Models with Qualitative Dependent Variables

Homework

Estimation returns to education in the Russian Federation

Krylov Ivan
Master, MEEP222
Julieta Egiazaryan
Master, MEEP222

Moscow 2023

INTRODUCTION

In today's rapidly evolving world knowledge keeps gaining more importance day by day, and thus more people tend to comprehend the necessity of getting decent education, as the skills and experience they acquire in universities determine their future in many different ways. One of these ways is their career prospects, including the company, their position and salary.

Hence it presents scientific interest to analyze how getting a higher education influences wage, people get at the work and to find out whether there is a remuneration gained on this basis. To adequately assess the remuneration, it is necessary to compare the salary of an individual, who is subject to the availability of higher education and without it. This is our research question.

Since the beginning of the 21st century, there has been an increase in the number of young people pursuing higher education. This is confirmed by Rosstat data, according to which the ratio of the number of young people (up to 17 years old) to those enrolled in universities has increased from 50.1% to 89.2% over the past decade. According to official data, after graduating from high school, approximately 70% of graduates tried to enroll in universities. What could be the reason for the growing demand for higher education among Russian youth?

Sociological research and surveys link the growing interest in higher education with the expectation of getting higher wages and obtaining high social status.

Admission to the university is preceded by a competitive selection. Some of those who pass it may subsequently be eliminated due to poor academic performance. As a result, the most motivated and capable individuals receive a diploma of higher education. Thus they have an advantage on the labor market (for example, diligence or knowledge) over those who do not complete their studies at the university. What is more, they have an opportunity to signal to the employers about their specific characteristics, like diligence, patience, accuracy. Consequently, the difference in wages of these two social categories will be caused not only by higher education itself, but also by the initial advantage of those who were able to secure the opportunity to get a higher education.

Accordingly, the purpose of the paper is to find out and analyze the influence of getting a higher education on the level of received wage.

In order to achieve the purpose we pose the following research tasks:

- to analyze and describe dataset;
- to apply the chosen model on the dataset;
- to estimate results obtained and to present key findings.

To solve the paper issue we use the Mincer earnings function, the most popular model to address the particular problem in modern economic research. The Mincer equation is a model representing wages as a function of completed schooling and experience. In its basic form, the logarithm of wages is represented as the sum of the number of years of schooling and a quadratic function of the number of years of work experience and represented by the following formula:

$$\ln w = f(s, x) = \ln w_0 + \rho s + \beta_1 x + \beta_2 x^2$$

where: w means earnings (the intercept w0 is the earnings of someone with no education and no experience); s stands for years of schooling; x is years of potential labor market experience. The parameters ρ, and β1, β2 can be interpreted as the returns to schooling and experience, respectively.

DATASET

In this paper we use data from the Russian Longitudinal Monitoring Survey (RLMS). RLMS is a survey conducted simultaneously for households and individuals, covering various aspects of their economical situation and health. We use 2021 data, which includes 13,187 observations of people aged 13 to 97 (5,403 males and 7,780 females). Within the framework of this work all the data were processed and put into proper form. Variables were chosen on the basis of scientific literature.

From many of the characteristics presented in the survey, we use the following:

- age – age of respondent;
- male – respondent's gender (male=1 if the respondent is male);
- exp – total work experience;
- special_exp – current job experience;
- wage – one month's salary of the individual;
- higher_educ – the presence or absence of higher education (higher_educ = 1, if the individual graduated from the university);
- employed – employment status (employed = 0 if the individual is unemployed, otherwise employed = 1);
- married – a dummy-variable for marital status of the respondent;
- child – a dummy-variable responsible for having children;

- PGT, town, city_center, MSC_SPB are dummy-variables responsible for the type of city the respondent lives in. If each of the variables are zero, then the individual lives in a village, if PGT=1, then the individual lives in an urban town, if town=1, then the individual lives in a city, city_center=1 – the individual lives in a regional center, MSC_SPB=1 – the individual lives in Moscow or St. Petersburg;
- health_1, health_2, health_3, health_4 - dummy-variables responsible for the self-assessment of the respondent's health. If each of the variables are zero, the health of the individual is very poor; if health_1=1, the health of the individual is poor, if health_2=1, the health of the individual is average, health_3=1 – the health of the individual is good, health_4=1 – the health of the individual is very good.

