Secondary vocational education and earnings in Brazil

Marina Aguas (IBGE, Rio de Janeiro, Brazil, e-mail: mariffaguas@gmail.com)
Danielle Carusi Machado (UFF, Niterói, Brazil, e-mail: daniellecarusi@vm.uff.br)

Abstract

Vocational education is a type of human capital investment that aims to provide workers with knowledge, know-how, skills, specific competences and practices in a specific area, which is expected to improve their performance in the labor market. Investing in vocational education can also reduce the barriers to an individual's entry in their first-time job, as well as improve the quality of the job-worker match. Consequently, more skilled and productive workers can benefit from higher wages, promoting positive impacts for the entire economy. This paper investigates whether secondary vocational education is related to labor earnings in Brazil. Making use of data from the 2007 PNAD, the empirical analysis is conducted through three different methodologies: OLS, Treatment Effect and Propensity Score Matching. Evidence shows that attainment of a secondary vocational education is associated with an increase in labor earnings between 20% and 24% in all these methods. Evidence also suggests that the decision to attend such programs does not seem to be correlated with unobserved productive characteristics of workers.

Keywords: labor market, income, vocational training.
JEL: C21, J24, J31

Resumo

Educação profissional é um tipo de investimento em capital humano que tem como objetivo prover os trabalhadores com conhecimentos e competências/práticas específicas para determinada área, sendo esperado que isto melhore sua performance no mercado de trabalho. Investir em educação professional pode igualmente reduzir as barreiras de entrada ao primeiro emprego e, também melhorar o casamento entre trabalhadores/empregadores. Consequentemente, trabalhadores mais produtivos e com melhores conhecimentos ganham salários mais altos, impactando positivamente a economia. O objetivo deste artigo é investigar como a educação profissional se relaciona com os rendimentos do trabalho no Brasil. Usaremos os dados da PNAD de 2007 e adotaremos três metodologias empíricas: OLS, efeito tratamento e PSM. Mostramos que o impacto da educação profissional é de aumentar os ganhos de 20% a 24% em todos métodos e que a decisão de realizar estes cursos não parece ser correlacionada com características produtivas não observadas.

Palavras chaves: mercado de trabalho, educação profissional, renda do trabalho.

Área 2: Economia social e do trabalho

Secondary vocational education and earnings in Brazil

1. Introduction

Vocational education is a type of human capital investment that aims to provide workers with knowledge, know-how, skills, specific competences and practices in a specific area, which is expected to improve their performance in the labor market. Investing in vocational education can also reduce the barriers to an individual's entry in their first-time job, as well as improve the quality of the job-worker match. Consequently, more skilled and productive workers can benefit from higher wages, promoting positive impacts for the entire economy.

In Brazil, the current education law (# 9394/1996) defines three types of vocational education: vocational training, secondary vocational education and tertiary vocational education. Each one of these categories is structured for a specific target, with specific prerequisites and objectives. These programs usually have a shorter duration than regular education courses. Vocational education can be seen as a faster way of qualifying the workforce, especially in countries with a large population characterized by low levels of education.

The aim of this paper is to investigate the relationship between earnings and vocational education. Specifically, we chose the secondary vocational education (SVE) to do our analysis, because it is considered an intermediate level of the vocational education, it has guidelines and rules for all courses defined by Brazilian law and it is not a tertiary education course. Therefore, in order to receive the secondary vocational education degree, students must have finished the secondary school, as well as attend and successfully complete vocational courses recognized and structured by the Brazilian government. Thus, this segment differs from both vocational training courses, which are not regulated and have no educational prerequisites, and tertiary vocational education, which is part of the undergraduate courses in Brazil.

The empirical analysis is inspired by Mincer (1974), which defines an earnings equation as a function of educational level, experience and individual characteristics. After some adjustments, our analysis incorporates the SVE as an explanatory variable in the educational context. We use the data from the Brazilian National Household Survey (PNAD) for the year of 2007, as this year was held an additional questionnaire about vocational education. We chose a sub-sample of people aged between 25 and 55 years old, prime age group for labor market analysis, that live in metropolitan regions or self-representative areas in Brazil.

The estimation of the earnings equation is performed through three different methods: Ordinary Least Squares (OLS), Treatment-Effect Model (TE) and Propensity Score Matching (PSM). These last two methods are proposed to fix the possible selection bias present in the fact that educational choices are not random. This correction is based on observable factors, to the extent that these two methods cannot capture the observed selectivity which is the classical problem of the relationship between education and earnings. The OLS is the usual method and the results are compared with other two methods.

