Wasted years: Informality and returns to education in Bolivia

Años perdidos: informalidad y retornos a la educación en Bolivia

Pablo Mendieta Ossio*

Abstract

During the last years some estimates showed a low average return on schooling in Bolivia. With the use of two databases, the standard annual household survey, and the novel demand for skills survey, I found a non-linear shape of schooling returns. Even more, I improved the estimation with the use of a segmented labor market approach. Joint a matching technique, I estimated models which support the hypothesis that this pattern is due to a widespread informality issue. I explore the ways this result could be enhanced with a skill connection and a general-equilibrium segmented search labor model.

Keywords: Schooling returns; informality; Bolivia.

Resumen

En los últimos años, algunas estimaciones mostraron un bajo retorno a la educación en Bolivia. Con el uso de dos bases de datos, la encuesta anual de hogares y una base novedosa de demanda de habilidades, se encuentra una forma no lineal de los retornos de la educación.

*   Ph.D. Student at UPB Graduate Economics Program 2017-2021 and Deputy Manager of the Bolivian Center of
     Economics at CAINCO.
     Contact: pablo.mendieta@cainco.org.bo

Incluso, se mejora la estimación con el uso de un enfoque de mercados laborales segmentados. Junto a una técnica de pareo se estiman modelos que apoyan la hipótesis de que este patrón se debe al alto grado de informalidad. Finalmente se discute las formas en las cuales esta relación puede ser mejorada con la inclusión de habilidades y un enfoque de equilibrio general de búsqueda de empleo en mercados segmentados.

Palabras clave: Retornos a la educación; informalidad; Bolivia.

Classification/clasificación JEL: C51, I26, J24

1.   Introduction

> Do not waste your time always searching for those wasted years[1]

During the last years, some Bolivian authors have highlighted low and decreasing returns to education in Bolivia using the traditional `Mincerian approach' developed first by Mincer (1974) with the use of information of annual household survey conducted by National Statistics Office (INE due its acronym in Spanish).

For instance, Villarroel and Hernani (2011) found that returns to education have decreased during the first decade of this century from 10% to around 5%. They do not hypothesize about the causes of this trend, although they discard education supply roots. Later and in a broad study focused on the roots of the declining poverty and inequality, Vargas and Garriga (2015) also found this pattern using both Mincer approach and with a correction of the omission bias using the approach proposed by Heckman (1979). Moreover, they include a set of controls that helps to isolate education returns.

Using more actual data from the same source, Andersen (2016) also found a similar drop from 11% in 1999 to nearly 4% in 2014. She highlighted the fact that returns were statistically above 0% even for the low return sector (transportation). Besides, she first suggested that linearity in the Mincer equation was absent in 2014. More interestingly, she provided some possible reasons for this decline. On the supply side, she attributed it to the expansion of

---

1   Portion of the song 'Wasted years' from Iron Maiden in the album *Somewhere in Time* (1986)

skilled people mainly in the last years, while on the demand side she considered increased demand of unskilled people given the boom in some commodities' markets.

In what follows I will analyze the returns to education using both the standard 2015 household survey and the Demand for Skills Survey carried by the IADB in similar months. I found that even the standard specification is linear in the (transformed) variables, returns could vary according to the particular year of schooling and some milestones related to the end of specifics education stages. I also use the theoretical framework developed by Bobba *et al.* (2018) to set the econometric estimation of a multinomial logit model with four choices to estimate earnings equation for self-employed, informal workers and formal workers.

In that sense, I found evidence supporting a Mincer equation with non-linearities in the Bolivian case as was suggested for the US case by Card and Lemieux (2001) and Heckman *et al.* (2006).

The paper is organized as follows. After this introduction, I will discuss the data used in this analysis focusing on the characteristics related to the estimation. Then I will present some Mincerian estimates to discuss the underlying non-linearity. Finally, I will try to explain the results focusing on the role of informality and segmented labor markets.

2. Data and methodology

Most of the econometric analysis carried in this paper rest on the the 2015 annual household survey (best known as 'Encuesta de Hogares 2015' in Spanish) which is a yearly assessment carried by INE[2]. I will discuss its main characteristics, specially the ones related to the econometric estimation.

Regarding education, the survey includes years of education and the highest level attained. But one of the measurement problems is that Bolivia has experienced at least three partial reforms in educational levels in the last 20 years. So definition of primary and secondary education vary. Moreover, there is not a technical analysis of the differences in the syllabus of every reform.

---

2   More technical details of this survey are found at a the INE: Household Metadata. Even though INE surveys after 2016 are available, I will use the 2015 version because it was collected in similar months as the IADB survey that I will describe later in this section.

Concerning occupation, it is worth noting that INE considers a person employed if he/she worked at least one hour in the week before the survey. It also covers people aged 7 years and older given the extent of child work, mainly in rural areas.

For our purposes, it is important to mention that omission bias could be avoided directly from the survey because an unemployed person is asked for reasons not to look for a job, including unfruitful previous search, perceived and efective discouragement and expectations of better offers.

It also includes questions on weekly days and daily hours in the job. Unfortunately, it does not have a question on total years of experience, just current job experience. Because of a large extent of informal sector and self-employment, it has a lot of questions on primary and secondary occupation.

The vast size of informal sector, around 60% according to Velasco (2015), represents a challenge for a Mincerian estimation. In fact, this kind of estimation is used as wage equation, while in the Bolivian case it involves both wage and 'petty entrepreneurs' income, as it is noted by Villarroel *et al.* (2012).

In fact, 2015 survey includes 14,630 observations on labor income. But there were just 6,882 employees in the strictus sensus, because the remaining cases belong to entrepreneurs, self-employed people or even apprentices. And if we consider those who contribute to the pension system there are just 2,802 observations.

Thus the results of a Mincerian equation of the whole sample must be taken with caution, because they do not represent a wage equation, but a reduced-form earnings equation.

