substantially higher in São Paulo, because there are many beneficiaries in the State. In short, the figures in Tables 5 and 6 show that from a collective perspective, even if graduation in a specific degree (and/or location) has a small individual impact, from a collective perspective this impact can be greater in the face of a large number of beneficiaries.

As for other covariates x_{li} of the salaries, it should be noted that the average age in the treatment and control groups is 23,5 and 25,5 years old, respectively – this difference is statistically different from zero, and shows that the dropouts (controls) tend to be older. Men represent 58,5% of the observations, and earn on average 20,6% more than women, regardless of whether they belong to the treatment or control group. The number of scholarship students with disabilities is less than 0,2%; about 3/4 of the scholarships are of full value; close to 2/3 of the beneficiaries studied on the night shift; about half of the scholarship students went to universities, more so than in colleges; and 1/10 of the courses are ODL. Although the database offers some other explanatory variables to potentially apply in Equation (2), these were the ones that repeatedly proved statistically significant in the econometric exercises discussed below.

Table 6: Average Wage (w in monthly MWs) per selected FU, between treatment groups (d=1) and control (d=0), and preliminary estimate of the value potentially created by the program (B)

FU	w average		Difference	Graduates	Preview of B
	d=1 (a)	d=0 (b)	(c) = (a)-(b)	(d)	(c) × (d)
São Paulo	3,87	2,79	1,08	96.823	104.307
Minas Gerais	3,39	2,56	0,83	38.781	32.377
Rio Grande do Sul	3,36	2,64	0,72	28.280	20.496
Paraná	3,43	2,63	0,80	22.780	18.137
Rio de Janeiro	3,37	2,55	0,82	16.109	13.143
Distrito Federal	4,30	3,05	1,25	6.016	7.505
Goiás	3,42	2,65	0,77	8.774	6.769
Bahia	2,99	2,59	0,41	13.904	5.632
Amazonas	3,17	2,31	0,86	5.298	4.570
Mato Grosso	3,53	2,75	0,78	5.262	4.112
Santa Catarina	3,26	2,60	0,66	6.024	4.000
Pernambuco	3,20	2,59	0,61	5.539	3.356
Espírito Santo	3,28	2,51	0,77	4.230	3.248
Mato Grosso do Sul	3,55	2,54	1,00	3.146	3.156
Ceará	3,21	2,51	0,70	4.082	2.850
Rondônia	3,29	2,36	0,93	3.038	2.834
Pará	3,09	2,53	0,56	4.316	2.409
Piauí	2,76	2,28	0,48	2.300	1.107

Rio Grande do Norte	2,14	1,92	0,22	4.785	1.040

Source: MEC and RAIS. Elaborated by the authors.

5. Econometric results

Table 7 presents the estimated parameters for Equation (1) and auxiliary regressions. In addition to the treatment/control indicator d (or $\hat{d}$ when d is instrumentalized in a first stage), the covariates x_{li} are: dummies for cohorts, courses and the individual's FU (purposely omitted to spare a tedious discussion, which would have little to contribute to the analysis); age of the person and their square, as proxy of unobserved experiences; dummies (=1) for men, receiving full scholarships, ODL courses, night shift in face-to-face courses, course from a IES classified as a University (vis-à-vis a college or similar institution) and record of physical disability. For the first stage of 2SLS, we add dummies for the IGC grades (of 2, 3, 4 or 5) and the age of the IES; and in the case of Heckit, we add IMR (Results Measuring Instrument).

In column (i) are the OLS results using all available observations, where the ATT estimate is $\hat{\gamma}^{OLS} = 0{,}279$. Considering that since the salary is in logarithm and that d is a dummy, this value represents a semi-elasticity. Therefore, on average (controlled by covariates, ceteris paribus), the interpretation is that a graduate earns 27,9% more than a non-graduate – close to the 30% previously discussed in descriptive statistics.