The results show that people who attended a SVE course have a positive and significant impact on their salary per hour, regardless of the estimation techniques used. For the OLS estimation, this effect increases wage per hour in 22%. In the TE method, the expected gain in hourly wage between those who have the training is 21% higher than those who did not. In this method, the results suggest the absence of sample selection bias. Finally, the PSM method shows salary increases ranging between 22% and 24%, according to the matching method chosen. These values are similar to those found in the OLS method and support the idea of the absence of selectivity in the observable characteristics of the sample.

2. Data and descriptive statistics

From the PNAD, of 2007, we selected a sample of employed people, with positive earnings, aged between 25 and 55 years old and with a high school diploma, only. This group should live in metropolitan regions or self-representative areas in Brazil. These choices were made to reduce heterogeneity problems between individuals and regions. The restriction to analyzing people who completed high school but did not attend college, aims to evaluate the return of the training between people who might benefit from it and that have similar characteristics. This sub-sample corresponds to 45% of the population between 25 and 55 years old, in 2007.

With the vocational education questionnaire in PNAD of 2007, it was possible to identify people who attended or completed a SVE course as well as their individual characteristics (sex, age, race, place of residence). Additionally, as part of the TE model, we created two identification variables that are based on observable characteristics of individuals in the sample to try to correct the selectivity bias of the data. These variables attempt to capture if the place of residence, and its surroundings, influence people's decision to have a training course. Table 1 presents a description of these variables.

Table 1: Variables

Variables	Description
Attended SVE course	Equal to 1 if the person held a SVE course, 0 otherwise.
Complete SVE course	Equal to 1 if the person complete a SVE course, 0 otherwise.
Percentage of people who have completed a SVE course in the stratum of residence	For every person i, aged J and resident in stratum T, calculate the number of people, who are residents in stratum T, aged equal to or greater than J, excluding itself, which concluded a SVE course, divided by the number of people, aged equal to or greater than J, excluding i itself, resident in stratum T.
Percentage of people who did not attend a SVE course because there was no provision of such courses in the stratum of residence	For every person i, aged J and resident in stratum T, calculate the number of people, who are residents in stratum T, aged equal to or greater than J, excluding itself, which did not attend a SVE course because there was no provision of such courses in the stratum of residence, divided by the number of people, aged equal to or greater than J, excluding i itself, resident in stratum T.

With respect to characteristics of the sample, Tables 2 and 3 show the individuals' profiles. Note that 14% of people aged 25-55 years old and with completed secondary school had attended a SVE course, and 12.7% completed this course. Among people who completed SVE course, 35% have never worked in the professional field of training. With respect to individuals without SVE course, it is important to emphasize that most part had no interest in doing them and a minority failed to do so due to the absence of courses.

The individual characteristics of people with and without SVE courses, in Table 3, reveal themselves very close: the percentage of men and women is about the same, the percentage of whites and the average age is slightly larger for those who attended SVE courses. With respect to labor market indicators, we notice some differences. The proportion of employed people with positive labor earnings is 10 percentage points higher for those who attended SVE courses compared to those without such training. The average hourly income for people who did not attend a SVE course is approximately 70% of the average hourly income of those who attended.

Table 2: individual characteristics

Variável	Brasil
People between 25-55 years old who are employed and with positive hourly earnings (people in 1000)	31,417
Attended SVE course (% of people)	13.9
Completed the course	91.4
Completed the course and work or have worked in the training area	64.0
Reason for not attended SVE course: (% of people):	
Is attending now any vocational education course or did a different segment	34.9
no provision of such courses	2.4
Lack of interest	36.8
Other reasons	11.9

Source: own elaboration, using PNAD 2007/IBGE.

Table 3: Characteristic of employed people with positive hourly income by attendance a SVE course

Variables	Attended a SVE course	Did not Attedent a SVE course
Women	47.3	45.8
White people	57.3	53.5
Avarage age	37.5	36.4
% of people between 25-55 years old who are employed with positive hourly earnings in the total population of this age group	80.5	70.9
Average hourly income (in BRL from 2007)	8.82	6.21

Source: own elaboration, using PNAD 2007/IBGE.

