Expansion of Tertiary Education, Employment and Wages: Evidence from the Russian Transition

Natalia Kyui *

November 2011

Abstract

This paper analyzes the effects of the educational system expansion on labor market outcomes. It explores the expansion of tertiary (i.e. post-secondary) education in the Russian Federation over the past 15 years, as a natural experiment. Regional changes in the number of university slots, as a result of educational reforms, provide an exogenous variation in access to higher education. Using simultaneous equation models, the influence of education on employment and wages is estimated for those who improved their educational attainment due to increases in educational opportunities. The estimation results, which are robust to changes in model specifications, suggest strong positive returns to education in terms of wages and employment. Considering this gradual increase in access to universities, the paper further estimates heterogeneous returns to education for individuals who were exposed to different degrees of expansion within higher education. The results reveal decreasing returns to education for those who subsequently complied with pursuing higher education, as access to the educational system became easier and easier. Moreover, a non-parametric estimation of the model with essential heterogeneity is undertaken, in order to identify marginal returns to higher education. Returns to education are found to be decreasing for lower levels of individual unobserved characteristics, which positively influence higher education attainment. Therefore, the expansion of the higher education system significantly increased the wages of those who were exposed to growing numbers of university slots. Nonetheless, this increase was smaller than the returns to education for those who would have pursued higher education anyway.

Key words: educational choice, returns to education, heterogeneity

JEL: J24, I20

---

* Paris School of Economics, University Paris-I Panthéon–Sorbonne.
Centre d'Économie de la Sorbonne, 106-112 boulevard de l'Hôpital, 75013 Paris, France.
E-mail: natalia.kyui@univ-paris1.fr.

This paper was written in part during a research visit at Columbia University.
I gratefully acknowledge the kind support and helpful discussions of Leigh L. Linden, Peter Arcidiacono, and Natalia Radtchenko. I would like to thank the Russian Longitudinal Monitoring Survey Phase 2, funded by the USAID and NIH (R01-HD38700), Higher School of Economics and Pension Fund of Russia, and provided by the Carolina Population Center and Russian Institute of Sociology, for making these data available.

Contents

1 Introduction

The influence of education on earnings has been widely analyzed in the economic literature for both developed and developing economies. Studies show that workers with higher educational levels have consequently higher wages. Papers, which analyze developed economies, find strong positive returns to education, while research for developing economies, and especially transition economies, highlights the growing trends in the returns to education during recent decades. However, the persistent question is the identification of a true causal effect of education on earnings, separated from any selection biases (selection on ability, preferences, gains, etc.). The main purpose of its identification is to answer a question: whether an additionally provided education for those who do not obtain it now would increase their wages and would generate a positive economic return for a country. Such inquiry is especially relevant for developing economies, for which the increase in the educational level of the population is one of the main objectives of economic development.

The current study explores a natural experiment - expansion of the higher education system in the Russian Federation during the period 1990-2000, which doubled the number of slots in colleges, and thus, doubled the number of college graduates. It identifies the returns to higher education in Russia by using this expansion as exogenous variation in access to higher education. Additionally, the paper quantifies the effects of the expansion of the educational system on further labor market outcomes, namely employment and wages. The gradual changes in access to colleges during the analyzed period allow us to estimate the heterogeneous returns to a higher education degree for those who have been exposed to different degrees of expansion within higher education.

Several studies have analyzed the consequences of educational policies on educational attainment and labor market outcomes. They have identified the returns to education using the changes in the educational system (see Card (1999) and Belzil (2006) for an extensive analysis of other methods for estimation of the returns to education). We can distinguish two main directions in this economic literature.

The first set of papers analyzes compulsory schooling laws and their changes, focusing thus mainly on secondary education. The analysis of compulsory schooling laws and their effects on educational attainment and wages was pioneered by the work of Angrist and Krueger (1991) for USA data, and followed by works conducted for other countries (see Oreopoulos (2006b) for UK, Pischke and von Wachter (2008) for Germany, Oreopoulos (2006a) for Canada). Overall, these studies find a positive effect of compulsory schooling legislation on educational attainment and future wages, though the magnitude of this effect varies among countries (for example, almost zero returns in Germany).

The second set of papers analyzes the access to education and its changes. The term "access to education" in this direction of the literature consists of three aspects:

(1) financial access: studies analyzing the influence of tuition fees and financial-aid policies on educational enrollment and further wages, for example Kane and Rouse (1993), Card and Lemieux (2001b), Arcidiacono (2005);

(2) convenience of physical access: papers focusing on the distance to high schools

or colleges, presence of high schools or colleges in the district, for example Card (1995), Conneely and Uusitalo (1997);

(3) infrastructure and limitation of access due to capacities: studies focusing on the capacity of the educational system and its changes. Duflo (2001) uses the exposure to school construction in Indonesia (major governmental project) to identify its effects on educational attainment and future wages. Walker and Zhu (2008) analyze the expansion of higher education in UK during the period 1994-2006, comparing the returns to education between different cohorts of people. They find no changes in the returns to higher education for men (except for the bottom quartile of the distribution) and a significant rise in returns for women. The current paper is the closest in spirit to this direction of the economic literature.

This study analyzes the labor market consequences of major expansion of the higher education system in the Russian Federation from 1990 to 2000. It is focused on employment opportunities and monetary returns to higher education. Using IV technique and simultaneous equation models, we identify the returns to education for youths who have been exposed to the expansion of the higher education system. The gradual expansion allows us to estimate the heterogeneous returns to a higher education degree for youths who have benefited from these reforms at different levels of expansion of the higher education system. Additionally, we use the recent nonparametric method to identify the marginal returns to education - the heterogeneous returns to a higher education degree, which vary with the level of unobserved individual characteristics. These two approaches to the estimation of heterogeneous returns to higher education give similar results. Therefore, this study provides the strong evidence of the decreasing returns to education for youths who obtain higher education due to increasing access to college, in other words, who comply with pursuing higher education, as access to the educational system becomes easier and easier. However, these returns are positive and large in magnitude (60%-80% wage increase for a higher education degree). Therefore, the expansion of the higher education system has increased significantly the wages of those who have been exposed to the growing number of slots in universities. However, this increase is smaller than the returns to higher education for those who would have obtained it anyway.

The next two sections describe the institutional context of the Russian educational reforms and the data. *Section 4* discusses the estimation results of the returns to education. *Section 5* provides four robustness checks. *Sections 6* and *7* analyze the heterogeneity in the returns to education. *Section 8* concludes. *Technical Appendix* provides details on the estimation of the simultaneous equations models and the models with essential heterogeneity.

2 Institutional Context

The Educational System in Russia consists of four levels: primary and general education (8 years at general schools); secondary education (additional 2 years at general or specialized schools); tertiary (i.e. any post-secondary) education; and post-graduate education (3-6 years of graduate education). Tertiary education is presented by two levels (in the current study we also refer to them as the 1st and 2nd levels of tertiary education: 1TE and 2TE). 1st-level tertiary education is post-secondary professional education or vocational education, which consists of 2-3 years of study at technical schools or specialized schools (military, medical, musical). 2nd-level tertiary education is higher (professional) education: 4-6 years after secondary education at universities and colleges.

During the Soviet Union period, the government regulated the tertiary education sector completely. Education was financed from the country's budget and was free for students. The Soviet government determined the number of slots in tertiary education by majors according to the economy's predicted needs in professionals for the following years. Therefore, the number of slots in tertiary education was limited. The growing access to secondary education (universal secondary education was the main priority for the Soviet government since 1965) increased the number of applicants for tertiary education programs. Potential students obtained their admission on a competitive basis. Admission tests selected high-ability candidates for studying among those who obtained a secondary education degree. There was a slight expansion of the tertiary education system during the Soviet period, however the capacity of the tertiary education system was almost stable until the transitional period.

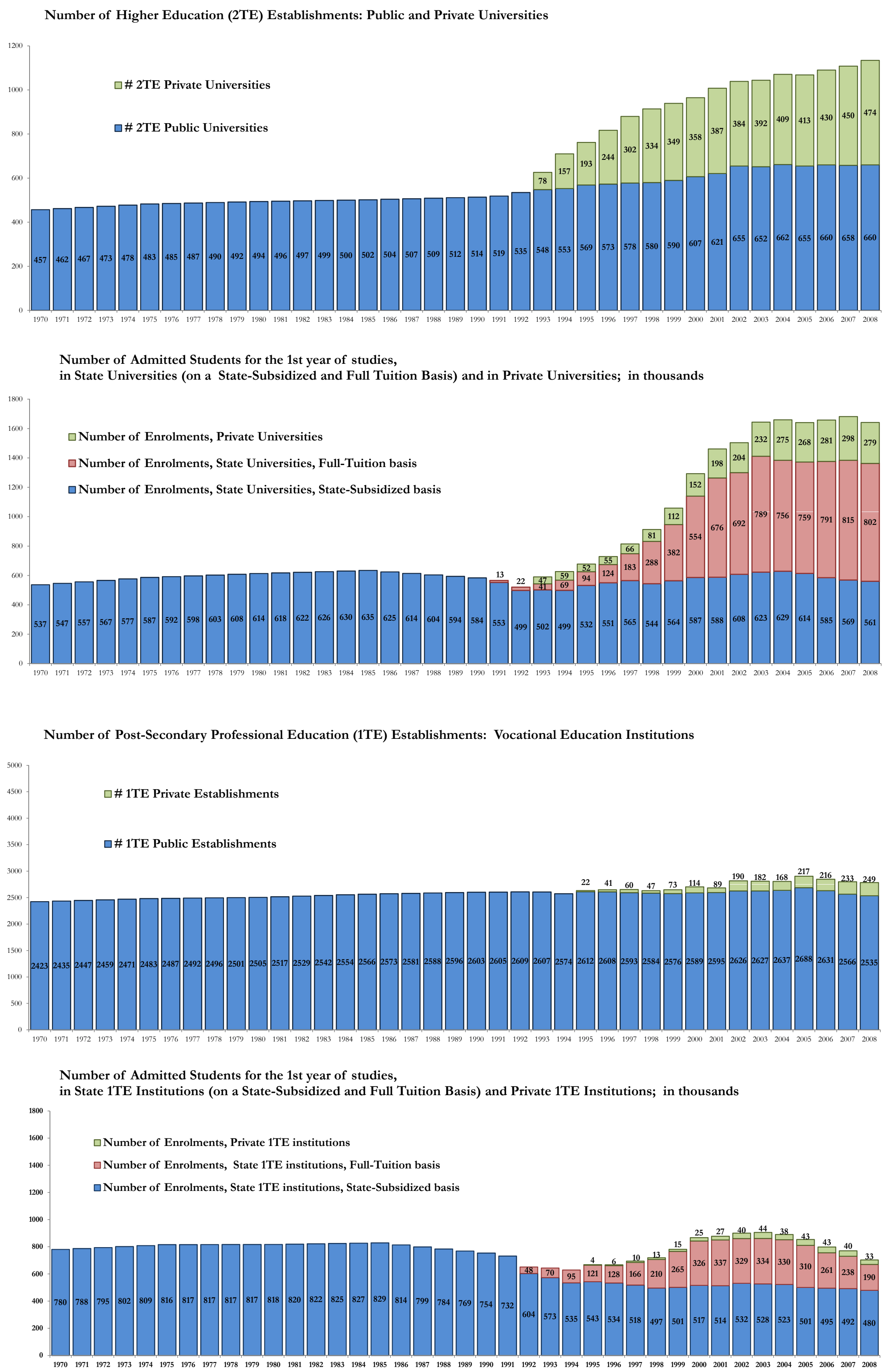
With the beginning of transition, in 1992, the Russian government passed the law "*About Education*", which marked a starting point for major changes in the tertiary education system. The most important amendment changed state-subsidized tertiary education into mixed forms: both on a state-subsidized and tuition basis. First, the government authorized the creation and operation of non-public tertiary education institutions (colleges, institutes, universities). They should be run as not-for-profit organizations, providing educational services to the population on a full-tuition basis. Second, the government authorized public tertiary education institutions to provide some paid educational services, in other words, to admit some students on a full-tuition basis in addition to the state-subsidized slots. The government continues to finance the state-subsidized slots in tertiary education. These changes, along with a large persistent demand for tertiary education (to a large extent because of the positive returns to education and massive secondary education), led to the major expansion of the tertiary education system in Russia. During the following decade, the number of slots in the higher education system more than doubled.

This expansion of the post-secondary education system in Russia happened only due to the creation and expansion of tertiary education on a full-tuition basis, both in public and private educational establishments. *Figure 1* describes these changes. The major increase happened in the 2nd-level tertiary education (higher education), while the 1st-level tertiary education (vocational tertiary education) kept the same levels. The first two graphs in *Figure 1* depict changes in the 2nd-level tertiary education, the third and fourth graphs - in the 1st-level tertiary education. The number of public universities slightly increased during the 1990-2008 years (+28%), and the number of private universities grew from zero up to 72% of the number of public universities (*1st graph* in *Figure 1*). However, in terms of the number of admitted students, the main increase happened in the full-tuition slots in public universities (2nd graph). Overall, the students' admission to the universities increased by 181% during 1990-2008, among which 137% because of the full-tuition slots in public universities.

Even though expansion of the higher education system has been driven by the full-tuition education, Kyui (2011) shows that youths with all family background have benefited from these reforms in terms of educational attainment.

Figures A-1 in *Appendix* depict these changes in the educational system over the period 1960-2008. The first graph shows the number of admitted students for the 1st year of studies in the 2nd-level tertiary education by categories of studies: in the state (public) universities on a state-subsidized or full-tuition basis and in the private universities. The second graph describes the changes in the size of the cohort of youths, graduating from secondary school or being 17 years old in the corresponding years. The third graph shows the number of admitted students for the 1st year of studies in the 1st-level tertiary education by categories of studies. Even though the size of the cohort of youth was growing over the period 1990-2003, the expansion of the higher education system was significantly larger. The increase in the number of admitted students led to the corresponding increase in the number of college graduates, as presented by *Figures A-2*. As the third graph suggests, the majority of the students finish 5-years programs, and, thus, obtain a specialist's degree. Therefore, the increase in the number of graduates corresponds to the increase in the number of slots 5 years before (as shown by the first and second graphs).
Figures A-3 in *Appendix* show the increase in the educational opportunities corrected by the increase in the corresponding cohort of 17 year-old youths. The proportion of the number of slots to the size of 17-y.o. cohorts is adjusted by the share of 17 year-olds among the students admitted to the 1st-year studies. As it is shown by *Figures A-4*, 17 year-olds are in the majority among the new entrants in the universities. Therefore, the access to higher education increased from 17-20% up to the 40% after the beginning of the transition to the market economy.
Figures A-5 in *Appendix* describe the situation in the 1st-level tertiary education from the same perspectives. National statistics suggest that there were no significant changes in the opportunities for the 1st-level tertiary education attainment, though a portion of the slots started to operate on a full-tuition basis.

Figure 1: Expansion of the Tertiary Education System in Russia: 1970-2008



Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

Nevertheless, the government still significantly controls the tertiary education sector.

First, in order to perform an educational activity, all educational institutions (the requirements are the same for public and private institutions) must obtain a license for providing the educational services. Licensing decisions are based on the full information about educational programs and available resources of the establishments: available buildings, faculty, other educational facilities (such as libraries, etc.), and non-educational facilities (cafeterias, etc.). The obtained licences strictly determine the maximum number of students allowed to be enrolled in an educational institution. Therefore, the government still controls the number of slots in tertiary education: 1) direct control through the number of state-subsidized slots and 2) indirect control through the licensing of educational activity. It is possible for educational establishments to enlarge this allowed maximum number of students. In order to do that, educational institutions have to go through the official procedure of changing the parameters of their educational license, which also includes the analysis of the organization's resources, educational activity, etc. Therefore, even though the educational establishments are free to choose the number of students to admit, they have to obtain an authorization in advance, and thus to prove their financial, physical and human resources. These official limitations could be the main explanation of the gradual expansion of the tertiary education system during the past decade.

Second, the government provides quality control through licensing and accreditation of educational establishments. The accreditation of an institution permits to deliver the state diplomas of tertiary education.

Third, the government has a direct control over the shares of majors in tertiary education on a state-subsidized basis, as well as the indirect control over them for full-tuition slots through licensing the educational activity (for all private and public educational organizations).

Figures 2 and *3* illustrate the tertiary education expansion within Russian regions (*Figure 2* - for 1st-level tertiary education, and *Figure 3* - for 2nd-level tertiary education). The regions are the seven Russian Federal Districts.

The first graphs of these two groups of figures show the number of admitted students for the first year of studies. The second graphs illustrate the size of the cohort of 17-year-old youths in the corresponding years. The third graphs depict the shares of the admitted students to the cohorts. The fourth graphs present the adjusted shares that were corrected by the proportion of the current year secondary school graduates among the admitted students for the first year of studies.

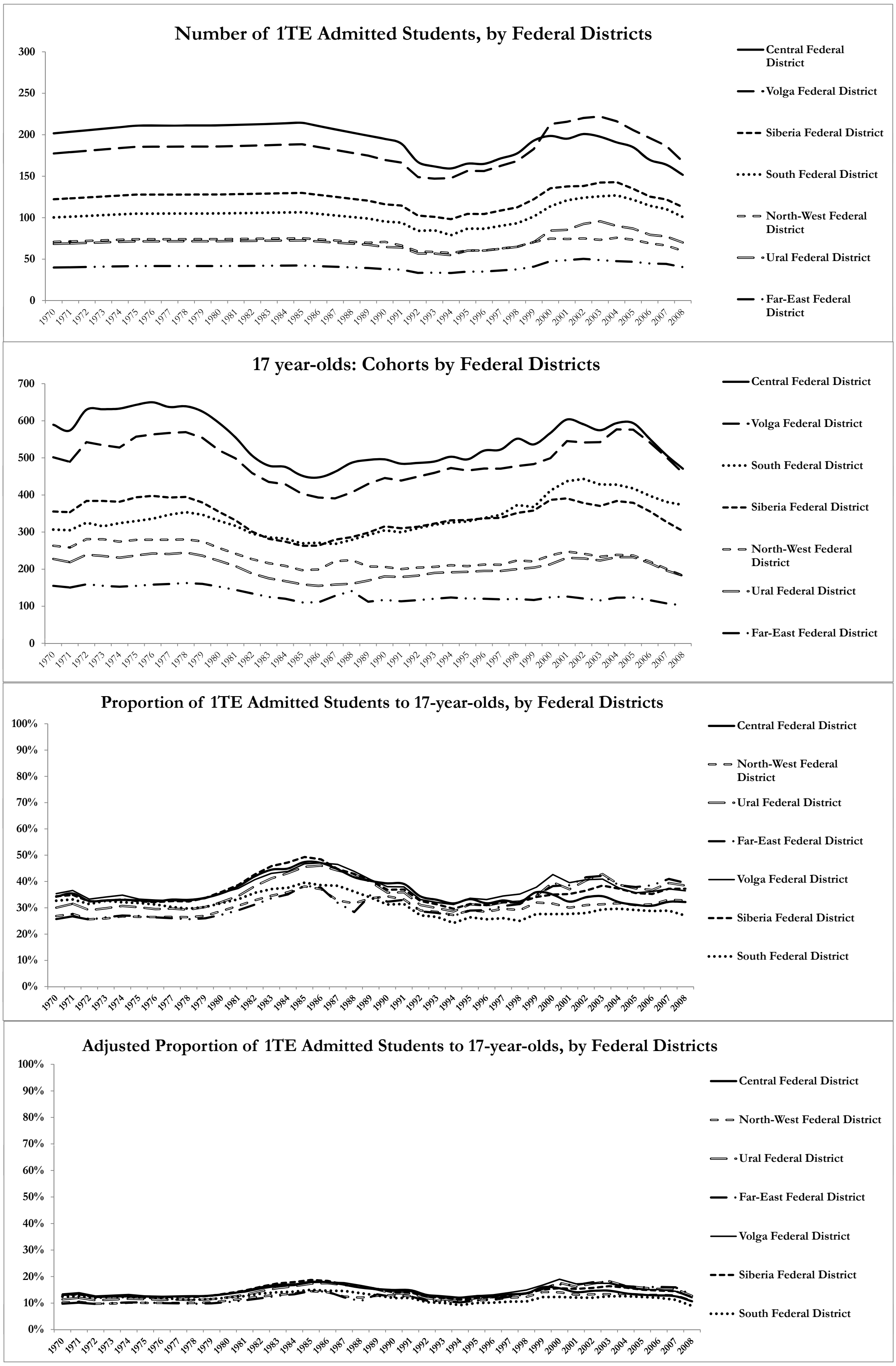
These data suggest that historically there were regions with larger and lower access to tertiary education, both general (2TE) and vocational (1TE). As soon as the number of slots is adjusted by the size of the population, these differences among regions become very small for the 1st-level tertiary education and still important for the 2nd-level tertiary education. The Central Federal District (which also includes Moscow) is historically the region with the highest access to tertiary education.

The expansion of tertiary education was slightly different among regions. The regions with the highest number of slots in the tertiary education system have kept their leaderships. These figures and the history of expansion, which we discussed earlier, suggest that the educational resources in the regions at the moment of the beginning of transition along with the governmental control over the licensing could be an explanation of these regional differences in the trajectories of expansion.

In the paper, we define the access to tertiary education as a number of available slots and as a proportion of slots relative to the youth cohorts. In fact, this expansion is a result not only of governmental actions, but also of the choices of educational institutions that were enlarging their sizes over the time. Nevertheless, the licensing restrictions, which are imposed by the government, have determined the gradual nature of the expansion. Therefore, these time changes in access to education within regions were plausibly exogenous for youths and their families.

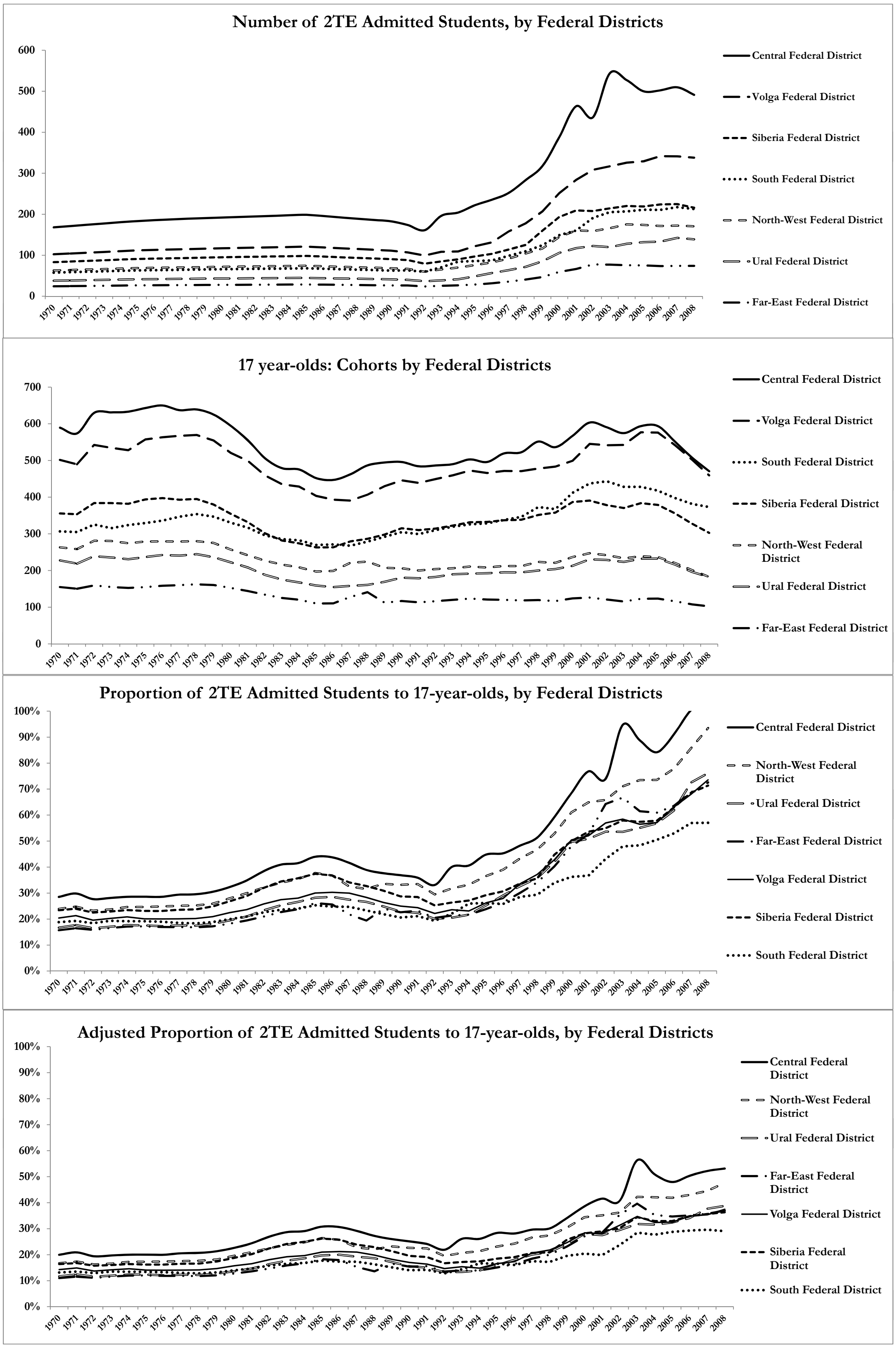
Empirical studies show increasing returns to education during the transitional period in Russia. Studies conducted for the Soviet period show a small or even negative value of the returns to education. Gregory and Kohlhase (1988) found the returns to the higher education equals 5% and 2.3% for secondary education. Belokonnaya et al. (2007) by estimation of Mincer's equation show that the returns to higher education in 2005 are positive both for men and women, nevertheless larger for female employees (27% and 40% in comparison to secondary education). Returns to incomplete higher education and post-secondary professional education are also positive and significant, but are much smaller than the returns to higher education. In all these studies for the returns to education in Russia, educational attainment was treated as exogenous. The current paper takes into account the endogenous nature of educational choices.

Figure 2: Regional Expansion of the 1st-Level Tertiary Education System in Russia: 1970-2008



Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

Figure 3: Regional Expansion of the 2nd-Level Tertiary Education System in Russia: 1970-2008



Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

3 Data Description

We use the data from the Russian Longitudinal Monitoring Survey (RLMS). It is a series of nationally representative surveys designed to monitor the effects of Russian reforms on the health and economic welfare of households and individuals in the Russian Federation. The RLMS data are described in Swafford et al. (1999a) and Swafford et al. (1999b).

