



BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2013-45

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

by Natalia Kyui

Bank of Canada Working Paper 2013-45

December 2013

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

by

Natalia Kyui

Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
nkyui@bankofcanada.ca

Bank of Canada working papers are theoretical or empirical works-in-progress on subjects in economics and finance. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

Acknowledgements

I gratefully acknowledge the helpful discussions and kind support of Peter Arcidiacono, Paul Beaudry, Katya Kartashova, Guy Lacroix, Thomas Lemieux, Leigh L. Linden, Teodora Paligorova, Natalia Radtchenko, and Michel Sollogoub. I also thank audience participants at several conferences and seminars, where the earlier version of this paper was presented, such as EEA-ESEM'2011, JMA'2012, AEFP'2013, SCSE'2013, CLSRN-CEA'2013, ESPE'2013, EALE'2013, RECODE'2013, seminars at Laval University, Carleton University, Bank of Canada, and Centre d'Etudes de l'Emploi.

For making these data available, 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.

Abstract

This paper analyzes the effects of an educational system expansion on labour market outcomes, drawing upon a 15-year natural experiment in the Russian Federation. Regional increases in student intake capacities in Russian universities, a result of educational reforms, provide a plausibly exogenous variation in access to higher education. Additionally, the gradual nature of this expansion allows for estimation of heterogeneous returns to education for individuals who successfully took advantage of increasing educational opportunities. Using simultaneous equations models and a non-parametric model with essential heterogeneity, the paper identifies strong positive returns to education in terms of employment and wages. Marginal returns to higher education are estimated to decline for lower levels of individual unobserved characteristics that positively influence higher education attainment. Finally, the returns to higher education are found to decrease for those who, as a result of the reforms, increasingly pursued higher education.

JEL classification: J24, I20

Bank classification: Labour markets; Development economics

Résumé

Cette étude analyse les effets d'un élargissement du système éducatif sur les salaires et les perspectives d'emploi des diplômés. Elle se fonde sur l'expérience naturelle de l'expansion de l'accès à l'enseignement supérieur dans la Fédération de Russie au cours des 15 dernières années. L'accroissement du nombre de places disponibles dans les universités au niveau régional, du fait des réformes de l'éducation, fournit une variation vraisemblablement exogène de l'accès à l'enseignement supérieur. Le caractère graduel de cet élargissement permet en outre d'estimer les rendements hétérogènes de l'éducation pour les individus qui ont bénéficié de l'amélioration de l'accès à l'enseignement supérieur. S'appuyant sur des résultats obtenus à l'aide de modèles à équations simultanées et d'un modèle non paramétrique à hétérogénéité essentielle, l'étude relève que les rendements de l'enseignement supérieur sont élevés sur le plan des salaires et de l'emploi. D'après les résultats de l'estimation, les rendements marginaux de l'enseignement supérieur diminuent lorsque les caractéristiques individuelles non observables qui influencent positivement la probabilité d'accéder à l'enseignement supérieur sont plus faibles. Enfin, il apparaît que les rendements de l'enseignement supérieur sont plus faibles pour ceux qui ont pu entreprendre des études supérieures à la suite des réformes que pour ceux qui poursuivaient des études supérieures indépendamment de l'élargissement du système éducatif.

Classification JEL : J24, I20

Classification de la Banque : Marchés du travail; Économie du développement

1 Introduction

This study quantifies the effects of an educational system expansion on labour market outcomes and identifies the distribution of the returns to education, by exploring a natural experiment: a fifteen-year expansion of the higher education system in the Russian Federation. During 1990–2005, the country, referred to here as Russia, saw a doubling in student intake capacities in universities and, associated with this, a doubling in the number of university graduates. This expansion, government-regulated to a large extent, provides a plausibly exogenous variation in access to higher education at the regional level, and thus allows identification of the returns to higher education in Russia. The gradual nature of this increase in educational opportunities makes it possible to estimate returns to education for individuals who were exposed to different degrees of higher education expansion and who successfully took advantage of that. Finally, the paper evaluates the influence of this expansion on labour market outcomes, namely on employment and wages.

Whether and to what extent additional education does increase an individual's wages and employment prospects, as well as generate a positive economic return for a country – these are important questions both in the literature and for policy-makers. Such an inquiry is especially relevant for those economies aiming to increase the educational level of their population. Numerous studies show that workers with higher educational levels have correspondingly higher wages. A current leading focus of the literature, however, is identification of a causal effect of education on labour market outcomes, such as employment and wages, and thus its separation from potential selection biases.

This paper contributes primarily to the literature exploring changes in an educational system for the identification of the causal effects of education on labour market outcomes, in particular, on earnings.¹ Within this literature are two main directions of research: analysis of changes in compulsory schooling laws and of changes in access to education.

Using the first approach, researchers find a positive effect of compulsory schooling legislation on educational attainment and future wages. Yet the magnitude of this effect significantly varies among countries (see Angrist and Krueger (1991) for the United States, Oreopoulos (2006b) for the United Kingdom, Pischke and von Wachter (2008) for Germany, and Oreopoulos (2006a) for Canada). Because of the nature of compulsory schooling legislation, this approach allows identification of the effects of additional years of secondary schooling; these studies often use the total length of schooling as an educational variable, and thus assume a constant return to a year of schooling for different educational levels.

The second approach, analysis of the changes in access to education on educational attainment and labour market outcomes, encompasses studies focusing on the financial side of the educational system (tuition fees and financial-aid policies: Kane and Rouse (1993), Card and Lemieux (2001b), Arcidiacono (2005)), on the physical access to educational institutions (distance to high schools and colleges, or their presence in the district: Card

¹See Card (1999) and Belzil (2006) for an extensive analysis of other approaches to identifying the causal effect of education on earnings.

(1995), Conneely and Uusitalo (1997)), and on the capacity of an educational system. This last group of studies focuses on educational infrastructure and capacities. Duflo (2001) uses the exposure to school construction in Indonesia, a major governmental project, to identify its effects on educational attainment and future wages. Walker and Zhu (2008) analyze the expansion of higher education in the United Kingdom during the period 1994–2006, comparing the returns to education among different cohorts. They find no changes in the returns to higher education for men, with one exception for the bottom quartile of the distribution, and a significant rise in returns for women.

This is the first paper to our knowledge to identify the returns to higher education by exploring the educational reforms in Russia, and the first to quantify the influence of the expansion of the higher education system, caused by these reforms, on labour market outcomes. Though several studies have analyzed the returns to education in Russia, they treat educational attainment as exogenous, thus not correcting for the associated selection bias (the exception being Cheidvasser and Benitez-Silva (2007), discussed below). Studies conducted for the Soviet period² show a small effect of education on earnings (however, non-pecuniary returns might also have been important during this period).³ Gregory and Kohlhase (1988) estimate the returns to education of a sample of Russian migrants to Israel; they find positive and significant returns to completion of higher education for white-collar workers (12.79%–22.38%) and insignificant returns for blue-collar workers. The work of Cheidvasser and Benitez-Silva (2007) provides an analysis of the returns to education in the Russian Federation for the years 1992–1999. Using a linear regression estimation of Mincer’s equation, they show that the returns to an additional year of schooling are not higher than 5% during the analyzed period. 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, though larger for female employees (27% and 40%, respectively, in comparison to secondary education). Regarding the correction for the endogeneity of the educational variable, Cheidvasser and Benitez-Silva (2007) use the instrumental variable methodology, similarly to studies for other countries, by exploring the changes in compulsory schooling laws. Because the majority of the Russian population complete secondary education, the current paper instead focuses on higher education attainment and the returns to a higher education degree.

The major expansion of the Russian higher education system from 1990 to 2005, a result of educational reforms, led Russia to become a leader, among OECD and G-20 countries,

²The Union of Soviet Socialist Republics (USSR), or Soviet Union, was a constitutionally socialist state that existed in Eurasia from 1922 to 1991. The current study refers to this time as the Soviet period.

³Studies conducted with the data for the Soviet period (Gregory and Kohlhase (1988), Katz (1999), Ofer and Vinokur (1992)) report different results on the returns to education, mainly due to the significantly different population samples. The selectivity problem is thus important in these studies, but the limited information on wages for this period does not provide more reliable and representative results. Nevertheless, it is not the purpose of the current study to comment on the evaluation of the returns to education in the Soviet period and after the collapse of the Soviet Union. These studies’ results are mentioned simply to provide the reader with some background information about wages before the transition period.

in tertiary education attainment of the population.⁴ Figure 1 illustrates the expansion of the tertiary education system in Russia, in comparison to other countries. The lines show the proportion of students in tertiary education institutions relative to the size of the population (the number of tertiary students per 1,000 people).⁵

This expansion is characterized by three main features that allow reliance on it as an exogenous and non-anticipated change in educational opportunities for the population. First, near-universal secondary education in Russia by the beginning of these reforms led to a significantly larger demand for higher education than the existing capacities. Second, the reforms allowed for increases in student intake capacities in the higher education system, and thus affected only the supply side of the educational market. Therefore, student enrolment over this period was increased because of the weakening of the system's supply constraints. Finally, the yearly incremental increase in the number of university slots, and thus student intake, was regulated by the government through the licensing system. More details on these features are provided in the next section.