3. Methodology

Research on the impact of the SVE on labor earnings follows the approach implemented by Mincer (1974) in which labor earnings is represented as a function of education and experience. We try to capture the related effect of carrying out a SVE course on wages including a dummy variable indicating whether the individual attended or not this kind of education in the Mincer equation. Therefore, the model to be estimated is:

$$\mathbf{y_i} = \mathbf{lnR_i} = \mathbf{f}(\mathbf{nt_i}, \mathbf{x_i}, \mathbf{z_i}) + \mathbf{u_i}, \ i=1, 2, ..., n \qquad \text{(eq.1)}$$

Where lnR_i indicates the natural logarithm of hourly labor earnings of individual i, nt_i is the binary variable indicating whether the person i did or did not a SVE, x_i is her labor experience, $z_i$are other individual characteristics, like gender, place of residence, etc, and u_i is a stochastic error.

The model is estimated by three methods: ordinary least squares (OLS), treatment-effect (TE) and propensity score matching (PSM). A concern about OLS estimates is the fact that individuals can choose to do or not to do a SVE course and this choice can be a source of bias if some unobservable individual characteristics, such as skills talent, affect wages and also the probability of doing a SVE course. Thus, an individual with more skill talent may receive a higher salary and may also have a higher probability to engage in a vocational course. Therefore, it is possible that higher earnings for those who are trained are due in fact to their greater ability.

In order to dealing with the sample selection problem mentioned before, we use two others methods: TE and PSM, which are very useful in the economic literature of impact evaluation of public policies and social programs. These techniques allow researchers to measure the causal effect of a "treatment" or "program" on a generic outcome variable. In our case, the SVE can be considered a "program" or "treatment" and the amount of hourly labor earnings as the outcome variable. The better way to evaluate the impact of SVE would be to calculate the difference between the individual`s labor earnings that result from participating and not in training. We do not observe, however, the wages of individuals who receive training if they had not received it, and we do not observe wages for individuals who were not trained if they were trained. In order to estimate the casual effect of the treatment we should reconstruct the outcomes that are not observable, making use of counterfactuals. The TE and PSM methods propose different ways to estimate counterfactuals outcomes.

3.1. Treatment-effect model[1]

This method estimates the effect of nt_i on the continuous variable observed y_i, hourly labor earnings, conditional on exogenous independent variables x_i and w_i, taking into account the decision process of doing or not doing a SVE course. This process is modeled as a result of a latent unobserved variable, nt_i^*, which is supposed to be a linear function of the exogenous explanatory variables w_i, and a random error component e_i. That is,

$$y_i = x_i\,\beta + \delta nt_i + u_i \quad \text{(eq. 2)} \qquad\qquad nt_i^* = w_i\,\gamma + e_i \quad \text{(eq. 3)}$$

$$nt_i = \begin{cases} 1, & if\ nt_i^* > 0 \\ 0, & otherwise \end{cases} \qquad \text{(eq. 4)}$$

Where, e and u have a bivariate normal distribution with zero mean and correlation equal ρ.

Although this method has a parametric specification and, therefore, the identification can be achieved even if W = X, in practice, the estimator will not perform well in the absence of an exclusion restriction in W. The set of variables in X are: sex, age, age squared, race, place of residence and household situation for each individual. The set of variables in W includes those in X and the identification variables shown in the data section.

Identification variables are intended to represent the perceived benefit associated with attending or not a SVE course in the vicinity of the residence of each individual i. We assume that the higher the proportion of persons aged greater than or equal to i who have completed a SVE course, the greater their influence on younger workers decision regarding attending this modality of education. Similarly, higher the percentage of people who did not attained such courses because there was no provision of this type of education in their neighborhood diminishes the younger individuals incentives to attend a SVE course.

Considering the sample selection problem, the wage equation for treated individuals is represented by (5):

$$E(y_{1i}|nt_i = 1) = x_i\,\beta + \delta + E(u_i|nt_i = 1) = x_i\,\beta + \delta + \rho\sigma_u\lambda(w_i\,\gamma) \qquad \text{(eq.5)}$$

And the wage equation for non treated individuals is represented by eq. (6):

[1] For more information, see Maddala (1983), Barnowet al. (1981) and Cong&Drukker (2000).

$$E(y_{0i}|nt_i = 0) = x_i\,\beta + \rho\sigma_u \left[\frac{-\phi(w_i\,\gamma)}{1-\Phi(w_i\,\gamma)}\right] \qquad \text{(eq.6)}$$

Where λ is the inverse Mills ratio and the parameters will be estimated in two stages.

The difference in expected hourly labor earnings between participants and non-participants can be written as:

$$E(y_{1i}|nt_i = 1) - E(y_{0i}|nt_i = 0) = \delta + \rho\sigma_u \left[\frac{\phi(w_i\,\gamma)}{\Phi(w_i\,\gamma)\{1-\Phi(w_i\,\gamma)\}}\right] \qquad \text{(eq.7)}$$

Where ϕ and Φ are density and cumulative distribution functions of a normal distribution. If the correlation between error terms, ρ, is zero, the difference reduces to δ, and we could estimate it using OLS. However, it is believed that ρ is positive, since there are unobserved factors that influence the decision to do a SVE course. Thus, the OLS method should overestimate the treatment effect, δ.