We use the sample of 24-47 year olds, observed in 2000-2008 years. We analyze the dataset as repeated cross-section data. The observed people were taking their decisions about tertiary education attainment, when they were 17 years old, in 1970-2001, thus, before and during the expansion of tertiary education in Russia. We have 41585 observations overall, among them 10288 are unemployed, 31297 - employed. However, among the employed population we observe wages only in 91.5% cases. Thus, the final sample, which we use for our analysis consists of 28622 employed and 10288 unemployed people.

We use three variables to describe education obtained by individuals:

1. *Number of Years of Schooling* (including all educational levels) - continuous variable.

2. *Number of Years of Higher Education (2TE) Studies* (in the 2nd-level tertiary education establishments) - continuous variable.

3. *Higher Education Degree* attainment - dummy variable, which takes value 1 if a person has a higher education degree, 0 - otherwise.

The variable *Number of Years of Schooling* varies from 9 to 15 years. 15 years of education corresponds to a higher education degree. 10 years of education corresponds to completed secondary school. 13 and 14 years of education correspond to complete / incomplete vocational education or to incomplete higher education. *Table 1* shows the proportions of the population by the number of years of schooling.

The variable *Number of Years of Higher Education Studies* takes the value 0 years for those who did not start higher education, 3 years for incomplete higher education degree, and 5 years for complete higher education degree. *Table 2* shows the proportions of the population with different number of years of higher education studies.

Finally, the variable *Higher Education Degree* has the value 1 for those who have finished higher education, and 0 - otherwise. 23.26% of the whole population and 25.87% of the working population with observed wages have higher education degrees (these figures correspond to the number of years of schooling being equal 15, or number of years of higher education studies being equal to 5).

An individual is considered to be employed if he or she declares to have a job. The individual average monthly income from working activity is used as a measure of wages. Wages are adjusted for inflation by the Consumer Price Index. In order to account for the

Table 1: Number of Years of Schooling, 24-47 y.o., 2000-2008 years.

Years of Schooling	**Employed, 24-47 y.o. with Observed Wages**	**Employed 24-47 y.o.**	**Unemployed 24-47 y.o.**	**ALL 24-47 y.o.**
9	9.24%	9.4%	16.03%	11.04%
10	36.78%	37.09%	45.32%	39.12%
13	21.78%	21.44%	17.49%	20.46%
14	6.32%	6.31%	5.53%	6.12%
15	25.87%	25.77%	15.63%	23.26%
Total	28 622	31 297	10 288	41 585

Source: RLMS, 2000-2008, Author's calculations

Table 2: Number of Years of 2TE Studies, 24-47 y.o., 2000-2008 years.

Years of 2TE Studies	**Employed, 24-47 y.o. with Observed Wages**	**Employed 24-47 y.o.**	**Unemployed 24-47 y.o.**	**ALL 24-47 y.o.**
0	67.81%	67.92%	78.84%	70.62%
3	6.32%	6.31%	5.53%	6.12%
5	25.87%	25.77%	15.63%	23.26%
Total	28 622	31 297	10 288	41 585

Source: RLMS, 2000-2008, Author's calculations

regional differences in prices and to make regional wages comparable, wages are additionally corrected by the regional price of standard product set.

The sample consists of 24-47 year-old people observed in 2000-2008. Therefore, they were 17 years old in 1970-2001. Thus, in the sample there are as the cohorts of people studying before the expansion, as well as the cohorts of people studying during the gradual expansion (1990-2001) of the higher education system.

4 Returns to Education: Local Average Treatment Effects

This section discusses the identification and estimation of the returns to education using the variation in the opportunities of obtaining tertiary education (thus, variation in the number of available slots in the tertiary education) as an instrument for the educational levels obtained. The exposure of an individual to the expansion of the higher education system is identified by the year when he or she was 17 years old (thus, graduating from secondary school and taking a decision about further tertiary education) and by the Federal District of residency. Along with the control for Federal Districts fixed effects and for year of birth fixed effects, the changes in the capacity of the higher education system provides a plausibly exogenous variation in access to higher education.

4.1 1st Stage Equation. Instruments' Choice and Quality.

Table 3 reports the results for the "first-stage" equation: influence of the changes in the educational system (increasing number of slots) on educational attainment. Estimations are conducted for the 24-47 year-old working population.

In the six specifications of the first stage, the following variables that determine the expansion of the higher education system are used:
(1) The number of slots in the higher education system at Federal Districts levels and the Size of the corresponding cohort of youth.
(2) The number of slots in the higher education system at Federal Districts levels.
(3) The number of slots in the higher education system at the Russian level and the Size of the corresponding youth cohorts.
(4) The number of slots in the higher education system at the Russian level.
(5) The proportion of slots in the higher education system relative to the corresponding cohort of youths at Federal Districts levels.
(6) The proportion of slots in the higher education system relative to the corresponding cohort of youths at the Russian level.

Looking at the F-Statistics for the test on exclusion of the instruments from the first-stage equation, we can conclude that the variables describing the number of slots have a higher explanation power than the variables describing the ratio of the number of slots to the size of the corresponding youth cohorts. The variable for the size of the cohort of youths is insignificant in almost all specifications. Moreover, the variables at Federal Districts levels have a higher explanation power, than the variables at the Russian level. This fact is due to the additional variation in the size of the higher education system and due to the different patterns of its expansion among regions. Therefore, for the further estimations the number of slots in the 2nd-level tertiary education at Federal Districts levels is used as the instrument for educational attainment (with or without control for the

Table 3: Influence of the Higher Education System Expansion on Educational Attainment, 24-47 y.o., 2000-2008 years.

Variables	**Educational Attainment** (1)	(2)	(3)	(4)	(5)	(6)
Education = Number of Years of Studies; OLS Model.						
♯ of 2TE slots, by Districts	0.320***	0.335***				
Size of 17y.o. Cohort, by Districts	0.045					
♯ of 2TE slots, in RF			0.064***	0.070***		
Size of 17y.o. Cohort, in RF			0.012			
Proportion of 2TE slots, by Districts					0.888**	
Proportion of 2TE slots, in RF						1.260***
F-statistics Exclusion	19.07	37.06	10.47	20.44	6.74	11.00
Education = Number of 2TE years of Studies; OLS Model.						
♯ of 2TE slots, by Districts	0.228***	0.263***				
Size of 17y.o. Cohort, by Districts	0.105**					
♯ of 2TE slots, in RF			0.033**	0.050***		
Size of 17y.o. Cohort, in RF			0.031***			
Proportion of 2TE slots, by Districts					0.314	
Proportion of 2TE slots, in RF						0.458
F-statistics Exclusion	16.00	26.93	11.91	15.82	1.23	1.95
Education = 2TE Degree; Probit Model.						
♯ of 2TE slots, by Districts	0.101***	0.123***				
Size of 17y.o. Cohort, by Districts	0.067**					
♯ of 2TE slots, in RF			0.014*	0.024***		
Size of 17y.o. Cohort, in RF			0.017***			
Proportion of 2TE slots, by Districts					0.065	
Proportion of 2TE slots, in RF						0.204
LR test, χ^2	18.91	10.77	19.46	8.59	0.14	1.07
LR test, P-Value	[0.000]	[0.001]	[0.000]	[0.003]	[0.712]	[0.301]
Year & Districts F.E.	X	X	X	X	X	X
Sex, Age, Age^2	X	X	X	X	X	X
*Districts * Year Clusters*	X	X	X	X	X	X
Observations: 28622						

Source: RLMS, 2000-2008, Author's calculations

size of youth cohorts). This variable passes the test of the instrument weakness (according to Angrist and Pischke (2009), the critical value for F-statistics is 10, below which we can argue that the instrument is weak).

The main potential problem with this instrument is the fact that it is perfectly correlated with region-cohort effects. This correlation occurs because of the identification of the exposure of an individual to the tertiary education expansion by his year of birth and district of residency. That means that there is no variations of the instrument within group of people born in one year and living in the same federal district. Therefore, it is necessary to assume that if there were any unobserved changes affecting wages, they were either occurring for all regions at the same time, or they were not correlated with the tertiary education system expansion in the regions over time. Unfortunately we have no possibility to test this assumption or to use another base of variation for evaluation of the influence of tertiary education system expansion on educational attainment. All the specifications account for the years and regions fixed effects, thus capturing the unobserved variations of other characteristics by regions and over the analyzed period of time. However, it is still necessary to assume that there were no other changes correlated with changes of the educational system on a region-year basis, which are not captured by years and federal districts fixed effects. Additionally, we control for the cohort year of birth fixed effects and region-specific linear cohort trends, when performing robustness checks. We also add the interactions between the number of available slots and sex. First, this helps us to overcome the described above correlation of the instrument with region-cohort-effect, because in this case the instrument varies per region, year, and sex. Second, this specification allows us to take into account that the higher education expansion have affected differently male and female youths. Additional discussion on the within-regions and over-time variation of the instruments is presented in *Appendix B*.

4.2 Returns to Education: Wages. OLS and IV estimations.

First, we discuss the OLS estimation results of the wage equation for the employed population 24-47 years old. *Table 4* shows the summary of estimated results for the instrument *Number of 2TE slots in the Educational System.* There are three specification with different measures of educational attainment. Column named (1) shows the regression of wages (in logarithms) on educational attainment, column named (2) shows the regressions of wages on educational attainment and the instrument, column named (3) - regressions only on the instrument (reduced form estimation of influence of the higher education system expansion on wages).

OLS results suggest that the returns to one year of schooling (any schooling or years of higher education studies) is equal to a 10% increase in wages; a higher education degree provides a 58% increase in wages (exp(0.456)=1.578), which is approximately 5 times higher than the influence of an increase in one year of higher education studies. Therefore, all our measures of educational attainment lead to similar results. Furthermore, these results suggest that the expansion of the tertiary education system has a positive effect on wages (see column (3) for results). At the same time, its influence becomes insignificant once we control for educational attainment (see column (2)). Therefore, the influence of the higher education expansion on wages occurs mainly through the changes in educational attainment and does not have a direct influence on future wages of the corresponding cohorts. In one of the next sections, the potential General Equilibrium effects of the higher education expansion on wages are also discussed and tested. Here, the main focus is the individual returns to education.

Second, we analyze the estimation results of the instrumental variable estimations of the wage equation. The described above expansion of the higher education system at regional levels in Russia is used as the instrument for educational attainment. The estimation procedures for the joint models of educational choice and wages are presented in *Technical Appendix*: *section 1.1* - for a number of years of schooling, *section 1.2* - for a discreet variable of higher education attainment.

Table 5 presents the estimation results for the number of years of schooling and higher education studies, and *Table 6* shows the results for the variable a higher education degree.

In columns (1) and (2) we present OLS estimations for comparison; in the column (2) we allow returns to education to vary by sex. Columns (3) show the IV estimation results. Columns (4) and (5) present the estimation results for the Maximum Likelihood estimations of the simultaneous equations model for the educational choice and wages. The similarity of results in the column (3) and (4) in *Table 5* suggests that the assumption about joint normality of residuals, which is necessary for the ML estimations, does not significantly influence the estimation results. At the same time, standard errors are significantly lower for ML estimators. However, such assumption allows us to estimate the differences in returns to education for the male and female population, which are presented in column (5) of both tables, and is crucial for further estimations.

Estimations of the returns to education, which are obtained using IV or MLE, are higher than the estimations by OLS. Returns to one year of schooling is estimated to be around 15% (exp(0.15)-1 = 16%), and returns to a higher education degree are approximately 80%. Returns to 2TE degree are higher for female population by 4%, however this difference is relatively small compared to the difference in wages between males and females estimated as 62%.

IV estimations of the returns to schooling in this case report the increase in wages for those who have obtained a higher education degree because of the higher education expansion and would not have gone to tertiary education otherwise (Angrist et al. (1996)). The results suggest that individuals who have benefited from the expansion of the higher education system in terms of educational attainment have also gained in terms of wages.

Table 4: Reduced Form Estimations of the Returns to Education, OLS. 24-47 y.o. The Employed Population, 2000-2008.

Variables	**(1) ln(Wage)**	**(2) ln(Wage)**	**(3) ln(Wage)**
1st Specification:			
Years of Schooling	0.098***	0.098***	
	(0.003)	(0.003)	
♯ of 2TE slots, by Districts		0.017	0.050**
		(0.022)	(0.023)
R^2	0.289	0.289	0.235
2nd Specification:			
Years of 2TE Studies	0.097***	0.097***	
	(0.002)	(0.002)	
♯ of 2TE slots, by Districts		0.025	0.050**
		(0.020)	(0.023)
R^2	0.286	0.286	0.235
3rd Specification:			
*2TE_Degree**	0.456***	0.455***	
	(0.011)	(0.011)	
♯ of 2TE slots, by Districts		0.030	0.050**
		(0.020)	(0.023)
R^2	0.280	0.280	0.235
Sex, Age, Age^2	X	X	X
Federal Districts & Years Fixed Effects	X	X	X
*Federal Districts*Year Clusters*	X	X	X
Observations	28622	28622	28622

Source: RLMS, 2000-2008.

Table 5: OLS, IV and MLE Estimations of the Returns to Education. 24-47 y.o. Employed Population, 2000-2008.

Variables	**(1)**	**(2)**	**(3)**		**(4)**		**(5)**	
	ln(Wage)	**ln(Wage)**	**Education**	**ln(Wage)**	**Education**	**ln(Wage)**	**Education**	**ln(Wage)**
1st Specification:								
♯ *Years of Schooling*	0.098^{***}	0.089^{***}		0.150^{**}		0.150^{***}		0.140^{***}
	(0.003)	(0.004)		(0.066)		(0.016)		(0.006)
♯ *Years of Schooling · Female**		0.019^{***}						0.019^{***}
		(0.005)						(0.005)
*Male**	0.501^{***}	0.727^{***}	-0.854^{***}	0.545^{***}	-0.854^{***}	0.545^{***}	-0.854^{***}	0.770^{***}
	(0.010)	(0.062)	(0.020)	(0.056)	(0.020)	(0.015)	(0.020)	(0.061)
♯ *of 2TE slots*			0.335^{***}		0.335^{***}		0.335^{***}	
			(0.055)		(0.054)		(0.054)	
$\rho(\varepsilon_1, \varepsilon_2)$						-0.143^{***}		-0.142^{***}
F test of excluded instruments:			37.06	[0.000]				
2nd Specification:								
♯ *Years of 2TE Studies*	0.097^{***}	0.090^{***}		0.191^{***}		0.190^{***}		0.183^{***}
	(0.002)	(0.004)		(0.073)		(0.013)		(0.013)
♯ *Years of 2TE Studies · Female**		0.013^{***}						0.013^{**}
		(0.005)						(0.005)
*Male**	0.459^{***}	0.478^{***}	-0.437^{***}	0.500^{***}	-0.437^{***}	0.500^{***}	-0.437^{***}	0.519^{***}
	(0.010)	(0.015)	(0.031)	(0.031)	(0.031)	(0.009)	(0.031)	(0.017)
♯ *of 2TE slots*			0.263^{***}		0.263^{***}		0.263^{***}	
			(0.051)		(0.053)		(0.054)	
$\rho(\varepsilon_1, \varepsilon_2)$						-0.250^{***}		-0.250^{***}
F test of excluded instruments:			26.93	[0.000]				
Sex, Age, Age^2	X	X	X	X	X	X	X	X
Districts & Years F.E.	X	X	X	X	X	X	X	X
*Districts*Year Clusters*	X	X	X	X	X	X	X	X
Observations	28622	28622	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008.

Table 6: OLS, IV and MLE Estimations of the Returns to Education. 24-47 y.o. The Employed Population, 2000-2008.

Variables	**(1)**	**(2)**	**(4)**		**(5)**	
	ln(Wage)	**ln(Wage)**	**Education**	**ln(Wage)**	**Education**	**ln(Wage)**
3rd Specification:						
*2TE Degree**	0.456***	0.430***		0.774***		0.749***
	(0.011)	(0.019)		(0.054)		(0.057)
2TE Degree · Female**		0.046*				0.036
		(0.024)				(0.025)
*Male**	0.455***	0.467***	−0.266***	0.482***	−0.383***	0.617***
	(0.010)	(0.014)	(0.017)	(0.010)	(0.083)	(0.061)
♯ of 2TE slots			0.128***		0.126***	
			(0.036)		(0.035)	
$\rho(\varepsilon_1, \varepsilon_2)$				−0.235***		−0.230***
Sex, Age, Age^2	X	X	X	X	X	X
Districts & Years F.E.	X	X	X	X	X	X
*Districts*Year Clusters*	X	X	X	X	X	X
Observations	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008.

4.3 Returns to Education: Employment and Wages.

In the previous section, the wages of the employed population were analyzed. However, education also influences employment probability. In this section, the empirical models account for the influence of education on employment and for the selection into employment.

Table 7 presents the estimation results for the effects of education on employment using Probit and Simultaneous Equations Model of Education and Employment (similar to the bivariate probit model in the 3rd case). The estimation procedure is described in *Technical Appendix*: *section 2.1* - for a number of years of schooling, *section 2.2* - for a higher education degree attainment. The estimation results suggest that education (measured as years of schooling or higher education degree attainment) significantly increases the probability of being employed. The IV-Probit estimations are similar to the Probit results, and the estimated correlations between random terms in educational choice and employment equations are not significantly different from zero in all specifications. One additional year of tertiary education increases the probability of being employed by 3% (marginal effect).

Furthermore, the joint three-equations model of educational choice, employment probability, and wages is estimated. Estimations are conducted by the Maximum Likelihood method. The estimation procedure is described in *Technical Appendix*: *section 3.1* - for a number of years of schooling, *section 3.2* - for a higher education degree attainment. The regional unemployment rates are used as the exclusion restriction for the employment equation. This specification allows for the correlations between unobservable components of these equations. *Table 8* lists the estimation results for returns to education, taking into account the selection into employment.

Correction for the selection into employment does not significantly change the estimated returns to education, but increases the estimated effects of education on employment. The correlations between unobservable terms in education, employment and wage equations are all negative and statistically significant. This fact suggests that those who have higher probabilities of tertiary education attainment would have lower wages and lower employment probabilities if they obtain a lower educational degree, compared to the low-educated workers. By taking into account the correlations of random terms, the model therefore controls for the self-selection of workers into education, based on their unobservable characteristics, which also affect further employment probability and wages.

Table 7: Probit and ML Estimations of the Influence of Education on Employment.
24-47 y.o., All Population, 2000-2008.

Variables	**(1) Probit Work**	**(2) Probit Work**	**(4) MLE Education**	**(4) MLE Work**	**(5) MLE Education**	**(5) MLE Work**
1st Specification:						
♯ Years of Schooling	0.100***	0.101***		0.084**		0.085***
	(0.003)	(0.005)		(0.040)		(0.028)
*♯ Years of Schooling · Female**		−0.001				−0.001
		(0.007)				(0.007)
*Male**	0.334***	0.322***	−0.780***	0.321***	−0.780***	0.309***
	(0.015)	(0.077)	(0.022)	(0.035)	(0.022)	(0.081)
♯ of 2TE slots			0.312***		0.312***	
			(0.054)		(0.052)	
$\rho(\varepsilon_1, \varepsilon_2)$			0.0355		0.0353	
χ^2			0.16	[0.689]	0.33	[0.565]
2nd Specification:						
♯ Years of 2TE Studies	0.080***	0.087***		0.092***		0.099***
	(0.003)	(0.006)		(0.030)		(0.028)
*♯ Years of 2TE Studies · Female**		−0.012				−0.012
		(0.007)				(0.007)
*Male**	0.278***	0.266***	−0.352***	0.283***	−0.352***	0.270***
	(0.014)	(0.016)	(0.021)	(0.019)	(0.021)	(0.019)
♯ of 2TE slots			0.277***		0.277***	
			(0.050)		(0.051)	
$\rho(\varepsilon_1, \varepsilon_2)$			−0.026		−0.024	
χ^2			0.18	[0.670]	0.17	[0.682]
3rd Specification:						
*2TE Degree**	0.389***	0.413***		0.389***		0.413***
	(0.018)	(0.030)		(0.019)		(0.026)
2TE Degree · Female**		−0.037				−0.037
		(0.037)				(0.038)
*Male**	0.276***	0.269***	−0.224***	0.276***	−0.224***	0.269***
	(0.014)	(0.016)	(0.020)	(0.027)	(0.018)	(0.025)
♯ of 2TE slots				0.129***		0.129***
				(0.033)		(0.032)
$\rho(\varepsilon_1, \varepsilon_2)$			−0.000		−0.000	
χ^2			0.001	[0.98]	0.001	[0.98]
Sex, Age, Age^2	*X*	*X*	*X*	*X*	*X*	*X*
Districts & Years F.E.	*X*	*X*	*X*	*X*	*X*	*X*
*Districts*Year Clusters*	*X*	*X*	*X*	*X*	*X*	*X*
Observations	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008.

Table 8: Maximum Likelihood Estimations of the Influence of Education on Employment and Wages: 3-equations Model. 24-47 y.o., All Population, 2000-2008.

Variables	**(1) 3 equations: MLE**			**(2) 3 equations:MLE**		
	Education	**Work**	**Wages**	**Education**	**Work**	**Wages**
			1st Specification:			
♯ *Years of Schooling*		0.159***	0.119***		0.157***	0.110***
		(0.032)	(0.041)		(0.024)	(0.025)
♯ *Years of Schooling · Female**					0.009	0.013***
					(0.006)	(0.005)
*Male**	−0.780***	0.380***	0.437***	−0.780***	0.052	0.584***
	(0.022)	(0.027)	(0.034)	(0.022)	(0.085)	(0.061)
♯ *of 2TE slots*	0.312***			0.312***		
	(0.053)			(0.052)		
Covariance Matrix:	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
ε_1	4.878***			4.878***		
ε_2	−0.357**	1		−0.388***	1	
ε_3	−0.253	−0.657***	0.833***	−0.233*	−0.634***	0.811***
χ^2: $\rho(\varepsilon_1, \varepsilon_2)$	4.74	P-Value: [0.029]		11.09	P-Value: [0.000]	
χ^2: $\rho(\varepsilon_1, \varepsilon_3)$	1.69	P-Value: [0.194]		3.76	P-Value: [0.053]	
χ^2: $\rho(\varepsilon_2, \varepsilon_3)$	851.9	P-Value: [0.000]		849.3	P-Value: [0.000]	
			2nd Specification:			
♯ *Years of 2TE Studies*		0.172***	0.146***		0.183***	0.136***
		(0.014)	(0.042)		(0.028)	(0.049)
♯ *Years of 2TE Studies · Female**					−0.001	0.012***
					(0.007)	(0.005)
*Male**	−0.352***	0.306***	0.393***	−0.352***	−0.112***	0.405***
	(0.021)	(0.014)	(0.018)	(0.021)	(0.040)	(0.021)
♯ *of 2TE slots*	0.276***			0.276***		
	(0.049)			(0.050)		
Covariance Matrix:	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
ε_1	4.449***			4.449***		
ε_2	−0.470***	1		−0.527***	1	
ε_3	−0.328*	−0.657***	0.862***	−0.310	−0.628***	0.839***
χ^2: $\rho(\varepsilon_1, \varepsilon_2)$	52.17	P-Value: [0.000]		16.54	P-Value: [0.000]	
χ^2: $\rho(\varepsilon_1, \varepsilon_3)$	3.25	P-Value: [0.071]		2.18	P-Value: [0.140]	
χ^2: $\rho(\varepsilon_2, \varepsilon_3)$	437.8	P-Value: [0.000]		264.28	P-Value: [0.000]	

Continued on next page

Table 8 – continued from previous page

Variables	(1) 3 equations: MLE Education	Work	Wages	(2) 3 equations: MLE Education	Work	Wages
			3rd Specification:			
*2TE Degree**		0.887^{***}	0.820^{***}		0.892^{***}	0.797^{***}
		(0.053)	(0.045)		(0.049)	(0.049)
2TE Degree · Female**					0.005	0.025
					(0.037)	(0.027)
*Male**	-0.226^{***}	0.476^{***}	0.388^{***}	-0.224^{***}	0.476^{***}	0.394^{***}
	(0.016)	(0.032)	(0.013)	(0.017)	(0.035)	(0.018)
♯ of 2TE slots	0.134^{***}			0.134^{***}		
	(0.034)			(0.033)		
Covariance Matrix:	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
ε_1	1			1		
ε_2	-0.330^{***}	1		-0.336^{***}	1	
ε_3	-0.280^{***}	-0.670^{***}	0.900^{***}	-0.275^{***}	-0.670^{***}	0.898^{***}
χ^2: $\rho(\varepsilon_1, \varepsilon_2)$	111.92	P-Value: [0.000]		130.83	P-Value: [0.000]	
χ^2: $\rho(\varepsilon_1, \varepsilon_3)$	145.36	P-Value: [0.000]		134.04	P-Value: [0.000]	
χ^2: $\rho(\varepsilon_2, \varepsilon_3)$	1327.38	P-Value: [0.000]		2052.57	P-Value: [0.000]	
Unemployment · Sex		*X*			*X*	
Sex, Age, Age^2	*X*	*X*	*X*	*X*	*X*	*X*
Districts & Years F.E.	*X*	*X*	*X*	*X*	*X*	*X*
*Districts*Year Clusters*	*X*	*X*	*X*	*X*	*X*	*X*
Observations	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008.

5 Robustness Check

In this section four robustness checks are presented. The goal is to verify if the fact that the instrument (number of slots in the higher education system) is inseparable from cohort effect on a regional basis could potentially bias the results presented in the previous sections. Because of this inseparability (perfect correlation) our results may be biased in the following cases:

1. Different cohorts come to the labor market at different time (even at the regional levels). Therefore, different cohorts could experience different labor market conditions during the first period in the labor market. First job, first experience, beginning wages, and other characteristics could significantly affect wages and employment prospective later. If these effects of the labor-market-entry-time for individuals are correlated with the number of slots in the higher education system in the year when they turned 17, the returns to education will be overestimated or underestimated. In order to account for this problem, we additionally include the dummy variables for the years of birth into all equations (educational choice, employment, and wages). Moreover, we include district-specific linear cohort time trends in order to account for unobserved changes within regions over time.