This paper uses two approaches to identify the returns to education and the effects of the expansion of the higher education system on the labour market. First, using the instrumental variable (IV) technique and simultaneous equations models, the paper identifies the returns to education for youths who took advantage of this expansion. Thus, the gradual nature of the expansion allows for estimation of the returns to education for youths who benefited from these reforms at different stages of the expansion. The estimation results by this method are shown to be robust to various changes in the model specification. Second, the paper uses the recently developed non-parametric model with essential heterogeneity to identify marginal returns to education – the heterogeneous returns that vary with the level of unobserved individual characteristics. This estimation is based on the research of Carneiro et al. (2001), Heckman et al. (2006), and Heckman and Vytlacil (2007), who develop the theoretical and empirical framework for the marginal treatment effects (MTE) estimations. The main advantage of using this model is that both the heterogeneity of the returns to education and the selection of individuals based on these heterogeneous returns can be accounted for.

These two approaches yield similar, though complementary, results. The study provides evidence of decreasing returns to education for youths who obtain higher education due to increasingly easy access to universities. However, these returns remain positive and large, associated with 30%–90% wage increases for a higher education degree. The marginal returns to higher education are found to decrease for lower levels of individual unobserved

⁴See OECD (2013) for recent information on tertiary education attainment.

⁵Tertiary education corresponds to categories 5A, 5B and 6 of the International Standard Classification of Education (ISCED) before its revisions in 2011 (ISCED (2006)). Level 5A is defined as “tertiary programmes that are largely theoretically based and are intended to provide sufficient qualifications for gaining entry into advanced research programmes and professions with high skills requirements.” Level 5B programs focus on “occupationally specific skills geared for entry into the labour market.” Level 6 corresponds to “programmes which lead to the award of an advanced research qualification.”

characteristics, which positively influence higher education attainment. Overall, the results suggest that this expansion of the higher education system has significantly increased the wages of those who took advantage of its growing capacity. However, this increase is smaller than the returns to education for those who would have pursued higher education regardless.

The next two sections describe the institutional context of the Russian educational reforms and the data. Section 4 discusses the estimation results using the IV approach of the returns to education and provides several robustness checks. Section 5 analyzes the heterogeneity in the returns to education, using a non-parametric approach. Section 6 concludes. The Technical Appendix provides details on the estimation of the simultaneous equations models and the model with essential heterogeneity.

2 Institutional Context

This section describes in detail the educational system in Russia and the recent reforms on which this paper focuses. The Russian system, inherited from the Soviet period, consists of four levels: primary and general education (8 years at schools); secondary education (an additional 2 years at general or specialized schools); tertiary education (2-6 years at a post-secondary institution); and post-graduate education (3-6 years at universities with PhD programs). Tertiary education is in turn divided into two streams that are not consecutive levels. The first, corresponding to ISCED level 5B, is post-secondary technical or vocational education; it consists of 2-3 years of study at technical or specialized schools, such as military, medical or musical. The second type, corresponding to ISCED level 5A, is higher education; it requires 4-6 years of studies at universities after secondary education. This last educational level is the main focus of the current study.

During the Soviet period, the government completely regulated the educational sector. Education, both secondary and tertiary, was financed from the country's budget and was free for students (thus referred to here as state-subsidized education).⁶ One of the main achievements of the educational policy during the Soviet period was a large expansion of secondary education: the share of the population 15 years and older without a secondary education degree decreased from almost 61% in 1959 to less than 20% in 1989.⁷ As a result, the share of the Russian working-age population with at least a secondary level of

⁶The development of education became a priority for the Russian government at the beginning of the twentieth century, and even now education is a high priority for the country. After the Revolution of 1917, the Russian government started major reforms in the educational sector in order to provide universal public education (Lukov (2005)). The tuition fees in secondary and especially tertiary education were completely eliminated by 1954. The 1958 law introducing compulsory 8-year schooling determined some general principles of the connections between work and education, framing vocational schooling and prioritizing not only education for youths, but also for experienced workers going back to school.

⁷Cited numbers are from the Russian census of 1959 and 1989. Universal secondary education was the main educational priority of the Soviet government from 1965 on.

education increased to more than 90% and placed Russia among the countries with the highest level of secondary education attainment of population in the world.⁸ Regarding tertiary education, the Soviet government determined the number of slots according to the economy's predicted needs for the following years. The number of slots in tertiary education was limited and fully state-subsidized. Aspiring students were admitted on a competitive basis and admission tests selected high-ability candidates. The growing access to secondary education increased the number of applicants for tertiary education programs, especially to universities.

With the beginning of the transition, in 1992, the Russian government passed the law "On Education," marking the start of major changes in the tertiary education system. The most important amendment introduced tuition-based, in addition to state-subsidized, tertiary education. First, the government authorized the creation and operation of non-public (private) tertiary education institutions with the proviso that they be run as not-for-profit organizations, albeit providing educational services to the population on a full-tuition basis. Second, the government authorized public tertiary education institutions to admit students on a full-tuition basis in addition to the state-subsidized slots, which continue to be financed by the government. These changes on the supply side, along with a large and persistent demand for tertiary education (primarily due to universal secondary education and perceived positive returns to tertiary education), led to a major expansion of the tertiary education system in Russia. During the following decade, the number of slots in the tertiary education system more than doubled. Figures 2 and 3 show these changes.

The major increase happened in higher education, while vocational tertiary education kept the same levels. Figure 2 depicts changes in higher education, and Figure 3 those in vocational tertiary education. The number of public universities grew during the years 1990–2005 (up 27%), and private universities increased their share in the total number of higher education institutions from zero to 39%. However, in terms of the number of admitted students, the main increase happened in the full-tuition slots in public universities. Overall, admission to all universities increased by 181% during 1990–2005, among which 130% was because of the full-tuition slots in public universities. The increase in the number of admitted students led to a corresponding increase in the number of university graduates. The size of the youth cohorts grew slightly over the same period; however, the expansion of the higher education system was significantly greater (further details on this point are provided below).

Even though the expansion of the higher education system was driven by full-tuition education, Kyui (2011) shows that youths with all family backgrounds (with better- or lower-educated parents, as well as from high- or low-income families) benefited from these reforms in terms of educational attainment. Certainly, the expansion of the higher educa-

⁸Secondary or higher levels of education in Russia correspond to the third and higher educational categories in the ISCED classification. The share of people among 25- to 34-year-olds with at least a secondary education level in 2004 was 92.3%, while among 25- to 64-year-olds it was 90.10% (sources: Gohberg et al. (2007), Russian Federal Committee for Statistics - www.gks.ru).

tion system is limited by the demand for higher education; in other words, by the number of youths willing to enter the higher education institutions. Until recent years, admission selection processes were locally organized by educational institutions, which is why the information is available on the ratios of students applying to those admitted.⁹ As shown in Figure 4, there are more applicants than admitted students, especially for prestigious majors such as economics and law. These summary statistics provide evidence that the demand during this expansion still exceeded the student intake capacities, and, therefore, student enrolment over this period increased because of and in keeping with the weakening of the system's capacity constraints.

This increase in student intake capacities of the higher education system, as well as its gradual nature and regional variation, was to a large extent regulated by the government. This is the main reason why this expansion can be considered as an exogenous and non-anticipated change in educational opportunities for the population. In particular, the government regulates the size and quality of higher education through licensing and accreditation of educational establishments. In order to perform an educational activity, all educational institutions, both public and private, must obtain a licence. Licensing decisions are based on full information about an establishment's educational programs and available resources: buildings, faculty, and other educational and non-educational facilities. The licences strictly determine the maximum number of students allowed to be enrolled. It is possible for educational establishments to change the parameters of their licences and thus to increase the number of students allowed for admission, yet they must obtain authorization in advance, and thus prove their financial, physical and human resources. Such implementing of the reforms suggests that the government to a large extent regulates the yearly incremental increase in the student intake capacities, and thus the gradual expansion of the higher education system at the regional level.

Figure 5 illustrates the expansion of the higher education system within seven Russian federal districts. The first graph shows the number of admitted students for the first year of studies. The second illustrates the size of the cohort of 17-year-old youths in the corresponding years. The third graph depicts the proportions of the admitted students to the cohorts.¹⁰ These data suggest that, historically, there are regions with greater and lesser access to higher education. Particularly, the Central federal district (which includes Moscow) is historically the region with the highest access. Even adjusting for the size of

⁹Admission tests aimed at selecting the best students among applicants. The laws "About Education" (1992) and "About Higher and Post-Graduate Education" (1996) state that admission must be guaranteed for better-performing students. Students can apply to several institutions, but since they must pass admission exams in each institution separately, their choices are limited by their time, health and skills constraints. Students can also apply to both state-subsidized and full-tuition slots, though some institutions may restrict the choice in terms of majors and types of programs. There are no unified rules for this matter among educational institutions. Recently, the unified secondary education graduation exam was introduced, which replaces the admission exams in the selection processes of tertiary education. However, these last years are not included in the current analysis.