The sign of the parameters associated with age is consistent with the expected for a Mincer equation: the salary increases with the accumulation of experience up to a certain age, and then decreases. As for the parameters of the other covariates, we notice that, on average (controlled by the covariates, ceteris paribus): men earn 23,3% more; those who attended ODL earn 7,6% less; and those who studied on the night shift and at a university earn 2,2% and 4,6% more, respectively. Finally, in this regression, R^2 was 0,121.

In column (ii) we can see the first stage results of 2SLS, regressing d with the covariates of the OLS plus the dummies for IGC levels and IES age. This way, we can see that men, older people and those who studied in ODL have lower chances of graduation; and those who received a full scholarship have higher chances. In addition to that, higher levels of quality and the Institution's age are associated with higher chances of graduation, which stands in favor of their use as instruments – although, of course, the absence of correlation of these variables with non-observed abilities cannot be evaluated.

From this first stage, we define $\hat{d}$ as the linear projection of d on the instruments and other covariates. Thus, $\hat{d}$ can be interpreted as an estimate of the "linear probability of graduation". In these terms, column (iii) shows the second stage results of the 2SLS, in which d is replaced by $\hat{d}$ on the right side of the equality of Equation (1). Consequently, $\hat{\gamma}^{2SLS} = 0{,}741$ represents the difference in expectations E(ln w | $\hat{d}$+1, covariadas) - E(ln w | $\hat{d}$, covariadas), and is not directly considered to mean semi-elasticity.

To compare the magnitude of the OLS and 2SLS estimates, an alternative is to compute E(ln w | $\hat{d}^{máx}$,covariadas) - E(ln w | $\hat{d}^{mín}$,covariadas) = $\hat{\gamma}^{2SLS} \times (\hat{d}^{máx} - \hat{d}^{mín}) = 0{,}453$, for the maximum and minimum values of $\hat{d}$ [and recalling that in the case of semi-elasticity this would be equivalent to defining $\hat{d}^{máx} = 1$ e $\hat{d}^{mín} = 0$]. Seeing as this value is greater than $\hat{\gamma}^{OLS}$, the OLS result for the ATT must be underestimated. In fact, in the Durbin-Wu-Hausman test, whose statistic was 335,26 for 127 covariates, the null hypothesis of the same estimates generated by OLS and 2SLS is rejected, indicating the endogeneity of d and suggesting that OLS underestimates the ATT – which is a coherent result, for example, with the discussion presented by Barbosa-Filho e Pessoa (2010).

Column (iii) establishes a dummy emp=1 for those who are employed (i.e., with w observed in RAIS), as dependent variable on a regressed Probit in the covariates of the OLS, in order to estimate the IMR to be applied in the Heckit – following the procedure detailed in Mullahy (1998) or Wooldridge (2010, p.563-564). From that we understand that the chance of being employed increases up to a certain age, and then decreases; and, that it is more (less) likely to observe someone employed if this person has studied at night (with disability). Thus, columns (iv) and (v) show the results for Heckit using d and $\hat{d}$, respectively, where essentially the same results of OLS and 2SLS are observed.

Table 7: Results of estimated parameters for Equation (2) and auxiliary regressions