2. There could be a significant correlation between unobserved cohort characteristics (such as ability, preferences for education, etc.) and the number of slots in the higher education system at the region-year level. If this correlation is positive, the returns to education will be overestimated. The significance of these effects can be decreased by introducing the interaction variables between the characteristics of the system and sex.

3. It is important to take into account possible General Equilibrium effects of the higher education system expansion. For example, in the year of an increasing number of university graduates coming to the labor market, we can observe a general equilibrium effect of the increasing supply of college graduates - decrease in their wages. First, we account for the cohorts' effects on general equilibrium of the labor market by introducing the year of birth dummy variables. Second, we estimate the changes in the returns to education over the analyzed period.

5.1 Instrument and Cohort Effects

There is no any measures of individual ability available in our database, and it is difficult to account for all labor market conditions that could potentially affect wages and are correlated with the number of slots in the higher education system. The variables describing unemployment, GDP growth, and other macro characteristics cannot account completely for all labor-market conditions at the moment of entry. Therefore, we conduct a test by including the dummy variables for years of birth - cohorts fixed effects. Thus, we include 30 dummy variables for the 1953-1983 years of birth of the cohorts. This modification allows us to account for the described above first and partially third cases.

Table 9 presents the estimation results for the 1st stage of the model - educational choices.

The estimated coefficients are shown for two cases: with control for the year-of-birth fixed effects [(1), (3), (5), (7)] and without [(2), (4), (6), (8)]. Estimations are conducted for the three educational outcomes - number of years of schooling [(1)-(2)], number of years of 2TE studies [(3)-(4)], and higher education (2TE) degree attainment [(5)-(8)]. The estimated coefficients for the influence of the higher education system capacity on educational attainment are similar and significant. However, the explanatory power of the instrument decreases by half.

To explore how it might affect the estimations of the returns to education, we include the years of birth fixed effects in all equations: for educational choice, employment and wages. The estimation results for two-equations model (educational choice and wages) are presented in *Table 10*, for three-equations model (educational choice, employment and wages) - in *Table 11.* Estimations (1) and (3) do not account for the year of birth fixed effects, while the models (2) and (4) include 30 dummy variables for the year of birth in all equations. Additionally, models (3) and (4) take into account the possible heterogeneity in the returns to education among the male and female populations. The estimation results of the returns to education with the control for the year of birth fixed effects are similar to the results without control for the years of birth (thus, to the estimation results in the previous sections). Therefore, the correlation of our instrument with cohort effects does not seem to bias the estimation results, because they are robust to the control for the year of birth fixed effects. In the third section of Robustness Checks, the model, which also accounts for the region-specific linear cohort trends, is estimated.

Table 9: Influence of the Higher Education System Expansion on Educational Attainment, 24-47 y.o., 2000-2008 years. Additional Controls - Cohort Year-Birth Fixed Effects.

Variables	(1) ~ OLS **Years of Schooling**	(2) ~ OLS **Years of Schooling**	(3) ~ OLS **Years of 2TE**	(4) ~ OLS **Years of 2TE**
♯ *of 2TE slots, by Districts*	0.202	0.335***	0.245***	0.263***
	(0.129)	(0.055)	(0.089)	(0.051)
Year & Districts F.E.	X	X	X	X
Cohort Year-Birth Fixed Effects	X		X	
F-statistics Exclusion	2.45	37.06	7.63	26.93
P-Value	[0.123]	[0.000]	[0.008]	[0.000]

Variables	(5) ~ Probit **2TE Degree**	(6) ~ Probit **2TE Degree**	(7) ~ OLS **2TE Degree**	(8) ~ OLS **2TE Degree**
♯ *of 2TE slots, by Districts*	0.115**	0.123***	0.043**	0.043***
	(0.059)	(0.034)	(0.019)	(0.011)
Year & Districts F.E.	X	X	X	X
Cohort Year-Birth Fixed Effects	X		X	
F-statistics Exclusion			5.09	15.76
P-Value			[0.028]	[0.000]
LR test, χ^2	2.56	10.77		
LR test, P-Value	[0.109]	[0.001]		
Observations:	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 10: ML Estimations of the Joint Model of Educational Attainment and Wages, 24-47 y.o., 2000-2008 years.
Additional Controls - Cohort Year-Birth Fixed Effects.

Variables	(1) ∼ ML **2TE Degree**	(1) ∼ ML **Wage**	(2) ∼ ML **2TE Degree**	(2) ∼ ML **Wage**
*2TE Degree**		0.774^{***}		0.776^{***}
		(0.054)		(0.039)
♯ of 2TE slots, by Districts	0.128^{***}		0.114^{*}	
	(0.036)		(0.061)	
*Male**	-0.266^{***}	0.482^{***}	-0.268^{***}	0.484^{***}
	(0.017)	(0.010)	(0.017)	(0.011)
$\rho(\varepsilon_1, \varepsilon_2)$	-0.235^{***}		-0.238^{***}	
chi^2, P-Value	46.51	[0.000]	83.20	[0.000]
Birth Cohort Fixed Effects			X	X
Districts & Year F.E.	X	X	X	X
Sex, Age, Age^2	X	X	X	X
Districts · Year Clusters	X	X	X	X

Variables	(3) ∼ ML **2TE Degree**	(3) ∼ ML **Wage**	(4) ∼ ML **2TE Degree**	(4) ∼ ML **Wage**
*2TE Degree**		0.749^{***}		0.744^{***}
		(0.057)		(0.054)
2TE Degree · Female**		0.036		0.040
		(0.025)		(0.025)
♯ of 2TE slots, by Districts	0.126^{***}		0.115^{**}	
	(0.035)		(0.058)	
*Male**	-0.383^{***}	0.617^{***}	-0.374^{***}	0.623^{***}
	(0.083)	(0.061)	(0.083)	(0.061)
$\rho(\varepsilon_1, \varepsilon_2)$	-0.230^{***}		-0.230^{***}	
chi^2, P-Value	53.05	[0.000]	56.60	[0.000]
Birth Cohort Fixed Effects			X	X
Districts & Year F.E.	X	X	X	X
Sex, Age, Age^2	X	X	X	X
Sex · Age, Sex · Age^2	X	X	X	X
Districts · Year Clusters	X	X	X	X
Observations:	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 11: ML Estimations of the Joint Model of Educational Attainment, Employment and Wages, 24-47 y.o., 2000-2008 years. Additional Controls - Cohort Year-Birth Fixed Effects.

Variables	(1) ∼ ML 2TE Degree	(1) ∼ ML Work	(1) ∼ ML Wage	(2) ∼ ML 2TE Degree	(2) ∼ ML Work	(2) ∼ ML Wage
*2TE Degree**		0.887***	0.820***		0.888***	0.820***
		(0.053)	(0.045)		(0.051)	(0.056)
*Male**	−0.226***	0.476***	0.388***	−0.229***	0.479***	0.389***
	(0.016)	(0.032)	(0.013)	(0.016)	(0.032)	(0.013)
♯ of 2TE slots, by Districts	0.134***			0.134**		
	(0.034)			(0.056)		
$\rho(\varepsilon_1, \varepsilon_2)$	−0.330***			−0.328***		
$\rho(\varepsilon_1, \varepsilon_3)$	−0.295***			−0.297***		
$\rho(\varepsilon_2, \varepsilon_3)$	−0.707***			−0.708***		
χ^2: $\rho(\varepsilon_1, \varepsilon_2)$	111.92	P-Value:	[0.000]	122.40	P-Value:	[0.000]
χ^2: $\rho(\varepsilon_1, \varepsilon_3)$	145.36	P-Value:	[0.000]	106.22	P-Value:	[0.000]
χ^2: $\rho(\varepsilon_2, \varepsilon_3)$	1327.38	P-Value:	[0.000]	1351.87	P-Value:	[0.000]
Birth Cohort Fixed Effects				*X*	*X*	*X*
Districts & Year F.E.	*X*	*X*	*X*	*X*	*X*	*X*
Sex, Age, Age^2	*X*	*X*	*X*	*X*	*X*	*X*
Districts · Year Clusters	*X*	*X*	*X*	*X*	*X*	*X*

Variables	(3) ∼ ML 2TE Degree	(3) ∼ ML Work	(3) ∼ ML Wage	(4) ∼ ML 2TE Degree	(4) ∼ ML Work	(4) ∼ ML Wage
*2TE Degree**		0.892***	0.797***		0.892***	0.795***
		(0.049)	(0.049)		(0.055)	(0.044)
2TE Degree · Female**		0.005	0.025		0.006	0.027
		(0.037)	(0.027)		(0.039)	(0.027)
*Male**	−0.224***	0.476***	0.394***	−0.227***	0.479***	0.395***
	(0.017)	(0.035)	(0.018)	(0.017)	(0.034)	(0.018)
♯ of 2TE slots, by Districts	0.134***			0.134**		
	(0.033)			(0.055)		
$\rho(\varepsilon_1, \varepsilon_2)$	−0.336***			−0.334***		
$\rho(\varepsilon_1, \varepsilon_3)$	−0.290***			−0.291***		
$\rho(\varepsilon_2, \varepsilon_3)$	−0.707***			−0.708***		
χ^2: $\rho(\varepsilon_1, \varepsilon_2)$	130.83	P-Value:	[0.000]	172.41	P-Value:	[0.000]
χ^2: $\rho(\varepsilon_1, \varepsilon_3)$	134.04	P-Value:	[0.000]	268.32	P-Value:	[0.000]
χ^2: $\rho(\varepsilon_2, \varepsilon_3)$	2052.57	P-Value:	[0.000]	1343.99	P-Value:	[0.000]
Birth Cohort Fixed Effects				*X*	*X*	*X*
Districts & Year F.E.	*X*	*X*	*X*	*X*	*X*	*X*
Sex, Age, Age^2*, Sex · Age*	*X*	*X*	*X*	*X*	*X*	*X*
Districts · Year Clusters	*X*	*X*	*X*	*X*	*X*	*X*
Observations:	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008, Author's calculations

5.2 Increasing Instrument's Explanation Power: Introducing Interactions with Other Characteristics

The increasing number of slots in the higher education system affects the male and female populations differently. The evidence of this heterogeneous influence is analyzed by Kyui (2011). Taking into account the different influences of the tertiary education expansion on educational attainment also allows to increase the explanatory power of the 1st equation in the models (educational choice equation).

Table 12 lists the estimated coefficients for the 1st stage equation: educational attainment under the assumption of the different influences of the higher education expansion for the male and female populations. Columns (1)-(2) list the results for the number of years of schooling, (3)-(4) - for the number of years of 2TE studies, and finally, (5)-(8) for a higher education degree attainment. These results suggest that by introducing interactions between sex and the number of slots, the explanatory power of the 1st-stage equations increases. Moreover, even with the control for the year of birth fixed effects, the instruments (the number of slots and the number of slots interacted with the dummy variable for female population) pass the instrument-weakness test discussed above. However, there is no any changes in the explanatory power for the equation of the number of years of schooling, thus there is no heterogeneous influence of the higher education expansion for this variable. For two other measures of educational attainment, the number of slots had a higher influence for female educational attainment.

Furthermore, we estimate the two-equations model (educational attaintment and wages) and the three-equations model (educational attainment, employment and wages) for the variable higher education attainment controlling for the interactions between the instrument and sex in the 1st-stage equation. The estimation results are presented in *Table 13* and *Table 14* accordingly. The magnitude of the returns to education is lower when models account for the different influence of the system expansion on male and female educational choices, but the estimated coefficients are all around 80%. Differences in returns to education between male and female workers are not statistically significant.

Table 12: Influence of the Higher Education System Expansion on Educational Attainment, 24-47 y.o., 2000-2008 years. Cohort Year-Birth Fixed Effects & ♯ of 2TE slots Interactions with Sex.

Variables	(1) ∼ OLS **Years of Schooling**	(2) ∼ OLS **Years of Schooling**	(3) ∼ OLS **Years of 2TE**	(4) ∼ OLS **Years of 2TE**
♯ of 2TE slots, by Districts	0.202	0.338***	0.174*	0.184***
	(0.129)	(0.060)	(0.093)	(0.056)
*♯ of 2TE slots · Female**	0.000	−0.005	0.190***	0.182***
	(0.042)	(0.041)	(0.043)	(0.042)
*Male**	−0.852***	−0.860***	−0.216***	−0.222***
	(0.053)	(0.052)	(0.057)	(0.057)
Year & Districts F.E.	X	X	X	X
Cohort Year-Birth Fixed Effects	X		X	
F-statistics Exclusion	1.23	18.80	16.86	27.70
P-Value	[0.300]	[0.000]	[0.000]	[0.000]

Variables	(5) ∼ Probit **2TE Degree**	(6) ∼ Probit **2TE Degree**	(7) ∼ OLS **2TE Degree**	(8) ∼ OLS **2TE Degree**
♯ of 2TE slots, by Districts	0.076	0.078**	0.028	0.026**
	(0.061)	(0.035)	(0.020)	(0.012)
*♯ of 2TE slots · Female**	0.102***	0.096***	0.041***	0.039***
	(0.022)	(0.021)	(0.007)	(0.007)
*Male**	−0.143***	−0.148***	−0.037***	−0.039***
	(0.033)	(0.034)	(0.011)	(0.011)
Year & Districts F.E.	X	X	X	X
Cohort Year-Birth Fixed Effects	X		X	
F-statistics Exclusion			22.05	25.12
P-Value			[0.000]	[0.000]
LR test, χ^2	17.98	24.66		
LR test, P-Value	[0.000]	[0.000]		
Observations:	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 13: ML Estimations of the Joint Model of Educational Attainment and Wages, 24-47 y.o., 2000-2008 years.
Interactions of ♯ 2TE Slots with Sex, Cohort Year-Birth Fixed Effects.

Variables	(1) ∼ ML **2TE Degree**	(1) ∼ ML **Wage**	(2) ∼ ML **2TE Degree**	(2) ∼ ML **Wage**
*2TE Degree**		0.757***		0.758***
		(0.043)		(0.048)
♯ of 2TE slots, by Districts	0.086**		0.078	
	(0.037)		(0.059)	
*♯ of 2TE slots · Female**	0.089***		0.095***	
	(0.021)		(0.022)	
*Male**	−0.158***	0.481***	−0.154***	0.483***
	(0.033)	(0.009)	(0.034)	(0.010)
$\rho(\varepsilon_1, \varepsilon_2)$	−0.223***		−0.225***	
chi^2*, P-Value*	65.77	[0.000]	55.26	[0.000]
Birth Cohort Fixed Effects			X	X
Districts & Year F.E.	X	X	X	X
Sex, Age, Age^2	X	X	X	X
Districts · Year Clusters	X	X	X	X

Variables	(3) ∼ ML **2TE Degree**	(3) ∼ ML **Wage**	(4) ∼ ML **2TE Degree**	(4) ∼ ML **Wage**
*2TE Degree**		0.733***		0.730***
		(0.098)		(0.101)
2TE Degree · Female**		0.036		0.040
		(0.027)		(0.026)
♯ of 2TE slots, by Districts	0.085**		0.076	
	(0.037)		(0.063)	
*♯ of 2TE slots · Female**	0.093***		0.099***	
	(0.021)		(0.022)	
*Male**	−0.152***	0.610***	−0.148***	0.618***
	(0.033)	(0.063)	(0.033)	(0.060)
$\rho(\varepsilon_1, \varepsilon_2)$	−0.219***		−0.219***	
chi^2*, P-Value*	13.34	[0.000]	11.93	[0.001]
Birth Cohort Fixed Effects			X	X
Districts & Year F.E.	X	X	X	X
Sex, Age, Age^2	X	X	X	X
Sex · Age, Sex · Age^2	X	X	X	X
Districts · Year Clusters	X	X	X	X
Observations:	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 14: ML Estimations of the Joint Model of Educational Attainment, Employment and Wages, 24-47 y.o., 2000-2008 years. Cohort Year-Birth Fixed Effects & ♯ of 2TE slots Interactions with Sex.

Variables	(1) ∼ ML			(2) ∼ ML		
	2TE Degree	**Work**	**Wage**	**2TE Degree**	**Work**	**Wage**
*2TE Degree**		0.868***	0.830***		0.867***	0.831***
		(0.047)	(0.037)		(0.052)	(0.033)
*Male**	−0.124***	0.487***	0.389***	−0.122***	0.491***	0.390***
	(0.030)	(0.033)	(0.014)	(0.029)	(0.032)	(0.014)
♯ of 2TE slots, by Districts	0.086***			0.087		
	(0.033)			(0.057)		
*♯ of 2TE slots · Female**	0.086***			0.090***		
	(0.018)			(0.018)		
$\rho(\varepsilon_1, \varepsilon_2)$	−0.318***			−0.315***		
$\rho(\varepsilon_1, \varepsilon_3)$	−0.302***			−0.303***		
$\rho(\varepsilon_2, \varepsilon_3)$	−0.710***			−0.711***		
Birth Cohort Fixed Effects				X	X	X
Districts & Year F.E.	X	X	X	X	X	X
Sex, Age, Age^2	X	X	X	X	X	X
Districts · Year Clusters	X	X	X	X	X	X
Variables	**(3) ∼ ML**			**(4) ∼ ML**		
	2TE Degree	**Work**	**Wage**	**2TE Degree**	**Work**	**Wage**
*2TE Degree**		0.937***	0.776***		0.929***	0.778***
		(0.047)	(0.046)		(0.052)	(0.059)
2TE Degree · Female**		0.008	0.022		0.010	0.023
		(0.037)	(0.026)		(0.037)	(0.026)
*Male**	0.000	1.273***	0.123*	0.019	1.279***	0.128*
	(0.083)	(0.077)	(0.069)	(0.084)	(0.076)	(0.069)
♯ of 2TE slots, by Districts	0.078**			0.078		
	(0.033)			(0.057)		
*♯ of 2TE slots · Female**	0.101***			0.105***		
	(0.019)			(0.019)		
$\rho(\varepsilon_1, \varepsilon_2)$	−0.356***			−0.352***		
$\rho(\varepsilon_1, \varepsilon_3)$	−0.277***			−0.280***		
$\rho(\varepsilon_2, \varepsilon_3)$	−0.696***			−0.698***		
Birth Cohort Fixed Effects				X	X	X
Districts & Year F.E.	X	X	X	X	X	X
Sex, Age, Age^2*, Sex · Age*	X	X	X	X	X	X
Districts · Year Clusters	X	X	X	X	X	X
Observations:	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008, Author's calculations

5.3 Region-Specific Cohort Trends

In addition to federal districts and cohort fixed effects, we account for the district-specific linear cohort trends. The regional linear cohort trends are included in all equations: educational choice and wages. The estimation results of the joint model of higher education attainment and wages are presented in *Table 15.* In comparison to (1) and (2), specifications (3) and (4) include different returns to education for males and females, as well as different responses of the male and female populations to the expansion of the tertiary education system. Models (2) and (4) include the district-specific linear cohort trends, according to the seven Russian federal districts, in both equations - higher education attainment and wages. The explanation power of the instrument decreases when the regional cohorts' time-trends are controlled for. However, the estimated returns to education are similar in the cases without and with the control for the cohorts' trends within the regions, nevertheless, they are slightly higher in the second cases. Therefore, such specification accounts for the unobserved linear time changes within regions that were occurring at the same time as the expansion of the higher education system (for example, changes in the labor market). The estimation results are robust to the inclusion of region specific linear cohort trends.

Table 15: ML Estimations of the Joint Model of Educational Attainment and Wages, Cohort Fixed Effects & District-Specific Linear Cohort Trends.

Variables	(1) ∼ ML		(2) ∼ ML	
	2TE	**Wage**	**2TE**	**Wage**
*2TE Degree**		0.779^{***}		0.783^{***}
		(0.063)		(0.047)
*Male**	-0.268^{***}	0.484^{***}	-0.268^{***}	0.484^{***}
	(0.018)	(0.011)	(0.017)	(0.010)
♯ of 2TE slots, by Districts	0.112^{*}		0.092	
	(0.060)		(0.069)	
$\rho(\varepsilon_1, \varepsilon_2)$		-0.238^{***}		-0.241^{***}
Districts, Year & Cohort F.E.	X	X	X	X
Regional Linear Cohort Trend			X	X
Variables	(3) ∼ ML		(4) ∼ ML	
	2TE	**Wage**	**2TE**	**Wage**
*2TE Degree**		0.722^{***}		0.728^{***}
		(0.109)		(0.090)
2TE Degree · Female**		0.037		0.036
		(0.025)		(0.024)
*Male**	-0.153^{***}	0.491^{***}	-0.155^{***}	0.491^{***}
	(0.033)	(0.014)	(0.033)	(0.014)
♯ of 2TE slots, by Districts	0.076		0.047	
	(0.061)		(0.075)	
*♯ of 2TE slots · Female**	0.094^{***}		0.092^{***}	
	(0.022)		(0.021)	
$\rho(\varepsilon_1, \varepsilon_2)$		-0.212^{***}		-0.216^{***}
Districts, Year & Cohort F.E.	X	X	X	X
Regional Linear Cohort Trend			X	X
Observations:	28622	28622	28622	28622

Source: RLMS, 2000-2008, 24-47 y.o., Author's calculations

5.4 Changes in the Returns to Education over time

This section examines possible general equilibrium effects of the tertiary education expansion on wages for all workers. Previous studies argue that the variations in the returns to education significantly depend on the variations in the supply of college graduates: Katz and Murphy (1992), Card and Lemieux (2001a). Increasing supply of college graduates decreases their wages in the labor market equilibrium, except the two following cases: 1) perfectly elastic demand for higher education graduates in the labor market; 2) increasing demand for higher education graduates over the same period of time. The following empirical models control for the changes in wages and returns to education over the analyzed period of time 2000-2008.

First, *Table 16* presents the estimation results of the wage equation, controlling for the time varying returns to education and wages. The wage equation is estimated for the same measures of education (variable *"Education"*): number of years of schooling (*I*), number of years of higher education studies (*II*), and higher education attainment (*III*). The model includes a linear trend in the returns to education (columns *(1)* - variable *"Education"* · *Time-Trend*) and interaction variables of education and years, for which wages are observed (columns *(2)*). Correspondingly, control variables include linear time-changes in wages for all workers and years fixed effects. The estimated coefficients for the time variation in the returns to education are not statistically significant. These coefficients are low in magnitude and any tendencies for negative or positive changes in the returns to education can not be revealed. However, there is a positive linear trend in wages for all workers over the analyzed period of time - approximately 12% of yearly increase in wages.

Second, *Table 17* lists the estimation results for the time varying influence of education on employment. Similarly, there is no statistically significant changes in this influence during the analyzed period.

Finally, we estimate these time varying returns to education using the instrument - number of slots in the higher education system in a year when an individual turned 17. The estimations are conducted by the Maximum Likelihood method. The 1st equation for educational attainment is in the form of linear equation for the first two measures of education, and in the form of Probit for higher education degree attainment. The estimation results are presented in *Table 18*. These results also suggest that there were no significant changes in the returns to education over 2000-2008. Notably, the coefficients for linear and non-linear variations in the returns to education are not statistically significant. Moreover, estimated coefficients have a low magnitude and randomly changing signs over the analyzed period of time.

Therefore, the estimation results did not reveal the general equilibrium effects of the higher education expansion on wages. In average, all wages were growing from 2000 to 2008, however returns to education were stable. These results suggest that the effects of the growing supply of college graduates on wages are compensated by the growing demand for highly-qualified workers.

Table 16: Time-Varying Returns to Education, OLS, 24-47 y.o., 2000-2008 years.

Variables	**I. "Education"= Years of Schooling**		**II. "Education"= Years of 2TE Studies**		**III. "Education"= 2TE Degree**	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
	ln(Wage)	**ln(Wage)**	**ln(Wage)**	**ln(Wage)**	**ln(Wage)**	**ln(Wage)**
"Education"	0.099***	0.104***	0.093***	0.096***	0.452***	0.456***
	(0.007)	(0.010)	(0.006)	(0.008)	(0.025)	(0.032)
"Education" · Time-Trend	−0.000		0.001		0.001	
	(0.001)		(0.001)		(0.004)	
"Education" · 2001		−0.016		−0.010		−0.039
		(0.015)		(0.012)		(0.052)
"Education" · 2002		−0.005		−0.003		−0.001
		(0.012)		(0.010)		(0.045)
"Education" · 2003		0.002		0.008		0.032
		(0.012)		(0.009)		(0.036)
"Education" · 2004		−0.016		−0.005		−0.030
		(0.013)		(0.010)		(0.042)
"Education" · 2005		−0.004		0.002		0.006
		(0.013)		(0.011)		(0.052)
"Education" · 2006		−0.003		0.009		0.041
		(0.014)		(0.011)		(0.048)
"Education" · 2007		−0.003		0.006		0.018
		(0.013)		(0.010)		(0.040)
"Education" · 2008		−0.010		−0.003		−0.036
		(0.012)		(0.010)		(0.040)
Time-Trend-2000-2008	0.129***		0.125***		0.127***	
	(0.014)		(0.004)		(0.004)	
Year-2001		0.431**		0.254***		0.251***
		(0.170)		(0.034)		(0.038)
Year-2002		0.454***		0.398***		0.397***
		(0.130)		(0.025)		(0.028)
Year-2003		0.455***		0.465***		0.472***
		(0.138)		(0.026)		(0.028)
Year-2004		0.776***		0.595***		0.599***
		(0.154)		(0.028)		(0.030)
Year-2005		0.766***		0.718***		0.724***
		(0.156)		(0.030)		(0.032)
Year-2006		0.899***		0.848***		0.853***
		(0.158)		(0.031)		(0.032)
Year-2007		0.980***		0.933***		0.939***
		(0.153)		(0.029)		(0.029)
Year-2008		1.222***		1.096***		1.110***
		(0.133)		(0.028)		(0.030)
Districts Fixed Effects	*X*	*X*	*X*	*X*	*X*	*X*
Sex, Age, Age^2, Sex · Age	*X*	*X*	*X*	*X*	*X*	*X*
Districts · Year Clusters	*X*	*X*	*X*	*X*	*X*	*X*
Observations:	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 17: Time-Varying Influence of Education on Employment, Probit, 24-47, 2000-08.