Descriptive statistics of the sample are presented in Table 1.

According to Table 1, we have different numbers of observations of regressors, which leads us to the fact that we have to deal with censored data. Most of the regressors are presented as dummy variables.

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	6,329	33865.83	22466.97	1200	375000
exp	11,149	23.80569	13.88963	0	66
special_exp	6,329	8.680676	9.062081	0	53
male	13,187	.4097217	.4918009	0	1
higher_educ	13,187	.2756503	.4468583	0	1
PGT	13,187	.0693865	.2541199	0	1
town	13,187	.2574505	.4372462	0	1
city_center	13,187	.3068932	.4612222	0	1
MSC_SPB	13,187	.1215591	.3267883	0	1
employed	13,187	.4799424	.4996165	0	1
age	13,187	47.57496	19.12626	13	97
married	13,187	.5667703	.4955405	0	1
child	13,187	.2800485	.4490397	0	1
health_1	13,140	.1052511	.3068884	0	1
health_2	13,140	.4838661	.4997586	0	1
health_3	13,140	.3745053	.4840133	0	1
health_4	13,140	.0265601	.1608	0	1

Table 1. Description of statistical data

MODEL

To solve the research tasks we decided to use in our study the Heckman model. It includes two components: a binary choice model for participation equation and a linear model for the "intensity" of participation. This determines the advantage of Heckman model over Tobit model: the fact that factors, affecting the models, generally speaking, may be different. So, the exact model specification we estimate in our research is:

$$\log\_wage_i = a * higher\_educ_i + \beta_1 exp + \beta_2 exp\_square_i + x_i'\beta + \varepsilon_i$$

$$employed_i^* = z_i'\gamma + u_i$$

$$employed_i = \begin{cases} 1, \ if \ employed_i^* \geq 0 \\ 0, \ if \ employed_i^* < 0 \end{cases}$$

$$\forall \begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}$$

where X'_i is a set of variables: special_exp, male, PGT, town, city_center, MSC_SPB;

Z'_i is a set of variables: age, married, child, PGT, town, city_center, MSC_SPB, health_1, health_2, health_3, health_4.

In our case "log_wage" is the dependent variable for the outcome equation, whereas "employed" is the dependent variable for the selection equation. Thus we have the opportunity to observe the influence of different regressors on the chosen two latent variables. This is one of the reasons why we have decided to use the Heckman model in this case.

The presence of two equations in the work allows us, firstly, to identify which of the respondents are employed (participation equation), then to reveal the level of their wages (outcome equation). Thus we divide our sample on the basis of being currently employed or not, which makes our analisis more adequate and rational.

Education is set as an exogenous regressor along with other factors. Other regressors for outcome equation and selection equation were chosen on the basis of scientific literature.

In our work we decided to assess our model by maximum likelihood method.

ESTIMATION RESULTS

After the Stata operations we get our model. Estimation results are presented in table 2. Model is adequate, as Prob > chi2 = 0.0000. We have 13,140 observations, 6,858 of which are censored. All of the variables are significant at 5% level of significance. Also the null hypothesis (H0) of the likelihood ratio (LR) test, which examines the independence of equations, is rejected at a significance level of 1%, therefore we can conclude that the errors are interdependent, indicating the presence of selection.

As for Heckman's lambda, that exhibits bias occurred because of the non-randomly selected observations, in our case it is significant too and the results show that use of the Heckman model is justified.

In the matter of interpretation, first we should mention that only results of the outcome equation are subject to interpretation. What is more, we may interpret them as in linear regression model. Nevertheless we may not do so with the results of selection model, only state the direction of the connection. To solve this problem we should analyze marginal effects.