Covariate \ Dependent Variable [and procedure]	ln w	d	ln w	emp	ln w	ln w	ln w	ln w	ln w	ln w
	[using the entire database]						[using only paired observations]			
	(i) OLS	(ii) 2SLS 1º stage	(iii) 2SLS 2º stage	(iv) Probit to estimate IMR	(v) Heckit with d	(vi) Heckit with dˆ	(vii) OLS	(viii) 2SLS 2º stage	(ix) Heckit with d	(x) Heckit with dˆ
d ou dˆ	0,279*** (0,012)		0,741*** (0,183)		0,278*** (0,011)	0,718*** (0,181)	0,294*** (0,017)	0,729*** (0,013)	0,225*** (0,013)	0,721*** (0,016)
Age	0,067*** (0,006)	-0,034*** (0,003)	0,084*** (0,016)	0,004*** (0,001)	0,069*** (0,006)	0,085*** (0,015)	0,087*** (0,016)	0,095*** (0,019)	0,091*** (0,016)	0,098*** (0,015)
Age^2	-0,022*** (0,001)	0,005 (0,004)	-0,024*** (0,003)	-0,040*** (0,002)	-0,043*** (0,004)	-0,045*** (0,004)	-0,027*** (0,004)	-0,028*** (0,003)	-0,053*** (0,004)	-0,052*** (0,003)
Male	0,233*** (0,007)	-0,050*** (0,002)	0,255*** (0,016)	0,005 (0,006)	0,230*** (0,007)	0,251*** (0,017)	0,253*** (0,018)	0,262*** (0,027)	0,250*** (0,019)	0,258*** (0,025)
Full Scholarship	0,021 (0,012)	0,036*** (0,009)	0,003 (0,012)	0,002 (0,006)	0,022 (0,012)	0,005 (0,012)	-0,001 (0,012)	0,002 (0,013)	0,001 (0,014)	0,002 (0,016)
ODL	-0,076* (0,035)	-0,115** (0,041)	-0,033 (0,025)	0,049 (0,041)	-0,045 (0,038)	-0,004 (0,023)	-0,004* (0,001)	-0,004 (0,024)	-0,003 (0,023)	-0,004 (0,022)
Night Shift	0,022** (0,008)	0,004 (0,004)	0,020** (0,006)	0,166*** (0,007)	0,095*** (0,017)	0,096*** (0,017)	0,026*** (0,014)	0,024*** (0,013)	0,017*** (0,014)	0,011*** (0,014)
University	0,046** (0,016)	-0,001 (0,010)	0,040 (0,021)	0,016** (0,006)	0,052** (0,017)	0,046* (0,021)	0,033 (0,021)	0,024 (0,022)	0,041 (0,021)	0,029 (0,022)
Disability	-0,013 (0,031)	0,001 (0,015)	-0,013 (0,031)	-0,187*** (0,049)	-0,098** (0,034)	-0,103** (0,031)	-0,044 (0,041)	-0,038 (0,037)	-0,153** (0,044)	-0,143** (0,045)
IGC 2		0,051*** (0,014)								
IGC 3		0,191*** (0,013)								
IGC 4		0,207*** (0,020)								
IGC 5		0,219** (0,041)								
Age of the IES		0,010*** (0,001)								
IMR (Results Measuring Instrument)					1,151*** (0,224)	1,212*** (0,263)			1,476*** (0,287)	1,399*** (0,290)
			(constants and dummies for cohorts, courses and FU have been added, but purposely omitted here)							
R^2 (♦Pseudo-R^2)	0,121	0,047	0,111	0,073♦	0,122	0,112	0,113	0,086	0,114	0,087

Robust standard deviation in parentheses; *** p < 0,01 , ** p < 0,05 and * p < 0,10 .

Next, columns (vii), (viii), (ix) and (x) show the estimated results following the same procedures described above, but using only the observations paired by PSM. The pairing protocol followed the steps described in Gertler et al. (2018) (and the estimated results are purposely omitted, as in previous cases, to spare a tedious discussion, which would have little to contribute to the analysis): a Probit is estimated to be d the dependent variable, using the covariates of the OLS; the propensity scores are computed; a common support is defined and treatment/control units similar in propensity score are selected, using the criterion of the nearest neighbor; a balancing test is performed to evaluate the quality of the pairing; and then unpaired units are discarded. In the end, essentially the same results estimated with the other specifications are observed.

Due to this last conclusion, it repeats the estimation of the OLS using all of the units of observation restricted to the course, where the measured ATT [in square brackets] can be displayed from highest to lowest, in the following terms: Pharmacy, [0,50], Medicine [0,49], Information Systems [0,44]; Nursing [0,43]; Civil Engineering, [0,41], Mechanical Engineering, [0,39], Economics [0,39]; Computing Science, [0,37]; Nutrition [0,36], Production Engineering, [0,35]; Accounting [0,33]; Electrical Engineering, [0,31], Management [0,27]; Social Communication [0,26], Architecture [0,25]; Marketing [0,25]; Law [0,19]; Psychology [0,19]; Biology [0,19]; and Physiotherapy [0,10]. In the other courses, the estimated ATT was less than 0,10 or statistically zero.