Variables	**I. "Education"= Years of Schooling**		**II. "Education"= Years of 2TE Studies**		**III. "Education"= 2TE Degree**	
	(1) Work	**(2) Work**	**(1) Work**	**(2) Work**	**(1) Work**	**(2) Work**
"Education"	0.105***	0.099***	0.087***	0.087***	0.444***	0.424***
	(0.010)	(0.015)	(0.009)	(0.010)	(0.044)	(0.055)
"Education" · Time-Trend	−0.001		−0.001		−0.010	
	(0.002)		(0.001)		(0.008)	
"Education" · 2001		0.002		−0.003		0.017
		(0.018)		(0.014)		(0.075)
"Education" · 2002		0.000		−0.013		−0.064
		(0.020)		(0.017)		(0.081)
"Education" · 2003		−0.004		−0.013		−0.042
		(0.020)		(0.015)		(0.074)
"Education" · 2004		0.017		0.003		0.027
		(0.021)		(0.017)		(0.090)
"Education" · 2005		0.012		0.002		0.016
		(0.019)		(0.016)		(0.072)
"Education" · 2006		0.002		−0.001		−0.038
		(0.016)		(0.012)		(0.062)
"Education" · 2007		−0.010		−0.024		−0.125
		(0.024)		(0.016)		(0.092)
"Education" · 2008		−0.008		−0.014		−0.070
		(0.020)		(0.015)		(0.080)
Time-Trend-2000-2008	0.037*		0.026***		0.026***	
	(0.022)		(0.003)		(0.003)	
Year-2001		−0.042		−0.015		−0.019
		(0.213)		(0.027)		(0.027)
Year-2002		−0.013		0.005		0.005
		(0.246)		(0.035)		(0.033)
Year-2003		0.091		0.057**		0.054**
		(0.240)		(0.026)		(0.024)
Year-2004		−0.116		0.078***		0.080***
		(0.246)		(0.028)		(0.027)
Year-2005		−0.054		0.072**		0.075***
		(0.226)		(0.031)		(0.028)
Year-2006		0.107		0.125***		0.132***
		(0.184)		(0.021)		(0.022)
Year-2007		0.282		0.188***		0.187***
		(0.281)		(0.031)		(0.029)
Year-2008		0.252		0.164***		0.169***
		(0.239)		(0.032)		(0.033)
Districts Fixed Effects	X	X	X	X	X	X
Sex, Age, Age^2, Sex · Age	X	X	X	X	X	X
Districts · Year Clusters	X	X	X	X	X	X
Observations:	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008, Author's calculations

Table 18: Time-Varying Returns to Education, MLE, 24-47 y.o., 2000-2008 years.

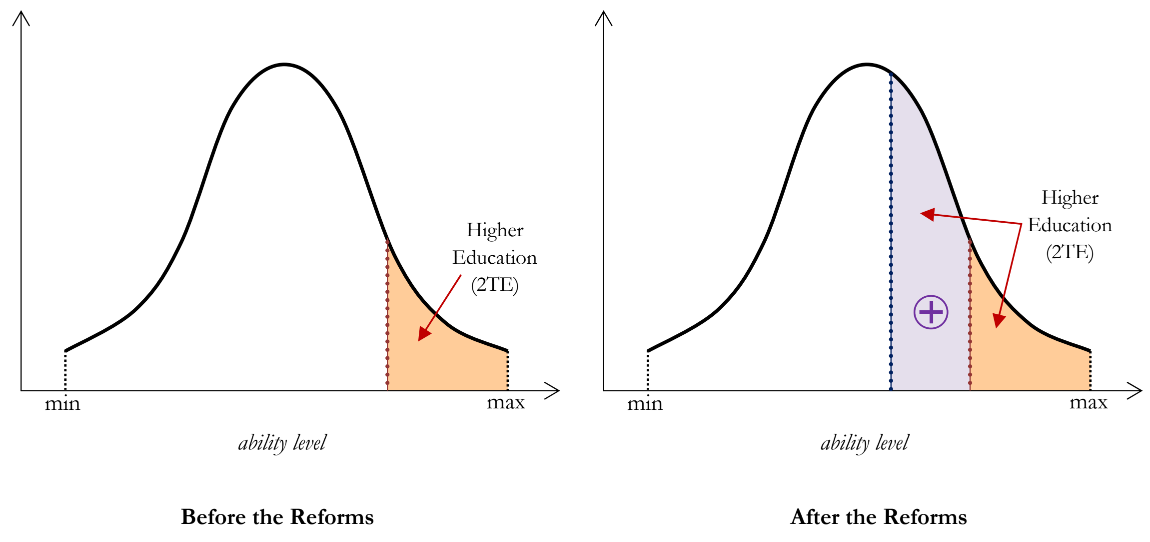
Variables	**I. "Education"= Years of Schooling (1)**	**(2)**	**II. "Education"= Years of 2TE Studies (1)**	**(2)**	**III. "Education"= 2TE Degree (1)**	**(2)**
2nd Equation:	**ln(Wage)**					
"Education"	0.150***	0.157***	0.183***	0.190***	0.774***	0.782***
	(0.032)	(0.017)	(0.006)	(0.009)	(0.032)	(0.066)
"Education" · Time-Trend	−0.000		0.001		0.000	
	(0.001)		(0.001)		(0.004)	
"Education" · 2001		−0.016		−0.010		−0.043
		(0.015)		(0.012)		(0.053)
"Education" · 2002		−0.005		−0.003		−0.003
		(0.012)		(0.010)		(0.047)
"Education" · 2003		0.002		0.008		0.029
		(0.012)		(0.009)		(0.039)
"Education" · 2004		−0.016		−0.005		−0.035
		(0.013)		(0.010)		(0.044)
"Education" · 2005		−0.004		0.002		0.001
		(0.013)		(0.011)		(0.053)
"Education" · 2006		−0.003		0.009		0.033
		(0.014)		(0.011)		(0.049)
"Education" · 2007		−0.003		0.006		0.010
		(0.013)		(0.010)		(0.042)
"Education" · 2008		−0.010		−0.004		−0.047
		(0.012)		(0.010)		(0.042)
1st Equation:	**Education**					
♯ of 2TE slots, by Districts	0.335***	0.335***	0.263***	0.263***	0.128***	0.128***
	(0.055)	(0.054)	(0.054)	(0.053)	(0.035)	(0.035)
$\rho(\varepsilon_1, \varepsilon_2)$	−0.143*	−0.147***	−0.244***	−0.252***	−0.236***	−0.235***
χ^2	3.00	12.67	517.19	275.77	43.75	361.61
P-Value	[0.084]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Controls for Equations:						
Year Fixed Effects		X		X		X
Districts Fixed Effects	X	X	X	X	X	X
Sex, Age, Age^2, Sex · Age	X	X	X	X	X	X
Districts · Year Clusters	X	X	X	X	X	X
Observations:	38910	38910	38910	38910	38910	38910

Source: RLMS, 2000-2008, Author's calculations

6 Returns to Higher Education Before and After Educational System Reforms

Expansion of the higher education system in Russia provided an opportunity to increase educational attainment for a part of the population. *Figure 4* illustrates these changes in educational attainment. The levels of ability (observed by universities) are assumed to be distributed according to the black lines in the graphs between the *min* and *max* values (the shape of the curve is taken as an example, it does not affect any of the following conclusions). Assuming that universities select the best performing students, the following sorting occurs. Before the reforms, top 17-20% of the students (with the highest ability levels) were admitted to the universities, as it is illustrated by the painted fragment in the left figure. After the expansion of the higher education system, the percentage of youths admitted to the universities reached 40-45%. This expansion is illustrated in the right graph of *Figure 4*. The zone, signaled as ⊕, depicts students who obtain a higher education degree after the reforms and would not have a possibility to be admitted to the universities if the system had not expanded. These people (⊕) are "switchers" ("compliers") in the IV terminology (Angrist et al. (1996)).

Figure 4: Sorting of Students to Higher Education by the Level of Ability, Before and After Reforms.



IV estimation of the wage equation provides a measure of the returns to education for these "switchers" ("compliers"). However, as there were also some changes in the number of slots in the higher education system during the Soviet period, by IV estimation we measure the average returns to education for everyone who switched to obtaining higher education degrees because of the changes in the educational opportunities. Therefore, it is worthwhile to analyze whether individuals who have lower ability levels, thus switching later, have the same/higher/lower returns to education. There could be two possible reasons for the differences in the returns for people switching earlier or later in time. First, less able people may have lower returns to education, if the education production function and wages depend on individual abilities. Second, expansion of the tertiary education system makes both groups worse in terms of the average level of ability - people with higher education degrees and people without higher education degrees. Therefore, the signal of ability through educational attainment becomes more fuzzy, and employees could pay lower wages because of this. The IV strategy here allows us to estimate the returns to education for these individuals on the border of access to higher education, thus to estimate the difference in wages between these two groups when the level of ability in both groups has decreased. This section discusses the estimation results of the returns to education depending on the level of ability. The time of switch - before or after the higher education expansion - proxies unobserved ability levels.

First, we control for the time of switch by introducing a discreet variable for before and after reforms cohorts. Then, we control for the time of switch during the continuous gradual changes in educational opportunities.

The variable "*17yo-after-1992*" takes value *1* if a person was 17 years old in 1993 or later, and *0* otherwise. The split of the employed population according to this variable is presented in *Table 19*.

Table 19: Number of Individuals who was 17 years old before and after 1992.

	Total	**17yo before 1993**	**17yo after 1992**	**17yo before 1993, %**	**17yo after 1992, %**
Employed:					
2000-2008	28622	22041	6581	77.01%	22.99%
With HE degree	7404	5556	1848	75.04%	24.96%
in % to employed	25.87%	25.21%	28.08%		
Without HE degree	21218	16485	4733	77.69%	22.31%
in % to employed	74.13%	74.79%	71.92%		
Employed:					
2000	2526	2431	95	96.24%	3.76%
2001	2781	2550	231	91.69%	8.31%
2002	2901	2530	371	87.21%	12.79%
2003	3038	2529	509	83.25%	16.75%
2004	3188	2511	677	78.76%	21.24%
2005	3076	2266	810	73.67%	26.33%
2006	3738	2616	1122	69.98%	30.02%
2007	3739	2407	1332	64.38%	35.62%
2008	3635	2201	1434	60.55%	39.45%

Source: RLMS 2000-2008 years, Author's calculation.

Table 20 presents the estimation results for the variable number of years of schooling, *Table 21* - for the variable number of higher education (2TE) studies, and *Table 22* - for higher education attainment. The estimations of the two equations model (educational attainment and wages) are conducted by the Maximum Likelihood method.

Columns (1) and (2) show the OLS estimations of the returns to education, where column (2) additionally controls for the difference in the returns to education among male and female workers. Column (3) shows the estimation results for the joint model of educational attainment and wages, controlling for the differences in the returns to education for those who have switched earlier or later to obtain a higher education degree (interactions of educational variable with "*17yo-after-1992*"). Column (4) add the differences in educational attainment (because of the changing number of slots) between males and females to the 1st-stage-equation. Columns (5) and (6) measure different returns to education for men and women.

The estimation results suggest that the returns to education are lower for people who have switched to obtaining higher education degrees after the educational system expansion: by 2-3% for a year of schooling and by 13% for a higher education degree. Additionally, this decrease in the returns to education is higher for the male population. Therefore, while the average returns to a higher education degree equal to 80%, their decrease for women and men who have switched to higher education after the beginning of the transitional period equals to 4% and 15% correspondingly. Those who have switched later are also younger. We account for this by controlling for the potential experience (=*age-18*) and by introducing dummy variable for these younger cohorts. The results show that in average the wages are higher for the younger cohorts, however the differences between highly and low educated workers are smaller for the post-reforms cohorts.

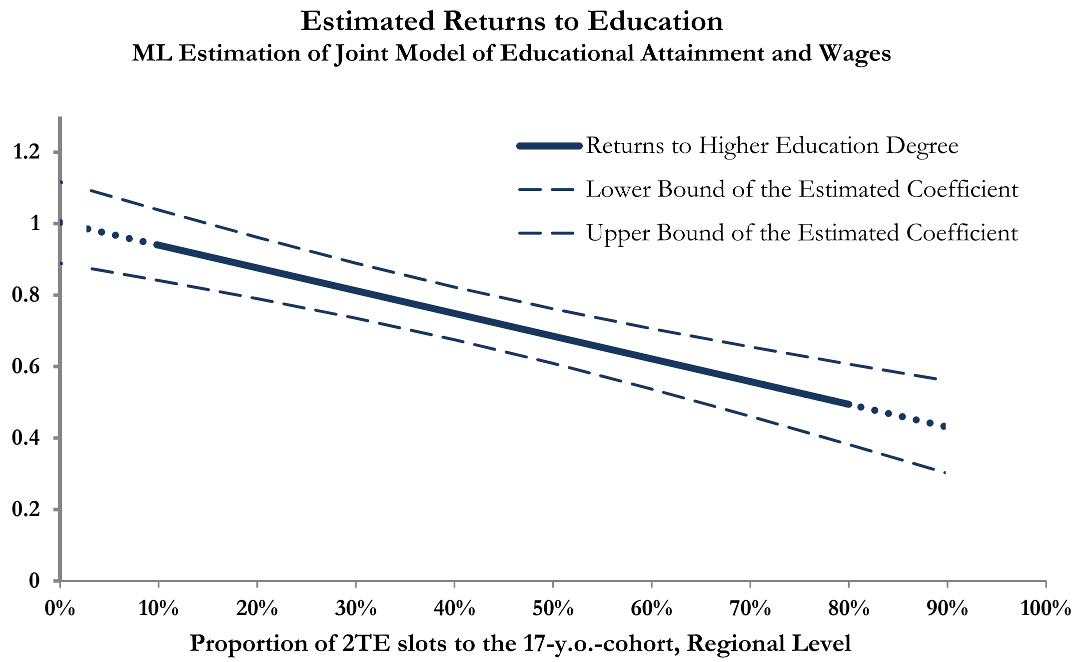
Additionally, we control for different returns to experience for people with the different educational levels. The estimation results are listed in *Table 23* for two models: with [(3)-(4)] and without [(1)-(2)] interaction variables for a higher education degree and experience. The table lists the estimation results for the OLS [(1), (3)] and the joint model of educational attainment and wages [(2), (4)]. Including the interaction variables for education and experience separates the direct and indirect (through an increase in the returns to experience) effects of education on wages. Thus, a higher education degree increases the returns to each year of experience by 2.2%. For the male population, we observe an increase in their wages during the reforms, however the returns to higher education are lower by 7% for those who have switched to obtaining a higher education degree after the beginning of transition. There is a significant decrease in wages for the female population with lower levels of education. However, the returns to education for the female population are 9% higher for those who have obtained this degree after the reforms. Nevertheless, by introducing the interaction variables of education and experience we separate the returns to education into two channels (direct effect on wages and indirect effect through an increase in the returns to experience). As the potential experience for those who are younger is lower (by construction), and the proportion of people with higher education degree is greater for younger workers, especially for the female population, these facts together could cause the observed variation in the returns for the female population. Without controlling for the different returns to the experience by educational levels, we find decreasing (but statistically insignificant) decline in returns to education for female workers and large decline in returns for male workers.

Finally, the setting of the reforms allows us to estimate the returns to education for those who have obtained higher education degrees during the periods of different levels of access to higher education, when the number of slots in the tertiary education system covered higher or lower percentage of youth cohorts. As it is shown in the previous section, those who have switched to higher education degrees later have lower returns to education (especially, male workers). Moreover, we interact higher education attainment with the proportion of slots in a region (relative to youth cohorts) when a person was 17 years old. *Table 24* lists the estimation results for the wage equation and the joint model of educational attainment and wages. Variable *Share of 2TE slots* represents the proportion of the number of slots in universities to the size of cohort of 17 year-olds in a region.

The estimation results reveal that the higher the proportion of youths obtaining higher education degree in a cohort is, the lower the returns to higher education for this cohort are. Additionally, the increasing share of the higher education graduates in a cohort is negatively correlated with the wages of the female workers without higher education degrees. This fact means that if we compare the wages of female workers without higher education degrees who have finished secondary school in the periods of smaller or larger access to higher education, wages are lower for those who have experienced an increase in the number of slots. This result could be explained by the lower average level of ability for workers who do not obtain a higher education degree when access is larger (see *Figure 4*). We do not find a significant decrease in wages for male workers without higher education degrees if they have experienced larger access to higher education. Additionally, both male and female workers with lower levels of ability and with higher education degrees (thus, obtaining higher education under larger access to universities) have lower returns to education, than graduates who have obtained higher education during tougher access to universities. Such decrease in the returns to education is higher for male workers. Even though the difference between male and female workers is not statistically significant, the magnitude of the coefficient for the decline in returns is 50% higher for men.

The variable *Share of 2TE slots* is a continuous variable, which take values from 0.15 to 0.77 in the dataset. That is why, it is more useful to show a visual representation of these estimation results. *Figure 5* depicts the estimated coefficients for the returns to a higher education degree depending on the proportion of the higher education slots to the cohorts at the moment when an individual was finishing secondary school and was taking a decision about higher education enrollment.

Figure 5: Estimated Returns to Education Conditional on the Access to Higher Education.



The estimation results suggest the decreasing returns to education for those who switch to higher education degree attainment in the periods of larger, thus easier, access to the university system. This fact could be explained by the following patterns in the labor market:

1) In the framework of "Education Increasing the Productivity" theory: the returns to higher education could be lower for people with lower ability levels. The periods with larger access to higher education are characterized by the fact that the average level of ability of college entrants is lower (see *Figure 4*).

2) In the framework of "Education as a Signal of Ability" theory: when a larger cohort of higher education graduates come to the labor market, the signal of the ability is more vague compared with the periods of lower access to higher education. The labor market rewards higher education with lower wages when the signal of ability is less precise.

3) The expansion of higher education have probably occurred to a different extent for high-quality and low-quality colleges. The current study does not analyze the changes in the average quality of educational institutions and its influence on the returns to education. However, if we look at the changes in the number of students per teacher ratio and the size of educational surfaces per students ratio over the analyzed period of time (these characteristics could serve as the rough proxies for the quality of higher education institutions in Russia), there was a small decline in these measures, especially in the beginning of expansion - *Figure A-6*.

Therefore, this section reveals the heterogenous returns to education depending on the individual exposure to higher education expansion in Russia. The returns to education are lower for workers who increase their educational attainment due to expansion of the educational system, however these returns are positive, statistically significant, and large in magnitude. Therefore, the estimations take into account the heterogeneity in the returns to education, which is determined by the observable characteristics, namely, time of switch. Additionally, it allows us to derive some conclusions about the varying returns to education with the unobservable levels of ability. It is possible to inference them because of the fact that larger access to higher education means a lower average ability level of college entrants and graduates.

The following section further explores the heterogeneity in the returns to higher education, by allowing them to vary with the unobservable individual characteristics that affect the decision to pursue higher education.

Table 20: Returns to Education (Years of Schooling) Obtained Before and After Expansion, MLE, 24-47 y.o., 2000-2008 years.

Variables	(1) OLS	(2) OLS	(3) MLE		(4) MLE		(5) MLE		(6) MLE	
	ln(Wage)	ln(Wage)	Education	ln(Wage)	Education	ln(Wage)	Education	ln(Wage)	Education	ln(Wage)
♯ *Years of Schooling*	0.105***	0.110***		0.160***		0.161***		0.165***		0.162***
	(0.002)	(0.004)		(0.004)		(0.017)		(0.023)		(0.022)
♯ *Years of Schooling · Male**		−0.013**						−0.013**		−0.013**
		(0.006)						(0.006)		(0.006)
♯ *Schooling · 17yo-after-1993**	−0.026***	−0.010		−0.026***		−0.026***		−0.010		−0.010
	(0.005)	(0.007)		(0.004)		(0.004)		(0.007)		(0.007)
♯ *Schooling · 17yo-after-1993* · Male**		−0.022**						−0.022**		−0.022**
		(0.010)						(0.010)		(0.010)
*17yo-after-1993**	0.355***	0.143		0.356***		0.356***		0.145		0.145
	(0.060)	(0.094)		(0.050)		(0.051)		(0.095)		(0.095)
17yo-after-1993 · Male**		0.293**						0.291**		0.291**
		(0.126)						(0.127)		(0.128)
*Male**	0.501***	0.694***	−0.854***	0.548***	−0.852***	0.549***	−1.043***	0.752***	−1.063***	0.748***
	(0.010)	(0.124)	(0.020)	(0.011)	(0.054)	(0.015)	(0.174)	(0.125)	(0.184)	(0.129)
♯ *of 2TE slots*			0.335***		0.335***		0.335***		0.340***	
			(0.054)		(0.060)		(0.053)		(0.060)	
♯ *of 2TE slots · Female**					0.001				−0.010	
					(0.042)				(0.042)	
$\rho(\varepsilon_1, \varepsilon_2)$			−0.154***		−0.155***		−0.153**		−0.143**	
$\rho(\varepsilon_1, \varepsilon_2)$: χ^2, *[P-Value]*			168.71	[0.000]	12.00	[0.001]	5.62	[0.018]	5.44	[0.020]
Age, Age^2	X	X	X	X	X	X	X	X	X	X
Age · Sex, Age^2 *· Sex*		X					X	X	X	X
Years & Districts F.E.	X	X	X	X	X	X	X	X	X	X
*Districts*Years Clusters*	X	X	X	X	X	X	X	X	X	X
Observations	28622	28622	28622	28622	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 21: Returns to Education (Years of 2TE Studies) Obtained Before and After Expansion, MLE, 24-47 y.o., 2000-2008 years.

Variables	(1) OLS ln(Wage)	(2) OLS ln(Wage)	(3) MLE Education	ln(Wage)	(4) MLE Education	ln(Wage)	(5) MLE Education	ln(Wage)	(6) MLE Education	ln(Wage)
♯ *Years of 2TE Studies*	0.102***	0.105***		0.203***		0.137***		0.206***		0.175***
	(0.003)	(0.003)		(0.004)		(0.018)		(0.013)		(0.015)
♯ *Years of 2TE Studies · Male**		−0.007						−0.007		−0.007
		(0.006)						(0.006)		(0.006)
♯ *2TE St-s · 17yo-after-1993**	−0.022***	−0.005		−0.022***		−0.022***		−0.005		−0.005
	(0.004)	(0.007)		(0.004)		(0.004)		(0.007)		(0.007)
♯ *2TE St-s · 17yo-after-1993* · Male**		−0.028***						−0.028***		−0.028***
		(0.009)						(0.009)		(0.009)
*17yo-after-1993**	0.064***	0.010		0.061***		0.063***		0.007		0.007
	(0.018)	(0.030)		(0.018)		(0.018)		(0.030)		(0.030)
17yo-after-1993 · Male**		0.088**						0.087**		0.090**
		(0.040)						(0.040)		(0.040)
*Male**	0.458***	0.504***	−0.437***	0.503***	−0.225***	0.474***	−0.785***	0.584***	−0.484***	0.554***
	(0.010)	(0.100)	(0.031)	(0.009)	(0.058)	(0.012)	(0.147)	(0.100)	(0.177)	(0.101)
♯ *of 2TE slots*			0.263***		0.190***		0.260***		0.196***	
			(0.052)		(0.058)		(0.051)		(0.057)	
♯ *of 2TE slots · Female**					0.179***				0.157***	
					(0.043)				(0.043)	
$\rho(\varepsilon_1, \varepsilon_2)$			−0.268***		−0.096**		−0.270***		−0.190***	
$\rho(\varepsilon_1, \varepsilon_2)$: χ^2, *[P-Value]*			412.75	[0.000]	3.83	[0.050]	76.54	[0.000]	23.70	[0.000]
Age, Age^2	X	X	X	X	X	X	X	X	X	X
Age · Sex, Age^2 *· Sex*		X					X	X	X	X
Years & Districts F.E.	X	X	X	X	X	X	X	X	X	X
*Districts*Years Clusters*	X	X	X	X	X	X	X	X	X	X
Observations	28622	28622	28622	28622	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 22: Returns to Education (2TE Degree) Obtained Before and After Expansion, MLE, 24-47 y.o., 2000-2008 years.

Variables	(1) OLS ln(Wage)	(2) OLS ln(Wage)	(3) MLE 2TE	(3) MLE ln(Wage)	(4) MLE 2TE	(4) MLE ln(Wage)	(5) MLE 2TE	(5) MLE ln(Wage)	(6) MLE 2TE	(6) MLE ln(Wage)
*2TE Degree**	0.485***	0.488***		0.820***		0.803***		0.809***		0.800***
	(0.012)	(0.016)		(0.047)		(0.049)		(0.031)		(0.069)
2TE Degree · 17yo-after-1992**	−0.124***	−0.038		−0.128***		−0.129***		−0.045		−0.045
	(0.024)	(0.036)		(0.024)		(0.024)		(0.035)		(0.036)
2TE Degree · Male**		−0.008						0.006		0.006
		(0.030)						(0.031)		(0.031)
2TE D. · 17yo-after-1992* · Male**		−0.156***						−0.152***		−0.152***
		(0.046)						(0.047)		(0.047)
*17yo-after-1992**	0.059***	0.012		0.059***		0.059***		0.012		0.011
	(0.017)	(0.028)		(0.017)		(0.017)		(0.027)		(0.028)
17yo-after-1992 · Male**		0.079**						0.079**		0.081**
		(0.038)						(0.038)		(0.038)
*Male**	0.455***	0.480***	−0.266***	0.483***	−0.158***	0.482***	−0.373***	0.513***	−0.212**	0.508***
	(0.010)	(0.101)	(0.017)	(0.009)	(0.033)	(0.010)	(0.083)	(0.101)	(0.106)	(0.102)
♯ of 2TE slots			0.128***		0.087**		0.126***		0.088**	
			(0.035)		(0.036)		(0.034)		(0.037)	
*♯ of 2TE slots · Female**					0.089***				0.082***	
					(0.021)				(0.022)	
$\rho(\varepsilon_1, \varepsilon_2)$			−0.245***		−0.234***		−0.240***		−0.234***	
$\rho(\varepsilon_1, \varepsilon_2)$: χ^2, *[P-Value]*			62.81	[0.000]	55.66	[0.000]	176.53	[0.000]	25.36	[0.000]
Age, Age^2	X	X	X	X	X	X	X	X	X	X
Age · Sex, Age^2 *· Sex*		X					X	X	X	X
Years & Districts F.E.	X	X	X	X	X	X	X	X	X	X
*Districts*Years Clusters*	X	X	X	X	X	X	X	X	X	X
Observations	28622	28622	28622	28622	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 23: Returns to Education (2TE Degree) Obtained Before and After Expansion, MLE, 24-47 y.o., 2000-2008 years.
With Education-Occupation Interaction Variables.