3.2 Matching Model

The Matching model is a non-parametric approach that seeks to reproduce an experiment ex post. The method consists in selecting, among those in the untreated group, a sample of individuals as similar as possible to the treated group, called control group. This method departs from the assumption that all relevant differences between treated and untreated persons are captured by the observed characteristics. Under this assumption, all relevant outcome differences between treated and non-treated individuals are captured by their observable attributes; the only remaining difference between the two groups is the treatment status. In our case, the treatment is the attainment of a SVE course.

The matching on the propensity score is implemented as follows. The method proposes to summarize in a single variable index, called propensity score, the individual characteristics pre-treatment. The propensity score (PS), defined by Rosenbaum and Rubin (1983), is the probability of receiving treatment given the set of observed characteristics jointly affecting treatment status and outcomes: $p(X) \equiv Pr(nt = 1|X) = E(nt|X)$ (eq.8)

Since treated and non-treated individuals with the same value of the PS have the same distribution of the vector of regressors X, and adopting the assumption that treatment is randomly assigned within the groups defined by X, Rosenbaum e Rubin (1983) show that treatment is also randomized within groups defined by p(X). As a result, if the PS of an individual i is known, p(Xi), so the average treatment effect on the treated (ATT) is given by:

$$ATT \equiv E[Y_{1i} - Y_{0i}|nt_i = 1] = E\{E[Y_{1i} - Y_{0i}|nt_i = 1, p(X_i)]\}$$

$$= E\{E[Y_{1i}|nt_i = 1, p(X_i)] - E[Y_{0i}|nt_i = 0, p(X_i)] \,|nt_i = 1\} \qquad \text{(eq.9)}$$

Where Y_{1i} and Y_{0i} are the potential results of the two contrafactuals situations of treated and non-treated. So, conditioning on the propensity score allows us to obtain unbiased estimates of average treatment effects.

The first step to evaluate SVE effect on earnings is to estimate Pr(nti=1|Xi). We estimate the propensity score using a probit regression model, in which treatment status is regressed on observed baseline characteristics. The estimated PS is the predicted probability of treatment derived from the fitted regression model. As it is too difficult to observe two individuals with the same value of p(X) to do the matching, the match for each individual in treated group to someone in the control group depends on the distance between their propensity scores. We compare four different methods to do this matching: (i) the nearest neighbor matching, which selects the non-participant for whom the value of Pj is closest to Pi, (ii) the radius matching, where a match for individual i is selected only if the propensity score falls within a radius r from Pi, (iii) the kernel matching, where the outcome of treated unit i is matched to a weighted average of the outcomes of more (possibly all) non-treated units, where the weight given to non-treated unit j is a function of the closeness of the characteristics of i and j, and (iv) the stratification method, where the sample is stratified into K groups according to PS intervals. For each group, these covariates are balanced, that is, for each covariate there is no significant difference between the groups.

4. Results

Table 4 shows the OLS estimates results of eq. 1, considering individuals who attended or completed a VSE course. The results show that attending a SVE course means getting an hourly labor earnings about 21% higher than that received by those who have not attended such a course, even with all the prerequisites of doing it. The estimated coefficients are significant at 1% level and vary little when geographical controls are included. For people who finished the course, a slightly higher and significant effect on hourly labor earnings is generated (22%), with our without geographical controls.

Table 4: Results – OLS estimates

Variables	**Ln (hourly labor earnings)**

	OLS		OLS with geographical controls	
Attended a SVE course	0.211*** (0.016)		0.213*** (0.017)	
Completed a SVE course		0.221*** (0.017)		0.224*** (0.017)
Observations	22,262	22,262	22,262	22,262
R^2	0.187	0.187	0.212	0.212

Robust standard errors in brackets. *** p<0,01, ** p<0,05, * p<0,1. Source: microdata PNAD 2007 (IBGE).

Table 5 shows EF two-stage estimate results for eq. (2) and (3). For the first stage, coefficients of the identification variables are significant and exhibit the expected sign. The decision to make a SVE course is positively associated with the percentage of people who have completed it in the neighborhood, and negatively related to the percentage of people who did not this type of course because there was no provision of that in the neighborhood. The second stage estimates the wage equation taking into account all the explanatory variables and the selectivity term from the first stage (lambda variable). The coefficients associated with the participation or conclusion of a SVE course are significant and positive. But the selectivity correction term (lambda) is negative and not significant for both specifications. With our two identification variables, sample selection does not seem to exist.