To avoid these risks, I add more control variables to the main equation in order to isolate the effect of education on 'labor' income, following the approach of Vargas and Garriga (2015).

Additionally, I employed data from a survey carried out on skills. In 2012 the Inter-American Development Bank released a report titled 'Disconnected: Skills, Education, and Employment in Latin America' in Bassi *et al.* (2012). The main concern of this volume was to

show the disconnection between the labor market and the educational and training system in the region.

One of the chapters of that book was devoted to understand the concernings of the employers regarding skills and general and specific abilities of the workers. So, the IADB conducted a survey in Argentina, Brazil and Chile from around 1,200 firms.

The survey was a 'Demand for Skills Survey' (DSS), a standard instrument for a better understanding of labor demand, mainly regarding some characteristics considered as useful for firms in order to have higher productivity. Besides, the IADB could estimate a skill gap between supply and demand in the labor market.

Five years later, the Bolivian government agreed to carry out this survey in the main three cities of the country (La Paz, Cochabamba and Santa Cruz). A general discussion about the results can be found in Urquidi (2015).

Bolivian DSS was conducted between January and April of 2015 to 1,831 firms, most of them SME, while a quarter of the sample was devoted to the medium and big firms. It was conducted following the guidelines of the IADB to have comparable results with the 2012 report. It was in charge of the Information and Statistic Collection Center at the Private University of Bolivia (CEGIE).

The questionnaire comprises around 90 questions, including a) detailed list of main jobs specifying gender, age, experience and wage for each item; b) effects of labor regulations; c) skills assessment of the three main non-administrative jobs; and d) training.

Given the length of the survey, just 460 medium and big firms answered the section devoted to the skills assessment. Thus the results are biased to them. However, given that twenty jobs inside the firm are assessed we have potential answers for more than 4,000 employees.

Regarding the methodology, I begin with some basic explorations on the nature of the relationship among schooling and earnings, showing some signs of a non-linear association.

Then I estimate a reduced form of earnings equation controling for other factors as ethnicity, gender, urbanization, among others in two forms: simple OLS technique and omision bias corrected regression.

After that, I employ a segmented labor approach to understad in which segment (unemployed, self-employed, informarl worker, or formal) this non-linear relationship is more relevant.

Finally, I use the IADB data corrected with a matching algorithm to estimate this earnings equation with information from firms rather than workers. This procedure give us a detailed view of the main determinants of earnings and the effect of informality[3].

3. A simple and preliminary exploration of effects of education on earnings

Before the estimation, I made an analysis of years of experience according to the highest educational level attained by individuals. To simplify the analysis and avoid the wide range of earnings, I use the median income for each year of experience.

As Figure 1 shows, there is a similar pattern among median income and experience for each educational level, a non-linear shape that is clearer for the whole data. Then the quadratic form of experience seems appropriate for this set of data.

Following a similar approach, I calculated some statistics of (log) hourly earnings according to educational levels to find out whether different returns to education emerge previous to regression analysis (Table 1). It seems obvious that a higher educational level is associated to a higher income although high variability of earnings for each level do not allow us to reject the alternate hypothesis of different mean earnings between reported levels.

---

3   These steps are discussed more deeply in subsequent sections.

**Figure 1: Median income for each year of experience for people
with secondary, tertiary, technical education and whole data**
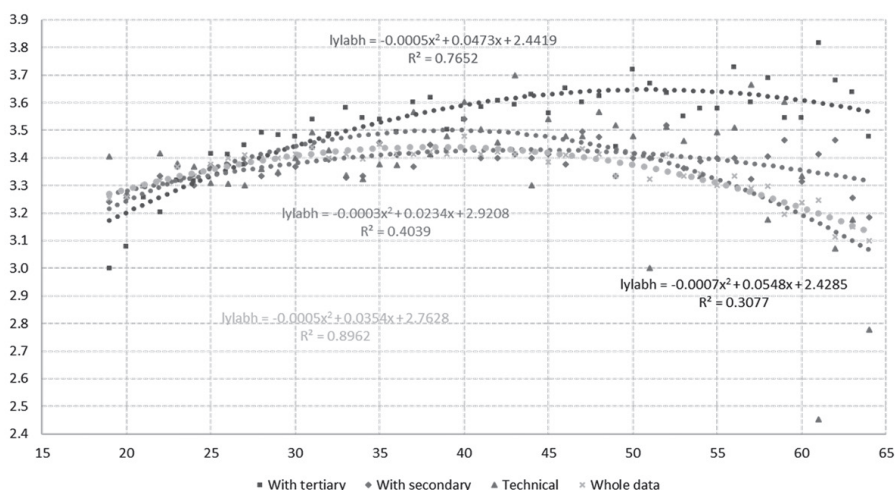
Source: Own calculations based on Encuesta de Hogares 2015.

**Table 1
Log of hourly earnings by educational level**

Educational level	Mean	Std. Dev	Freq. (weighted)	Relative Freq
None	1.0548	1.6709	215,265	5.1%
Incomplete primary	1.3650	1.9006	35,990	0.9%
Complete primary	1.7789	1.3053	1,127,916	26.9%
Incomplete secondary	2.1549	1.15593	543,141	12.9%
Complete secondary	2.2912	1.1005	1,206,323	28.7%
Tertiary	2.7912	1.1545	1,071,442	25.5%
Other	1.8298	1.3268	541	-

Source: Own calculations based on Encuesta de Hogares 2015.

Letting aside this detail, it could be inferred that primary education provides an annual average return of 12% compared to no schooling. Following similar criterion, secondary education adds an annual mean return of 8.5%. Finally, tertiary education represents an average return of 10%. It must be noted that this analysis does not consider the experience factor or other relevant control variables.

