Central European University
Advanced Labor Economics
Replication Paper

The Impact of Education on the Employment of Young Males in Kyrgyzstan

by

Chyngyz Davletov

Professors: John S. Earle and Almos M. Telegdy

April 16, 2016

ABSTRACT

This paper analyses the effect of education on the employment of young males in Kyrgyzstan. The data was taken from the "Life in Kyrgyzstan" panel survey that covers 2010–2012 years and almost 14,000 households. Heckman probit estimation was applied in order to control for the possible selection bias in the decision to study. Expected lifetime earnings and opportunity cost of education were used as instruments in estimating the selection equation of Heckman probit. Results suggest the absence of statistically significant impact of education on employment.

CONTENTS

LIST OF TABLES

1 INTRODUCTION

The purpose of this work is to examine the effect of education on the employment of young males living in a transitional economy. Kyrgyzstan surely belongs to such countries. After the collapse of the Soviet Union the country has experienced difficult times both in terms of politics and economics. Government debt keeps growing despite the help from many countries and international organizations facilitating economic growth and reduction of poverty such as, for example, IMF or EBRD. The issues that the country keeps facing have a direct impact on the life of its population, especially young males, on behalf of whom lies the hope for the recovery and bright future for the coming new generations. In the light of the recent global crisis youth un/employment became a problem even in much healthier countries. This fact is reflected in a number of recent research involving investigations in this field. For example, Marelli and Vakulenko (2014) performed a structural analysis of youth unemployment in Italy, Russia and analysed aggregate trends in some other EU countries involving both micro- and macroeconomic determinants of unemployment such as personal characteristics, living in urban/rural areas, regional unemployment rate, and others, and applying Heckman probit estimation procedure. They found *inter alia* that the probability to be unemployed decreases the most with the household disposable income, age and marital status (switching from single to married). The role of vocational training is extensively analysed in the

paper of Biavaschi et al. (2012), where the authors suggest that vocational training may play a crucial role in the young people's transition into work. The importance of this variable in determining employment status of youth is taken into account in the research of Pastore (2012), which had inspired the current research. However, it was the methodology used in the work of Pastore that attracted the author of this text. In particular, the author used interesting instruments in the Heckman probit model to estimate the effect of education on the employment of young Poles. Those will be discussed in the next section with references to the source. The rest of the paper is organized as follows: section 2 introduces the data and discusses the methodology, section 3 presents the results, and section 4 concludes and makes objectives for further research.

2 DESCRIPTION OF DATA AND METHODOLOGY

2.1 Data and Variables

The data was taken from the "Life in Kyrgyzstan" panel survey of households under the research project "Economic Transformation, Behavior and Living Standards of Households in Central Asia: the Case of Kyrgyzstan" prepared by German Institute for Economic Research DIW Berlin (2016). This data set covers 2010–2012 years and almost 14,000 households in Kyrgyzstan. In order to avoid the numerous issues associated with the use of the panel data, the cross-section for the year 2012, being the most recent, was used in estimations.

The choice of the variables and sampling procedures in this work was aimed on getting the final sample that would be as similar as possible to that of Pastore (2012) keeping in mind the differences between the data sets. The set of variables and their description is given in Table 1.

In the original paper, Pastore defined employment in the following way: individuals are considered employed if they declare to have some type of paid work during the reference week, independent of whether they are also students or not; specific question identified whether an individual is a student, and the rest were considered jobless, where joblessness included both unemployment and inactivity. The author argues that unemployed and inactive individuals can be pooled together because in the case of young people they are often alike. In the current paper the same sampling was used. While for defining whether an individual is a student or not there was a single specific question, in order to define the employment state four questions, asking whether an individual did some work for governmental office, private business, farm, or other entity, were used, and an individual was considered employed if at least for the one of those the answer was yes. Also, the main sample included individuals aged between 15 and 30 to be consistent with that of Pastore.

It should be noted that, due to technical issues, the effect of the relationship of an individual to the head of the household was not included in the estimations. Besides the data set, this was another distinction between this and the replicated paper.

For determining the marital status of an individual the survey has 7 choices: 1 – married, 2 – divorced, 3 – living together, 4 – separated, 5 – widowed, 6 – single, and 7 – other. From this 7-fold definition a conventional binary marital status (see Table 1) was defined.