6. Financial Mathematics

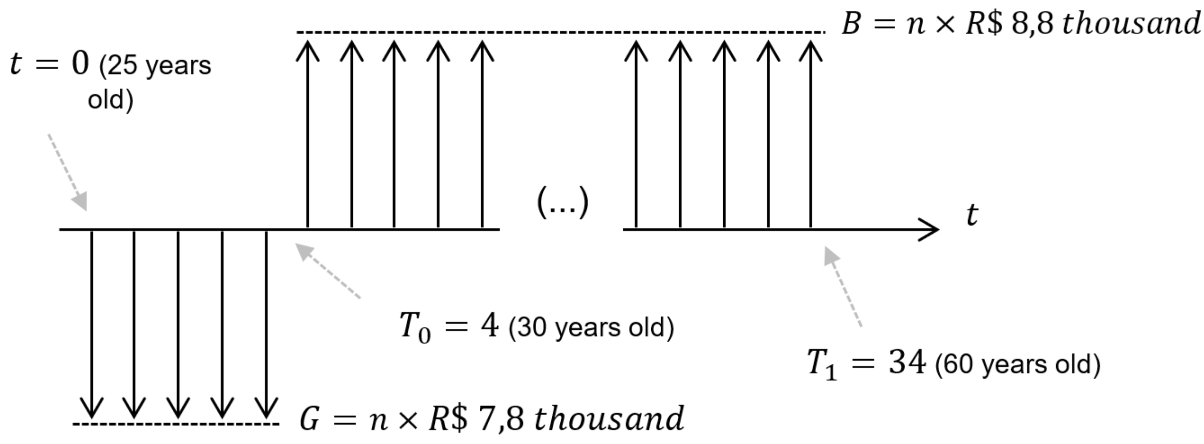
Based on what has been discussed so far, the idea of NVP^{social} expressed in Formula (2) is resumed, and γ=0,2 is considered. This ATT in appreciation is the OLS estimate rounded down, although the 2SLS numbers suggest that this represents an underestimated value. Therefore, we start from a conservative hypothesis for the magnitude of ProUni's impact in the context being analyzed.

Considering $r = \exp(\gamma) - 1$, it occurs that $B \cong (r/(1+r)) \times \Sigma_i w_{1i} = 18\% \times \Sigma_i w_{1i}$, where $\Sigma_i w_{1i}$ is the wage mass of graduates. That is, we estimate that 18% of this mass would be consequence of ProUni, and 82% would exist even in the absence of the program. So B represents the collective benefit of the policy, in that it would be money created through the mechanism of scholarships, to be consumed, invested, taxed, etc.

As previously seen, the average salary of graduates in the first post-training years is 3,50 MW/month. So, for n graduates, $\Sigma_i w_{1i}/n = 3{,}5 \times$ R\$ $1.045 \times 13{,}3 =$ R\$ 48,6 thousand/year occurs, considering the MW of 2020 (R\$ 1.045) and 13,3 annual salary (i.e., twelve months plus the thirteenth and the vacation additional). Consequently, $B = n \times$ R\$ 8,8 thousand/year.

We also know that annual tax expenditure to generate a scholarship is about R\$ 6 thousand, in 2020 values. But a scholarship is potentially taken by more than one beneficiary, since there are close to 30% cases of evasion or disconnection from the program (for lack of frequency and/or excess of failed classes). Therefore, it is more realistic to consider something like $G=n\times1{,}3\times$ R\$ 6 thousand/year$=n\times$R\$ 7,8 thousand/year. Because of the friction in allocating scholarships, it is also more reasonable to consider five years of tax exemption to generate a diploma, implicating that $T_0=4$, even if the representative training is four years.