Variables	Without *Experience* · *Education* (1) OLS Wage	(2) MLE 2TE Degree	(2) MLE Wage	With *Experience* · *Education* (3) OLS Wage	(4) MLE 2TE Degree	(4) MLE Wage
*2TE Degree**	0.486***		0.795***	0.179*		0.497***
	(0.016)		(0.039)	(0.103)		(0.106)
2TE · 17yo-after-1993**	−0.037		−0.042	0.099*		0.094*
	(0.036)		(0.036)	(0.051)		(0.051)
2TE · Male**	−0.007		0.008	−0.002		0.013
	(0.030)		(0.030)	(0.030)		(0.031)
2TE · 17yo-after-1993* · Male**	−0.157***		−0.156***	−0.163***		−0.163***
	(0.046)		(0.046)	(0.047)		(0.046)
*17yo-after-1993**	−0.021		−0.023	−0.055**		−0.058**
	(0.022)		(0.022)	(0.025)		(0.025)
17yo-after-1993 · Male**	0.139***		0.143***	0.142***		0.145***
	(0.026)		(0.026)	(0.026)		(0.026)
Age-18	0.032***	0.010	0.032***	0.027***	0.010	0.027***
	(0.006)	(0.009)	(0.006)	(0.007)	(0.009)	(0.007)
(Age-18)^2	−0.001***	−0.000	−0.001***	−0.001***	−0.000	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
*(Age-18) · 2TE Degree**				0.023**		0.022**
				(0.010)		(0.010)
(Age-18)^2 *· 2TE Degree**				−0.000		−0.000
				(0.000)		(0.000)
*Male**	0.435***	−0.153***	0.457***	0.432***	−0.154***	0.455***
	(0.016)	(0.034)	(0.016)	(0.017)	(0.033)	(0.016)
♯ of 2TE slots		0.085**			0.084**	
		(0.040)			(0.036)	
*♯ of 2TE slots · Female**		0.093***			0.093***	
		(0.022)			(0.022)	
$\rho(\varepsilon_1, \varepsilon_2)$		−0.232***			−0.235***	
χ^2, *[P-Value]*		86.31	[0.000]		105.24	[0.000]
Districts & Year F.E.	X	X	X	X	X	X
Districts · Year Clusters	X	X	X	X	X	X
Observations:	28622	28622	28622	28622	28622	28622

Source: RLMS, 2000-2008, Author's calculations

Table 24: Returns to 2TE Degree Obtained at Different Expansion Level.

Variables	**(1) OLS**	**(2) MLE**		**(3) OLS**	**(4) MLE**	
	Wage	**2TE Degree**	**Wage**	**Wage**	**2TE Degree**	**Wage**
Without Education-Occupation Interaction Variables						
*2TE Degree**	0.628***		1.003***	0.614***		0.972***
	(0.039)		(0.058)	(0.041)		(0.197)
2TE Degree · Male**				0.014		0.028
				(0.079)		(0.081)
*Share of 2TE Slots · 2TE**	−0.578***		−0.636***	−0.453***		−0.508***
	(0.110)		(0.110)	(0.126)		(0.129)
Share of 2TE Slots · 2TE · Male**				−0.223		−0.221
				(0.218)		(0.217)
Share of 2TE Slots	−0.162		−0.196	−0.315		−0.386*
	(0.211)		(0.208)	(0.214)		(0.204)
*Share of 2TE Slots · Male**				0.256*		0.327**
				(0.139)		(0.136)
♯ of 2TE slots		0.130**			0.089	
		(0.054)			(0.058)	
*♯ of 2TE slots · Female**					0.103***	
					(0.022)	
$\rho(\varepsilon_1, \varepsilon_2)$		−0.263***			−0.256**	
χ^2, *[P-Value]*		140.73	[0.000]		3.93	[0.048]
With Education-Occupation Interaction Variables						
*2TE Degree**	0.295***		0.674***	0.287***		0.647***
	(0.087)		(0.104)	(0.089)		(0.113)
2TE Degree · Male**				0.023		0.037
				(0.079)		(0.080)
*Share of 2TE Slots · 2TE**	−0.264*		−0.326**	−0.143		−0.198
	(0.132)		(0.138)	(0.154)		(0.161)
Share of 2TE Slots · 2TE · Male**				−0.229		−0.228
				(0.220)		(0.221)
Share of 2TE Slots	−0.237		−0.270	−0.391*		−0.463**
	(0.212)		(0.228)	(0.215)		(0.214)
*Share of 2TE Slots · Male**				0.261*		0.333**
				(0.139)		(0.138)
♯ of 2TE slots		0.129**			0.088	
		(0.059)			(0.059)	
*♯ of 2TE slots · Female**					0.103***	
					(0.022)	
$\rho(\varepsilon_1, \varepsilon_2)$		−0.261***			−0.256***	
χ^2, *[P-Value]*		73.19	[0.000]		43.24	[0.000]
Districts, Year, Cohort F.E.	*X*	*X*	*X*	*X*	*X*	*X*
Districts · Year Clusters	*X*	*X*	*X*	*X*	*X*	*X*
Observations:	28622	28622	28622	28622	28622	28622

Source: RLMS, 24-47 y.o., 2000-2008, Author's calculations

7 Marginal Treatment Effects: Self-Selection, Sorting on Gains and Heterogenous Returns

In this section, relying on the recent non-parametric models with essential heterogeneity, we identify the heterogeneous returns to education for the Russian population. We analyze the variation in the returns to education based on the unobservable characteristics. In all previous sections, we were in the framework of instrumental variables. IV provides an estimation of the Local Average Treatment Effect - the returns to higher education for those who have switched to higher education because of the changes in the instrument (in our case, because of the growing opportunities in access to higher education), and who would not pursue higher education otherwise. Heckman (1997), Heckman et al. (2003), Carneiro et al. (2001) question the OLS and IV approaches to the estimation of the Mincer's Wage Equation. They show that in the presence of heterogeneity in the returns to education and selection on the gains (students take into account their heterogeneous returns while choosing their educational attainment), OLS and IV are not consistent estimators of the mean returns to education. In the previous section, we show the evidence of the varying returns to education for people obtaining higher education degree in the periods characterized by lower or higher access to higher education institutions. Students may observe (expect) their different returns to higher education, and therefore, self-select themselves to obtaining higher education based on their individual returns to higher education. Under certain assumptions, it is possible to identify the heterogeneous returns to education with Marginal Treatment Effects estimation via the method of Local Instrumental Variables. In this section, we base on the previous research of Carneiro et al. (2001), Heckman et al. (2006), Heckman and Vytlacil (2007) on the theoretical and empirical framework of the marginal treatment effects estimations, which account for the heterogeneity and selection in the estimations of treatment effects.

Marginal Treatment Effect (*MTE*) for the returns to education estimation is the average return to schooling for individuals who are indifferent to obtaining education at different levels of unobservable characteristics, which influence this educational choice along with other observable characteristics we can account for. The concept of MTE was first introduced by Bjorklund and Moffitt (1987). Carneiro et al. (2001), Heckman et al. (2006), Heckman and Vytlacil (2007) develop the theoretical framework of the MTE estimation for the returns to schooling, identification of ATE, TT, TUT effects from MTE, and provide the empirical applications for the USA data. Heckman and Li (2003) and Wang et al. (2007) apply these methods for the estimation of the returns to college in China.

7.1 Returns to Education: Model with Essential Heterogeneity

We assume only one educational choice: obtaining higher education degree ($S_i = 1$) or not ($S_i = 0$). We estimate the model with heterogeneous returns to education θ_i, which vary for the population:

$$lnW_i = \alpha + \beta \cdot X_i + \theta_i \cdot S_i + U_i \tag{1}$$

In the literature, such model is referred to a category of "random coefficient models" or "heterogeneous treatment effect models" (Heckman et al. (2006)). Therefore, we have two potential wage outcomes (for workers with and without higher education degree).

$$\begin{cases} lnW_{1,i} = \alpha_1 + \beta_1 \cdot X_i + U_{1,i}, \text{ if } S_i = 1, \\ lnW_{0,i} = \alpha_0 + \beta_0 \cdot X_i + U_{0,i}, \text{ if } S_i = 0. \end{cases} \tag{2}$$

In such a specification we assume that the influence of other observable characteristics may vary according to the educational level obtained ($\beta_0 \neq \beta_1$). For example, this assumption allows us to account for different returns to education for male and female populations. Additionally, $U_{1,i}$ and $U_{0,i}$ are random shocks for wage equations, where $E(U_{1,i}|X_i) = 0$ and $E(U_{0,i}|X_i) = 0$. The random shocks also vary for the different educational levels.

The educational choice equation we define as:

$$S_i = \begin{cases} 1, \text{ if } S_i^* \geq 0 \\ 0, \text{ if } S_i^* < 0 \end{cases}, \ where \ S_i^* = \gamma \cdot Z_i - V_i \tag{3}$$

Here, S_i^* is a latent variable representing the utility of an individual i for obtaining a higher education degree. This utility is determined by observed and unobserved characteristics: Z_i and V_i correspondingly. Thus, V_i is the unobserved heterogeneity of individual i in the educational choice equation. The higher the value of the unobserved parameter V_i is, the less likely an individual i would obtain a higher education degree.

$U_{1,i}$, $U_{0,i}$ and V_i - are correlated. We do not specify their correlation or their joint distribution.

We could not observe both wage outcomes (with and without a higher education degree) for each individual. Therefore, observed wages could be expressed as:

$$\begin{aligned} lnW_i &= S_i \cdot lnW_{1,i} + (1 - S_i) \cdot lnW_{0,i} = \\ &= S_i \cdot (\alpha_1 + \beta_1 \cdot X_i + U_{1,i}) + (1 - S_i) \cdot (\alpha_0 + \beta_0 \cdot X_i + U_{0,i}) = \\ &= \alpha_0 + \beta_0 \cdot X_i + [(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot X_i + (U_{1,i} - U_{0,i})] \cdot S_i + U_{0,i} \end{aligned} \tag{4}$$

Therefore, returns to higher education degree are expressed as:

$$\theta_i = [(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot X_i + (U_{1,i} - U_{0,i})] \tag{5}$$

Where, $(\beta_1 - \beta_0) \cdot X_i$ represents the returns to education varying with the observable characteristics X_i, and $(U_{1,i} - U_{0,i})$ stands for the variation in returns to education based on unobservable characteristics.

If we estimate the wage equation (4) by OLS, we therefore assume the homogeneous returns to education θ:

$$lnW_i = \alpha + \beta \cdot X_i + \theta \cdot S_i + U_i \tag{6}$$

Or heterogeneous returns to education based on observable characteristics:

$$lnW_i = \alpha + \beta_0 \cdot X_i + (\beta_1 - \beta_0) \cdot X_i \cdot S_i + \theta \cdot S_i + U_i \tag{7}$$

By OLS we obtain the following estimator:

$$\begin{aligned}
\hat{\theta}^{OLS}(x_i) &= E(lnW_i|X_i = x_i, S_i = 1) - E(lnW_i|X_i = x_i, S_i = 0) = \\
&= E(\alpha_1 + \beta_1 \cdot X_i + U_{1,i}|X_i = x_i, S_i = 1) - E(\alpha_0 + \beta_0 \cdot X_i + U_{0,i}|X_i = x_i, S_i = 0) = \\
&= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot x_i + E(U_{1,i}|S_i = 1) - E(U_{0,i}|S_i = 0) = \\
&= E(\theta_i|X_i = x_i) + E(U_{1,i}|S_i = 1) - E(U_{0,i}|S_i = 0) = \\
&= ATE(x_i) + E(U_{1,i}|S_i = 1) - E(U_{0,i}|S_i = 1) + E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0) = \\
&= ATE(x_i) + \{E(U_{1,i} - U_{0,i}|S_i = 1)\} + \{E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0)\} = \\
&= ATE(x_i) + Sorting\ on\ Gains\ Effect^U_{1,i} + Selection\ Bias_{1\to 0,i} = \\
&= ATE(x_i) + SGE^U_{1,i} + SB_{1\to 0,i}
\end{aligned} \tag{8}$$

Or:

$$\begin{aligned}
&= ATE(x_i) + \{E(U_{1,i} - U_{0,i}|S_i = 0)\} + \{E(U_{1,i}|S_i = 1) - E(U_{1,i}|S_i = 0)\} = \\
&= ATE(x_i) + Sorting\ on\ Gains\ Effect^U_{0,i} + Selection\ Bias_{0\to 1,i} = \\
&= ATE(x_i) + SGE^U_{0,i} + SB_{0\to 1,i}
\end{aligned} \tag{9}$$

Therefore, the bias of OLS estimator can be decomposed into two components: Sorting on the Gains Effect and Selection Bias.
The Selection Bias $SB_{1\to 0,i} = E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0)$ describes the fact that unobservable factors, which influence the decisions to obtain a higher education degree, affect wages. It shows the difference in wages between those who obtain higher education and not in the case if nobody of them has a higher education degree. For example, if $SB_{1\to 0,i}$ is positive, people who self-select themselves to the higher education may have higher ability levels and would have higher wages anyway, than those who choose not to obtain higher education degree.
Sorting on the Gains Effect $SGE^U_{1,i} = E(U_{1,i} - U_{0,i}|S_i = 1)$ represents the mean gain in unobservable components of wage equation for people who choose higher education degree $S_i = 1$. The non-zero value of the Sorting on the Gains Effect means that individuals self-select themselves to higher education based on their wage returns to unobservable characteristics in the case of obtaining or not a higher education degree. The positive sign for the $SGE^U_{1,i}$ means that individuals with a higher education degree have unobservable characteristics that are better paid if they obtain this level of education.

The Average Treatment Effect (ATE) means the average return to higher education degree for a randomly assigned person, and is usually the main parameter of interest in the treatment evaluation. As we have seen, according to the equations (8), the OLS estimator of the returns to education is a biased estimator of the ATE, if we have a selection and/or sorting on the gains effects.

By using IV framework we obtain the following estimator:

$$\begin{aligned}
\hat{\theta}^{IV}(x_i) &= \frac{Cov(I_i, lnW_i)}{Cov(I_i, S_i)} = \\
&= \frac{Cov(I_i, \alpha_0 + \beta_0 \cdot x_i + [(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot x_i + (U_{1,i} - U_{0,i})] \cdot S_i + U_{0,i})}{Cov(I_i, S_i)} = \\
&= \frac{Cov(I_i, \alpha_0 + \beta_0 \cdot x_i + [(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot x_i] \cdot S_i) + Cov(I_i, U_{0,i} + (U_{1,i} - U_{0,i}) \cdot S_i)}{Cov(I_i, S_i)} = \\
&= ATE(x_i) \cdot \frac{Cov(I_i, S_i)}{Cov(I_i, S_i)} + \frac{Cov(I_i, U_{0,i} + (U_{1,i} - U_{0,i}) \cdot S_i)}{Cov(I_i, S_i)} = \\
&= ATE(x_i) + \frac{Cov(I_i, (U_{1,i} - U_{0,i}) \cdot S_i)}{Cov(I_i, S_i)} = \\
&= ATE(x_i) + \frac{Cov(I_i, (U_{1,i} - U_{0,i}) | S_i = 1) \cdot Pr(S_i = 1 | x_i, I_i)}{Cov(I_i, S_i)} \qquad (10)
\end{aligned}$$

Where I_i is the instrument for the selection equation, therefore it is correlated to the V_i but is not correlated to the U_i, $U_{1,i}$, $U_{0,i}$.

Therefore, the IV estimator would be equal to the ATE only if either (1) $U_{1,i} - U_{0,i} = 0$ - no heterogeneity in the returns to education, and, thus, no sorting on gain or (2) $U_{1,i} - U_{0,i}$ is independent of S_i, thus no sorting on gains. In the absence of essential heterogeneity, i.e. in the case when people sort themselves based on the gains on unobservables, the IV is a consistent estimator of the ATE (Average Treatment Effect). However, in the presence of sorting on the gains, IV may overestimate the ATE by putting more weights on the returns for treated individuals (Heckman (1997)).

In the frame of the model with essential heterogeneity, we can write the Effects of Treatment on Treated (TT) and the Effects of Treatment on Untreated (TUT) in the following way:

$$\begin{aligned}
TT(x_i) &= E(lnW_{1,i} - lnW_{0,i} | X_i = x_i, S_i = 1) = \\
&= E(\alpha_1 + \beta_1 \cdot X_i + U_{1,i} - \alpha_0 - \beta_0 \cdot X_i - U_{0,i} | X_i = x_i, S_i = 1) = \\
&= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot x_i + E(U_{1,i} - U_{0,i} | S_i = 1) = \\
&= ATE(x_i) + Sorting\ on\ Gains\ Effect^U_{1,i} = ATE(x_i) + SGE^U_{1,i} \qquad (11)
\end{aligned}$$

$$\begin{aligned}
TUT(x_i) &= E(lnW_{1,i} - lnW_{0,i} | X_i = x_i, S_i = 0) = \\
&= E(\alpha_1 + \beta_1 \cdot X_i + U_{1,i} - \alpha_0 - \beta_0 \cdot X_i - U_{0,i} | X_i = x_i, S_i = 0) = \\
&= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot x_i + E(U_{1,i} - U_{0,i} | S_i = 0) = \\
&= ATE(x_i) + Sorting\ on\ Gains\ Effect^U_{0,i} = ATE(x_i) + SGE^U_{0,i} \qquad (12)
\end{aligned}$$

Where, $SGE^U_{1,i}$ stays for the Sorting on the Gains Effect for those who obtain a higher education degree, and $SGE^U_{0,i}$ - the Sorting on the Gains Effect for those who do not obtain higher education (therefore, their returns to unobservable characteristics if they obtain a higher education degree).

The described above effects are conditional on the individual characteristics $X_i = x_i$. The average effects in the population could be written as:

$$ATE = \int_X ATE(x_i)\ dF_X(x) \tag{13}$$

$$TT = \int_{X|S=1} TT(x_i)\ dF_{X|S=1}(x) \tag{14}$$

$$TUT = \int_{X|S=0} TUT(x_i)\ dF_{X|S=0}(x) \tag{15}$$

$$\hat{\theta}^{OLS} = ATE + SGE_1 + SB_{1\to 0} \tag{16}$$

It is worthwhile to point out that moving to the population effects, we can have the Effect of the Sorting on Gains based on the observable characteristics, if $E(X|S=1) \neq E(X|S=0) \neq E(X)$, in other words, if people sort themselves based on the observable characteristics. For example, if female workers have higher returns to education they may obtain higher education more often than male workers. Therefore, the Sorting on the Gains Effects for the population will be written as the followings:

$$SGE_1 = SGE^U_1 + (\beta_1 - \beta_0) \cdot (E(X|S=1) - E(X)) = TT - ATE \tag{17}$$

$$SGE_0 = SGE^U_0 + (\beta_1 - \beta_0) \cdot (E(X|S=0) - E(X)) = TUT - ATE \tag{18}$$

and

$$TT = ATE + SGE^U_1 + (\beta_1 - \beta_0) \cdot (E(X|S=1) - E(X)) \tag{19}$$

$$TUT = ATE + SGE^U_0 + (\beta_1 - \beta_0) \cdot (E(X|S=0) - E(X)) \tag{20}$$

In order to estimate all these parameters properly, we rely on the Marginal Treatment Effects (MTE) estimation:

$$\begin{aligned} MTE(X_i = x, U_{S,i} = u_s = p) &= E(\theta_i | X_i = x, U_{S,i} = u_s) = \\ &= E(lnW_{1,i} - lnW_{0,i} | X_i = x, U_{S,i} = u_s) = \\ &= \left. \frac{\partial E(lnW_i | X_i = x, P(Z_i) = p)}{\partial p} \right|_{p=u_s} \end{aligned} \tag{21}$$

Where, $U_{S,i}$ is the uniformal transformation of the random term of the educational equation V_i: $U_S = F_V(V)$, $U_S \rightsquigarrow Unif[0,1]$. Therefore, we can rewrite the educational choice equation (3) as the following:

$$\begin{aligned} S_i &= \begin{cases} 1, \text{ if } \gamma \cdot Z_i - V_i \geq 0 \\ 0, \text{ if } \gamma \cdot Z_i - V_i < 0 \end{cases} = \begin{cases} 1, \text{ if } F_V(\gamma \cdot Z_i) \geq F_V(V_i) \\ 0, \text{ if } F_V(\gamma \cdot Z_i) < F_V(V_i) \end{cases} = \\ &= \begin{cases} 1, \text{ if } P(Z_i) \geq U_{S,i} \\ 0, \text{ if } P(Z_i) < U_{S,i} \end{cases} \end{aligned} \tag{22}$$

The Marginal Treatment Effect measures the returns to education (average treatment effect) at the points u_s of marginal changes in U_S (in other words, at the points of marginal changes of V - unobservable component in the educational choice equation). As we measure them at the points of $P(Z_i) = p = u_s$, it means that at these values of unobservable component in the educational choice equation, individuals are indifferent between obtaining a higher education degree or not. Therefore, the MTE is the marginal willingness to pay for the $lnW_{1,i}$ versus $lnW_{0,i}$, given the observed characteristics X_i and the level of unobserved characteristics $U_{S,i} = u_s$. The small values of u_s (close to 0) means the small values of V (large values of $-V$), therefore are associated with people who are more likely to obtain a higher education degree, given their unobservable characteristics. Correspondingly, MTE measured for large values of u_s (close to 1) show the returns to education for people less likely to obtain higher education, based on their unobservable characteristics.

Heckman and Vytlacil (1999), Heckman et al. (2006), and Heckman and Vytlacil (2007) show that ATE, TT, and TUT could be determined as the weighted averages of MTE, according to the following formulas:

$$ATE(x_i) = \int_0^1 MTE(x_i, u_s) \cdot h_{ATE}(x_i, u_s)\ du_s \tag{23}$$

$$TT(x_i) = \int_0^1 MTE(x_i, u_s) \cdot h_{TT}(x_i, u_s)\ du_s \tag{24}$$

$$TUT(x_i) = \int_0^1 MTE(x_i, u_s) \cdot h_{TUT}(x_i, u_s)\ du_s \tag{25}$$

Where the weights are:

$$h_{ATE}(x_i, u_s) = 1$$

$$h_{TT}(x_i, u_s) = \frac{Pr(P(Z) > u_s|X = x_i)}{\int_0^1 Pr(P(Z) > u_s|X = x_i)\ du_s}$$

$$h_{TUT}(x_i, u_s) = \frac{Pr(P(Z) < u_s|X = x_i)}{\int_0^1 Pr(P(Z) < u_s|X = x_i)\ du_s}$$

Therefore, the weights are constructed in the manner that for TT effect the individuals who have higher probabilities of being treated given u_s (to obtain a higher education degree) contribute with larger weights, while for TUT effect the higher weights are giving for the individuals with lower probabilities of being treated.

The estimation procedure is described in *Technical Appendix*: *Section 4*.

7.2 Estimation Results

The method of the marginal treatment effects estimation relies on the maximum full support of predicted propensity scores. The major predictors of tertiary education attainment are the educational background of parents and family income. In the data we observe information only on parents' educational background and only for the year 2006. If we estimate the probability of obtaining a higher education degree without information on parents' educational background for the years 2000-2008, we obtain the propensity scores p being in the interval of $\{0.05; 0.4\text{-}0.5\}$. If we add the variables describing parents' educational background as a predictor of higher education attainment, we get the estimated propensity scores covering almost all support from 0 to 1 (more precisely, $\{0.05; 0.8\}$). That is why, in this section we conduct the estimations only for the employed individuals observed in 2006: 24-52 years old. We work with 4326 observations of workers, for whom we have information about their parents' educational background (96.3% out of 4491 employed people in the dataset).

First, we estimate the educational choice equation as a Probit model. The explanatory variables includes: parents' educational background, cohort year of birth dummy variables, number of slots in the higher education system in a year when an individual was 17 years old, and the number of slots interacted with parent's educational background.

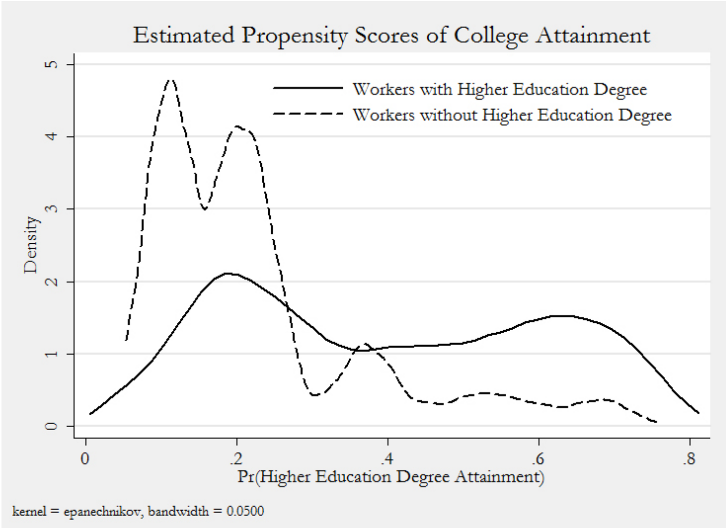
Table 25 shows the estimated coefficients of the educational choice equations. Predicted propensity scores are presented in *Figure 6*.

Table 25: Estimated Coefficients for the Educational Choice Equation (Probit), 24-52 y.o., 2006.

Variables	**Coefficients**
♯ of 2TE Slots	0.112^{*}
	(0.067)
*♯ of 2TE Slots · Parents w 1TE**	-0.070
	(0.094)
*♯ of 2TE Slots · Parents w SE**	-0.078
	(0.086)
*Parents w 1TE**	0.538^{***}
	(0.119)
*Parents w 2TE**	1.230^{***}
	(0.120)
*Male**	-0.388^{***}
	(0.044)

Control Variables: Sex, Age, Years of Birth fixed effects.
Observations: 4326. Source: RLMS 2006, Author's calculations.

Figure 6: Estimated Propensity Scores of Higher Education Attainment, 24-52 y.o., 2006.



Then, we estimate the coefficients of the wage equation β_0 and $(\beta_1 - \beta_0)$. The estimation results are presented in the *Table 26* below.

Table 26: Estimated Coefficients for the Wage Equation (Local Linear Regression), 24-52 y.o., 2006.