The next step is to proceed with the estimation of a Mincerian earning equation. In its simple form, Mincer equation is a relationship between hourly earnings, years of education and a quadratic expression related to experience. The result of the simplest case is the following:[4]

$$\widehat{\text{lylabh}} = \varphi + \underset{(0.0033)}{0.0775} \times edu + \underset{(0.0028)}{0.0340} \times exp - \underset{(6e\text{-}005)}{0.0006} \times exp^2$$

$$N = 13397 \quad \overline{R}^2 = 0.118 \quad \hat{\sigma} = 1.1665 \tag{1}$$

$$(\text{standard errors in parentheses})$$

So, this simplest form suggests an annual mean return of schooling of 7.8%, above other estimates, although the exclusion of other control variables clearly gives biased estimates of this parameter. In the case of experience, this initial estimation suggests a concave function of years of experience. Earnings increase until they reach a maximum at 27 years of experience independent of schooling years.

Also, I will estimate as a preliminary exercise a polynomial version of the previous equation to look for non-linearity on returns of education. As I will explain in the next section, I am using a 4th degree polynomial:

$$\widehat{\text{lylabh}} = \varphi + \underset{(0.0412)}{0.16815} \times edu + \underset{(0.0031)}{0.0334} \times exp - \underset{(0.0006)}{0.0007} \times exp^2$$

$$- \underset{(0.0072)}{0.0243} \times edu^2 + \underset{(0.0005)}{0.0018} \times edu^3 - \underset{(0.00001)}{0.0007} \times edu^4 \tag{2}$$

$$N = 12969 \quad \overline{R}^2 = 0.1249 \quad \hat{\sigma} = 1.1565$$

$$(\text{standard errors in parentheses})$$

To obtain the average return of education for the whole sample, consider a vector x_j with a sequence from 0 to 25, where the element $x_k \in [0, 25]$ is a specific year of education, $\hat{\alpha}$ is the vector of coefficients showed in the former equation and w is a vector of

---

populational weights that ponders people with k years of schooling. Additionally, the vector $k = \begin{pmatrix} k & k^2 & k^3 & k^4 \end{pmatrix}$ is part of the matrix $K = k' \otimes I$

Then the average return is given by equation (3):

$$\hat{r} = \sum_{k=0}^{24} w_k \times \hat{\alpha}^k k = w'\left(\hat{\alpha}^k K\right) \tag{3}$$

Using this equation, the weighted average return of education is 7.9%, which does not differ from the linear form. However, we will see later that it implies different rates for each educational level.

4. Standard single equation estimation of earnings equations

To isolate with more accuracy the effect of education, I have regressed the hourly earnings according to this specification:

$$\ln\left(\frac{y_i}{h_i}\right) = \alpha + \beta sch_i + \gamma_0 \exp_i + \gamma_1 \left(\exp_i\right)^2 + \psi' x_i + u_i \tag{4}$$

Ergo, I have added many other variables that could affect earnings levels with results shown in Table 2. Most of them are included as a 'slope effect' or as a dummy variable multiplied by schooling, rather than just a dichotomous variable for intercept effect. I will explain the most relevant regarding the purpose of the estimation.

One key variable is informality. In fact, formal/informal categories were defined as follows: formal with long-term social security (pension system) and short-term (medical insurance); two categories of semi-formal with or long-term or short-term social security; and informal to those who do not have neither long nor short term social security.

To avoid omission bias I started with the inclusion of a large set of variables given more than enough available degrees of freedom. Then I have left just the significant variables. Using the previous approach, I calculated that the average return to education is around 2%.

Table 2
Estimate of linear multiple control Mincer equation
OLS estimates ($n = 13862$)
Dependent variable: **lylabh**
(Heteroskedasticity-robust standard errors)

|	Coefficient	Std. Error	t-ratio	p-value
Intercept	1.89577	0.0739969	25.62	0.0000
Education				
Education (linear form)	0.0217956	0.00425206	5.126	0.0000
Work experience				
Experience	0.0298097	0.00256651	11.61	0.00000
Experience squared	−0.000462078	3.94455e−005	−11.71	0.0000
Occupational category				
Board	0.385090	0.0557273	6.910	0.0000
Professional	0.330586	0.0353976	9.339	0.0000
Technician	0.109476	0.0462356	2.368	0.0179
Agricultural	−0.861809	0.0436431	−19.75	0.0000
Gender				
Slope effect	−0.0120659	0.00395049	−3.054	0.0023
Intercept effect	0.359479	0.0474868	7.570	0.0000
Formality (slope)				
Formal	0.0215399	0.00205583	10.48	0.0000
Semi formal (social security)	0.00831508	0.00253728	3.277	0.0011
Informal	−0.00938036	0.00233991	−4.009	0.0001
Area				
Urban	−0.141810	0.0303211	−4.677	0.0000
Economic sector (slope)				
Mining	0.0240757	0.00404876	5.946	0.0000
Industry	−0.0101920	0.00213115	−4.782	0.0000
Commerce	−0.0229826	0.00252443	−9.104	0.0000
Region (slope effect)				
Chuquisaca	−0.0111043	0.00420093	−2.643	0.0082
La Paz	−0.00543014	0.00169869	−3.197	0.0014
Oruro	−0.0172941	0.00394328	−4.386	0.0000
Potosi	−0.0212310	0.00410886	−5.167	0.0000
Etnicity				
Indigenous (slope)	0.0224272	0.00478657	4.685	0.0000
Indigenous (intercept)	−0.310689	0.0483729	−6.423	0.0000
Other				
Migration to SCZ (slope)	0.0140921	0.00561538	2.510	0.0121

Source: Own calculations based on Encuesta de Hogares 2015.