¹⁰These shares are adjusted by the proportion of the current-year secondary school graduates among the first-year admitted students.

the population, these differences in the capacity of the higher education system are still important among regions, and thus the expansion of higher education varied among regions. The regions with the highest number of slots in the higher education system have kept their leadership positions. Therefore, the varying educational resources in the regions at the beginning of transition, along with governmental control over licensing, are a main explanation of these regional differences in the trajectories of expansion. Overall, these factors strongly suggest that the yearly changes in access to higher education within regions were plausibly exogenous and non-anticipated for youths and their families.

3 Data Description

Data are taken from the Russian Longitudinal Monitoring Survey (RLMS), 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,b).

The samples of 24- to 47-year-olds, interviewed in the years 2000–2008, are used as the repeated cross-sections for this analysis. Those observed were making 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.¹¹ Overall, there are 41,585 observations; among them, 10,288 are unemployed and 31,297 are employed. However, among the employed population the wages are observed only in 91.5% of cases. Thus, the final sample used for the analysis consists of 28,622 employed and 10,288 unemployed people.

To describe individuals' educational attainment, the dummy variable *Higher Education Degree Attainment* is used, which takes a value of 1 if a person has a higher education degree, and 0 otherwise. In the sample, 23.26% of the whole population and 25.87% of the working population with observed wages have higher education degrees (Table 1). Moreover, the continuous variable representing the total number of years of education (*Years of Education*) is used to test the robustness of the results to the model's specification

¹¹The majority of graduates from secondary school are 17 years old. Thus, in 2007, 92.03% of 15-year-olds and 84.02% of 16-year-olds were in the secondary education system; these numbers decline to 52.88% and 31.73% for 17- and 18-year-olds, respectively (source: <http://stat.edu.ru>). However, some students finish later and some do not go directly to higher education (for example, because of a failure during admission exams) but attend later. Thus, among newcomers to higher education in 1995, 63% were the corresponding year's secondary school graduates and 8% were the corresponding year's technical school graduates, while 27% graduated from secondary education in earlier years, and 2% had already obtained a tertiary education degree. The share of the same year's general or technical secondary education graduates among the newcomers to higher education decreased during the analyzed period, from 71% in 1995 to 70% in 2000, and to 65% in 2008 (source: www.gks.ru). However, there is value in the consistency of the proposed approach to the instrumental modelling of educational choices, mainly because it still allows the capture of the variation in access to tertiary education. Indeed, the factors resulting in the delay of one's tertiary education attendance (especially failure during admission exams) could also reflect the rank of such an individual in a population's distribution of abilities, and thus this instrument also captures the changes in educational attainment for such individuals.

(see section 4.4 for details).

Table 1: Higher Education Degree Attainment, 24- to 47-year-olds, 2000–2008

Higher Education Attainment	Employed, 24–47-y.o. with Observed Wages	Employed 24–47-y.o.	Unemployed 24–47-y.o.	ALL 24–47-y.o.
0	74.13%	74.23%	84.37%	76.74%
1	25.87%	25.77%	15.63%	23.26%
Total	28,622	31,297	10,288	41,585

Sources: RLMS, 2000–2008, author’s calculations

An individual who self-identifies as having a job is considered to be employed. The individual’s average monthly income from working activity is used as a measure of wages.¹² Wages are adjusted for inflation by the consumer price index. In order to account for the regional differences in prices, and to make regional wages comparable, wages are additionally corrected by the regional price of a standard product set. The exposure of an individual to the expansion of the higher education system is identified by the year when the individual turned 17 years old (thus graduating from secondary school and deciding on further tertiary education) and by the federal district of residency. Along with the control for year and federal district fixed effects, the changes in the capacity of the higher education system provide a plausibly exogenous variation in access to higher education.

4 Returns to Education: Local Average Treatment Effects

This section discusses the identification and estimation of the returns to higher education and the effects of the educational system expansion on employment and wages. Variation in the opportunities of obtaining higher education (i.e., variation in the number of available slots in universities) is used as the instrument for higher education attainment. First, the section describes the explanatory power of this instrument. Second, it shows the estimation results for the returns to education in terms of employment and wages. Finally, several robustness checks are described. Section 5 discusses the estimation of heterogeneity in the returns to education.

4.1 Instrument choice and characteristics

Table 2 reports the results for the “first-stage” equation: the influence of the changes in the educational system (increasing number of slots) on higher education attainment.

¹²Information on the number of hours worked contains many missing and out-of-normal-range values; for this reason, the hourly wage is not used for the estimations.

Four specifications of the first stage are tested (Table 2), using the following variables to determine the expansion of the higher education system:

- (1) The number of slots in the higher education system at the federal district levels (“# of HE slots, by districts”).
- (2) The number of slots in the higher education system at the national level (“# of HE slots, in RF”).
- (3) The proportion of slots in the higher education system relative to the corresponding cohort of youths at the federal district levels (“Proportion of HE slots, by districts”).
- (4) The proportion of slots in the higher education system relative to the corresponding cohort of youths at the national level (“Proportion of HE slots, in RF”).

All the specifications control for the corresponding size of the cohort of 17-year-old youths (at the federal district level and at the national level, respectively). The estimation results based both on the whole population and on only the employed population are listed.

Table 2: Influence of the Higher Education System Expansion on Higher Education Degree Attainment, 24- to 47-year-olds, 2000–2008

Variables	Higher Education Degree Attainment			
	1	2	3	4
Whole Population				
# of HE slots, by districts	0.175*** (0.031)			
# of HE slots, in RF		0.032*** (0.007)		
Proportion of HE slots, by districts			0.639*** (0.145)	
Proportion of HE slots, in RF				0.749*** (0.169)
LR test, χ^2	32.45	24.35	19.49	24.02
LR test, p-value	[0.000]	[0.000]	[0.000]	[0.000]
Employed Population				
# of HE slots, by districts	0.154*** (0.030)			
# of HE slots, in RF		0.029*** (0.007)		
Proportion of HE slots, by districts			0.554*** (0.149)	
Proportion of HE slots, in RF				0.685*** (0.173)
LR test, χ^2	18.81	14.9	10.91	14.57
LR test, p-value	[0.000]	[0.000]	[0.001]	[0.000]

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels).

Controls: size of 17-year-old cohorts by districts / RF, respectively, district fixed effects, gender (male, dummy variable for males), age, age², age·male, age²·male. Sources: RLMS, 2000–2008, author’s calculations.

Observations: 38,910 (whole population) and 28,622, respectively.

Statistics for the test on exclusion of the instruments from the first-stage equation suggest that the variables describing the number of slots have a higher explanatory power than those describing the ratio of the number of slots to the size of the corresponding youth cohorts. Moreover, the variables at federal district levels have a higher explanatory power than those at the Russian level. This result was expected, and it is due both to the additional variation in the size of the higher education system and to the different patterns of its expansion among regions. Therefore, for further estimations, the number of slots in the higher education at federal district levels is used as the instrument for the higher education degree attainment (with a control for the size of 17-year-old cohorts at the federal district level). This variable passes the instrument-weakness test (see also section 4.4 for details using the continuous variable to describe educational attainment).¹³

The main potential problem with this instrument is the fact that it is perfectly correlated with region-cohort effects, which means that there are no variations of the instrument within a 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 not correlated with the tertiary education system expansion within the regions over time. In order to account for that, several steps are performed. First, all the specifications control for the year and region fixed effects, thus capturing the unobserved variations of other characteristics by regions and over the analyzed period. However, it is still necessary to assume that there were no changes, affecting wages and correlating with the educational system expansion within regions, other than those captured by the fixed effects for years and federal districts. Therefore, additional controls for the cohort year-of-birth fixed effects and region-specific linear cohort trends are included when performing robustness checks.