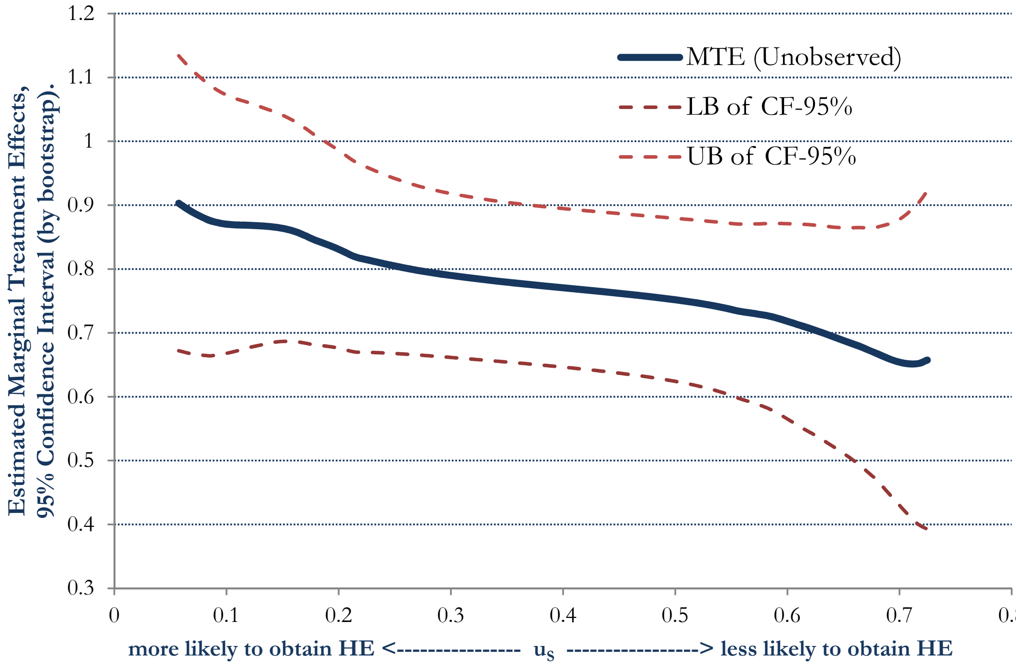
Variables	β_0	$\beta_0 - \beta_1$
*Male**	0.483***	−0.292
	(0.070)	(0.235)
(Age-18)	0.130*	
	(0.079)	
(Age-18)^2	−0.034*	
	(0.020)	

Control Variables: Sex, (Age-18), (Age-18)^2, City Type fixed effects.
Observations: 4326. Source: RLMS 2006, Author's calculations.

Further, we estimate MTE, ATE, TT, TUT according to the procedure described in the previous subsection.

Figure 7 presents the estimated marginal returns to education as a function of unobserved characteristics u_s. The estimation results suggest the declining returns to higher education with unobserved characteristics, which decrease the probability of higher education degree attainment. *Figure 8* depicts marginal treatment effects MTE^U and marginal treatment effects adjusted to the differences in the returns based on observable characteristics MTE (in our case male-female differences in the returns to education).

Figure 7: Estimated Marginal Treatment Effects (MTE^U), 24-52 y.o., 2006.



Note: Confidence Interval is calculated by bootstrap with 500 iterations.

Figure 8: Estimated Marginal Treatment Effects (MTE, MTE^U), 24-52 y.o., 2006.

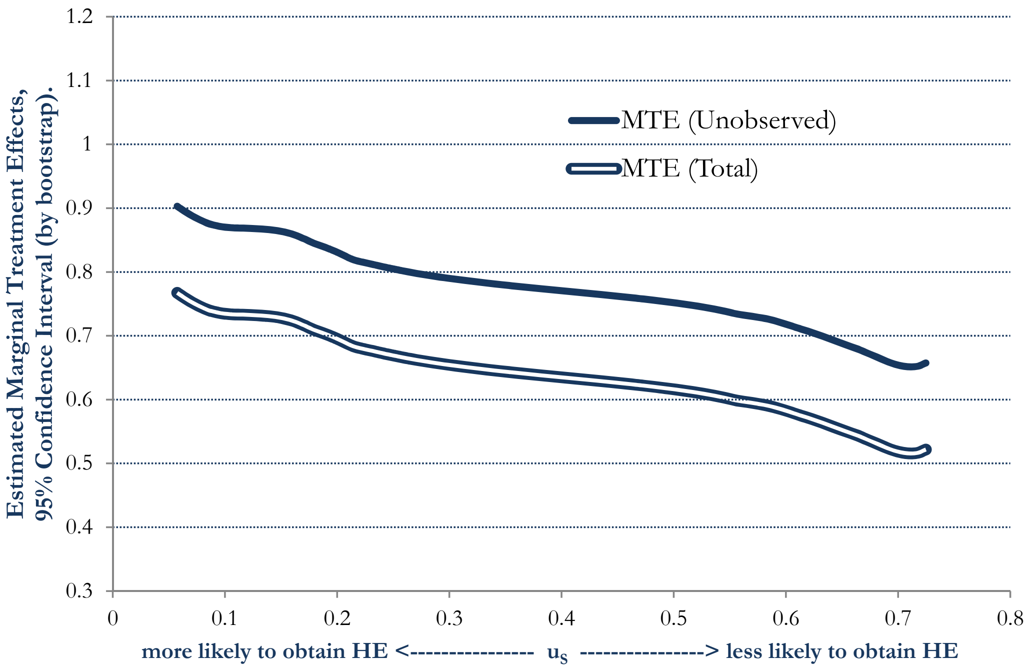


Table 27 presents the estimation results for ATE, TT and TUT effects. *Table 28* show the decompositions of this treatment effects, according to the formulas *46*, *47*, *48*, on the contributions of the returns to observed and unobserved characteristics.

Table 27: Estimated Effects: ATE, TT, TUT. 24-52 y.o., 2006.

Treatment Effects:	**Estimated Effects**	**Standard Errors**	**T-Statistics**
TT	0.734	0.083	8.79
ATE	0.676	0.070	9.61
TUT	0.635	0.063	10.01

Source: RLMS 2006, Author's calculations.

Note: Standard Errors are calculated using bootstrap with 500 iterations.

Table 28: Decomposition of the Estimated Effects: ATE, TT, TUT. 24-52 y.o., 2006.

Treatment Effects:	**Total**	**Contribution by Observed Characteristics** TT^{Obs}, ATE^{Obs}, TUT^{Obs}	**Contribution by Unobserved Characteristics** TT^{U}, ATE^{U}, TUT^{U}
TT	0.734	-0.108	0.842
ATE	0.676	-0.136	0.811
TUT	0.635	-0.146	0.781

Source: RLMS 2006, Author's calculations.

Notations 1: $TT^{Obs} = (\widehat{\beta_1 - \beta_0}) \cdot E(X|S=1)$, $ATE^{Obs} = (\widehat{\beta_1 - \beta_0}) \cdot E(X)$, $TUT^{Obs} = (\widehat{\beta_1 - \beta_0}) \cdot E(X|S=0)$

Notations 2: $TT^{U} = E(MTE^{U}|S=1)$, $ATE^{U} = E(MTE^{U})$, $TUT^{U} = E(MTE^{U}|S=0)$

We can calculate the Sorting on the Gains Effects for those who obtain a higher education degree and not.

$$SGE_1^U = E(U_{1,i} - U_{0,i}|S_i = 1) = TT^U - ATE^U = 0.031$$
$$SGE_0^U = E(U_{1,i} - U_{0,i}|S_i = 0) = TUT^U - ATE^U = -0.030$$
$$SGE_1 = TT - ATE = 0.734 - 0.676 = 0.058$$
$$SGE_0 = TUT - ATE = 0.635 - 0.676 = -0.041$$

We observe positive Sorting on the Gains for those who obtain higher education and not. Those who obtain higher education have higher returns to their unobservable characteristics if they have a higher education degree ($SGE_1^U = 0.031 > 0$). Those who do not obtain higher education have lower returns to their unobservable characteristics if they obtain a higher education degree (SGE_0^U = -0.030<0). The same patterns we observe for the Sorting on the Gains based on observable characteristics, in our case it is sex differences in the returns to education (SGE_1 - SGE_1^U = 0.058-0.031 = 0.027>0, SGE_0 - SGE_0^U = -0.041+0.030 = -0.011<0). For example, female workers have higher returns to education, and, thus, they are more likely to obtain a higher education degree. This Sorting on the Gains determine the differences between TT, ATE and TUT effects.

Selection Effects take the following value:

$$SB_{1\to 0,i} = E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0) = \hat{\theta}^{OLS} - TT = 0.413 - 0.734 = -0.321$$
$$SB_{0\to 1,i} = E(U_{1,i}|S_i = 1) - E(U_{1,i}|S_i = 0) = \hat{\theta}^{OLS} - TUT = 0.413 - 0.635 = -0.222$$

We observe negative Selection Effect for those who pursue higher education ($SB_{1\to 0,i}$). Therefore, those who follow higher education studies would have lower wage returns to their unobservable characteristics if they do not have a higher education degree, in comparison to those who do not obtain it. The fact that $SB_{0\to 1,i} - SB_{1\to 0,i} = TT - TUT = 0.099 > 0$ corresponds to the decreasing marginal returns to education, because if we provide the education to everybody the difference in wages between those who obtain it now and those who do not ($= -0.222$) would be lower in absolute value, than if we forbid higher education for all of them ($= -0.321$).

Overall, the presented in this section results are close to the results of the estimation of returns to education in the IV framework (previous sections). This fact is due to the low in magnitude Sorting on the Gains Effect, which biases the IV estimator, as we have seen in *Equation 10*. Heterogeneous returns to education, estimated by IV and MTE methods, also show very similar patterns (See *Figures 5* and *8* for a comparison of these estimations). Returns to a higher education degree are decreasing with the unobservable characteristics, which reduce the probability of obtaining higher education (as it is shown by MTE estimation, *Figure 8*). Those who switch to obtaining a higher education degree later in time, when access to the higher education system becomes easier, have lower returns to education (as it is shown by IV estimation, *Figure 5*). Those who increase their educational attainment with the increasing access to higher education, have thus higher values of unobservable characteristics, which decrease the probability of higher education attainment, than those who obtain it when access to higher education is tougher. That is why we obtain a similar patterns of the decrease in the returns to education, estimated by IV and MTE methods.

8 Conclusion

The current study has identified the returns to education for the Russian population and has quantified the effects of expansion of the higher education system in Russia on labor market outcomes, namely employment and wages.

We report the evidence of the large returns to higher education in terms of both employment and wages, which are higher for the female population. We have not revealed any significant general equilibrium effects of this expansion on wages.

Additionally, we have found the decreasing returns to education for youths entering the higher education system during the periods of easier (larger) access to colleges.

The estimation results of the marginal returns to education also suggest the declining returns to education in the population based on the unobservable characteristics negatively affecting higher education attainment. We have determined a light positive sorting on the gains, which means that youths self-select themselves to colleges based on their expected returns to education. However it is low in magnitude, and that is why we have obtained close estimation results for the returns to education using IV and MTE methods.

Therefore, expansion of the higher education system in Russia had positive effects on wages for those who increased their educational attainment. However, their returns to education were lower than of those who obtained a higher education degree during the periods of tougher access to colleges.

The large expansion of the higher education system in Russia provides an exogenous variation in access to higher education degrees, increasing significantly college enrollment. In the current paper, we have estimated the effects of such expansion on labor market outcomes. However, this natural experiment could potentially be used for the analysis of the influence of education on other life outcomes: occupational choices, work-life balance, marriages (incl. partner's "quality"), childbearing decisions (time of children birth, quantity of children, time dedicated to children education), educational outcomes of children of the next generation, health attitudes, risky behavior, etc. These questions are potential topics for future research, for which the transition reforms in the Russian Federation could serve as an identification strategy for the parameters that determine the importance of education.

References

Angrist, J. D. and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 1991, *106* (4), 979–1014.

Angrist, J.D. and J.-S. Pischke, "Mostly Harmless Econometrics: An Empiricists Companion.," *Princeton University Press*, 2009.

_ , G.W. Imbens, and D.B. Rubin, "Identification of Causal Effects Using Instrumental Variables.," *Journal of the American Statistical Association*, 1996, *91(434)*, 444–455.

Arcidiacono, Peter, "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?," *Econometrica*, 2005, *73* (5 (Sept.)), 1477–1524.

Belokonnaya, L., V. Gimpelson, T. Gorbacheva, O. Gihareva, R. Kapelushnikov, and A. Lukianova, "Wage formation: analysis in the light of work professions," *Wages in the Russian Federation: Evolution and Differentiation. Moscow: State University Higher School of Economics.*, 2007.

Belzil, Christian, "The Return to Schooling in Structural Dynamic Models: A Survey," *IZA Discussion Paper No. 2370*, 2006.

Bjorklund, A. and R. Moffitt, "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 1987, *69.*

Card, David, "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, by L.N. Christofides, E.K. Grant, R. Swidinsky*, 1995, pp. 201–222.

_ , "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 1999, *3*, 1801–1863.

_ **and Thomas Lemieux**, "Can Falling Supply Explain the Rising Return to College for Young Men," *Quarterly Journal of Economics*, 2001, *116*, 705–746.

_ **and** _ , "Education, Earnings , and the "Canadian G.I. Bill"," *Canadian Journal of Economics*, 2001, *34* (2), 313–344.

Carneiro, P., J.J. Heckman, and E. Vytlacil, "Estimating the Return to Education When It Varies Among Individuals," *Working Paper, University of Chicago*, 2001.

Conneely, Karen and Roope Uusitalo, "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model," *Discussion Paper, Industrial Relations Sections, Princeton University*, 1997.

Duflo, Esther, "Consequences of School Construction in Indonesia," *The American Economic Review*, 2001, *91* (4), 795–813.

Froelich, Markus and Blaise Melly, "Unconditional quantile treatment effects under endogeneity," *IZA Discussion Paper*, 2008, (3288).

_ **and** _ , "Estimation of quantile treatment effects with STATA," *Stata Journal*, 2010.

Gohberg, L.M., I.U. Zabaturina, N.V. Kovaleva, V.I. Kuznetsova, O.K. Ozerova, and L.A. Rosovetskaya, *Education in the Russian Federation*, Statistical Yearbook. SU-HSE, Moscow., 2007.

Gregory, P. and J. Kohlhase, "The Earnings of Soviet Workers: Evidence from The Soviet Interview Project," *The Review of Economics and Statistics*, 1988, *70* (1).

Heckman, J. J., "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 1997, *32* (3), 441–462.

Heckman, James J., Lance J. Lochner, and Petra E. Todd, "Fifty Years of Mincer Earnings Regressions," *NBER Working Paper*, 2003, (9732).

Heckman, J.J. and E. Vytlacil, "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 1999, *96*.

_ and _ , "Econometric Evaluation of Social Programs," *Handbook of Econometrics*, 2007.

Heckman, J.J and Xuesong Li, "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000," *IFAU Working Paper*, 2003, (2003:17).

Heckman, J.J., S. Urzua, and E. Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of economics and statistics*, 2006, *88* (3).

Kane, Thomas J. and Cecilia E. Rouse, "Labor Market Returns to Two- and Four-Year Colleges: is a Credit a Credit and Do Degrees Matter?," *NBER Working Paper*, 1993, (4268).

Katz, L. and K. Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 1992, *107*, 35–78.

Kyui, Natalia, "Tertiary Education Expansion, Family Background and Educational Choices," *Working Paper, Centre d'Economie de la Sorbonne*, 2011.

Oreopoulos, Philip, "The Compelling Effects of Compulsory Schooling: Evidence from Canada," *Canadian Journal of Economics*, 2006, *39* (1), 22–52.

_ , "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *The American Economic Review*, 2006, *96* (1), 152–175.

Pischke, Steve and Till von Wachter, "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation," *Review of Economics and Statistics*, 2008, *90* (3), 592–598.

Swafford, M., M. Kosolapov, and P. Kozyreva, "International standarts for evaluating survey research.," *Universe of Russia*, 1999, *1-2*, 281–302.

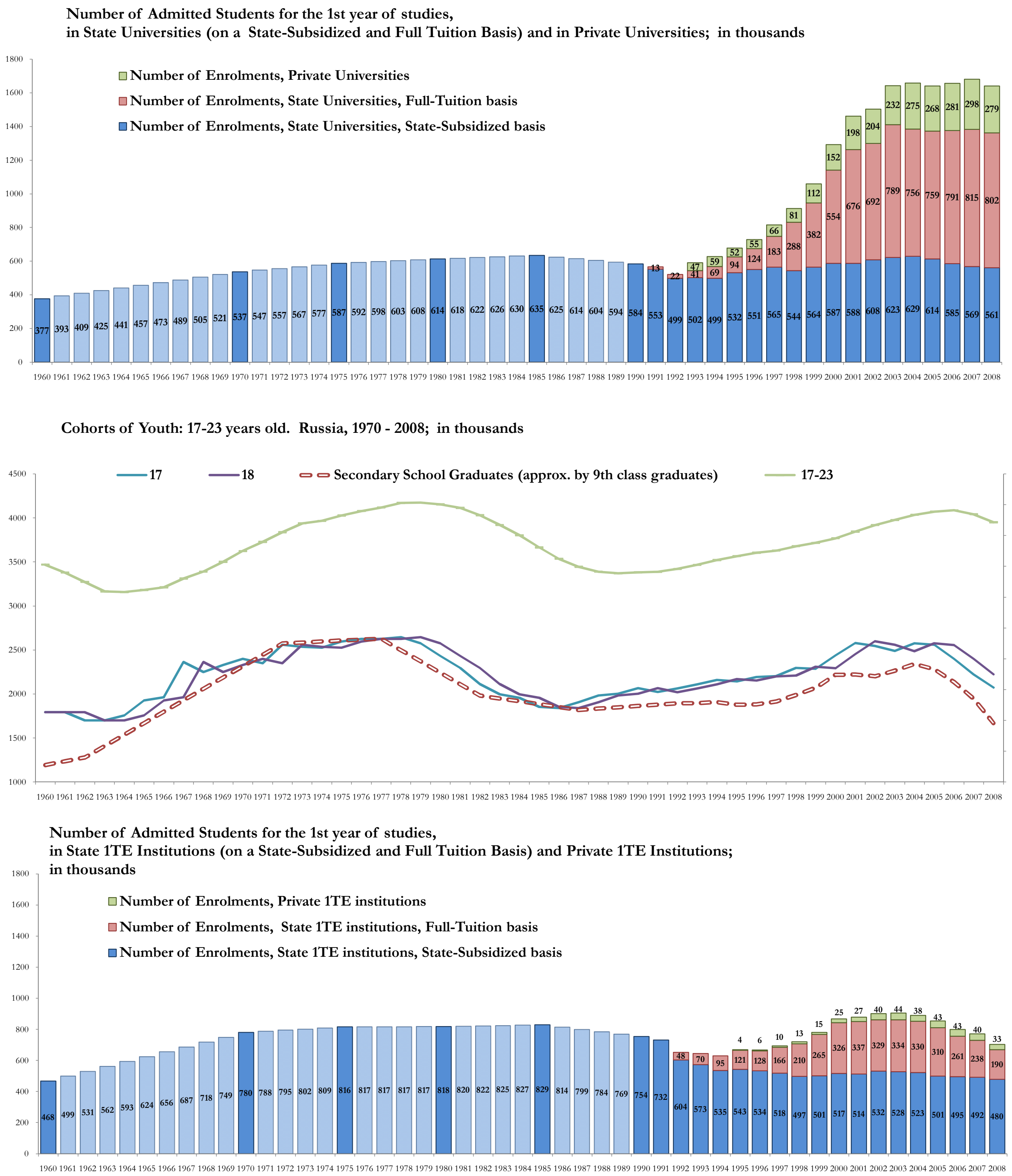
_ , _ , and _ , "The Russian Logitudinal Monitoring Survey: Measuring the Well-Being of Russians during the 1990s.," *Universe of Russia*, 1999, *3*, 153–172.

Walker, Ian and Yu Zhu, "The College Wage Premium and the Expansion of Higher Education in the UK," *Scandinavian Journal of Economics*, 2008, *110* (4), 695–709.

Wang, X., B. M. Fleisher, H. Li, and S. Li, "Access to Higher Education and Inequality: The Chinese Experiment," *IZA Working Paper*, 2007, (2823).

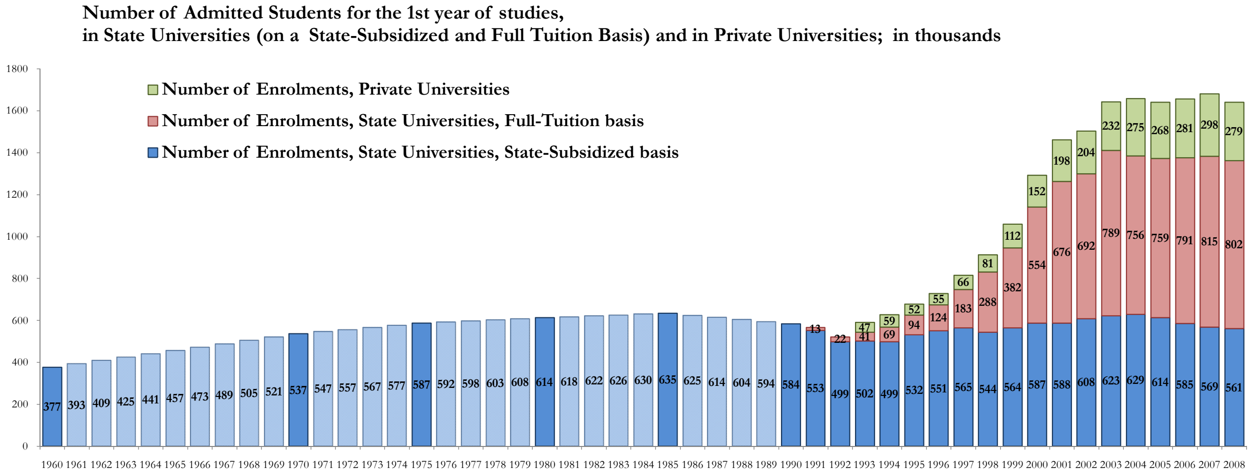
A Appendix: Figures

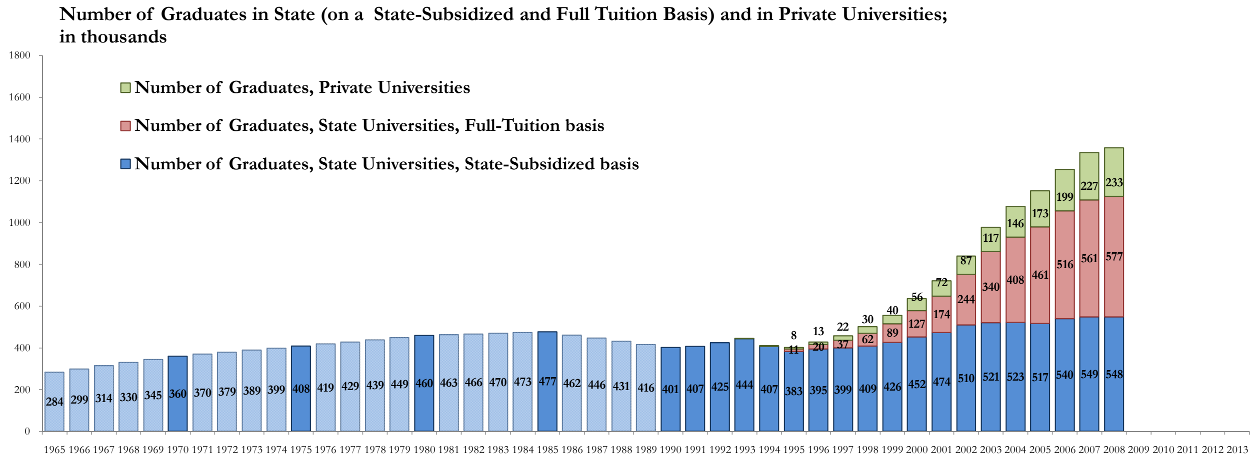
Figure A-1: 1st Level, 2nd Level Tertiary Education Admission & Cohort Size, by years. In thousands of students.

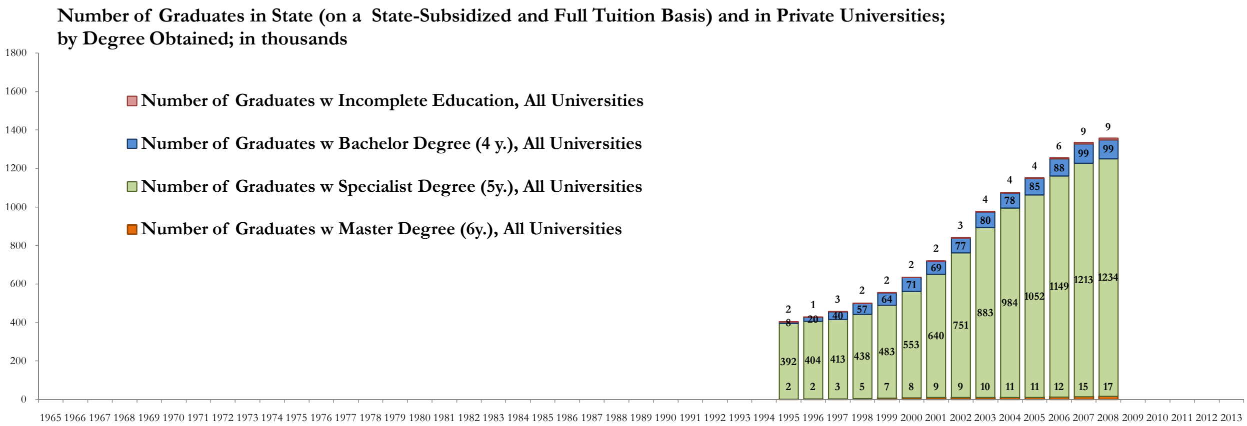


Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

Figure A-2: 2nd Level Tertiary Education Admission & Graduation, by years. In thousands of students.