To analyze with more detail the effect of education, I removed the effects of control variables except for education. Thus, I computed the conditional income. Then I calculated the mean for each year of education. A clear pattern of non-linearity also arose showing different rates of return for each year, shown in Figure 2.

$$\overline{\text{lylabh}} = \varphi + 0.945 \times \overline{edu} - 0.0947 \times \overline{edu}^2 + 0.0062 \times \overline{edu}^3 - 0.0001 \times \overline{edu}^4$$
$$N = 25 \quad \overline{R}^2 = 0.9775$$

(5)

Figure 2: Conditional mean income and years of education
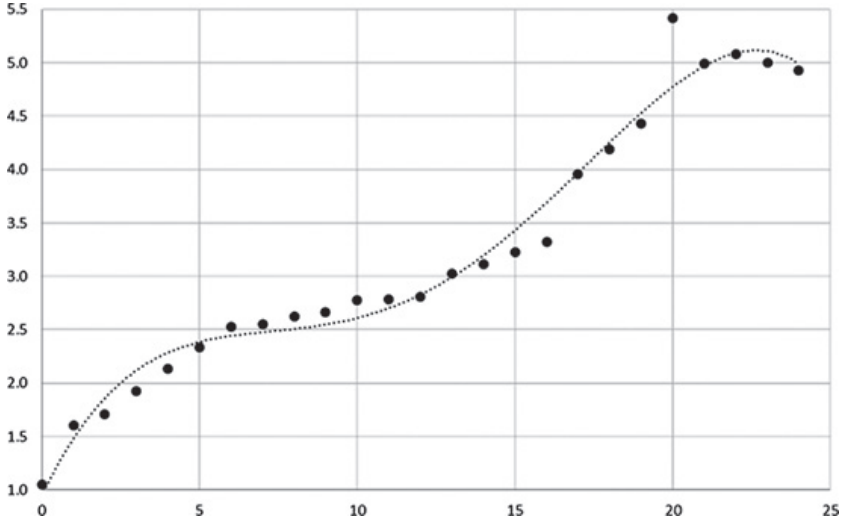
Source: Own calculations based on Encuesta de Hogares 2015.

Then I included this polynomial form in the whole equation also corrected by the weights provided by INE to extrapolate results to whole population, the results are reported in Table 3. These results are also consistent with an average return of 2.3%, slightly above regarding the linear effect of education reported in Table 2.

Nevertheless, the pattern differs according to each year of education. As it is showed in Figure 3, year 1 represents a marginal return of 7.2%, probably explained by the effects of writing and reading skills.

Then the marginal return decreases until the first year of secondary education, where there is an inflection point. So, returns to education gradually increase reaching their local maximal around the end of tertiary education. Finally, graduate education is just profitable on

average for short courses. Anyway, cumulative returns differ in statistical terms just for tertiary education given the variation within a year, even controlling for other variables.

Table 3
Estimate of multiple control non-linear Mincer equation

Weighted OLS estimates ($n = 13862$)
Dependent variable: **lylabh**
(Heteroskedasticity-robust standard errors)

	Coefficient	Std. Error	t-ratio	p-value
Intercept	1.89337	0.104263	18.16	0.0000
Education				
Education	0.0922082	0.0384059	2.401	0.0164
Education squared	−0.0154355	0.00681914	−2.264	0.0236
Education cubed	0.00111844	0.000476762	2.346	0.0190
Education to the four	−2.54782e−005	1.12915e−005	−2.256	0.0241
Work experience				
Experience	0.0270222	0.00379487	7.121	0.0000
Experience squared	−0.000424753	6.36239e−005	−6.676	0.0000
Occupational category				
Board	0.351987	0.0621866	5.660	0.0000
Professional	0.269781	0.0540348	4.993	0.0000
Agricultural	−0.880443	0.0490456	−17.95	0.0000
Gender				
Slope effect	−0.0157492	0.00512113	−3.075	0.0021
Intercept effect	0.412047	0.0613735	6.714	0.0000
Formality (slope)				
Formal	0.0221515	0.00247408	8.953	0.0000
Semi formal (social security)	0.00804239	0.00322162	2.496	0.0126
Informal	−0.00854888	0.00293539	−2.912	0.0036
Area				
Urban	−0.142061	0.0332578	−4.272	0.0000
Economic sector (slope)				
Mining	0.0231564	0.00473534	4.890	0.0000
Industry	−0.0105637	0.00246989	−4.277	0.0000
Commerce	−0.0229858	0.00298396	−7.703	0.0000
Region (slope effect)				
Chuquisaca	−0.0135982	0.00497800	−2.732	0.0063
La Paz	−0.00522771	0.00207051	−2.525	0.0116
Oruro	−0.0170713	0.00471737	−3.619	0.0003
Potosi	−0.0195647	0.00444608	−4.400	0.0000
Etnicity				
Indigenous (slope)	0.0300788	0.00565018	5.324	0.0000
Indigenous (intercept)	−0.418491	0.0568813	−7.357	0.0000
Other				
Migration to SCZ (slope)	0.0121918	0.00593281	2.055	0.0399

Source: Own calculations based on Encuesta de Hogares 2015.

Figure 3: Average and marginal return to education for each year of education
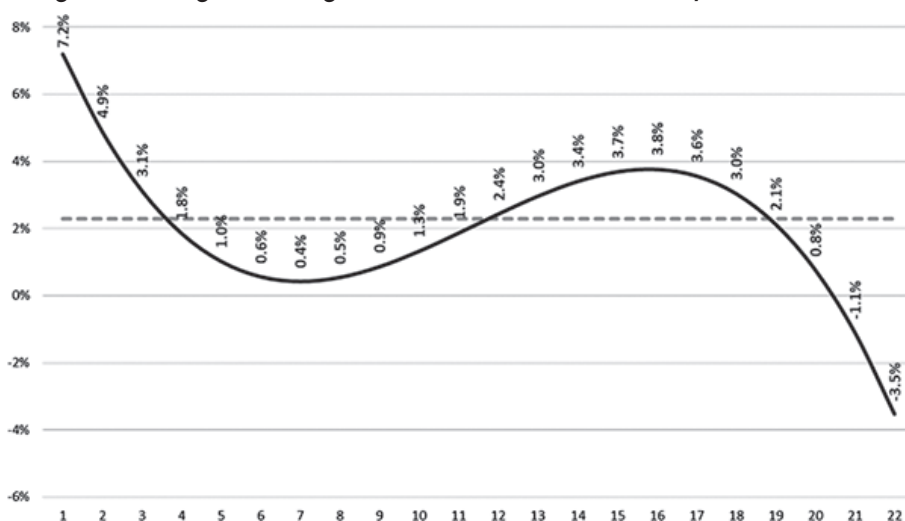
Source: Own calculations based on Encuesta de Hogares 2015.