Table 3 summarizes the reduced-form estimation results for the wage and employment equations (by ordinary least squares (OLS) and probit, respectively). Column 1 presents the regression of wages (in logarithms) on higher education attainment, column 2 shows the regression of wages on higher education attainment and the instrument (number of slots in higher education at the regional level), and column 3 lists the regression only on the instrument (a reduced-form estimation of the influence of higher education expansion on wages). Columns 4–6 show similar results for the employment equation. First, higher education attainment significantly affects employment and wages (columns 1 and 4); however, such a reduced-form estimation procedure does not account for the potential selection bias. Second, these estimation results suggest significant positive effects of the expansion of the higher education system on both employment and wages (columns 3 and 6). At the same time, its influence becomes insignificant and much lower in magnitude once the variable for higher education attainment is included (see columns 2 and 5). Therefore, the influence of the higher education expansion on employment and wages occurs mainly through the changes in educational attainment, and does not seem to have a direct influence on the future employment and wages of the corresponding cohorts. In the robustness check, the

¹³According to Angrist and Pischke (2009), the critical value for F-statistics is 10, below which one can argue that the instrument is weak.

potential equilibrium effects of the higher education expansion on wages are also tested and discussed (see section 4.4 for more details); here, the main focus is on the individual returns to education.

Table 3: Reduced-Form Estimations: Wages and Employment, 24- to 47-year-olds, Employed and Whole Population, 2000–2008

Variables	ln(Wage), by OLS			Employment, by probit		
	1	2	3	4	5	6
<i>HE degree</i> *	0.457*** (0.011)	0.457*** (0.011)		0.389*** (0.021)	0.388*** (0.021)	
<i># of HE slots, by districts</i>		0.021 (0.021)	0.052** (0.023)		0.038 (0.028)	0.065** (0.027)
<i>Observations</i>	28,622	28,622	28,622	38,910	38,910	38,910

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels).

Controls: year and federal district fixed effects, gender (male), age, age², age·male, age²·male, size of 17-year-old cohorts by districts, regional unemployment rates for the employment equation.

Sources: RLMS, 2000–2008, author’s calculations. The wage equation is estimated for the employed population using the OLS method, and the employment equation is estimated for the whole population using the probit model.

Observations: 28,622 (employed population) and 38,910 (whole population).

4.2 Returns to education: wages

The joint model of higher education degree attainment and wages is estimated using the maximum likelihood method; the estimation procedure is described in section 1 of the Technical Appendix. The above-mentioned expansion of the higher education system at Russian regional levels is used as the instrument for higher education degree attainment. This simultaneous equations model allows for control of the endogeneity of educational attainment, and thus of the self-selection of individuals into higher education. Table 4 lists the estimation results.

Estimated returns to education, using this joint model of higher education degree attainment and wages, are higher than those obtained by OLS (Table 3). Returns to a complete higher education degree are estimated at the level of 76%. The correlation between unobserved terms in educational attainment and the wage equation is estimated to be negative and significant. IV estimations of the returns to education report an increase in wages for those who obtained an education degree because of the instrument variation and who would not otherwise have received it (Angrist et al. (1996)). Therefore, the estimated returns to education represent the local average treatment effect: returns to education for those who have obtained a higher education degree because of the growing access to universities. The results suggest that individuals who benefited from the expansion of the higher education system in terms of educational attainment have also gained in terms of wages.

Table 4: Estimations of the Returns to Higher Education (by maximum likelihood), Employed Population, 24- to 47-year-olds, 2000–2008

Variables	Employed Population	
	Education	ln(Wage)
<i>HE degree*</i>		0.756*** (0.034)
<i># of HE slots, by districts</i>	0.158*** (0.030)	
$\rho(\varepsilon_1, \varepsilon_2)$		-0.221*** (0.020)
$\sigma^2(\varepsilon_2)$		0.642*** (0.024)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels).

Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender (male), age, age², age·male, age²·male.

Sources: RLMS, 2000–2008, author’s calculations. Observations: 28,622.

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, and selection into employment is not random. In this section, the empirical model accounts for both endogeneity of educational attainment and self-selection into employment. Therefore, the influence of education on employment and wages is estimated.

The joint three-equations model of educational choice, employment and wages is estimated. Estimations are conducted by the maximum likelihood method; the procedure is described in section 2 of the Technical Appendix. Regional unemployment rates are used as the exclusion restriction for the employment equation. Such a specification allows for correlations between unobservable components of these three equations.

Table 5 lists the estimation results for the returns to education, taking into account the selection into employment. These results suggest that higher education significantly increases the probability of being employed, as well as wages. Moreover, correction for the selection into employment does not significantly change the estimated wage returns to education. The correlations between unobservable terms in education, employment and wage equations are all negative and statistically significant. This fact suggests that those with an increased probability of higher education attainment would have lower wages and a lower employment probability if they were to obtain a lower educational degree, compared to the less-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 5: Influence of Higher Education on Employment and Wages: 3-Equations Model, 24- to 47-year-olds, Whole Population, 2000–2008

Variables	Whole Population		
	3 equations: MLE		
	Education	Work	Wages
<i>HE degree*</i>		0.955*** (0.031)	0.769*** (0.041)
<i># of HE slots, by districts</i>	0.181*** (0.032)		
<i>Covariance matrix:</i>	ε_1	ε_2	ε_3
ε_1	1		
ε_2	-0.377***	1	
ε_3	-0.247***	-0.646***	0.871***
$\rho(\varepsilon_1, \varepsilon_2)$	-0.377***	(0.019)	
$\rho(\varepsilon_1, \varepsilon_3)$	-0.264***	(0.023)	
$\rho(\varepsilon_2, \varepsilon_3)$	-0.693***	(0.020)	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels).

Controls: size of 17-year-old cohorts, year and federal district fixed effects, gender (male), age, age², age-male, age²-male; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2000–2008, author’s calculations. Observations: 38,910.

4.4 Robustness check

This section describes four robustness checks performed.

First, the estimation results are tested for sensitivity to the choice of variable characterizing educational attainment. The main estimations, described above, use a binary variable for a higher education degree. This choice is motivated by the fact that the educational reforms affected entry to higher education, and thus attainment of a higher education degree. In the first robustness check, the estimations are conducted using the continuous variable for the number of years of education. The results provide similar estimations to those obtained in the previous section.

Second, the following two robustness checks seek to verify whether the inseparability of the instrument variation (the number of slots in the higher education system at the regional level) from the within-region changes over time could potentially bias the results reported in the previous section. In order to account for this potential problem, additional dummy variables for the years of birth (i.e., cohort fixed effects) are included in all equations (educational choice, employment and wages). Moreover, district-specific linear cohort time trends are also included to control for unobserved changes within regions over time. The estimation results are robust to these tests.

Third, the interaction terms of the instrument with other individual characteristics are included in the first equation for educational attainment, particularly in interactions with gender and parental educational background (concurrently, the gender and parental education variables are also included in all other equations).

The fourth and final robustness test aims to evaluate the importance in the labour market of possible equilibrium effects of the expansion of the higher education system. For instance, an increasing supply of new university graduates could have an effect on all other university degree holders in the labour market, namely, a potential decrease in their wages. Certainly, the cohorts' effects on equilibrium of the labour market are accounted for by introducing the year-of-birth dummy variables in the second robustness test. Additionally, changes in the returns to education are evaluated during the analyzed period.

4.4.1 Continuous educational variable

This section reports the estimation results, using a continuous variable for the number of years of education instead of a binary variable for higher education degree attainment. In particular, the variable *Years of Education* describes the number of years of schooling and ranges from 9 to 15 years. Higher education degree attainment corresponds to 15 years of total schooling; 13 and 14 years of schooling correspond to complete/incomplete vocational education or to incomplete higher education; and 10 years of schooling correspond to secondary school completion.

Table 6 lists the estimation results for the educational attainment equation and thus the effect of the expansion of the higher education system on the number of years of education. The instrumental variable for educational attainment stays the same – the number of slots in the higher education system at the regional level; as previously, the variables for the size of the corresponding 17-year-old cohorts at the regional levels are included in all estimations. This instrumental variable shows a positive and significant effect on the total years of schooling. F-statistics for the exclusion of the instrument from the educational attainment equation suggest that the instrument is not weak. The reduced-form estimation results for the wage and employment equations are listed in Table 7, with conclusions similar to those using the binary variable for higher education degree attainment. The number of slots in higher education significantly affects both the wages and employment of the corresponding cohorts (columns 3 and 6). However, these effects disappear once the education variable is included, suggesting that the higher education expansion affects both wages and employment, not directly, but rather through educational attainment (columns 2 and 5).

Table 6: Influence of the Higher Education System Expansion on Years of Education, 24- to 47-year-olds, 2000–2008

Variables	Number of Years of Education	
	Whole Population	Employed Population
# of HE slots, by districts	0.284*** (0.062)	0.320*** (0.054)
F-statistics: instrument exclusion	20.66	35.69
F-statistics: p -value	[0.000]	[0.000]

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS estimations. Robust standard errors in parentheses (clustered at year-district levels). Controls: size of 17-year-old cohorts by districts, district fixed effects, gender, age, age², age·male, age²·male. Sources: RLMS, 2000–2008, author’s calculations. Observations: 38,910 (whole population) and 28,622 (employed population), respectively.