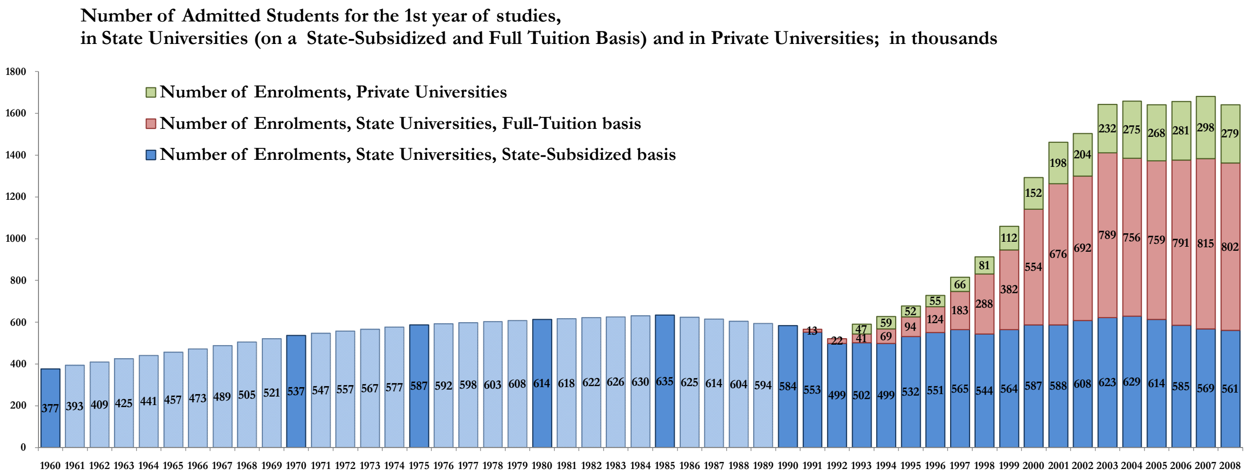


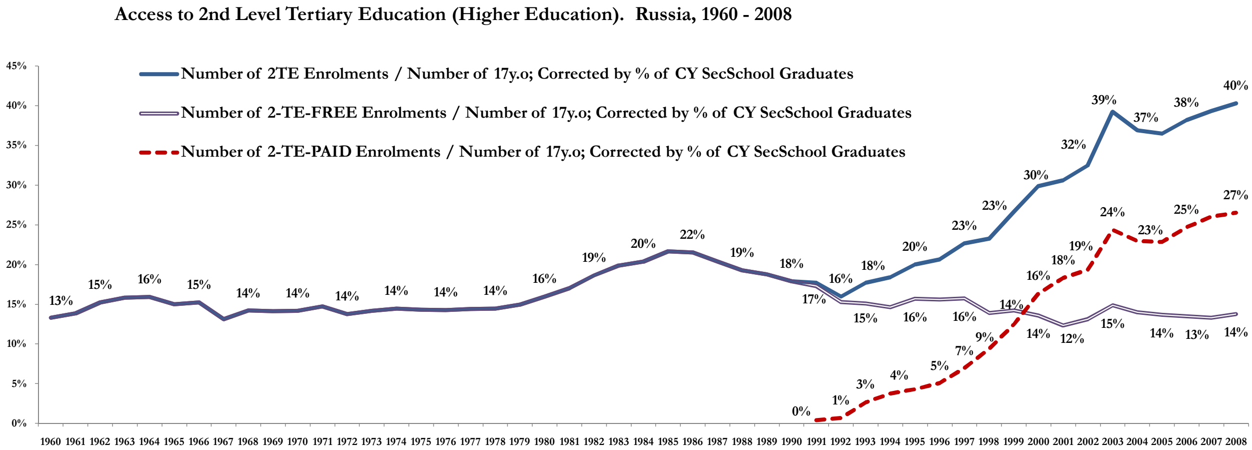




Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

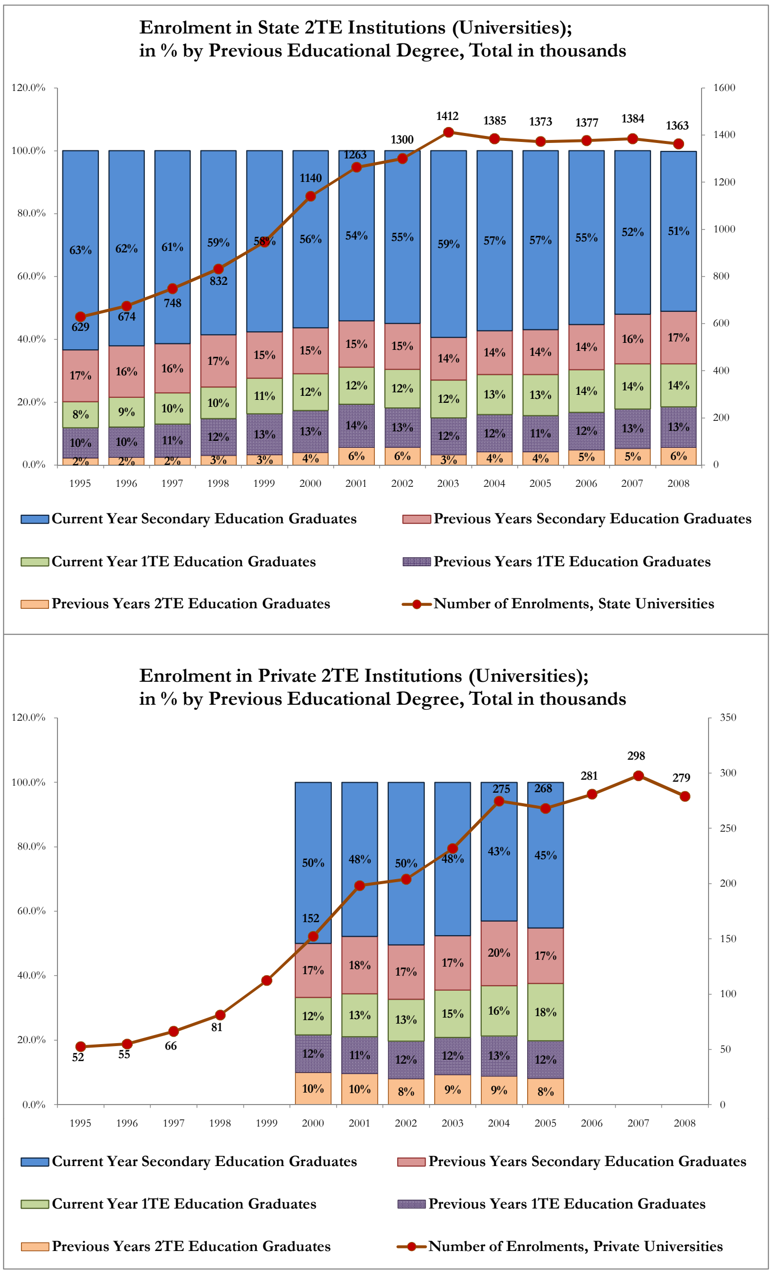
Figure A-3: Number of Students Admitted to the 2nd-level Tertiary Education, by years. In thousand students, and in % to the Corresponding Cohort of Youths.





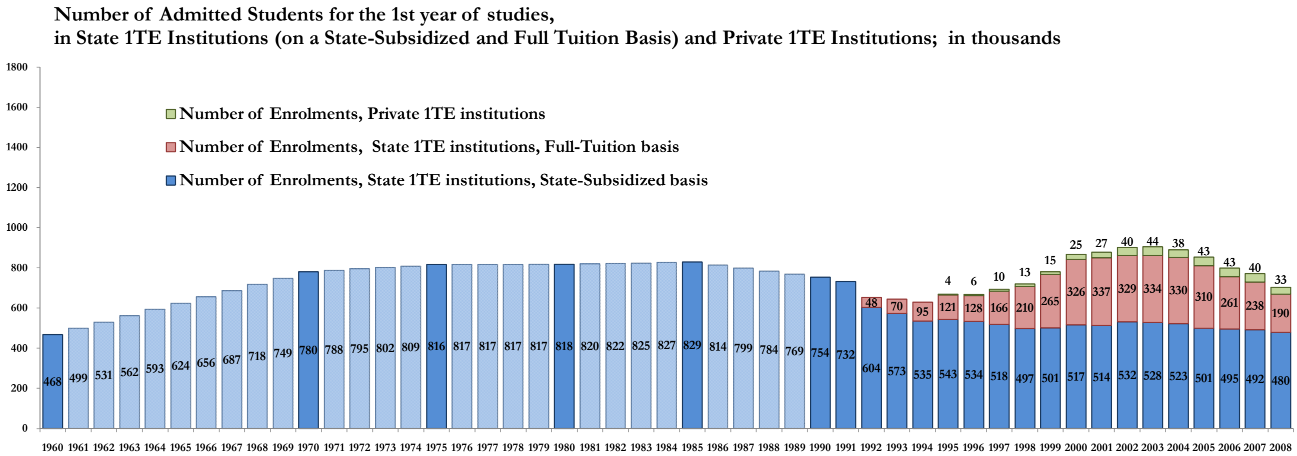
Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

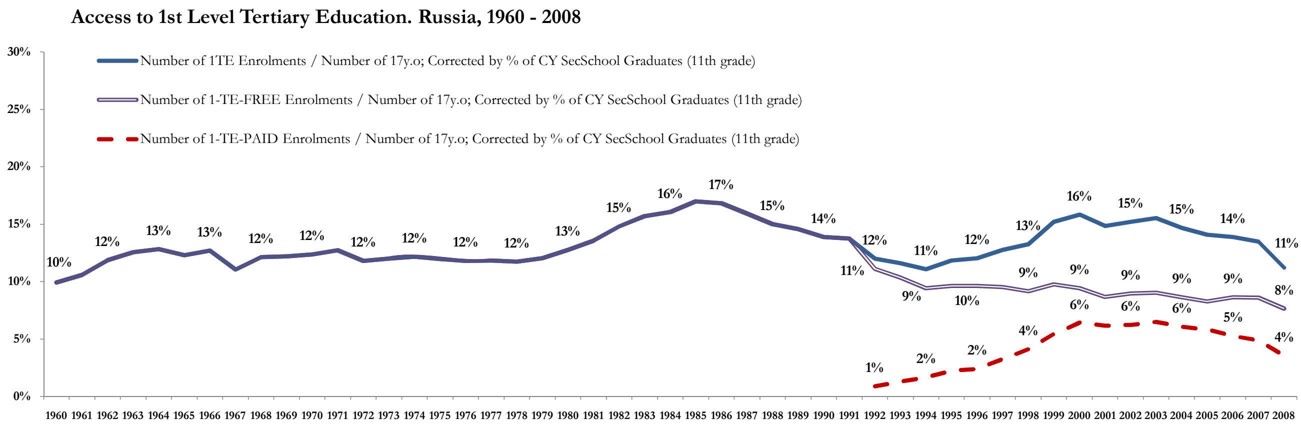
Figure A-4: 2nd Level Tertiary Education Admission, by years,
By Previous Level of Education Obtained.
In thousands of students.

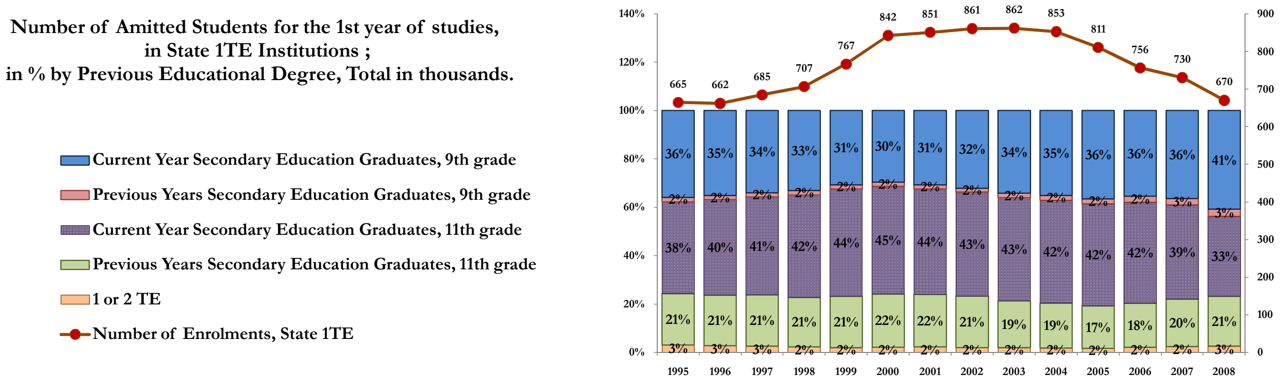


Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

Figure A-5: Number of Students Admitted to the 1st-level Tertiary Education, by years. In thousand students, and in % to the Corresponding Cohort of Youths.

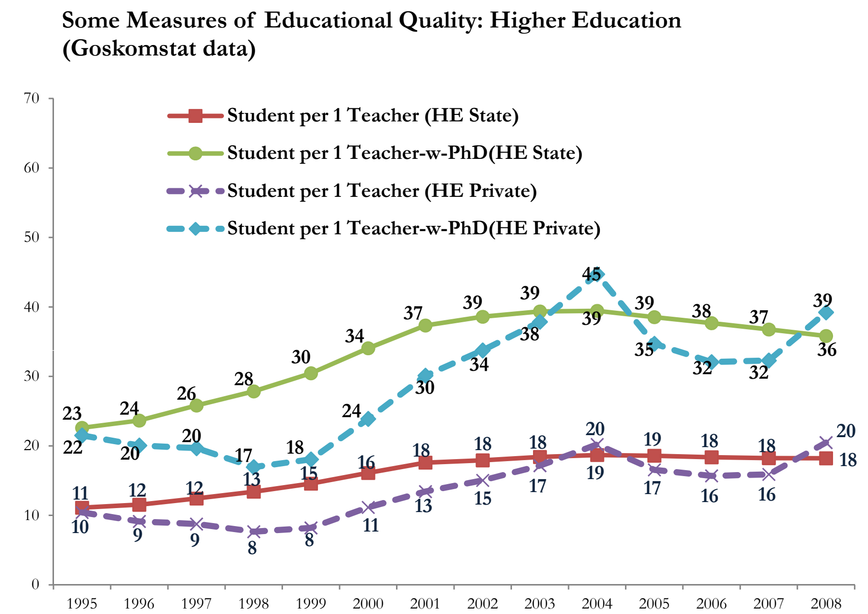


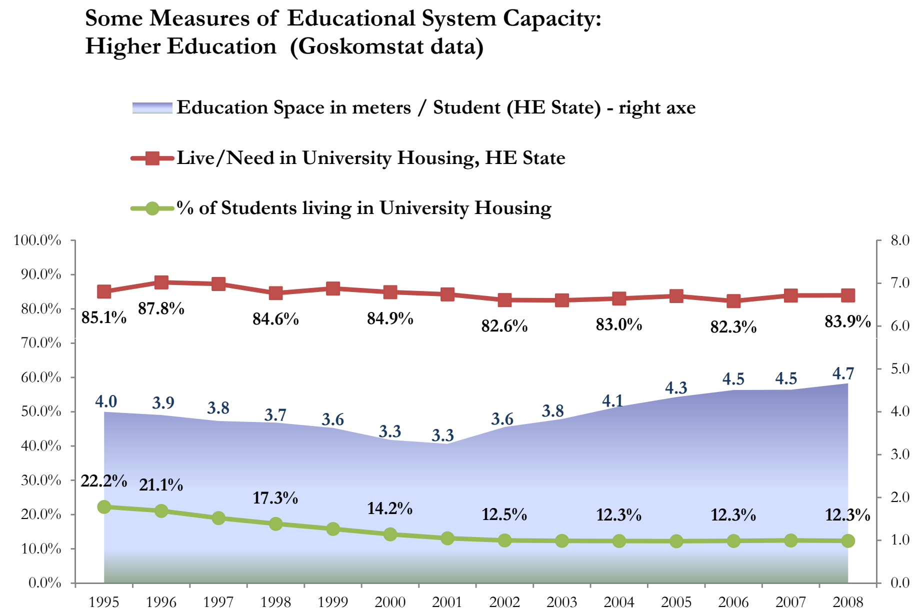




Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics.

Figure A-6: Changes in the Educational Quality of Universities, Russia, 1995-2008.





Source: RLMS data, 1995-2008, Author's calculations.

B Time and Within Region Instruments' Variation

In this section, an exercise is presented in order to show the variation of the instruments among regions and years. The idea is to show what we gain by looking at year-region variation and not just time or regional variations, and whether there is still the variation in our instruments if we control for the year and regions fixed effects. The regression is run using the regional level data of the number of slots in the higher education and proportion of slots in the higher education on the federal districts and years fixed effects. *Table B-1* presents the results.

Table B-1: Time-Region Variation in the Number of Slots in Higher Education

Variables	**♯ of 2TE Slots**			**♯ of 2TE Slots / Cohort Size**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Fixed Effects:						
Federal Districts	*X*	*X*		*X*	*X*	
Years	*X*		*X*	*X*		*X*
R^2	0.882	0.513	0.369	0.958	0.129	0.829
Observations	*273*	*273*	*273*	*273*	*273*	*273*

Observations: 7 Federal Districts * 39 time periods = 273

Source: Education in the Russian Federation (Gohberg et al. (2007)), Goskomstat (www.gks.ru) - Russian Federal Committee for Statistics., Author's calculations

This analysis also provides another explanation, why the number of slots seems to be a better instrument than the proportion of slots relative to the cohort. With the control for the regional and year of birth fixed effects, the number of slots has a higher variation than the proportion of slots. As we can see, the part of the variation of the "number of slots" on the regional level, which is explained by federal districts and years fixed effects, is equal to 88%, while the same part for the "proportions" is equal to 96%. 51% of the variation in the number of slots is explained by the permanent differences between regions, while 37% - by the over-time variations across all regions. Therefore, the remain of this variation in instruments' values is explored when the controls for regions and year of birth are included in the regression.

For all the analysis, the *Number of Slots in Higher Education (2TE)* is used as the instrument for educational attainment, because it which has showed the highest explanatory power. Controlling for the cohort size, however, does not significantly change the estimation results.

C Technical Appendix

In this section, we discuss the methods of estimation of econometric models, used in the paper. We present the joint models of educational choice and wages (Section 1: 1.1 and 1.2), educational choice and employment (Section 2: 2.1 and 2.2), as well as the joint models of educational choice, employment and wages (Section 3: 3.1 and 3.2). The Section 4 of this Appendix describes the estimation procedure for the Marginal Treatment Effects.

1. Educational Choice and Wages

1.1. Number of Years of Schooling and Wages

In this section, we describe the empirical model of wages with endogenous number of years of schooling.

$$\begin{aligned} Ed_i &= Education_i = \alpha_{ed} + X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\ W_i &= Wage_i = \quad \alpha_w + X_{w,i}\beta_w + \varepsilon_{2,i}; \\ \left\{ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \end{array} \right\} &\rightsquigarrow \quad N\left\{ E = \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \Sigma = \left(\begin{array}{cc} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{array} \right) \right\}. \end{aligned} \tag{26}$$

We note all parameters of the model, which we estimate, as:
$\theta = \{\alpha_{ed}; \beta_{ed}; \alpha_w; \beta_w; \sigma^2_{\varepsilon_1}; \sigma_{\varepsilon_1\varepsilon_2}; \sigma^2_{\varepsilon_2}\}$.
We note $X_i = \{X_{ed,i}; X_{w,i}\}$ - the set of all variables, which are used in the model. Note that Ed_i is included in $X_{w,i}$

This model is the case of two linear equations with correlated random terms.

The likelihood function for this model can be written as the following:

$$\begin{aligned} l(\theta, X_i) &= f(Ed_i, W_i) = \\ &= (\text{probability density function of bivariate normal distribution of } Ed_i \text{ and } W_i) \\ &= \frac{1}{2\pi\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}\sqrt{1-(\rho_{\varepsilon_1\varepsilon_2})^2}} \cdot exp\left(-\frac{Q}{2\cdot(1-(\rho_{\varepsilon_1\varepsilon_2})^2)}\right); \\ Q &= \left(\frac{Ed_i - (\alpha_{ed} + X_{ed,i}\beta_{ed})}{\sigma_{\varepsilon_1}}\right)^2 + \left(\frac{W_i - (\alpha_w + X_{w,i}\beta_w)}{\sigma_{\varepsilon_2}}\right)^2 \\ &\quad -2\rho_{\varepsilon_1\varepsilon_2}\frac{(Ed_i - (\alpha_{ed} + X_{ed,i}\beta_{ed}))(W_i - (\alpha_w + X_{w,i}\beta_w))}{\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}}; \\ \rho_{\varepsilon_1\varepsilon_2} &= \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sqrt{\sigma^2_{\varepsilon_1}\sigma^2_{\varepsilon_2}}}. \end{aligned} \tag{27}$$

Then,

$$L(\theta, X) = \prod_{i=1}^{N}(l(\theta, X_i)), \text{ and } ln(L(\theta, X)) = \sum_{i=1}^{N} ln(l(\theta, X_i)).$$

We use the Maximum Likelihood Method, by programming the likelihood function in STATA and then using *ml max* command.

In order to estimate the parameters of the covariance matrix: $\sigma^2_{\varepsilon_1}$, $\sigma_{\varepsilon_1\varepsilon_2}$, $\sigma^2_{\varepsilon_2}$, we use the Cholesky decomposition.

Cholesky decomposition states that: if A is a symmetric and positive definite matrix, then there exists a triangular operator matrix L such that $A = L \cdot L'$. Where L is a lower triangular matrix with strictly positive diagonal entries.

Therefore, we estimate three parameters l_{11}, l_{21}, l_{22} in the Cholesky matrix:

$$L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix}$$

Covariance matrix can be expressed as the following (Cholesky decomposition, thus, guarantees the matrix to by symmetric and positive definite):

$$\Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{pmatrix} = L \cdot L' = \begin{pmatrix} l^2_{11} & l_{11} \cdot l_{21} \\ l_{11} \cdot l_{21} & l^2_{21} + l^2_{22} \end{pmatrix}$$

As the maximization algorithm we use BHHH (Berndt, Hall, Hall, and Hausman).

The correctness of the program code has been verified with simulated data.

1.2. Higher Education (2TE) Attainment and Wages

In this section we describe the empirical model of wages (W_i) with endogenous education, which is described by the binary variable - higher education degree attainment (Ed_i).

$$\begin{aligned} Ed_i &= I(Y^*_{2TE,i} \geqslant 0), \\ \text{where} \quad & Y^*_{2TE,i} = X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\ W_i &= X_{w,i}\beta_w + \varepsilon_{2,i}; \end{aligned} \tag{28}$$

$$\left\{ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \end{array} \right\} \rightsquigarrow N \left\{ E = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{pmatrix} \right\}.$$

We have to normalize $\sigma^2_{\varepsilon_1} = 1$.

We note all parameters of the model, which we estimate, as:
$\theta = \{\beta_{ed}; \beta_w; \sigma_{\varepsilon_1\varepsilon_2}; \sigma^2_{\varepsilon_2}\}$.
We note $X_i = \{X_{ed,i}; X_{w,i}\}$ - the set of all variables, which are used in the model. Note

that Ed_i is included in $X_{w,i}$, and constant is included in both $X_{ed,i}$ and $X_{w,i}$.

This model is the case of two equations with correlated random terms: probit equation and linear equation.

The likelihood function for this model can be written as the following:

$$L(\theta, X) = \prod_{i=1}^{N} \left\{ (l_{Ed_i=1}(\theta, X_i))^{Ed_i} (l_{Ed_i=0}(\theta, X_i))^{1-Ed_i} \right\} \tag{29}$$

Where the parts of the likelihood function for individuals with higher education degree ($l_{Ed_i=1}(\theta, X_i)$) and without higher education degree ($l_{Ed_i=0}(\theta, X_i)$) can be expressed as the following:

$$\begin{aligned}
l_{Ed_i=1}(\theta, X_i) &= P(X_{ed,i}\beta_{ed} + \varepsilon_{1,i} \geqslant 0;\ W_i = X_{w,i}\beta_w + \varepsilon_{2,i}) = \\
&= f(W_i) \cdot P(X_{ed,i}\beta_{ed} + \varepsilon_{1,i} \geqslant 0 | W_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}} \phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_2}}\right) \cdot (1 - P(\varepsilon_{1,i} < -X_{ed,i}\beta_{ed} | W_i)) = \\
&= \frac{1}{\sigma_{\varepsilon_2}} \phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_2}}\right) \cdot \left(1 - \Phi\left(\frac{-X_{ed,i}\beta_{ed} - \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(W_i - X_{w,i}\beta_w)}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right)\right) = \\
&= \frac{1}{\sigma_{\varepsilon_2}} \phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_2}}\right) \cdot \Phi\left(\frac{X_{ed,i}\beta_{ed} + \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(W_i - X_{w,i}\beta_w)}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right) =
\end{aligned}$$

and

$$\begin{aligned}
l_{Ed_i=0}(\theta, X_i) &= P(X_{ed,i}\beta_{ed} + \varepsilon_{1,i} < 0;\ W_i = X_{w,i}\beta_w + \varepsilon_{2,i}) = \\
&= f(W_i) \cdot P(X_{ed,i}\beta_{ed} + \varepsilon_{1,i} < 0 | W_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}} \phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_2}}\right) \cdot P(\varepsilon_{1,i} < -X_{ed,i}\beta_{ed} | W_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}} \phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_2}}\right) \cdot \Phi\left(\frac{-X_{ed,i}\beta_{ed} - \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(W_i - X_{w,i}\beta_w)}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right) =
\end{aligned}$$

Where $\phi()$ is a standard normal probability density function, and $\Phi()$ is a cumulative distribution function of a standard normal distribution.

2. Educational Choice and Employment

2.1. Number of Years of Schooling and Employment

In this section we describe the empirical model of employment ($Empl_i$) with endogenous education, which is described by the linear variable - number of years of schooling (Ed_i).

$$\begin{aligned} Empl_i &= I(Empl_i^* \geqslant 0), \\ \text{where} \quad & Empl_i^* = X_{em,i}\beta_{em} + \varepsilon_{1,i}; \\ \\ Ed_i &= X_{ed,i}\beta_{ed} + \varepsilon_{2,i}; \end{aligned} \tag{30}$$

$$\left\{ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \end{array} \right\} \rightsquigarrow N \left\{ E = \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \Sigma = \left(\begin{array}{cc} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{array} \right) \right\}.$$

We have to normalize $\sigma^2_{\varepsilon_1} = 1$.

We note all parameters of the model, which we estimate, as:
$\theta = \{\beta_{em}; \beta_{ed}; \sigma_{\varepsilon_1\varepsilon_2}; \sigma^2_{\varepsilon_2}\}$.
We note $X_i = \{X_{em,i}; X_{ed,i}\}$ - the set of all variables, which are used in the model. Note that Ed_i is included in $X_{em,i}$, and constant is included in both $X_{em,i}$ and $X_{ed,i}$.

This model is the case of two equations with correlated random terms: probit equation and linear equation.