Other remarkable results related to schooling returns are:

- For each year of education, a formal worker has an additional 2% of return, a semi-formal 1% and an informal -1%.

- For each year of education, commerce represents 2% less of return, industry 1% less and mining 3% more than the average.

- Returns to education are lower than the national average in Chuquisaca (1%), Oruro (1%) and Potosí (2%) for each year of education.

- Migrants to Santa Cruz have 2% more of returns for every year of education.

Finally, I corrected the estimation with the 'selection bias' according to Heckman (1979). I employed a logit model of selection with gender, marital status, ethnicity, and the position of the worker in the family. Then the estimation included the inverse Mills ratio to deliver unbiased estimates, shown in Table 4. Compared with the OLS estimates, it implies an average return of 2.8%, mildly above the linear and non-linear effects of education reported previously.

Table 4
Sample selection corrected non-linear Mincer equation

Heckit estimates ($n = 18422$)
Dependent variable: **lylabh**
(QML standard errors)

	Coefficient	Std. Error	t-ratio	p-value
Intercept	2.05983	0.0996409	20.67	0.0000
Education				
Education	0.112756	0.0270288	4.172	0.0000
Education squared	−0.0178563	0.00488351	−3.656	0.0003
Education cubed	0.00124924	0.000339883	3.675	0.0002
Education to the four	−2.79493e−005	7.90184e−006	−3.537	0.0004
Work experience				
Experience	0.0241904	0.00308050	7.853	0.0000
Experience squared	−0.000390311	4.80731e−005	−8.119	0.0000
Occupational category				
Board	0.322287	0.0603760	5.338	0.0000
Professional	0.271593	0.0413589	6.567	0.0000
Agricultural	−0.849612	0.0369992	−22.96	0.0000
Gender				
Slope effect	−0.0126229	0.00379830	−3.323	0.0009
Intercept effect	0.233687	0.0531244	4.399	0.0000
Formality (slope)				
Formal	0.0216581	0.00276421	7.835	0.0000
Semi formal (social security)	0.00852058	0.00285439	2.985	0.0028
Informal	−0.00749973	0.00218753	−3.428	0.0006
Area				
Urban	−0.141334	0.0307461	−4.597	0.0000
Economic sector (slope)				
Mining	0.0227316	0.00635700	3.576	0.0003
Industry	−0.00966453	0.00256555	−3.767	0.0002
Commerce	−0.0229621	0.00226633	−10.13	0.0000
Region (slope effect)				
Chuquisaca	−0.0124761	0.00422717	−2.951	0.0032
La Paz	−0.00591073	0.00181014	−3.265	0.0011
Oruro	−0.0177939	0.00395394	−4.500	0.0000
Potosi	−0.0198026	0.00470370	−4.210	0.0000
Etnicity				
Indigenous (slope)	0.0169088	0.00484014	3.493	0.0005
Indigenous (intercept)	−0.260031	0.0471819	−5.511	0.0000
Sample selection parameter				
λ	−0.265455	0.0617428	−4.299	0.0000

Source: Own calculations based on Encuesta de Hogares 2015.

Besides, it points out that unobservable characteristics of workers imply higher probability of selection for employment if they are highly skilled, according to the interpretation suggested in Narayanan (2015).

The pattern of marginal effect of an additional year of schooling on earnings is like the non-linear estimate but with higher marginal effects at the basic levels of education, as shown in Figure 4.

Figure 4: Marginal return to education for each year of education
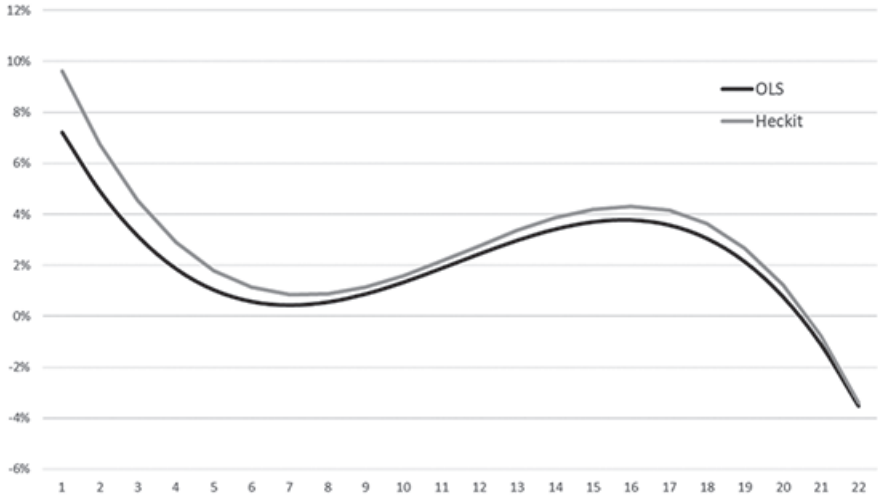
Source: Own calculations based on Encuesta de Hogares 2015.

To summarize this section, if we analyze returns to education with the standard approach of a single equation estimation, a non-linear approach seems better than the linear one controlling by other factors. The temporal pattern of marginal returns points at least two peaks: one associated to basic reading and writing skill and other to complete undergraduate studies. In the next section we will explore a more suitable approach for segmented markets.

5. A segmented labor market approach

Even the insights of the previous section and a more elaborated version with the sample selection technique could be misleading because Bolivian labor market is not similar to the ones of other developed countries.

In fact, and following Bobba *et al.* (2018), we can consider that the Bolivian labor market has four segments:

- Formal employees

- Informal employees

- Self-employed

- Unemployed

Then a broader approach must include both the matching and the search processes involved in this estimation procedure.