Table 7: Reduced-Form Estimations: Wages and Employment, 24- to 47-year-olds, Employed and Whole Population, 2000–2008

Variables	ln(Wage), by OLS			Employment, by probit		
	1	2	3	4	5	6
Years of education	0.098*** (0.003)	0.098*** (0.003)		0.098*** (0.004)	0.098*** (0.004)	
# of HE slots, by districts		0.006 (0.022)	0.052** (0.023)		0.024 (0.029)	0.065** (0.027)
Observations	28,622	28,622	28,622	38,910	38,910	38,910

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels).

Controls: year and federal district fixed effects, gender, age, age², age·male, age²·male, size of 17-year-old cohorts by districts, regional unemployment rates for the employment equation. Sources: RLMS, 2000–2008, author’s calculations. The wage equation is estimated for the employed population using the OLS method, and the employment equation is estimated for the whole population using the probit model. Observations: 28,622 (employed population) and 38,910 (whole population).

Table 8 lists the estimation results for the returns to education, using the number of higher education slots at the regional level as the instrument. The first section (IV) reports the estimation results using the instrumental variable two-stage least-squares (2SLS) approach. The second section (MLE) reports the estimation results using the maximum likelihood methods and thus assuming the joint normal distribution of random terms.¹⁴ The estimation results using 2SLS and MLE are similar, suggesting that the additional assumption about joint normal distribution of random terms does not significantly affect the results. As with the main estimation results, the returns to education are estimated to be higher than those without controlling for the selection into education. Table 9 lists the estimation results for the joint model of educational attainment (number of years of education), employment and wages.¹⁵ Overall, the estimated returns to education, in terms of

¹⁴The estimation procedure for the second approach is similar to that for the main model described in the Technical Appendix, with the exception that the equation for educational choice is now modelled as a linear equation. Detailed formulas can be provided upon request.

¹⁵The estimation procedure is similar to that for the main model, but using the linear equation instead of probit for the educational attainment. Detailed formulas are available upon request.

employment and wages, are found to be of the same magnitude as in the main estimations (note that higher education degree attainment corresponds to five years of studies). Therefore, the estimation results are robust to the form of the variable describing educational attainment.

Table 8: Estimations of the Returns to Higher Education (by IV and Maximum Likelihood), Employed Population, 24- to 47-year-olds, 2000–2008

Variables	Employed Population			
	IV		MLE	
	Education	ln(Wage)	Education	ln(Wage)
<i>Years of education</i>		0.116*** (0.070)		0.115*** (0.006)
<i># of HE slots, by districts</i>	0.320*** (0.054)		0.319*** (0.054)	
$\rho(\varepsilon_1, \varepsilon_2)$				-0.048*** (0.016)
$\sigma^2(\varepsilon_1), \sigma^2(\varepsilon_2)$			4.878*** (0.030)	0.619*** (0.023)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels). Estimations are conducted using IV and maximum likelihood methods, respectively.

Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender, age, age², age·male, age²·male. Sources: RLMS, 2000–2008, author’s calculations. Observations: 28,622.

Table 9: Influence of Higher Education on Employment and Wages: 3-Equations Model, 24- to 47-year-olds, Whole Population, 2000–2008

Variables	Whole Population		
	3 equations: MLE		
	Education	Work	Wages
<i>Years of education</i>		0.171*** (0.028)	0.165*** (0.028)
<i># of HE slots, by districts</i>	0.274*** (0.053)		
<i>Covariance matrix:</i>	ε_1	ε_2	ε_3
ε_1	4.878***		
ε_2	-0.415***	1	
ε_3	-0.463***	-0.587***	0.827***
$\rho(\varepsilon_1, \varepsilon_2)$	-0.188***	(0.063)	
$\rho(\varepsilon_1, \varepsilon_3)$	-0.231***	(0.063)	
$\rho(\varepsilon_2, \varepsilon_3)$	-0.646***	(0.037)	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels). Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender, age, age², age·male, age²·male; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2000–2008, author’s calculations. Observations: 38,910.

4.4.2 Cohort effects and regional time trends

As mentioned above, the instrument, the number of slots in the higher education system at the regional level, is inseparable from the region-cohort characteristics. Therefore, if any changes within regions were correlated with the expansion of the higher education system and were subsequently affecting labour market outcomes, the estimated returns to education could be biased. Even though the estimations described above suggest no direct effect of the system expansion on wages and employment beyond through educational attainment, this section tries to address the concern about other region-cohort changes that may affect labour market outcomes.

Different cohorts of youths come to the labour market at different times (even at regional levels), and thus may experience different conditions during their first period in the market, potentially significantly affecting wages and employment prospects later. It is difficult to account for all labour market conditions that could affect wages and that are correlated with the number of slots in the higher education system; similarly, there are no measures of individual ability available in the database. The variables describing unemployment, GDP growth and other macro characteristics cannot account completely for all labour market conditions at the moment of entry. If these effects of the entry time for individuals are correlated with the number of slots in the higher education system in the year the individuals turned 17, the returns to education will be over- or underestimated. Moreover, changes in the supply of university graduates in a particular year may affect the wages of other workers in the labour market (Katz and Murphy (1992), Card and Lemieux (2001a)). In order to account for these problems, additional dummy variables for the years of birth are included in all equations (educational choice, employment and wages) – 30 dummy variables for cohort birth years 1954–1984. Moreover, district-specific linear cohort time trends are also included to account for unobserved changes within regions over time.

The year-of-birth fixed effects (cohort fixed effects) and district-specific linear cohort trends (according to the seven Russian federal districts) are included in all equations for educational choice, employment and wages. The estimation results for the two-equations model (educational choice and wages) are reported in Table 10, and for the three-equations model (educational choice, employment and wages) in Table 11. The first specification in both tables includes only cohort fixed effects, while the second specification includes both cohort fixed effects and district-specific linear cohort trends. Therefore, such a specification accounts for the unobserved linear time changes within regions concurrent with the expansion of the higher education system (for example, changes in the labour market), as well as for the difference between different cohorts of youths. The estimation results of the returns to education with the control for the year-of-birth fixed effects and district-specific linear cohort trends are similar to the main estimation results, which suggests the robustness of the estimated results.

Table 10: Estimations of the Returns to Higher Education (by Maximum Likelihood),
Employed Population, 24- to 47-year-olds, 2000–2008

Variables	Employed Population			
	MLE (1)		MLE (2)	
	Education	ln(Wage)	Education	ln(Wage)
<i>HE degree</i> *		0.785*** (0.043)		0.783*** (0.039)
<i># of HE slots, by districts</i>	0.118** (0.055)		0.150* (0.079)	
$\rho(\varepsilon_1, \varepsilon_2)$		-0.243*** (0.026)		-0.241*** (0.023)
$\sigma^2(\varepsilon_2)$		0.644*** (0.026)		0.643*** (0.025)
<i>Cohort fixed effects</i>	X	X	X	X
<i>District-specific linear cohort trends</i>			X	X

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels). Estimations are conducted using IV and maximum likelihood methods, respectively.

Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender, age, age², age-male, age²-male. Sources: RLMS, 2000–2008, author’s calculations. Observations: 28,622.

Table 11: Estimations of the Returns to Higher Education (by Maximum Likelihood),
Whole Population, 24- to 47-year-olds, 2000–2008

Variables	Whole Population					
	3 equations: MLE (1)			3 equations: MLE (2)		
	Education	Work	Wages	Education	Work	Wages
<i>HE degree</i> *		0.946*** (0.037)	0.798*** (0.036)		0.942*** (0.061)	0.799*** (0.054)
<i># of HE slots, by districts</i>	0.141*** (0.050)			0.170** (0.067)		
<i>Covariance matrix:</i>	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
ε_1	1			1		
ε_2	-0.360***	1		-0.357***	1	
ε_3	-0.264***	-0.636***	0.863***	-0.265***	-0.639***	0.864***
$\rho(\varepsilon_1, \varepsilon_2)$	-0.360***	(0.019)		-0.357***	(0.035)	
$\rho(\varepsilon_1, \varepsilon_3)$	-0.284***	(0.021)		-0.285***	(0.029)	
$\rho(\varepsilon_2, \varepsilon_3)$	-0.685***	(0.023)		-0.687***	(0.022)	
<i>Cohort fixed effects</i>	X	X	X	X	X	X
<i>District-specific linear cohort trends</i>				X	X	X

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels). Estimations are conducted using IV and maximum likelihood methods, respectively.

Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender, age, age², age-male, age²-male; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2000–2008, author’s calculations. Observations: 38,910.

4.4.3 Introducing interactions with other characteristics

This section reports the estimation results, introducing additional interaction terms of the instrumental variable with other individual characteristics. In particular, the analysis is focused on the interactions with gender and parental educational background. First, such specifications differentiate the effects of the expansion of the higher education system on female and male students, as well as on students with different family backgrounds. Second, it allows for exploration of the within-region cohort variation of the instrument's effects. The increasing number of slots in the higher education system had a slightly different affect on the male and female populations, as well as on students with different parental educational background; evidence of these heterogeneous influences is analyzed in Kyui (2011). Finally, taking into account the heterogeneous influence of the expansion on educational attainment also increases the explanatory power of the first equation in the models (equation for educational choice).