Level of education was measured in the survey by a discrete variable that represents eight different levels in an ascending order as shown in Table 1. This variable was not modified or changed in the estimation of the main equation of Heckman probit. However, when the selection equation was estimated, in order to calculate the expected lifetime earnings (ELE) for each age, the definition of the variable was slightly different, which will be discussed later in the section describing methodology and calculation of instruments.

Variables age and age-squared were used as proxies for experience in accordance with the replicated paper. Variables urban, total assets and total income (both measured in Kyrgyz National currency - Soms)

Table 1: Description of the Variables

Variable	Description
employed	1 – individual is employed, 0 – jobless
age	The age of an individual
age-squared	The square of an individual's age
marital status	1 – married/living together, 0 – single/divorced/separated
education	1 – illiterate, 2 – elementary, 3 – incomplete secondary, 4 – secondary, 5 – primary, 6 – high school, 7 – university, 8 – PhD
student	1 – if male is enrolled at an educational institution, 0 – if not
urban	1 – if a household is situated in an urban area, 0 – rural
total assets	The total assets of the household measured in Soms
total income	The total income of the household measured in Soms
Instruments	
Expected lifetime earnings from high school diploma and university degree, ELE	$ELE_i = \sum_{k=31}^{65} \frac{EE_k}{(1-r)^{k-age}}$, where: $i = 1$ (university degree), 2 (high school diploma); $k = 31, \ldots, 65$ is the age; EE_k – are the means of predicted wages for certain age k; r – is the average annual interest rate for 2012 in Kyrgyzstan.
Opportunity cost of education, OCE	(Probability to find employment by age)×(average monthly wage) or $\Pr(\text{employment}_j) \times \tilde{w}_j$ for $j = 15, \ldots, 30$ (see section 2.2, page 5)

Note: Interest rates for 2012 were downloaded from National Bank of Kyrgyz Republic (2016) and the average, r, is equal to 6.85%

were added to the set of predictors as well to give credit to the existing literature and examine whether they are important in determining employment.

It is also worth noting that there is another predictor of employment which is not included in the current paper. In the original work, the author included measures of local unemployment rate, which is not present in the data set that is used for this research. This can be an important distinction because the variable was found to have an important impact on the employment.

Table 2 presents the descriptive statistics for the final sample. The average age of males in the sample is around 22 years, with the square of 197. The low mean for education signals about the level of literacy of young males in Kyrgyzstan. Moreover, the ratio of males who are currently involved (at the reference week) in studying at an educational institution is quite low – $\approx$ 15%. At the same time the ratio of employed is very high of 72%. These raw figures suggest that, unfortunately, young males do not value the numerous benefits of higher education.

2.2 Heckman Probit Model

The main estimated equation by Heckman (1979) probit is the following (assuming the latent equation):

$$\text{employed}_i(\text{employed}_i^\star > 0) = X_i\beta + \xi_i,\ \xi_i \sim N(0,1), \tag{1}$$

which is not always observed. The selection equation is intended to correct for selection bias in the decision to study by estimating:

$$\text{student}_i(\text{student}_i > 0) = Z_i\gamma + \epsilon_i,\ \epsilon_i \sim N(0,1), \tag{2}$$

Table 2: Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
age	2,122	22.1	4.41	15	30
age-squared	2,122	508	197.34	225	900
educ	1,316	4.26	1.196	1	7
marital	2,120	.29	.454	0	1
student	1,316	.147	.355	0	1
employed	1,316	.72	.45	0	1
urban	2,122	.30	.46	0	1
total assets	2,122	1,279,177	924,622.7	99	8,080,474
total income	2,122	20,602.56	17,216.7	0	262,499
ELE-hs	2,122	3,634,219	2,158	3,633,716	3,644,057
ELE-uni	2,122	4,309,101	1,771	4,308,689	4,317,176
OCE	2,122	5,904.5	1,288	3,209	7,179

where Z_i includes all the variables presented in Table 1 and X_i includes the same variables, except for the instruments, for individual i. If $\text{corr}(\epsilon_i, \xi_i) \neq 0$, then the standard probit is biased. In contrast, Heckman probit is believed to control for the bias.