Figure 3: Social Cash Flow Diagram of the Program using the estimated results



$$NPV^{social} = n \times \left(-\sum_{t=0}^{4} \frac{7{,}8}{(1+\tau)^t} + \sum_{t=5}^{34} \frac{8{,}8}{(1+\tau)^t}\right) \tag{3}$$

For the closing balance of the social cash flow, as shown in Figure 3, it is necessary to establish for how long a representative graduate can maintain the benefit. As a reminder, a scholarship student tends to start the degree at around 25 years of age. So it is fair to assume that the eventual graduation will take place near the age of 30. Taking into account that, currently, the minimum age for retirement is 57 years for women and 62 for men, we consider that a graduate will work representatively until at least 60 years of age. Consequently, it is reasonable to consider that the benefit would be maintained for something like 30 years, which implicates $T_1 = 34$, and the formula of NVP^{social} given by Equation (3).

That way, $NVP^{social} = 0$ implies an IRR such that τ=16%. Roughly speaking, this means that if the government has a policy alternative that generates socioeconomic impacts of return greater than 16% per year, and no extra resources are available, it is reasonable to consider the possibility of a reduction (or extinction) of ProUni in favor of this alternative.

To calculate the NVP^{social} properly, and a payback, it is necessary to stipulate a reference rate τ. In this sense, we consider τ=5%, equivalent to the current SELIC rate. The idea is that, instead of generating tax expenditure and creating scholarships, it would be possible to perform the tax collection of HEIs, and deposit this amount in a fund remunerated by SELIC. Thus, as an alternative to the mechanism of generating scholarships, the government could simply distribute these resources to the individuals of the target audience over time. So NVP^{social} would represent how much is gained from the existence of ProUni vis-à-vis this alternative, keeping the other premises constant.

With this configuration, $NVP^{social} = n \times$ R\$ 76 thousand is computed, with a 10-year payback; and, taking into account the estimated n>500 mil graduates that the program has already generated, we conclude that this investment has generated more than R\$ 38 billion since 2005. Nevertheless, it is clear that the values change in different scenarios, but based on what has been discussed so far, these values prove to be reasonably true to reality.

It Is also evident that the ATT can change considerably across different courses. In this sense, Table 8 shows the results of the Economic Analysis for the courses in which $\hat{\gamma}>0,1$, occurs, computing the components of NVP^{social} following the previous premises. In the first columns the estimates described above are reclaimed, and the fraction of the salary mass that can be attributed to ProUni is presented. We can see that, for example, this number would be 39 and 10% in Pharmacy and Physiotherapy, respectively.

Table 8: Results of the contrast between tax expenditure and the impact on the salary of graduates by selected course

Course	$\hat{\gamma}$	r/(1+r)	$(\Sigma_i \, _{1i})/n$	B/n	TIR	NVP^{social}/n	Payback
Pharmacy	0,50	0,39	4,93	25,45	39,25	204,40	6,43
Medicine	0,49	0,39	9,76	49,67	56,13	424,46	5,28
Sist. Info.	0,44	0,35	4,49	20,94	34,99	163,49	6,25
Nursing	0,43	0,35	4,13	19,02	32,98	145,99	6,94
Eng. Civil	0,41	0,33	4,34	19,08	33,05	146,54	6,93
Mech. Eng.	0,39	0,32	5,22	22,35	36,37	176,22	6,64
Economics	0,39	0,32	4,67	19,78	33,79	152,93	6,86
Computing	0,37	0,31	4,35	17,56	31,38	132,74	7,89
Nutrition	0,36	0,31	2,83	11,40	23,49	76,76	7,63
Prod. Eng.	0,35	0,30	4,53	17,70	31,54	134,02	7,91
Accounting	0,33	0,28	3,58	13,21	26,04	93,27	7,86
Electrical Eng.	0,31	0,27	5,56	19,72	33,73	152,41	6,87
Administration	0,27	0,24	3,31	10,27	21,78	66,53	8,78
Communication	0,26	0,23	3,11	9,37	20,33	58,31	9,81
Marketing	0,25	0,22	3,10	8,99	19,69	54,86	9,60
Architecture	0,25	0,22	2,78	8,09	18,13	46,70	9,96
Law	0,19	0,17	3,61	8,25	18,41	48,12	9,85
Psychology	0,19	0,17	2,75	6,13	14,31	28,91	11,69
Biology	0,19	0,17	2,66	6,05	14,13	28,15	9,61
Physiotherapy	0,10	0,10	2,39	3,12	6,55	1,52	23,92