Table 5
Multinomial logistic regression

	Multinomial logit			
	Dependent variable **Labor status**			
	(n=13297)			
	Unemployed	Self-employed	Informal employee	Formal employee
Age	0	0.359*	0.115	0.778***
	(.)	(1.98)	(0.81)	(3.65)
Schooling	0	-0.342	-0.146	-0.509*
	(.)	(-1.89)	(-1.03)	(-2.39)
Gender	0	0.540***	0.863***	0.559***
	(.)	(6.37)	(10.00)	(6.29)
Experience	0	-0.273	-0.0975	-0.697**
	(.)	(-1.50)	(-0.69)	(-3.26)
(Intercept)	0	-1.984	0.586	-8.045***
	(.)	(-1.83)	(0.69)	(-6.27)
Population	185,179	2.068,745	955,560	882,678
Sample	677	6,905	3,484	3,254

Source: Own calculations based on Encuesta de Hogares 2015.

Even the cited authors employ the Simulated Method of Moments to the complete model estimation, I follow their logic to estimate a multi equation model in two stages. So, I employed the multinomial or categorical response model described in a general approach in Cramer (2003) and with more details in Glewwe (1992) and implemented for a similar case in Gunther and Launov (2006), I estimated this model for these four categories (formal, informal, self-employed and unemployed) taking the last one as the baseline model. Results are shown in Table 5.

With these estimates, I calculated three earnings equations whose results are available in Table 6. Selection bias is only present in the informal earnings equation, where the implied negative relationship (*i.e.*, a positive sign of the correction term) could mean that a higher (lower) productive worker with a higher (lower) chance to be selected in this market has a lower (higher) wage. So, we have a plausible explanation of the low returns due an adverse selection phenomenon in informal market due to information asymmetry as pointed out by Narayanan (2015).

I must mention that the weighted average return of schooling is lower in the self-employed group (6.6%) than informal employees (10.1%) and formal ones (14.9%).

6. Returns according to labor demand from firms

Previous earnings estimates come from different kinds of households. On the other hand, IADB survey is focused in formal firms.

To address the compatibility between these databases I calculated the Mahalanobis distance between them. The results show that taking the whole INE database, IADB is more concentrated around the upper income and education segments. If we focus only on formal sector, there is a clear intersection among them as it is suggested in Figure 5.

To assess this feature, I employed the nearest neighbor matching algorithm based on the Mahalanobis distance taking into account some set of information as age, education, region and gender, leaving aside the income, following the guidelines of Caliendo and Kopeinig (2008). After picking the nearest neighbor, I compared the densities of (log) incomes and carried a Kruskal Wallis test, not rejecting the hypothesis that they come from a similar distribution (Figure 6).

Consequently, I used the micro-data of this survey to try to consider the effects of schooling and informality of labor demand in the Bolivian case but with primary data from the firms, rather than the usual approach of using the Bolivian annual household survey to have an earnings equation.

Table 6
Segmented labor market earnings equations

|	(1) Self-employed	(2) Informal employee	(3) Formal employee
Education	0.0661***	0.101***	0.416***
	(6.00)	(7.51)	(3.79)
Education (squared)			-0.0562***
			(-5.95)
Education (cubic)			0.00386***
			(6.33)
Education (at fourth)			-0.0000842***
			(-6.13)
Experience	0.0105	0.0486***	0.0469***
	(0.92)	(6.42)	(4.50)
Squared exp.	-0.000403***		-0.000524***
	(-4.19)		(-5.74)
Gender	0.0166***		0.00442
	(4.28)		(1.73)
λ_2	0.147		
	(1.94)		
λ_3		0.283***	
		(5.35)	
λ_4			-0.130
			(-1.17)
(Intercept)	1.916***	1.505***	-0.702
	(4.83)	(15.09)	(-0.38)
n	6563	3480	3115

Source: Own calculations based on Encuesta de Hogares 2015.

**Figure 5: Comparison of IADB and INE surveys in the
log(income) and education space**
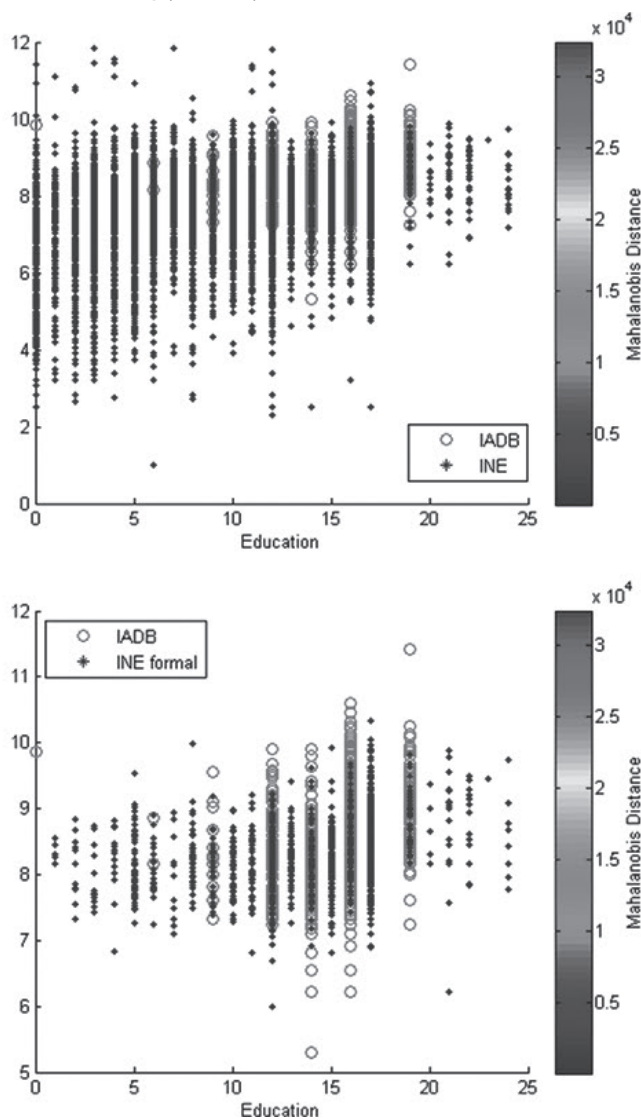
Source: Own calculations based on INE and IADB surveys.