Instruments As mentioned above, variables ELE and OCE were used as instruments to correct for the selection bias. ELE is calculated for university and high school graduates using a sample of males aged 31–65. Small change was introduced in the definition of the level of education in this calculation, that is, individuals with education levels that correspond to 5 and 6 from table 1 were treated as high school graduates, and individuals with education levels of 7 and 8 – university graduates. The reason for this modification is that it allows to estimate the values of ELE according to the formula (see Table 1) more precisely from a larger sample. In the original paper, the author uses more sophisticated estimation of ELE, whereas in the current work it is calculated in the following, perhaps naive, way:

$$w_i = \alpha_0 + \alpha_1 \text{ age}_i + \alpha_2 \text{ age}_i^2 + \alpha_3 L_i + \eta_i, \tag{3}$$

where w_i is the monthly wage in Soms, L is either high school or university degree for individual i. Equation (3) is estimated by OLS. Fitted values from this regression are treated as EE_k-s for each age of $k = 31, \ldots, 65$. In such a way expected lifetime earnings were calculated for each age from the main sample of males aged 15–30.[1]

Opportunity cost of education (OCE) was defined in the original paper as the product of the average hourly wage and the probability to find the employment by age for each male according to his age. In the current work, using monthly average predicted wage instead, OCE was calculated for each age in the main sample of males aged 15–30 in a way that is similar to the calculation of ELE. The probability to find employment was estimated with OLS:

$$\Pr(\text{employment}_j) = \text{employed}_i = \delta_0 + \delta_1 \text{ age}_i + \delta_2 \text{ age}_i^2 + \nu_i, \tag{4}$$

and the fitted values of employed_i were treated as the probability to find employment for each age 15–30. The average wage for each individual in the same sample of males aged 15–30 by age was obtained by

[1]This was done using R Studio, see Appendix for the details.

Table 3: Estimates of ELE and OCE

age	ELE-hs	ELE-uni	OCE
15	3,633,716	4,308,689	3,208.79
16	3,633,716	4,308,689	3,744.51
17	3,633,716	4,308,689	4,268.44
18	3,633,716	4,308,689	4,770.86
19	3,633,716	4,308,689	5,243.02
20	3,633,716	4,308,689	5,677.04
21	3,633,716	4,308,689	6,065.96
22	3,633,716	4,308,689	6,403.74
23	3,633,716	4,308,689	6,685.25
24	3,633,716	4,308,689	6,906.27
25	3,633,716	4,308,689	7,063.49
26	3,633,716	4,308,689	7,154.51
27	3,633,719	4,308,691	7,177.86
28	3,633,761	4,308,726	7,132.97
29	3,634,382	4,309,235	7,020.16
30	3,644,057	4,317,176	6,840.71

Note: All values for ELE-s and OCE are in Soms.

Table 4: Heckman Probit Estimated Effects

Variables X =	Main Equation Y = Employed	Selection Equation Y = Student
age	0.37 (0.61)	4.071 (5.15)
age-squared	−0.01 (0.0132)	−0.08 (0.095)
marital	0.45 (0.56)	−0.42*** (0.153)
education	0.15 (0.175)	−0.09* (0.054)
urban	−0.61 (0.58)	0.69*** (0.11)
total assets	−6.60e−08 (1.59e−07)	2.08e−07*** (5.43e−08)
total income	9.27e−06 (8.14e−06)	−4.12e−06 (3.08e−06)
ELE-hs	–	−0.000126 (0.000142)
OCE	–	−0.003 (0.0033)
Constant	−5.34 (6.76)	424.3 (504.1)
Observations	1 316	1 316
athrho	–	0.500 (0.88)

Notes: Standard errors are in parentheses.
Significance is denoted by asterisks: *** 1%, ** 5%, * 10%.

estimating the following equation with OLS:

$$\tilde{w}_j = w_i = \theta_0 + \theta_1 \text{ age}_i + \theta_2 \text{ age}_i^2 + \varsigma_i, \tag{5}$$

from which the calculation of OCE is straightforward: $\text{OCE}_j = \Pr(\text{employment}_j) \times \tilde{w}_j$ for each age $j = 15, \ldots, 30$.

The expected sign of ELE-s is negative, because the higher are the expected lifetime earnings, the less is the likelihood that a young male chooses studying than looking for a job. The expected sign for OCE is positive, because the higher is the opportunity cost of education, the more it is likely that a young male chooses searching for a job rather than education.

3 RESULTS

It was already apparent from Table 2 how dense are the distributions of both ELE-s by age. Table 3 presents just another evidence that both ELE-hs and ELE-uni are essentially constant across the ages of individuals in the sample. This is an expected outcome keeping in mind the formula according to which they were calculated (see Table 1).