The following column indicates the average salary observed among graduates by the respective course, $\Sigma_i w_{1i}/n$, in monthly MW. It shows that a person with a degree in Medicine earns almost twice as much as those who have a degree in Pharmacy, Information Systems or Nursing; and about four times more than those who have a degree in Physiotherapy, Biology or Psychology.

Connecting the previous values in accordance with what was described above, the next column shows the estimate of the average social benefit of investment in the scholarship program, B/n, in monthly MWs. We can see that a graduation in Medicine has double (or more) the value of a degree in Pharmacy, Information Systems or Nursing, as a direct consequence of the relatively higher salaries. That is reflected in higher IRR and average NPV, and lower payback.

The estimates expressed in Table 8 allow us to evaluate ProUni's scholarship policy from a couple of points. First, although the IES offer a variety of courses to beneficiaries, only a portion of them have a significant impact in terms of economic viability. In this sense, the option of courses with low social return should be rethought in favor of the adoption of scholarships for poor students in courses with greater possibility of inclusion in the labor market. On the other hand, contrasting the results expressed in this section with the figures previously presented in Table 3, the program has focused most of the scholarships in courses listed in Table 8, indicating that, in general, ProUni offered access and democratized education with positive economic return.

It should be noted that there is a considerable heterogeneity in employability, compensation and stability in the formal labor market, depending on the degree chosen by students. Therefore, from the information in Table 8, it is possible to visualize the existence of courses that have high individual impact with low social return (notably, Medicine) – since they contemplated a limited amount of students –, and also the existence of courses with lower individual impact (such as Administration) that, by encompassing a large contingent of beneficiaries, generate a rather high aggregate social impact.

With the possibility of redefining ProUni guidelines, and allocating scholarships preferably among courses with higher economic and social return, it would be of great importance to regularly monitor the distribution of these benefits, in view of the dilemma between an allocation that allows higher individual returns and lower social impacts, vis-à-vis courses that generate lower individual impact and, due to the large contingent, a large social impact.

7. Conclusion

Since 2005, when it was institutionalized, ProUni has cost more than a dozen billion reais in tax expenditures, and enabled the graduation of nearly half a million students. This monograph provided an unprecedented economic evaluation of the program, considering its impact on the salary of graduates, through the crossing of three microdata bases (SISPROUNI, CES and RAIS identified).

By defining graduates and dropouts as treatment and control groups, respectively, we estimate that, in the first years after graduation, on average, a graduate earns 30% more than their non-graduate counterparts. In a financial mathematics exercise with conservative hypotheses, we estimated that: ProUni's IRR would be 16% per year; the social NPV would already be greater than R$ 38 billion; and each generation of graduates would equate the tax expenditure with their cohort in about 10 years. Therefore, there is evidence that the program generates a collective benefit that would justify its existence.

On the other hand, in the analyses disaggregated by courses, it stands out that some trainings do not generate an impact on employability for the beneficiaries. Although courses such as Administration and Accounting Sciences have a smaller individual NPV than Pharmacy and Medicine, for example, we notice that the social NPV of the former is substantially higher due to the large number of beneficiaries in these courses. Moreover, we noticed that it is possible to improve the focus of the scholarships in order to potentialize the application of tax expenditure.