The results of marginal effects estimation are presented in table 3.

Heckman selection model
(regression model with sample selection)

Log likelihood = **-6430.648**

Number of obs	**13,140**
Censored obs	**6,858**
Uncensored obs	**6,282**
Wald chi2(**9**)	**2015.80**
Prob > chi2	**0.0000**

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
log_wage						
exp	.0060507	.0007598	7.96	0.000	.0045615	.0075399
exp_square	-.0001536	.000017	-9.03	0.000	-.0001869	-.0001203
special_exp	.0014894	.0003409	4.37	0.000	.0008213	.0021575
male	.1521567	.0054612	27.86	0.000	.141453	.1628604
higher_educ	.1186756	.0056654	20.95	0.000	.1075717	.1297796
PGT	.0910073	.0126022	7.22	0.000	.0663075	.1157071
town	.0340515	.0082272	4.14	0.000	.0179264	.0501766
city_center	.0722702	.0080336	9.00	0.000	.0565246	.0880158
MSC_SPB	.253808	.0100208	25.33	0.000	.2341677	.2734484
_cons	4.266935	.0124671	342.26	0.000	4.2425	4.29137
employed						
age	-.0122663	.0007964	-15.40	0.000	-.0138273	-.0107054
married	.5725838	.0270548	21.16	0.000	.5195574	.6256102
child	.7004496	.0304759	22.98	0.000	.640718	.7601812
PGT	.1600316	.0513285	3.12	0.002	.0594296	.2606335
town	.4826122	.0340611	14.17	0.000	.4158537	.5493706
city_center	.4422834	.0327402	13.51	0.000	.3781138	.5064531
MSC_SPB	.3226692	.0419115	7.70	0.000	.2405241	.4048142
health_1	.5330416	.2289734	2.33	0.020	.084262	.9818211
health_2	1.453835	.2248256	6.47	0.000	1.013185	1.894485
health_3	1.533674	.2257214	6.79	0.000	1.091268	1.976079
health_4	1.065519	.2365701	4.50	0.000	.6018496	1.529187
_cons	-1.696638	.2306743	-7.36	0.000	-2.148751	-1.244525
/athrho	-.3178129	.0436131	-7.29	0.000	-.4032929	-.2323329
/lnsigma	-1.532959	.011003	-139.32	0.000	-1.554524	-1.511393
rho	-.307528	.0394884			-.382763	-.2282409
sigma	.2158959	.0023755			.2112899	.2206024
lambda	-.066394	.0089739			-.0839826	-.0488055

LR test of indep. eqns. (rho = 0): chi2(**1**) = **53.99** Prob > chi2 = **0.0000**

Table 2. The results of the Heckman model estimation

Average marginal effects — Number of obs = 6,282
Model VCE : OIM

Expression : E(log_wage|Zg>0), predict(ycond)

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
exp	.0060507	.0007598	7.96	0.000	.0045615	.0075399
exp_square	-.0001536	.000017	-9.03	0.000	-.0001869	-.0001203
special_exp	.0014894	.0003409	4.37	0.000	.0008213	.0021575
male	.1521567	.0054612	27.86	0.000	.141453	.1628604
higher_educ	.1186756	.0056654	20.95	0.000	.1075717	.1297796
PGT	.0968203	.0124425	7.78	0.000	.0724335	.1212071
town	.0515818	.0077914	6.62	0.000	.0363109	.0668527
city_center	.0883356	.007639	11.56	0.000	.0733634	.1033078
MSC_SPB	.2655286	.0098061	27.08	0.000	.246309	.2847481
age	-.0004456	.0000677	-6.58	0.000	-.0005782	-.0003129
married	.0207984	.002847	7.31	0.000	.0152185	.0263784
child	.025443	.0035081	7.25	0.000	.0185673	.0323187
health_1	.0193621	.0086847	2.23	0.026	.0023404	.0363839
health_2	.0528089	.010744	4.92	0.000	.031751	.0738668
health_3	.0557089	.0110958	5.02	0.000	.0339615	.0774563
health_4	.0387037	.0100124	3.87	0.000	.0190798	.0583277

Table 3. The marginal effects of Heckman model

According to Table 3, almost all parameters have an impact on an individual's wages in Russia at any level of significance (the parameter heath_1 is also significant, but at the 5% level). Our main variable of interest, higher education, raises wages by 11.8% on average. This is one of the largest influences among all factors.