The final sample was obtained via merging the data from Table 3 with the main sample of males and the Heckman probit model was then estimated. Table 4 reports the estimated effects. It can be inferred that, unfortunately, neither instruments are strong, nor the impact of the education on the employment is significant. However, it might be the case that the weakness of instruments is the origin of the statistical insignificance of the coefficient on education in the second column (0.15). It is quite plausible that the strategy of calculating expected life earnings and opportunity cost of education that

both relied on the very primitive OLS predictions. The insignificance of athrho coefficient though shows that $\text{corr}(\epsilon_i, \xi_i) = 0$ meaning that there is no selection bias for young males. This result agrees with that of Pastore. However, this coefficient is also based on weak instruments. Although there is no statistical significance, the effect of education is quite small in magnitude as expected. There is no general agreement on the direction of the effects of instruments in the selection equation, however, as already mentioned, the magnitude of those effects is negligible. Finally, the impacts of total assets, total income, and urban variables are statistically indistinguishable from 0 as well, not favoring the existing literature.

4 CONCLUSION

This work aimed on estimating the impact of education on the employment outcome of males aged 15–30 in Kyrgyzstan. Estimation involved Heckman probit model in order to control for the selection bias caused by the choice of individuals whether to get enrolled into an educational institute or pursue employment. Results did not reveal any significant dependence of employment on education. The most probable reason is that the strategy for estimating instruments relied on the very primitive OLS predictions that could not correct the selection bias. The author of the replicated paper used more sophisticated methods of calculating instruments, and working on this is the top of mind for future work on this topic. There are other drawbacks of this paper that are associated with the more accurate choice and definitions of the variables in use. This applies to the inclusion of such factors as assets, income, location and relation to the household head. Working on these issues are the objective for future research.

REFERENCES

Biavaschi, C., Eichhorst, W., Giulietti, C., Kendzia, M. J., Muravyev, A., Pieters, J., Rodriguez-Planas, N., Schmidl, R., and Zimmermann, K. (2012). Youth unemployment and vocational training. IZA Discussion Papers 6890, Institute for the Study of Labor (IZA).

German Institute for Economic Research, DIW Berlin and Stockholm International Peace Research Institute (SIPRI) (2016). Life in Kyrgyzstan Panel Study, 2010-2013. IDSC of IZA.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Marelli, E. and Vakulenko, E. (2014). Youth Unemployment in Italy and Russia: Aggregate Trends and the Role of Individual Determinants. Working papers of the Department of Economics - University of Perugia (IT) 0001/2014, Università di Perugia, Dipartimento Economia.

National Bank of Kyrgyz Republic (2016). Official Interest Rates. http://www.nbkr.kg/index1.jsp.

Pastore, F. (2012). To study or to work? education and labor-market participation of young people in poland. *Eastern European Economics*, 50(3):49–78.

APPENDIX: R CODE THAT COMPUTES ELE-S AND OCE

```
setwd("your working directory")
numeric ← numeric(length(16))
a1 ← matrix(data = c(15:30), nrow =16, ncol = 2)
a1[,2] ← numeric
a2 ← a1
b1 ← read.csv("b_univ.csv")
b2 ← read.csv("b_hs.csv")
b1[,2] ← b1[,2]*12
b2[,2] ← b2[,2]*12
rate ← read.table("interestrate2012KG.txt")
r1 ← rate[,2]/100
r ← mean(r1) # mean interest rate for 2012 is 6.85%
c1 ← matrix(data = 0, nrow = nrow(b1), ncol = 16)
c2 ← matrix(data = 0, nrow = nrow(b2), ncol = 16)
for (i in 1:nrow(a1)){
   for (j in 1:nrow(b1)){
      c1[j,i] ← b1[j,2]/((1 - r^ (b1[j,1] - a1[i,1])))
   }
}
for (i in 1:nrow(a2)){
   for (j in 1:nrow(b2)){
      c2[j,i] ← sum(b2[j,2]/((1 - r^ (b2[j,1] - a2[i,1]))))
   }
}
for (i in 1:nrow(a1)){
   a1[i,2] ← sum(c1[,i])
}
for (i in 1:nrow(a2)){
   a2[i,2] ← sum(c2[,i])
}
write.csv(a1, "ELE_univ.csv") # import these two spreadsheets into stata for further mergings
write.csv(a2, "ELE_hs.csv")
```