The difference in expected earnings between those who attended the SVE course and those who did not attend it (eq. 7) is significant and equal to 0.207. For those who completed the course, compared to those who did not, the difference is 0.217. Results are slightly lower than those obtained by OLS.

Table 5: Results: TE estimates

Variables	**Ln (hourly labor earnings)**	**Attended a SVE course**	**hazard**	**Ln (hourly labor earnings)**	**Completed a SVE course**	**hazard**
Attended a SVE course	0.343** (0.133)					
Completed a SVE course				0.357** (0.145)		
Percentage of people who have completed a SVE course in the stratum of residence.		0.038*** (0.004)			0.036*** (0.004)	
Percentage of people who did not attend a SVE course because there was no provision of such courses in the stratum of residence		-0.051*** (0.012)			-0.048*** (0.012)	
Lambda			-0.072 (0.073)			-0.073 (0.078)
Constant	0.306*** (0.104)	-1.782*** (0.255)		0.386*** (0.108)	-2.888*** (0.266)	
Observations	22,262	22,262	22,262	22,262	22,262	22,262
Difference between treated and non-treated		0.2074*** (0.010)			0.2176*** (0.010)	

Robust standard errors in brackets. *** p<0,01, ** p<0,05, * p<0,1. Source: microdata PNAD 2007 (IBGE).

Table 6: Propensity Score Matching: estimates

Variables	ATT					
	Attended	**# treated**	**#control**	**Concluded**	**# treated**	**#control**
Nearest neighbor matching	0.221 (0.016)	3083	12793	0.231 (0.017)	2793	12443
Kernel	0.232 (0.014)	3083	19151	0.235 (0.015)	2793	19442
Stratification	0.209 (0.001)	3083	19151	0.219 (0.016)	2793	19442
Radiusmatching - radius=0,01	0.229 (0.015)	3083	19100	0.225 (0.015)	2793	19442
Radiusmatching - radius=0,001	0.228 (0.015)	3083	19100	0.220 (0.015)	2792	19372
Radiusmatching - radius=0,0001	0.234 (0.015)	3058	17482	0.229 (0.016)	2775	18020

Source: microdata PNAD 2007 (IBGE).

Table 6 shows the ATT results (eq. 9) using the PSM and the four possible techniques to do the matching between the treatment and the control group. The estimated ATT are positives and significant for all matching methods reported. Estimated coefficients for people who attended the SVE course range between 0.20 and 0.23. Therefore, a person who attended a SVE course has the hourly labor earnings 20% to 23% higher than another one who did not attended the training. The results for people who concluded a SVE course is a little bigger, between 22% to 23,5%.

The four matching methods used to make the correspondence between the treated group and the control group generated similar results. Moreover, results are close to the coefficients obtained by the OLS estimation, which does not consider the presence of sample selection, and also to the results obtained by TE method.

5. Conclusions

Without sample selection, OLS estimates to determine the effects of vocational secondary education on hourly labor earnings should be correct. The OLS results show that attended or concluded a vocational secondary course means getting hourly labor earnings about 21 to 22% higher than that received by those who have not attended such a course. Taking into account the sample selection, TE and PSM methods (and all four techniques to do matching), show similar results to OLS estimates. These estimated differences between treated and non-treated lie in the range 20 to 23%, so, no longer than OLS estimates and suggesting that there is no evidence of sample selection.

Despite this, we must keep in mind that the estimation methods that seek to correct the possible selection bias of the data are based on strong assumptions and their results should be interpreted with caution. In the case of TE method, there is the hypothesis that all heterogeneity among individuals regarding the decision to do a vocational secondary course is captured by the identification variables. In PSM method, conditional on the probability of participation, the average difference in results between the treatment group and the control is only explained by the participation in a vocational secondary education.

References

Barnow B., Cain G., Goldberger A. ,(1981). Issues in the Analysis of Selection Bias, in the Economics of Training, vol I, Ashenfelter and Lalonde eds., 1996.

Cong, R. and Drukker, D.M. (2000). Treatment Effects Model. *Stata Technical Bulletin*, 2001, vol. 10, issue 55

Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York.

Mincer, J. (1974) Schooling, Experience, and Earnings, National Bureau of Economic Research. Columbia University Press, New York.

Rosenbaum P, Rubin D. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70; 41-55.