Other variables that have an impact on wage levels are:

- Each additional year of experience increases an individual's income by 0.61% on average, but each additional year squared reduces it by 0.015%. It follows that people in Russia have a "career peak" after about 20 years of work experience;
- each additional year of work at the same company increases an individual's salary by 0.15% on average;
- men in Russia have, on average, 15.2% higher income than women;
- living in an urban town increases a resident's wages by 9.68% on average in comparison to residents of a village;
- living in a city increases a resident's wages by 5.15% on average in comparison to residents of a village;
- living in a regional center increases a resident's wages by 8.83% on average in comparison to residents of a village;

- living in Moscow or St. Petersburg increases a resident's wages 26.5% on average in comparison to residents of a village;
- if an individual gets one year older, their wages decrease on average by 0.04%;
- a married individual earns on average 2.08% more than a single person;
- an individual with a child earns, on average, 2.54% more than an individual without children;
- an individual who assesses their health as "poor" earns, on average, 1.94% more than an individual who assesses their health as "very poor"
- an individual who assesses their health as "neither good nor bad" earns on average 5.28% more than an individual who assesses their health as "very poor"
- an individual who assesses his/her health as "good" earns on average 5.57% more than an individual who assesses his/her health as "very poor"
- an individual who assesses his/her health as "very good" earns on average 3.87% more than an individual who assesses his/her health as "very poor"

CONCLUSIONS

Nowadays, education plays an important role on the road to success in life. Well-educated people are more likely to build a career in fields in which knowledge is a major factor of production. Therefore, the question of evaluating the return to education is particularly relevant. Based on data from the Russian Longitudinal Monitoring Survey for the year 2021, we estimated the actual value of the education premium in Russia. We used the Heckman model as our main model, which consists of two equations: the augmented Mincer equation and the selective equation, which measures the probability of an individual participating in the labor force. We obtained significant estimates of the effect of education on income: on average, having a university degree increases an individual's wages by 11.8%. In addition, other factors also have a significant positive effect on salary: work experience, having children and a spouse, good health, living in large population centers, and others. The results obtained in the work can be successfully used in the implementation of public policy measures in the field of education. However, in order to obtain more accurate estimates, it is necessary to use detailed data, including a more precise gradation of higher educational institutions and industry specifics of education, as well as to take into account non-random selection of individuals into different professions.

REFERENCES

1. Bondarenko I. (2020)Analysis of the relationship between income and return on professional education of graduates of russian universities // Imsit Bulletin - 3-10 p.
2. Dubin B.V., Gudkov L.D., Levinson A.G. (2004) Accessibility of higher education: social and institutional aspects // Accessibility of higher education / edited by S.V. Shishkin. – 246 p .
3. 71 percent of high school students of the Russian Federation seek to enter universities // RIA News. Available at: https://sn.ria.ru/20180417/1518825250.html
4. Mincer J. Schooling, Experience and Earnings // New York: National Bureau of Economic Research. 1974.
5. Andrushchak G.V., Kononova A.E. Where is it better to study: differentiation of earnings of university graduates // Bulletin of the Russian Monitoring of the economic situation and public health of the HSE (RLMS – HSE). 2012. C. 129-136.
6. Roshchin S.U., Rudakov V.N. The influence of the "quality" of the university on the salary of graduates // Questions of economics. 2016. No. 8. pp. 74-95