Figure 6: Densities of IADB and INE matched observations and Kruskal Wallis test
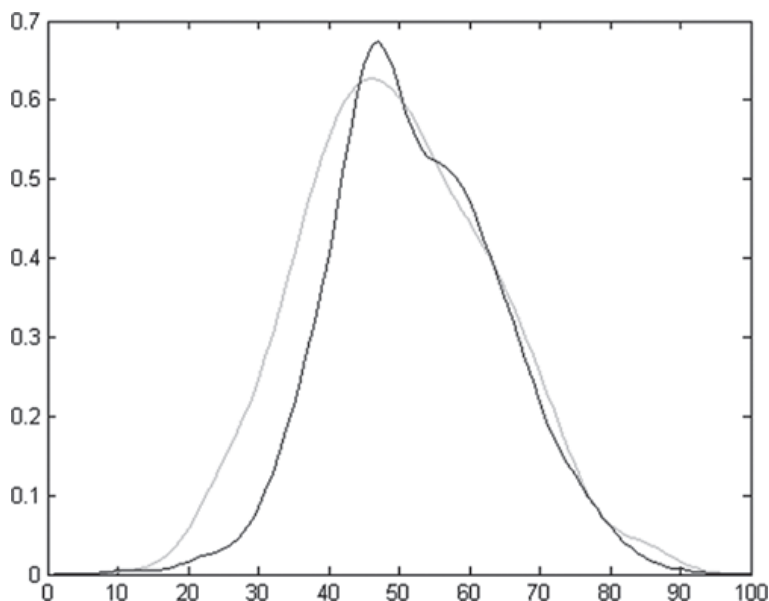
Source	SS	Df	MS	2	Prob>2
Columns	59282.4	1	59282.4	0.81	0.3684
Error	68599819.1	936	73290.4		
Total	68659101.5	937			

Source: Own calculations based on INE and IADB surveys.

Then I used the Heckman approach to estimate both linear and non-linear equations with IADB information (Table 7). Two clear patterns arose: i) the statistical no significance of the factor associated to the selection bias, and ii) a nonlinear shape of the returns to schooling in the formal sector[5]. Besides, weighted average is around 12%, close to the mean return of 15% of the multi-equation approach of the previous section.

To assess the gap between formal and informal workers, I must mention that the matching process described above delivered 72 cases (near 8%) of matched pairs between formal and

---

5   But the last insight must be taken with caution because less than one percent of workers have incomplete secondary. So the return of near 400% of the first degree of schooling seems unreliable in terms of a highly distorted non-linear pattern and the low degree of representativeness of low schooled people.

informal workers. A simple test about the average difference delivered that on average, a formal worker with similar characteristics of an informal has an income 41% higher than the last one.

Table 7
Selection sample correct estimation of formal labor demand

	(1) Linear		(2) Non-linear	
	(log) hourly income	Selection	(log) hourly income	Selection
Experience	0.0248***	-0.0156*	0.0239 ***	-0.0157*
	(8.82)	(-2.11)	(8.51)	(-2.12)
Experience (sqd)	-0.000267***		-0.000251 ***	
	(-4.26)		(-4.00)	
Gender	0.0728***		0.0670 ***	
	(4.26)		(3.92)	
Dummy CBA	0.133***		0.130 ***	
	(6.53)		(6.34)	
Dummy SCZ	0.260***		0.250 ***	
	(13.39)		(12.79)	
Professional	-0.517***	1.348***	-0.52 8***	1.347***
	(-23.84)	(4.51)	(-24.17)	(4.51)
Consultant	-0.792***	1.389***	-0.70 6***	1.389***
	(-27.72)	(4.18)	(-21.69)	(4.18)
Education	0.0878***	0.0817	4.163 **	0.0814
	(13.52)	(1.95)	(2.98)	(1.95)
Education squared			-0.531 **	
			(-3.27)	
Education cubic			0.0289 ***	
			(3.55)	
Education at fourth			-0.000562 ***	
			(-3.78)	
(Intercept)	3.233***	0.880	-7.541	0.885
	(28.54)	(1.16)	(-1.72)	(1.17)
ρ	0.0262		0.0230	
	(0.22)		(0.20)	
$log(\sigma)$		-0.741***		-0.745***
		(-66.37)		(-66.71)
n		4052		

Source: Own calculations based on IADB 2015.

To refine this result, I ran a regression with a slope effect with the following results:

$$\widehat{\text{lylab}} = \underset{(0.14320)}{7.47622} + \underset{(0.010112)}{0.0345848} \times \text{education} + \underset{(0.0047751)}{0.0252685} \times \text{formal gap}$$

$$n = 938 \quad \bar{R}^2 = 0.0715 \quad \hat{\sigma} = 0.58256 \tag{6}$$

(standard errors in parentheses)

As it is implied in Figure 7, the average return gap among formal and informal earnings varies between 2% and 4% for every additional year of schooling. This is coherent with this gap of near 40% mentioned previously