In particular, under the hypothesis that the appropriate legal provisions are properly analyzed and assessed, we propose a future investigation of an adjustment in the wording of Law No. 11.096

(which regulates the program). Which is: to create a mechanism that partially blocks the tax exemptions of IES that offer scholarships in courses where there is no evidence of increased employability for graduates; and, through this blocking, in a zero-sum game, finance the increase in the benefits of HEIs that offer scholarships in courses where there is strong evidence of increased employability of graduates – without generating extra costs for the Treasury. In this context, the research model presented here could be improved and regularly replicated by government agencies (e.g., INEP or IPEA), in order to encourage the creation and maintenance of this mechanism.

We also understand that this rule could be modified with some frequency, perhaps annually, based on what is most recently observed in RAIS. We believe that this would represent a type of management by results, which would enhance the collective benefit of ProUni.

8. References

Andriola, W. B e Barrozo-Filho, J. L. "Avaliação de políticas públicas para a educação superior: o caso do Programa Universidade Para Todos (ProUni)", Revista da Avaliação da Educação Superior 25 (2020): 594-621. https://doi.org/10.1590/S1414-40772020000300005

Barbosa, M. P.; Petterini, F. C.; Ferreira, R. T. "Expansion of Brazilian Federal Universities: is it possible to raise economic impacts?", Revista de Administração Contemporânea 24, (2019): 1-24. https://doi.org/10.1590/1982-7849rac2020190230

Barbosa-Filho, F. H.; Pessoa, S. A. "Educação e crescimento: o que a evidência empírica e teórica mostra?". Revista Economia 11 no.2 (2010): 265-303, (2010). http://www.anpec.org.br/revista/vol11/vol11n2p265_303.pdf

Becker, K. L.; Mendonça, M. J. "Avaliação de impacto do ProUni sobre a performance acadêmica dos estudantes". Texto para Discussão Ipea 2512: 1-38, (2019). https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_2512.pdf

Borgen, N. T. "College quality and hourly wages: Evidence from the self-revelation model, sibling models and instrumental variables". Social Science Research 48: 121-134, (2014). http://dx.doi.org/10.1016/j.ssresearch.2014.05.010

Brasil. "Relatório de Auditoria Operacional: Programa Universidade para Todos (ProUni) e Fundo de Financiamento ao Estudante do Ensino Superior (FIES)". Tribunal de Contas da União, (2009). http://portal2.tcu.gov.br/portal/page/portal/TCU/comunidades/programas_governo/auditorias

Brasil. "Relatório de Avaliação da Execução de Programas de Governo Nº 35: ProUni". Controladoria Geral da União (2015). https://eaud.cgu.gov.br/relatorios/

Corbucci, P. R. “Desafios da educação superior e desenvolvimento no Brasil”. Texto para Discussão Ipea 1287: 1-35, (2007). https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_1287.pdf

Corbucci, P. R. “Evolução do acesso de jovens à educação superior no Brasil”. Texto para Discussão Ipea 1950: 1-40, (2014). https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_1950.pdf

Corbucci, P. R; Kubota, L. C.; Meira, A. P. B. “Reconfiguração Estrutural da Educação Superior Privada no Brasil: nova fase da mercantilização do ensino”. Texto para Discussão Ipea 2256: 1-42, (2016). https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_2256.pdf

Dale, S. B.; Krueger, A. B; “Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables”. The Quartely Journal Of Economics 117 (4): 1491-1527 (2002) https://www.jstor.org/stable/4132484

Darolia, R. “Working (and studying) day and night: heterogeneous effects of working on the academic performance of full-time and part-time students”. Economics of Education Review 38: 38-50, (2014). https://doi.org/10.1016/j.econedurev.2013.10.004

De Negri, J. A.; Castro, P. F. D.; Souza, N. R. D.; Arbache, J. S. “Mercado Formal de Trabalho: Comparação entre os Microdados da RAIS e da PNAD”. Texto para Discussão Ipea 840: 1-29, (2001). http://repositorio.ipea.gov.br/bitstream/11058/2155/1/TD_840.pdf