The likelihood function for this model can be written as the following:

$$L(\theta, X) = \prod_{i=1}^{N} \left\{ (l_{Empl_i=1}(\theta, X_i))^{Empl_i} (l_{Empl_i=0}(\theta, X_i))^{1-Empl_i} \right\} \tag{31}$$

Where the parts of the likelihood function for employed ($l_{Empl_i=1}(\theta, X_i)$) and unemployed individuals ($l_{Empl_i=0}(\theta, X_i)$) can be expressed as the following:

$$
\begin{aligned}
l_{Empl_i=1}(\theta, X_i) &= P(X_{em,i}\beta_{em} + \varepsilon_{1,i} \geqslant 0;\ Ed_i = X_{ed,i}\beta_{ed} + \varepsilon_{2,i}) = \\
&= f(Ed_i) \cdot P(X_{em,i}\beta_{em} + \varepsilon_{1,i} \geqslant 0 | Ed_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_2}}\right) \cdot (1 - P(\varepsilon_{1,i} < -X_{em,i}\beta_{em} | Ed_i)) = \\
&= \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_2}}\right) \cdot \left(1 - \Phi\left(\frac{-X_{em,i}\beta_{em} - \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(Ed_i - X_{ed,i}\beta_{ed})}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right)\right) = \\
&= \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_2}}\right) \cdot \Phi\left(\frac{X_{em,i}\beta_{em} + \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(Ed_i - X_{ed,i}\beta_{ed})}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right) =
\end{aligned}
$$

and

$$
\begin{aligned}
l_{Empl_i=0}(\theta, X_i) &= P(X_{em,i}\beta_{em} + \varepsilon_{1,i} < 0;\ Ed_i = X_{ed,i}\beta_{ed} + \varepsilon_{2,i}) = \\
&= f(Ed_i) \cdot P(X_{em,i}\beta_{em} + \varepsilon_{1,i} < 0 | Ed_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_2}}\right) \cdot P(\varepsilon_{1,i} < -X_{em,i}\beta_{em} | Ed_i) = \\
&= \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_2}}\right) \cdot \Phi\left(\frac{-X_{em,i}\beta_{em} - \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_2}}(Ed_i - X_{ed,i}\beta_{ed})}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_2}}}}\right) =
\end{aligned}
$$

Where $\phi()$ is a standard normal probability density function, and $\Phi()$ is a cumulative distribution function of a standard normal distribution.

2.2. Higher Education (2TE) Attainment and Employment

In this section we describe the empirical model of employment ($Empl_i$) with endogenous education, which is described by the binary variable - higher education degree attainment (Ed_i).

$$
\begin{aligned}
Ed_i &= I(Y^*_{2TE,i} \geqslant 0), \\
\text{where} \quad & Y^*_{2TE,i} = X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\
Empl_i &= I(Empl^*_i \geqslant 0), \\
\text{where} \quad & Empl^*_i = X_{em,i}\beta_{em} + \varepsilon_{2,i};
\end{aligned} \tag{32}
$$

$$
\begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{Bmatrix} \rightsquigarrow N\left\{E = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{pmatrix}\right\}.
$$

We have to normalize $\sigma^2_{\varepsilon_1} = 1$ and $\sigma^2_{\varepsilon_2} = 1$.

We note all parameters of the model, which we estimate, as:
$\theta = \{\beta_{em}; \beta_{ed}; \sigma_{\varepsilon_1\varepsilon_2}\}$.

We note $X_i = \{X_{em,i}; X_{ed,i}\}$ - the set of all variables, which are used in the model. Note that Ed_i is included in $X_{em,i}$, and constant is included in both $X_{em,i}$ and $X_{ed,i}$.

This model is the case of two probit equations with correlated random terms. The estimation procedure is equivalent to bivariate probit in Stata (*biprobit*).

The likelihood function for this model can be written as the following:

$$L(\theta, X) = \prod_{i=1}^{N}\{(l_{Ed_i=1,Empl_i=1}(\theta, X_i))^{Empl_i \cdot Ed_i} \cdot (l_{Ed_i=0,Empl_i=0}(\theta, X_i))^{(1-Empl_i)(1-Ed_i)} \cdot \\ \cdot (l_{Ed_i=0,Empl_i=1}(\theta, X_i))^{Empl_i \cdot (1-Ed_i)} (l_{Ed_i=1,Empl_i=0}(\theta, X_i))^{(1-Empl_i)(Ed_i)}\} \quad (33)$$

Where the parts of the likelihood function can be expressed as the following:

For Unemployed People without Higher Education Degree:

$$l_{Ed_i=0,Empl_i=0}(\theta, X_i) = P(\varepsilon_1 < -X_{ed,i}\beta_{ed};\ \varepsilon_2 < -X_{em,i}\beta_{em}) = \\ = \Phi_2(-X_{ed,i}\beta_{ed};\ -X_{em,i}\beta_{em};\ \rho_{\varepsilon_1\varepsilon_2})$$

For Employed People without Higher Education Degree:

$$l_{Ed_i=0,Empl_i=1}(\theta, X_i) = P(\varepsilon_1 < -X_{ed,i}\beta_{ed};\ -\varepsilon_2 \leqslant X_{em,i}\beta_{em}) = \\ = \Phi_2(-X_{ed,i}\beta_{ed};\ X_{em,i}\beta_{em};\ -\rho_{\varepsilon_1\varepsilon_2})$$

For Unemployed People with Higher Education Degree:

$$l_{Ed_i=1,Empl_i=0}(\theta, X_i) = P(-\varepsilon_1 \leqslant X_{ed,i}\beta_{ed};\ \varepsilon_2 < -X_{em,i}\beta_{em}) = \\ = \Phi_2(X_{ed,i}\beta_{ed};\ -X_{em,i}\beta_{em};\ -\rho_{\varepsilon_1\varepsilon_2})$$

For Employed People with Higher Education Degree:

$$l_{Ed_i=1,Empl_i=1}(\theta, X_i) = P(-\varepsilon_1 \leqslant X_{ed,i}\beta_{ed};\ -\varepsilon_2 \leqslant X_{em,i}\beta_{em}) = \\ = \Phi_2(X_{ed,i}\beta_{ed};\ X_{em,i}\beta_{em};\ \rho_{\varepsilon_1\varepsilon_2})$$

Where $\rho_{\varepsilon_1\varepsilon_2} = \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sqrt{\sigma^2_{\varepsilon_1}\sigma^2_{\varepsilon_2}}} = \sigma_{\varepsilon_1\varepsilon_2}$.

$\Phi_2(a_1; a_2; \rho)$ is a joint cumulative distribution function for the two bivariate normally distributed variables with means = 0, variations = 1, and correlation ρ.

We use the Maximum Likelihood Method, by programming the likelihood function in STATA.

3. Educational Choice, Employment and Wages

3.1. Number of Years of Schooling, Employment and Wages

In this section we describe the empirical model, which jointly estimates educational choice (number of years of schooling - Ed_i), employment ($Empl_i$), and wages (W_i), where wages are observed only for the employed population.

$$
\begin{aligned}
Ed_i &= X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\
\\
Empl_i &= I(Empl_i^* \geqslant 0), \\
\text{where} \quad & Empl_i^* = X_{em,i}\beta_{em} + \varepsilon_{2,i}; \\
\\
W_i &= W_i^* \cdot (Empl_i), \\
\text{where} \quad & W_i^* = X_{w,i}\beta_w + \varepsilon_{3,i};
\end{aligned}
\tag{34}
$$

$$
\left\{ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{array} \right\} \rightsquigarrow N \left\{ E = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right), \Sigma = \left(\begin{array}{ccc} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} & \sigma_{\varepsilon_2\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_3} & \sigma_{\varepsilon_2\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{array} \right) \right\}.
$$

We have to normalize $\sigma^2_{\varepsilon_2} = 1$.

We note all parameters of the model, which we estimate, as:
$\theta = \{\beta_{ed}; \beta_{em}; \beta_w; \sigma^2_{\varepsilon_1}; \sigma^2_{\varepsilon_3}; \sigma_{\varepsilon_1\varepsilon_2}; \sigma_{\varepsilon_1\varepsilon_3}; \sigma_{\varepsilon_2\varepsilon_3}\}$.
We note $X_i = \{X_{ed,i}; X_{em,i}; X_{w,i}\}$ - the set of all variables, which are used in the model. Note that Ed_i is included in $X_{em,i}$ and $X_{w,i}$, and constant is included in all $X_{ed,i}$, $X_{em,i}$ and $X_{w,i}$.

This model is the case of three simultaneous equations with correlated random terms: linear, probit, and linear equations.

The likelihood function for this model can be written as the following:

$$
L(\theta, X) = \prod_{i=1}^{N} \left\{ (l_{Empl_i=0}(\theta, X_i))^{(1-Empl_i)} \cdot (l_{Empl_i=1}(\theta, X_i))^{Empl_i} \right\} \tag{35}
$$

Where the parts of the likelihood function for the employed ($l_{Empl_i=1}(\theta, X_i)$) and unemployed ($l_{Empl_i=0}(\theta, X_i)$) population can be expressed in the following way.

For the unemployed population:

$$\begin{aligned} l_{Empl_i=0}(\theta, X_i) &= P(Ed_i = X_{ed,i}\beta_{ed} + \varepsilon_{1,i};\; X_{em,i}\beta_{em} + \varepsilon_{2,i} < 0) = \\ &= f(Ed_i) \cdot P(X_{em,i}\beta_{em} + \varepsilon_{2,i} < 0 | Ed_i) = \\ &= \frac{1}{\sigma_{\varepsilon_1}} \phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_1}}\right) \cdot P(\varepsilon_{2,i} < -X_{em,i}\beta_{em} | Ed_i) = \\ &= \frac{1}{\sigma_{\varepsilon_1}} \phi\left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_1}}\right) \cdot \Phi\left(\frac{-X_{em,i}\beta_{em} - \frac{\sigma_{\varepsilon_1\varepsilon_2}}{\sigma^2_{\varepsilon_1}}(Ed_i - X_{ed,i}\beta_{ed})}{\sqrt{1 - \frac{(\sigma_{\varepsilon_1\varepsilon_2})^2}{\sigma^2_{\varepsilon_1}}}}\right) = \end{aligned}$$

For the employed population:

$$\begin{aligned} & l_{Empl_i=1}(\theta, X_i) = P(Ed_i = X_{ed,i}\beta_{ed} + \varepsilon_{1,i};\; X_{em,i}\beta_{em} + \varepsilon_{2,i} \geqslant 0;\; W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\ & = f(Ed_i,\; W_i) \cdot P(-\varepsilon_{2,i} \leqslant X_{em,i}\beta_{em} | Ed_i, W_i) = \\ & = f(Ed_i,\; W_i) \cdot \\ & \cdot\Phi\left(\frac{X_{em,i}\beta_{em} - \left[(-\sigma_{\varepsilon_1\varepsilon_2} \quad -\sigma_{\varepsilon_2\varepsilon_3}) \cdot \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix}^{-1} \cdot \begin{pmatrix} Ed_i - X_{ed,i}\beta_{ed} \\ W_i - X_{w,i}\beta_w \end{pmatrix}\right]}{\sqrt{\sigma^2_{\varepsilon_2} - \left[(-\sigma_{\varepsilon_1\varepsilon_2} \quad -\sigma_{\varepsilon_2\varepsilon_3}) \cdot \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix}^{-1} \cdot \begin{pmatrix} -\sigma_{\varepsilon_1\varepsilon_2} \\ -\sigma_{\varepsilon_2\varepsilon_3} \end{pmatrix}\right]}}\right) \end{aligned}$$

Where:

$$\begin{aligned} f(Ed_i, W_i) &= \frac{1}{2\pi\sigma_{\varepsilon_1}\sigma_{\varepsilon_3}\sqrt{1 - (\rho_{\varepsilon_1\varepsilon_3})^2}} \cdot exp\left(-\frac{Q}{2 \cdot (1 - (\rho_{\varepsilon_1\varepsilon_3})^2)}\right); \\ Q &= \left(\frac{Ed_i - X_{ed,i}\beta_{ed}}{\sigma_{\varepsilon_1}}\right)^2 + \left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_3}}\right)^2 \\ & -2\rho_{\varepsilon_1\varepsilon_3}\frac{(Ed_i - X_{ed,i}\beta_{ed})(W_i - X_{w,i}\beta_w)}{\sigma_{\varepsilon_1}\sigma_{\varepsilon_3}}; \\ \rho_{\varepsilon_1\varepsilon_3} &= \frac{\sigma_{\varepsilon_1\varepsilon_3}}{\sqrt{\sigma^2_{\varepsilon_1}\sigma^2_{\varepsilon_3}}}. \end{aligned}$$

Note that in [...] we place the operations with matrixes. $\Phi()$ is a cumulative distribution function of a standard normal distribution.

3.2. Higher Education (2TE) Attainment, Employment and Wages

In this section we describe the empirical model, which jointly estimates educational choice (higher education degree attainment - Ed_i), employment ($Empl_i$), and wages (W_i), where wages are observed only for the employed population.

$$\begin{aligned} Ed_i &= I(Y^*_{2TE,i} \geqslant 0), \\ \text{where} \quad & Y^*_{2TE,i} = X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\ \\ Empl_i &= I(Empl^*_i \geqslant 0), \\ \text{where} \quad & Empl^*_i = X_{em,i}\beta_{em} + \varepsilon_{2,i}; \\ \\ W_i &= W^*_i \cdot (Empl_i), \\ \text{where} \quad & W^*_i = X_{w,i}\beta_w + \varepsilon_{3,i}; \end{aligned} \tag{36}$$

$$\left\{ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{array} \right\} \rightsquigarrow N \left\{ E = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} & \sigma_{\varepsilon_2\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_3} & \sigma_{\varepsilon_2\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix} \right\}.$$

We have to normalize $\sigma^2_{\varepsilon_1} = 1$, $\sigma^2_{\varepsilon_2} = 1$.

We note all parameters of the model, which we estimate, as:
$\theta = \{\beta_{ed}; \beta_{em}; \beta_w; \sigma^2_{\varepsilon_3}; \sigma_{\varepsilon_1\varepsilon_2}; \sigma_{\varepsilon_1\varepsilon_3}; \sigma_{\varepsilon_2\varepsilon_3}\}$.
We note $X_i = \{X_{ed,i}; X_{em,i}; X_{w,i}\}$ - the set of all variables, which are used in the model. Note that Ed_i is included in $X_{em,i}$ and $X_{w,i}$, and constant is included in all $X_{ed,i}$, $X_{em,i}$ and $X_{w,i}$.

This model is the case of three simultaneous equations with correlated random terms: probit, probit, and linear equations.

The likelihood function for this model can be written as the following:

$$\begin{aligned} L(\theta, X) = & \prod_{i=1}^{N} \{ (l_{Ed_i=1, Empl_i=1}(\theta, X_i))^{Empl_i \cdot Ed_i} \cdot (l_{Ed_i=0, Empl_i=0}(\theta, X_i))^{(1-Empl_i)(1-Ed_i)} \cdot \\ & \cdot (l_{Ed_i=0, Empl_i=1}(\theta, X_i))^{Empl_i \cdot (1-Ed_i)} (l_{Ed_i=1, Empl_i=0}(\theta, X_i))^{(1-Empl_i)(Ed_i)} \} \end{aligned} \tag{37}$$

Where the parts of the likelihood function for the unemployed population can be expressed as the following:

For Unemployed People without Higher Education Degree:

$$l_{Ed_i=0,Empl_i=0}(\theta, X_i) = P(\varepsilon_1 < -X_{ed,i}\beta_{ed};\ \varepsilon_2 < -X_{em,i}\beta_{em}) = \\ = \Phi_2(-X_{ed,i}\beta_{ed};\ -X_{em,i}\beta_{em};\ \rho_{\varepsilon_1\varepsilon_2})$$

For Unemployed People with Higher Education Degree:

$$l_{Ed_i=1,Empl_i=0}(\theta, X_i) = P(-\varepsilon_1 \leqslant X_{ed,i}\beta_{ed};\ \varepsilon_2 < -X_{em,i}\beta_{em}) = \\ = \Phi_2(X_{ed,i}\beta_{ed};\ -X_{em,i}\beta_{em};\ -\rho_{\varepsilon_1\varepsilon_2})$$

And the parts of the likelihood function for the employed population can be expressed as the following:

For Employed People without Higher Education Degree:

$$l_{Ed_i=0,Empl_i=1}(\theta, X_i) = P(\varepsilon_1 < -X_{ed,i}\beta_{ed};\ -\varepsilon_2 < X_{em,i}\beta_{em};\ W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\ = f(W_i) \cdot P(\varepsilon_1 < -X_{ed,i}\beta_{ed};\ -\varepsilon_2 < X_{em,i}\beta_{em} |\ W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\ = \frac{1}{\sigma_{\varepsilon_3}}\phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_3}}\right) \cdot \\ \cdot\Phi_2\left(\frac{-X_{ed,i}\beta_{ed} - \frac{\sigma_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma^2_{\varepsilon_1} - \frac{\sigma^2_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}};\ \frac{X_{em,i}\beta_{em} + \frac{\sigma_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma^2_{\varepsilon_2} - \frac{\sigma^2_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}};\ -\rho_{(1,2|3)}\right),$$

$$\text{where: } \rho_{(1,2|3)} = \frac{\sigma_{\varepsilon_1\varepsilon_2} - \frac{\sigma_{\varepsilon_1\varepsilon_3}\cdot\sigma_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}{\sqrt{(\sigma^2_{\varepsilon_1} - \frac{\sigma^2_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}})(\sigma^2_{\varepsilon_2} - \frac{\sigma^2_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}})}}$$

For Employed People with Higher Education Degree:

$$l_{Ed_i=1,Empl_i=1}(\theta, X_i) = P(-\varepsilon_1 \leqslant X_{ed,i}\beta_{ed};\ -\varepsilon_2 < X_{em,i}\beta_{em};\ W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\ = f(W_i) \cdot P(-\varepsilon_1 \leqslant X_{ed,i}\beta_{ed};\ -\varepsilon_2 < X_{em,i}\beta_{em} |\ W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\ = \frac{1}{\sigma_{\varepsilon_3}}\phi\left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_3}}\right) \cdot \\ \cdot\Phi_2\left(\frac{X_{ed,i}\beta_{ed} + \frac{\sigma_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma^2_{\varepsilon_1} - \frac{\sigma^2_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}};\ \frac{X_{em,i}\beta_{em} + \frac{\sigma_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma^2_{\varepsilon_2} - \frac{\sigma^2_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}};\ \rho_{(1,2|3)}\right),$$

$$\text{where: } \rho_{(1,2|3)} = \frac{\sigma_{\varepsilon_1\varepsilon_2} - \frac{\sigma_{\varepsilon_1\varepsilon_3}\cdot\sigma_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}}}{\sqrt{(\sigma^2_{\varepsilon_1} - \frac{\sigma^2_{\varepsilon_1\varepsilon_3}}{\sigma^2_{\varepsilon_3}})(\sigma^2_{\varepsilon_2} - \frac{\sigma^2_{\varepsilon_2\varepsilon_3}}{\sigma^2_{\varepsilon_3}})}}$$

Where $\Phi_2(a_1; a_2; \rho)$ is a joint cumulative distribution function for the two bivariate normally distributed variables with means = 0, variations = 1, and correlation ρ.

4. Marginal Treatment Effects: Estimation Method

We follow the estimation strategy proposed in Carneiro et al. (2001) and Heckman et al. (2006): semi-parametric estimation using local instrumental variables technique.

1. We estimate the propensity scores (predicted probabilities for the higher education degree attainment) for the educational choice equation (3) by Probit: p. After that, we construct the common support for the propensity scores for workers with and without higher education degree: $p_{sup} = \{max(p_{0,min}, p_{1,min}); min(p_{0,max}, p_{1,max})\}$. Where $p_{1,min}$ and $p_{0,min}$ are the minimal values of propensity scores for the population with $(S = 1)$ and without $(S = 0)$ higher education degree , $p_{1,max}$ and $p_{0,max}$ are the maximal values of propensity scores for the same groups of workers.

2. Then we use a semi-parametric procedure with local polynomial regressions in order to estimate the marginal treatment effect (Carneiro et al. (2001)). This procedure consists of two steps.

   (a) β_0 and $(\beta_1 - \beta_0)$ are estimated by the double residual semi-parametric regression of the following equation, derived from the wage equation (4):

   $$lnW = \alpha_0 + \beta_0 \cdot X + [(\beta_1 - \beta_0) \cdot X] \cdot p + H(p) + \varepsilon_w \tag{38}$$

   Where $H(p)$ is a nonparametric function of p, and $E(\varepsilon_w|X, p) = 0$.

   In order to simplify the nonlinear component, we can rewrite this equation in the following way:

   $$E(lnW|p) = \beta_0 \cdot E(X|p) + (\beta_1 - \beta_0) \cdot E(X \cdot p|p) + H(p) \tag{39}$$

   $$lnW - E(lnW|p) = \beta_0 \cdot (X - E(X|p)) + (\beta_1 - \beta_0) \cdot (X \cdot p - E(X \cdot p|p)) + \varepsilon_w$$

   Therefore, we start by estimating $E(lnW|p)$, $E(X|p)$, and $E(X \cdot p|p)$ by the local linear regressions (package *locreg* in STATA: Froelich and Melly (2008), Froelich and Melly (2010)) of the lnW, X, and $X \cdot p$ on p. We use the cross-validation method to choose the smoothing parameter among the values 0.1÷0.9. Then, we calculate the residuals of these regressions: $lnW - (\widehat{lnW|p})$, $X - (\widehat{X|p})$, $X \cdot p - (\widehat{X \cdot p|p})$. Finally, by regressing (OLS) the first residuals on the second and third ones we get the estimated coefficients for β_0 and $(\beta_1 - \beta_0)$.

   (b) The nonparametric term $H(p)$ is determined from the residual $\tilde{W}$ of the equation (38):

   $$\tilde{W} = lnW - \hat{\beta}_0 \cdot X + [(\widehat{\beta_1 - \beta_0}) \cdot X] \cdot p = H(p) + \alpha_0 + \varepsilon_w \tag{40}$$

   $$H(p) + \alpha_0 = E(\tilde{W}|p) \tag{41}$$

   Therefore, we estimate $\frac{\partial H(p)}{\partial p}$ by local linear regression of $\tilde{W}$ on p.

Marginal Returns to education are, thus, determined as:

$$MTE(X = x, u_s) = \frac{\partial E(lnW|X, P(Z) = p)}{\partial p} = (\widehat{\beta_1 - \beta_0}) \cdot x + \frac{\partial H(p)}{\partial p}$$
$$= (\widehat{\beta_1 - \beta_0}) \cdot x + MTE^U(X = x, u_s) \quad (42)$$

To evaluate the Marginal Returns to higher education, we use the evenly spaced points of the set of values p_{sup}: 1÷99 centiles of p_{sup}.

3. The ATE, TT, TUT effects are calculated using formulas 13, 14, 15, 23, 24, 25, relying on the estimated values of MTE and MTE^U to calculate the differences between these effects based on observable and unobservable characteristics. Including Sorting on the Gains based on observable characteristics (for example, sex), therefore accounting for the fact that $E(X|S = 1) \neq E(X|S = 0) \neq E(X)$, the ATE, TT and TUT effects can be calculated according to the following formulas:

$$ATE(x_i) = \int_0^1 MTE(x_i, u_s) \cdot h_{ATE}(x_i, u_s)\, du_s =$$
$$= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{ATE}(x_i, u_s)\, du_s =$$
$$= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{ATE}(x_i, u_s)\, du_s \quad (43)$$
$$TT(x_i) = \int_0^1 MTE(x_i, u_s) \cdot h_{TT}(x_i, u_s)\, du_s =$$
$$= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{TT}(x_i, u_s)\, du_s =$$
$$= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{TT}(x_i, u_s)\, du_s \quad (44)$$
$$TUT(x_i) = \int_0^1 MTE(u_s) \cdot h_{TUT}(x_i, u_s)\, du_s =$$
$$= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{TUT}(x_i, u_s)\, du_s =$$
$$= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{TUT}(x_i, u_s)\, du_s \quad (45)$$

Where the weights are:

$$h_{ATE}(x_i, u_s) = 1$$
$$h_{TT}(x_i, u_s) = \frac{Pr(P(Z) > u_s|X = x_i)}{\int_0^1 Pr(P(Z) > u_s|X = x_i)\, du_s}$$
$$h_{TUT}(x_i, u_s) = \frac{Pr(P(Z) < u_s|X = x_i)}{\int_0^1 Pr(P(Z) < u_s|X = x_i)\, du_s}$$

$$\begin{aligned}
ATE &= \int_X ATE(x_i)\ dF_X(x) = \\
&= \int_X (\widehat{\beta_1 - \beta_0}) \cdot x_i\ dF_X(x) + \int_X \int_0^1 MTE^U(x_i, u_s) \cdot h_{ATE}(x_i, u_s)\ du_s\ dF_X(x) = \\
&= (\widehat{\beta_1 - \beta_0}) \cdot E(X) + E(MTE^U) \qquad (46) \\
TT &= \int_{X|S=1} TT(x_i)\ dF_{X|S=1}(x) = \\
&= \int_{X|S=1} (\widehat{\beta_1 - \beta_0}) \cdot x_i\ dF_{X|S=1}(x) + \\
&\quad + \int_{X|S=1} \int_0^1 MTE^U(x_i, u_s) \cdot h_{TT}(x_i, u_s)\ du_s\ dF_{X|S=1}(x) = \\
&= (\widehat{\beta_1 - \beta_0}) \cdot E(X|S=1) + E(MTE^U \cdot h_{TT}|S=1) \qquad (47) \\
TUT &= \int_{X|S=0} TUT(x_i)\ dF_{X|S=0}(x) = \\
&= \int_{X|S=0} (\widehat{\beta_1 - \beta_0}) \cdot x_i\ dF_{X|S=0}(x) + \\
&\quad + \int_{X|S=0} \int_0^1 MTE^U(x_i, u_s) \cdot h_{TUT}(x_i, u_s)\ du_s\ dF_{X|S=0}(x) = \\
&= (\widehat{\beta_1 - \beta_0}) \cdot E(X|S=0) + E(MTE^U \cdot h_{TUT}|S=0) \qquad (48)
\end{aligned}$$

4. Finally, bootstrap is used to calculate the variance and t-statistics for the estimated effects of ATE, TT, TUT, and 95% confidence interval for the MTE estimations.