Figure 7: Confidence ellipse between formal and informal matched workers
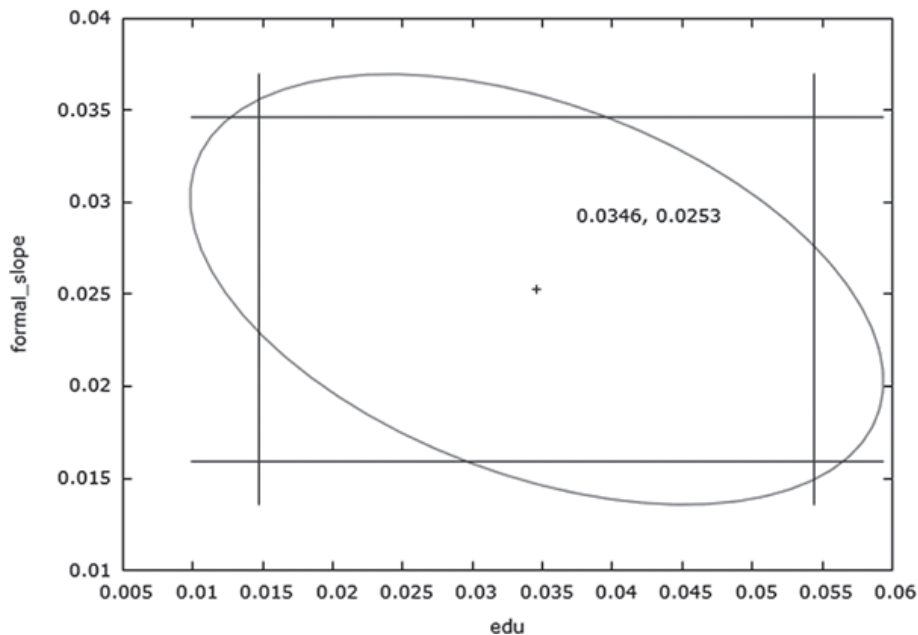
Source: Own calculations based on INE and IADB surveys.

7.   Concluding remarks

I showed that the puzzle of low average returns could be explained by non-linearity in the earnings equation. Specifically, it could be noted that acquiring a college degree or a university degree could give higher income on average.

But with a more careful analysis, I hypothesized that this pattern is due to a segmented labor market where informal and self-employed workers have lower returns than formal ones. Even more, people with similar age, education, gender, and sector have lower returns on schooling.

Even though this finding seems intuitive given the nature of informal markets, further research is needed. The simplest way is to run a non-parametric regression to capture better this nonlinear relationship, while an advanced one is to find the roots behind this earnings distribution and its relationship with schooling is found in Bobba *et al.* (2018).

They built an economic model with segmented markets, where the optimal schooling decision was made before entering the labor market. So, they discuss the distortions in that deliver different returns.

Future research could estimate the model for the Bolivian economy, as they for the Mexican economy. This would give more insights on the issue of why education could be wasted years for a huge part of Bolivian population.

References

1. Andersen, L. (2016, June 13). Does education pay in Bolivia? *Inesad Development Roast.* https://inesad.edu.bo/developmentroast/2016/06/does-education-pay-in-bolivia/

2. Bassi, M., Busso, M., Urzua, S. & Vargas, J. (2012) *Disconnected: Skills, Education, and Employment in Latin America.* Inter-American Development Bank. https://publications. iadb.org/en/disconnected-skills-education-and-employment-latin-america

3. Bobba, M., Flabbi, L. & Levy, S. (2018). *Labor Market Search, Informality and Schooling Investments* (IDB Working Paper Series Nº 863). https://publications.iadb.org/ publications/english/document/Labor-Market-Search-Informality-and-Schooling-Investments.pdf

4. Caliendo, M. & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys, 22(1),* 31-72. https://doi. org/10.1111/j.1467-6419.2007.00527.x

5. Card, D. & Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics,* 116*(2),* 705-746. https://www.jstor.org/stable/2696477

6. Cramer, J. S. (2003) *Logit Models from Economics and other Fields.* Cambridge University Press.

7. Glewwe, P. (1992). The Three-Choice Multinomial Probit With Selectivity Corrections. *Econometric Theory,* 8*(2),* 302-304. https://www.jstor.org/stable/3532448

8. Gunther, I. & Launov, A. (2006). *Competitive and Segmented Informal Labor Markets* (IZA Discussion Papers N° 2349). https://docs.iza.org/dp2349.pdf

9. Heckman, J., Lochner, L. & Todd. P. (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. Hanushek & F. Welch (eds.), *Handbook of the Economics of Education* (pp. 307-458). Elsevier. https://doi. org/10.1016/S1574-0692(06)01007-5

10. Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica,* 47*(1),* 153-161. https://www.jstor.org/stable/1912352

11. Mincer, J. (1974). *Schooling, Experience, and Earnings.* National Bureau of Economic Research. https://econpapers.repec.org/bookchap/nbrnberbk/minc74-1.htm

12. Narayanan, A. (2015). Informal employment in India: Voluntary choice or a result of labor market segmentation? *Indian Journal of Labour Economics,* 58*(1),* 119-167. https://doi.org/10.1007/s41027-015-0009-9

13. Solon, G., Haider, S. & Wooldridge, J. (2015). What Are We Weighting For? *Journal of Human Resources,* 50*(2),* 301-316. https://doi.org/10.3368/jhr.50.2.301

14. Urquidi, M. (2015). ¿Es suficiente estudiar para lograr un buen trabajo? *Factor Trabajo.* https://goo.gl/yk4yGM

15. Vargas, J.P.M. & Garriga, S. (2015). *Explaining inequality and poverty reduction in Bolivia* (IMF Working Papers 15/265). https://www.imf.org/external/pubs/ft/wp/2015/wp15265.pdf

16. Velasco, D. (2015). Informal Economy in Bolivia: Analysis, Evaluation and Quantification Based on Cash Monetary Effective Demand Approach (period 1994-2014). *Investigación & Desarrollo,* 15*(2),* 76-89. https://www.upb.edu/sites/default/files/adjuntos/7-I&D215-Velasco-Abs_0.pdf

17. Villarroel, P. & Hernani, W. (2011). *¿La educación todavía paga? La disminución de retornos a la educación en la Bolivia urbana* (Technical report). Fundación ARU.

18. Villarroel, P., Hernani, W. & Eid, A. (2012). *Entrepreneurship and economic mobility. a case of study of Bolivia* (Technical report). Fundación ARU.