Durham, E. Educação superior, pública e privada (1808-2000). In: BROCK, C., SCHWARTZMAN, S. (orgs.). “Os desafios da educação no Brasil”. Rio de Janeiro: Nova Fronteira 197-240 (2005). http://www.schwartzman.org.br/simon/desafios/7superior.pdf

Dynarski, S. “Does aid matter? Measuring the effect of student aid on college attendance and completion”. American Economic Review 93 no.1: 279-288, (2003). https://www.jstor.org/stable/3132174

Dynarski, S. “Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance,” National Tax Journal 53, no. 3: 629-662, (2000). https://www.journals.uchicago.edu/doi/10.17310/ntj.2000.3S.02

Faveri, D. B.; Petterini, F. C.; Barbosa, M. P. “Uma avaliação de impacto da política de expansão dos IFs nas economias dos municípios brasileiros”. Planejamento e Políticas Públicas 50: 125-147, (2018). https://www.ipea.gov.br/ppp/index.php/PPP/article/view/742/464

Gertler, P. J.; Martinez, S.; Premand, P.; Rawlings, L. B.; Vermeersch, C. M. “Impact evaluation in practice”. The World Bank, (2016). https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice

Haas, C. M.; Pardo, R. D. S. “Programa Universidade para Todos (ProUni): efeitos financeiros em uma instituição de educação superior privada”. Avaliação: Revista da Avaliação da Educação Superior (Campinas) 22, no.: 718-740, (2017). https://doi.org/10.1590/S1414-40772017000300008

Heckman, James J.; Lochner, Lance J.; Todd, Petra E. “Earnings functions, rates of return and treatment effects: The Mincer equation and beyond”. Handbook of the Economics of Education 1: 307-458, (2006). https://doi.org/10.1016/S1574-0692(06)01007-5

Katovich, E. S.; Maia, A. G. “A relação entre produtividade de trabalho e salário no Brasil: uma análise setorial”. Nova Economia 28, no. 1: 7-38, (2018). https://doi.org/10.1590/0103-6351/3943

Lépine, A. “Financial aid and student performance in college: evidence from Brazil.” Brazilian Review of Econometrics 38, no. 2: 221-261(2018). https://doi.org/10.12660/bre.v38n22018.75505

Mincer, Jacob. “Economic development, growth of human capital, and the dynamics of the wage structure”. Journal of Economic Growth 1, no. 1: 29-48, (1996). https://link.springer.com/content/pdf/10.1007/BF00163341.pdf

Mincer, Jacob. “The distribution of labor incomes: a survey with special reference to the human capital approach”. Journal of economic literature 8, no. 1: 1-26, (1970). https://www.jstor.org/stable/2720384

Mullahy, J. “Much ado about two: reconsidering retransformation and the two-part model in health econometrics”. Journal of health economics 17, no. 3: 247-281, (1998). https://doi.org/10.1016/S0167-6296(98)00030-7

Saraiva, Luiz Alex Silva et al. “A efetividade de programas sociais de acesso à educação superior: o caso do ProUni”. Revista de Administração Pública 45: 941-964, (2011). https://doi.org/10.1590/S0034-76122011000400003

Stinebrickner, R.; Stinebrickner, T. R. “Working during school and academic performance”. Journal of Labor Economics 21, no. 2: 473-491, (2003). https://doi.org/10.1086/345565

Taber, Christopher R. “The rising college premium in the eighties: Return to college or return to unobserved ability?” The Review of Economic Studies 68, no. 3: 665-691, (2001). https://www.jstor.org/stable/2695901

Tachibana, T. Y.; Menezes-Filho, N. A.; Komatsu, B. K. “Ensino superior no Brasil”. Policy Paper INSPER 4: 1-53, (2015). https://www.insper.edu.br/wp-content/uploads/2018/09/Ensino-superior-no-Brasil.pdf

Wooldridge, J. M. “Econometric analysis of cross section and panel data”. MIT press (2010). https://mitpress.mit.edu/books/econometric-analysis-cross-section-and-panel-data-second-edition