Table 12 lists the estimation results using the interaction of the instrumental variable with gender; a dummy variable for males is also included in all the equations. Column 1 corresponds to the two-equations model (educational attainment and wages); column 2 lists the results for the three-equations model (educational attainment, employment and wages). The number of slots has a greater influence on female educational attainment. The estimated returns to education are of the same magnitude as in the baseline estimations.

Table 12: Estimations of the Returns to Higher Education (by Maximum Likelihood), Employed and Whole Population, 24- to 47-year-olds, 2000–2008

Variables	Employed Population (1) 2 equations: MLE		Whole Population (2) 3 equations: MLE		
	HE*	Wages	HE*	Work	Wages
<i>HE degree*</i>		0.747*** (0.060)		0.921*** (0.053)	0.791*** (0.055)
<i># of HE slots, by districts · · Male *</i>	0.120*** (0.035)		0.137*** (0.032)		
<i># of HE slots, by districts · · Female *</i>	0.201*** (0.033)		0.216*** (0.035)		
<i>Covariance matrix:</i>	ε_1	ε_3	ε_1	ε_2	ε_3
ε_1	1		1		
ε_2			-0.355***	1	
ε_3	-0.172***	0.641***	-0.260***	-0.652***	0.875***
$\rho(\varepsilon_1, \varepsilon_2)$			-0.355***	(0.031)	
$\rho(\varepsilon_1, \varepsilon_3)$	-0.215***	(0.039)	-0.278***	(0.030)	
$\rho(\varepsilon_2, \varepsilon_3)$			-0.696***	(0.020)	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at district level). Estimations are conducted using the maximum likelihood method. Controls: size of 17-year-old cohorts by federal districts, year and federal district fixed effects, gender, age, age², age·male, age²·male; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2000–2008, author's calculations. Observations: 28,622 and 38,910, respectively.

Tables 13 and 14 show the estimation results using the interaction of the instrumental variable with the parents' educational background. The parents' education is described by one dummy variable, which takes a value of 1 if at least one of the parents has a higher education degree. The information about parental education is available only for the year 2006; both models are thus estimated only for this period. The control variables in all equations include the parental education variable (this variable is thus not used as the instrument, but only as the control), the full set of the cohort year-of-birth dummy variables (which also allow for control of the effect of age in the most flexible way) and the size of the corresponding 17-year-old cohorts at the regional level.

Table 13 lists the estimation results for the two-equations model (educational attainment and wages). Column 1 describes the results from the estimation in a similar setting as in the baseline model, but using data only from 2006. The estimated coefficient for the returns to education is of a similar magnitude as that using 2000–2008 data. Columns 2 and 3 include the control variable for the parents' education in both equations, while column 3 also includes the interaction term of the expansion of the higher education system with parental educational background in the first equation. The expansion of the higher education system affected to a slightly greater extent youths with highly educated parents. Moreover, controlling for the parental education in both educational attainment and wage equations decreases the estimated returns to education by approximately fifteen percentage points.

Table 13: Estimations of the Returns to Higher Education (by Maximum Likelihood), Employed Population, 24- to 47-year-olds, 2006

Variables	2 equations: MLE					
	HE*	1 Wages	HE*	2 Wages	HE*	3 Wages
<i>HE degree*</i>		0.738*** (0.057)		0.573*** (0.075)		0.589*** (0.048)
<i>‡ of HE slots, by districts</i>	0.213*** (0.071)		0.159*** (0.023)			
<i>‡ of HE slots, by districts · · Parents with HE *</i>					0.191*** (0.049)	
<i>‡ of HE slots, by districts · · Parents without HE *</i>					0.150*** (0.027)	
<i>Parents with HE *</i>			1.091*** (0.065)	0.126*** (0.048)	1.039*** (0.133)	0.119*** (0.041)
$\rho(\varepsilon_1, \varepsilon_2)$		-0.198*** (0.032)		-0.111* (0.062)		-0.124*** (0.041)
$\sigma^2(\varepsilon_2)$		0.540*** (0.028)		0.527*** (0.026)		0.528*** (0.025)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at district level). Estimations are conducted using the maximum likelihood method.

Controls: parental educational background (dummy variable for those who had at least one parent with a higher education degree), size of 17-year-old cohorts by federal districts, cohort year-of-birth fixed effects, federal district fixed effects, gender. Sources: RLMS, 2006, author's calculations. Observations: 3,738.

Table 14 lists the estimated results for the three-equations model (educational attainment, employment and wages). The results for the wage returns to education are similar to the two-equations model and smaller than in the baseline estimations; the estimated returns to a higher education degree, however, are smaller than in the main estimations.

Overall, these sensitivity tests suggest that the estimated returns to a higher education degree are robust, but accounting for the parents' education slightly reduces these returns.

Table 14: Estimations of the Returns to Higher Education (by Maximum Likelihood), Whole Population, 24- to 47-year-olds, 2006

Variables	3 equations: MLE					
	HE*	1 Work	Wages	HE*	2 Work	Wages
<i>HE degree</i> *		0.406** (0.188)	0.596*** (0.120)		0.405 (0.336)	0.603*** (0.097)
<i># of HE slots, by districts</i>	0.284*** (0.037)					
<i># of HE slots, by districts · · Parents with HE</i> *				0.295*** (0.071)		
<i># of HE slots, by districts · · Parents without HE</i> *				0.282*** (0.051)		
<i>Covariance matrix:</i>	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
ε_1	1			1		
ε_2	-0.027	1		-0.026	1	
ε_3	-0.157**	-0.637***	0.696***	-0.161***	-0.641***	0.700***
$\rho(\varepsilon_1, \varepsilon_2)$	-0.027	(0.116)		-0.026	(0.218)	
$\rho(\varepsilon_1, \varepsilon_3)$	-0.188**	(0.079)		-0.192***	(0.061)	
$\rho(\varepsilon_2, \varepsilon_3)$	-0.763***	(0.044)		-0.767***	(0.043)	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at district level). Estimations are conducted using the maximum likelihood method. Controls: parental educational background (dummy variable for those who had at least one parent with a higher education degree), size of 17-year-old cohorts by federal districts, cohort year-of-birth fixed effects, federal district fixed effects, gender; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2006, author's calculations. Observations: 4,964.

4.4.4 Changes in the returns to education over time

This section examines the overall equilibrium effects of the tertiary education expansion on wages for all workers (those for whom the access to higher education was both affected and unaffected by the reforms). Some studies argue that the variations in the returns to education significantly depend on the variations in the supply of university graduates, among them Katz and Murphy (1992), and Card and Lemieux (2001a). Increasing the supply of university graduates decreases their wages in the labour market equilibrium, except in the two following cases: (1) a perfectly elastic demand for higher education graduates in the labour market; (2) an increasing demand for higher education graduates over the same period of time. The following empirical models control for the changes in returns to education, both in terms of employment and wages, over the analyzed period of 2000–2008.

Table 15 lists the estimation results for the wage and employment equations, controlling for the time-varying returns to education and overall time-trends in both wages and employment. Models 1 and 3, for wages and employment, respectively, include a linear trend in the returns to education – $HE^* \cdot Time-trend$. Models 2 and 4, for wages and employment, include the interaction variables of a higher education degree attainment and years – $HE^* \cdot 200\sim^*$. Correspondingly, control variables include linear time-changes in wages for all workers and year fixed effects. The estimated coefficients for the time variation in the returns to education are not statistically significant for either the wage or employment equation. These coefficients are low in magnitude and any tendencies for negative or positive changes in the wage returns to education are not revealed in the analysis. However, there is a positive linear trend in wages for all workers over the analyzed period of time – an increase of approximately 13% yearly. Similarly, there are no statistically significant changes in the influence of higher education on employment probability during the analyzed period.

Therefore, the estimation results do not reveal the overall equilibrium changes of the returns to education during the analyzed period. On average, wages grew from 2000 to 2008 for workers both with and without a higher education degree, though returns to education remained stable. These results suggest that the effects of the growing supply of university graduates on wages were compensated by the growing demand for highly qualified workers.

Table 15: Time-Varying Returns to Education, 24- to 47-year-olds, 2000–2008

Variables	Wage Equation (OLS)		Employment Equation (probit)	
	Wage (1)	Wage (2)	Work (3)	Work (4)
$HE\ degree^*$	0.453*** (0.026)	0.456*** (0.033)	0.449*** (0.046)	0.431*** (0.060)
$HE^* \cdot Time-trend$	0.001 (0.004)		–0.011 (0.008)	
$Time-trend$	0.128*** (0.004)		0.000 (0.007)	
$HE^* \cdot 2001^*$		–0.039 (0.052)		0.014 (0.075)
$HE^* \cdot 2002^*$		–0.000 (0.046)		–0.077 (0.082)
$HE^* \cdot 2003^*$		0.034 (0.037)		–0.051 (0.088)
$HE^* \cdot 2004^*$		–0.028 (0.043)		0.019 (0.102)
$HE^* \cdot 2005^*$		0.008 (0.052)		0.023 (0.081)
$HE^* \cdot 2006^*$		0.042 (0.049)		–0.042 (0.068)
$HE^* \cdot 2007^*$		0.019 (0.041)		–0.128 (0.095)
$HE^* \cdot 2008^*$		–0.036 (0.041)		–0.093 (0.077)
$Year\ 2001^*$		0.252*** (0.038)		–0.110*** (0.040)
$Year\ 2002^*$		0.399*** (0.028)		–0.151*** (0.052)
$Year\ 2003^*$		0.475*** (0.028)		–0.033 (0.044)
$Year\ 2004^*$		0.603*** (0.031)		–0.036 (0.042)
$Year\ 2005^*$		0.729*** (0.032)		–0.086* (0.048)
$Year\ 2006^*$		0.860*** (0.032)		–0.053 (0.046)
$Year\ 2007^*$		0.945*** (0.030)		–0.084 (0.058)
$Year\ 2008^*$		1.117*** (0.031)		–0.074 (0.070)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses (clustered at year-district levels). Estimations are conducted using the OLS and probit methods, respectively. Controls: size of 17-year-old cohorts by federal districts, federal district fixed effects, gender, age, age², age·male, age²·male; regional unemployment rates and marital status in the employment equation. Sources: RLMS, 2000–2008, author’s calculations. Observations: 28,622 and 38,910, respectively.

5 Heterogeneous Returns to Higher Education

This section explores the heterogeneity in the returns to higher education, by allowing them to vary with both observable and unobservable individual characteristics that affect the decision to pursue higher education. This is done by estimating the marginal returns to education and using the model with essential heterogeneity.

The previous sections refer to the framework of instrumental variables. IV provides an estimation of the local average treatment effect – the returns to higher education for those who switched to higher education because of the changes in the instrument (in this study’s case, the growing opportunities in access to higher education), who would not otherwise have pursued higher education. However, the local average treatment effect does not necessarily correspond to the average treatment effect in the population, in other words, to the average returns to education in the population (the average treatment effect is usually the main parameter of interest in the treatment evaluation). Thus, Heckman (1997), Heckman et al. (2003), and Carneiro et al. (2001) show that, in the presence of heterogeneity in the returns to education and selection on the gains (namely, students taking 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 population. It is possible to identify, under certain assumptions, the heterogeneous returns to education with the marginal treatment effects (MTE) estimation via the method of local instrumental variables (LIV). The MTE for the returns to education estimation is the average return to schooling for individuals who are indifferent to accessing education at varying levels of unobservable characteristics, which influence this educational choice along with other observable characteristics that can be accounted for. The concept of MTE was first introduced by Bjorklund and Moffitt (1987). Carneiro et al. (2001), Heckman et al. (2006), and Heckman and Vytlacil (2007) develop the theoretical framework of MTE estimation for the returns to schooling; derive the identification of the average treatment effect (ATE), treatment on the treated (TT), and treatment on the untreated (TUT); and also provide the empirical applications of these methods for U.S. data. Heckman and Li (2003) and Wang et al. (2007) apply these methods to the estimation of the returns to education in China.

5.1 Returns to education: model with essential heterogeneity

The educational choice, obtaining a higher education degree, is expressed by the variable S_i , which takes a value of 1 for those with high education, and 0 otherwise. The returns to education θ_i vary in the population:

$$\ln W_i = \alpha + \beta \cdot X_i + \theta_i \cdot S_i + U_i. \quad (1)$$

In the literature, such a model is referred to as “random coefficient” or “heterogeneous treatment effect” models (Heckman et al. (2006)). There are two potential wage outcomes (for workers with and without higher education):

$$\begin{cases} \ln W_{1,i} = \alpha_1 + \beta_1 \cdot X_i + U_{1,i}, & \text{if } S_i = 1, \\ \ln W_{0,i} = \alpha_0 + \beta_0 \cdot X_i + U_{0,i}, & \text{if } S_i = 0. \end{cases} \quad (2)$$

Such a specification assumes 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 for consideration of 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 $E(U_{0,i}|X_i) = 0$. These random shocks also vary with educational levels.

The educational-choice equation is defined as

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

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 , the less likely an individual i would be to obtain a higher education degree. $U_{1,i}$, $U_{0,i}$ and V_i are correlated. This model does not require specification of their correlation or their joint distribution.

Observed wages thus could be expressed as

$$\begin{aligned} \ln W_i &= S_i \cdot \ln W_{1,i} + (1 - S_i) \cdot \ln W_{0,i} = \\ &= \alpha_0 + \beta_0 \cdot X_i + [(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot X_i + (U_{1,i} - U_{0,i})] \cdot S_i + U_{0,i}. \end{aligned} \quad (4)$$

Therefore, returns to a higher education degree are expressed as

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

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 the wage equation (4) is estimated by OLS with heterogeneous returns to education based on observable characteristics, then the estimator $\hat{\theta}^{OLS}$ is obtained:

$$\ln W_i = \alpha + \beta_0 \cdot X_i + (\beta_1 - \beta_0) \cdot X_i \cdot S_i + \theta \cdot S_i + U_i, \quad (6)$$

$$\begin{aligned} \hat{\theta}^{OLS}(x_i) &= E(\ln W_i | X_i = x_i, S_i = 1) - E(\ln W_i | X_i = x_i, S_i = 0) = \\ &= ATE(x_i) + \{E(U_{1,i} - U_{0,i} | S_i = 1)\} + \{E(U_{0,i} | S_i = 1) - E(U_{0,i} | S_i = 0)\} = \\ &= ATE(x_i) + SGE_{1,i}^U + SB_{1 \rightarrow 0,i}, \end{aligned} \quad (7)$$

or:

$$\begin{aligned} &= ATE(x_i) + \{E(U_{1,i} - U_{0,i} | S_i = 0)\} + \{E(U_{1,i} | S_i = 1) - E(U_{1,i} | S_i = 0)\} = \\ &= ATE(x_i) + SGE_{0,i}^U + SB_{0 \rightarrow 1,i}. \end{aligned} \quad (8)$$

Therefore, the bias of the OLS estimator in estimating the ATE can be decomposed into two components: sorting on the gains effect (SGE) and selection bias (SB).

The selection bias $SB_{1 \rightarrow 0,i} = E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0)$ describes the fact that unobservable factors, which influence the decision to obtain a higher education degree, affect wages. It shows the difference in wages between those who obtain higher education and those who do not in the case where neither of the two would have earned a higher education degree.

Sorting on the gains effect $SGE_{1,i}^U = E(U_{1,i} - U_{0,i}|S_i = 1)$ represents the mean gain in unobservable components of the wage equation for people who choose a higher education degree $S_i = 1$. The non-zero value of the sorting on the gains effect means that individuals self-select for higher education based on their wage returns to unobservable characteristics in the case of either obtaining, or not obtaining, a higher education degree. The positive sign for the $SGE_{1,i}^U$ means that individuals with a degree have unobservable characteristics, which will be better paid if they obtain this level of education.

Similarly, sorting on the gains effect causes a bias in the IV estimation of the ATE (using I_i as the instrument, which is correlated to the V_i but is not correlated to the U_i , $U_{1,i}$, $U_{0,i}$):

$$\begin{aligned}\hat{\theta}^{IV}(x_i) &= \frac{Cov(I_i, \ln W_i)}{Cov(I_i, S_i)} = \\ &= ATE(x_i) + \frac{Cov(I_i, (U_{1,i} - U_{0,i})|S_i = 1) \cdot Pr(S_i = 1|x_i, I_i)}{Cov(I_i, S_i)}.\end{aligned}\tag{9}$$

Consequently, the IV estimator would be equal to the ATE (thus $LATE = ATE$) only if either (1) $U_{1,i} - U_{0,i} = 0$, indicating no heterogeneity in the returns to education, or (2) $U_{1,i} - U_{0,i}$ is independent of S_i , meaning no sorting on gains. In the absence of essential heterogeneity (i.e., in cases of people not sorting themselves based on the gains on unobservables), the IV is a consistent estimator of the ATE. However, in the presence of sorting on the gains, IV may overestimate the ATE by putting more weight on the returns for treated individuals (Heckman (1997)).

In the frame of the model with essential heterogeneity, the effects of treatment on the treated (TT) and the effects of treatment on the untreated (TUT) can be written as follows:

$$TT(x_i) = E(\ln W_{1,i} - \ln W_{0,i}|X_i = x_i, S_i = 1) = ATE(x_i) + SGE_{1,i}^U,\tag{10}$$

$$TUT(x_i) = E(\ln W_{1,i} - \ln W_{0,i}|X_i = x_i, S_i = 0) = ATE(x_i) + SGE_{0,i}^U,\tag{11}$$

where $SGE_{1,i}^U$ is the sorting on the gains effect for those who obtain a higher education degree, and $SGE_{0,i}^U$ represents the sorting on the gains effect for those who do not (the latter representing their returns to unobservable characteristics if they would have obtained a degree).

The average effects in the population, therefore, could be written as¹⁶:

$$\begin{aligned} ATE &= \int_X ATE(x_i) dF_X(x); & TT &= \int_{X|S=1} TT(x_i) dF_{X|S=1}(x), \\ TUT &= \int_{X|S=0} TUT(x_i) dF_{X|S=0}(x); & \hat{\theta}^{OLS} &= ATE + SGE_1 + SB_{1 \rightarrow 0}. \end{aligned} \quad (13)$$

The MTE estimation is used in order to calculate all these parameters properly:

$$\begin{aligned} MTE(X_i = x, U_{S,i} = u_s = p) &= E(\ln W_{1,i} - \ln W_{0,i} | X_i = x, U_{S,i} = u_s) = \\ &= \left. \frac{\partial E(\ln W_i | X_i = x, P(Z_i) = p)}{\partial p} \right|_{p=u_s}, \end{aligned} \quad (14)$$

where, $U_{S,i}$ is the uniform transformation of the random term of the educational equation V_i : $U_S = F_V(V)$, $U_S \rightsquigarrow Unif[0, 1]$. Therefore, the educational-choice equation (3) can be rewritten as follows:

$$\begin{aligned} S_i &= \begin{cases} 1, & \text{if } \gamma \cdot Z_i - V_i \geq 0 \\ 0, & \text{if } \gamma \cdot Z_i - V_i < 0 \end{cases} = \begin{cases} 1, & \text{if } F_V(\gamma \cdot Z_i) \geq F_V(V_i) \\ 0, & \text{if } F_V(\gamma \cdot Z_i) < F_V(V_i) \end{cases} = \\ &= \begin{cases} 1, & \text{if } P(Z_i) \geq U_{S,i} \\ 0, & \text{if } P(Z_i) < U_{S,i} \end{cases}. \end{aligned} \quad (15)$$

The MTE 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 , the unobservable component in the educational-choice equation). Since these are measured at the points of $P(Z_i) = p = u_s$, at these values of the unobservable component in the educational-choice equation, individuals are indifferent about seeking a higher education degree. Therefore, the MTE is the marginal willingness to pay for the $\ln W_{1,i}$ versus $\ln W_{0,i}$, 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) mean that the small values of V (large values of $-V$) are therefore 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) shows 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 formulas (25), described in section 3 of the Technical Appendix. The weights

¹⁶Note that in considering the population effects, the effect of the sorting on gains can be 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 seek higher education more often than male workers. Therefore, the sorting on the gains effects for the population is written as

$$\begin{aligned} TT &= ATE + SGE_1^U + (\beta_1 - \beta_0) \cdot (E(X|S=1) - E(X)), \\ TUT &= ATE + SGE_0^U + (\beta_1 - \beta_0) \cdot (E(X|S=0) - E(X)). \end{aligned} \quad (12)$$

are constructed such that, for the 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 the TUT effects, the higher weights are given for the individuals with lower probabilities of being treated.

The estimation procedure is described in section 3 of the Technical Appendix.

5.2 Estimation results

The method of the MTE estimation relies on the full support of predicted propensity scores (predicted probabilities of a higher education degree attainment for those who have it and for those who do not). One of the major predictors of tertiary education attainment is the educational background of parents. In the data, information on parents' educational background is observed only for the year 2006. If the probability of obtaining a higher education degree is estimated without information on parents' educational background for the years 2000–2008, the propensity scores are obtained, being in the interval of $\{0.05; 0.4-0.5\}$. If the variables describing parents' educational background are added as a predictor of higher education attainment, the estimated propensity scores cover almost all support from 0 to 1 (more precisely, $\{0.08; 0.8\}$). For that reason, this section conducts estimations only for the employed individuals observed in 2006: 24–52 years old (4,491 observations).

First, the educational-choice equation is estimated as a probit model. The explanatory variables include parents' educational background, the number of slots in the higher education system in a year when an individual was 17 years old at the regional level and its interactions with parents' educational background, size of 17-year-old cohorts by federal districts, age and age², and federal district fixed effects. Table 16 shows the estimated coefficients of the educational-choice equation. Predicted propensity scores are shown in Figure 6.

Table 16: Estimated Coefficients for Higher Education Attainment (Probit)

Variables	Coefficient	S.E.
$\#$ of HE slots, by districts · Parents without HE*	0.191***	(0.040)
$\#$ of HE slots, by districts · Parents with HE*	0.220***	(0.080)
Parents with HE* · Male*	1.030***	(0.165)
Parents with HE* · Female*	1.124***	(0.114)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Control Variables: size of 17-year-old cohorts by federal districts, gender (female), age, age², age-female, age²-female, federal districts fixed effects. Observations: 4,491. Sources: RLMS 2006, author's calculations.

Further, marginal returns to education are estimated, and ATE, TT and TUT are calculated, according to the procedure described in section 3 of the Technical Appendix.

Figure 7 shows the estimated marginal returns to education as a function of unobserved characteristics u_s , which affect higher education attainment. The marginal returns to education are shown separately for the male and female population. The estimation

results suggest declining returns to higher education for individuals with higher levels of unobserved characteristics, which decrease the probability of higher education attainment.

Table 17 lists the estimation results for ATE, TT and TUT. Table 18 shows the decomposition of these treatment effects, according to formulas (26), (27), (28), on the contributions of the returns to observed (females versus males) and unobserved characteristics.

Table 17: Estimated Returns to Education in the Population: ATE, TT, TUT

Treatment Effects	Estimated Effects	Standard Errors	T-Statistics
TT	0.780	0.073	10.6
ATE	0.706	0.069	10.2
TUT	0.652	0.066	9.8

Note: standard errors are calculated using bootstrap with 1,000 iterations.

Table 18: Decomposition of the Estimated Returns to Education: ATE, TT, TUT

Treatment Effects	Total	Contribution by	Contribution by
		Observed Characteristics $TT^{Obs}, ATE^{Obs}, TUT^{Obs}$	Unobserved Characteristics TT^U, ATE^U, TUT^U
TT	0.780	0.236	0.544
ATE	0.706	0.199	0.507
TUT	0.652	0.185	0.467

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

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

The sorting on the gains effects can be calculated for those who do and do not obtain a higher education degree:

$$SGE_1 = TT - ATE = 0.780 - 0.706 = 0.074$$

$$SGE_0 = TUT - ATE = 0.652 - 0.706 = -0.054$$

$$SGE_1^U = E(U_{1,i} - U_{0,i}|S_i = 1) = TT^U - ATE^U = 0.544 - 0.507 = 0.038$$

$$SGE_0^U = E(U_{1,i} - U_{0,i}|S_i = 0) = TUT^U - ATE^U = 0.467 - 0.507 = -0.040.$$

Positive sorting on the gains effect is observed, both for those who access higher education and for those who do not, and both based on observable and unobservable characteristics. For those who pursue higher education, their unobservable characteristics would have higher returns in the case when they have a degree relative to when they do not ($SGE_1^U = 0.038 > 0$). Those who do not pursue higher education would have lower returns to their unobservable characteristics if they obtained a degree ($SGE_0^U = -0.040 < 0$). These facts correspond to the declining pattern of the marginal returns to higher education, based on the male population. Similarly, female workers have higher returns to education and are

more likely to attain a higher education degree ($SGE_1 - SGE_1^U = 0.074 - 0.038 = 0.036 > 0$). This sorting on the gains determines the differences between the TT, ATE and TUT effects.

Selection effects take the following value:

$$SB_{1 \rightarrow 0,i} = E(U_{0,i}|S_i = 1) - E(U_{0,i}|S_i = 0) = \hat{\theta}^{OLS} - TT^U = 0.327 - 0.544 = -0.217$$

$$SB_{0 \rightarrow 1,i} = E(U_{1,i}|S_i = 1) - E(U_{1,i}|S_i = 0) = \hat{\theta}^{OLS} - TUT^U = 0.327 - 0.467 = -0.140.$$

A negative selection effect can be observed for those who pursue higher education. Those who have a higher education degree would have lower wage returns to their unobservable characteristics if they did not have a higher education degree, in comparison to those who do not currently have a degree. If education was provided to everybody, the difference in wages between those who attain it now and those who do not would be lower in absolute value than if higher education were forbidden for all (0.140 versus 0.217).

Overall, the estimated average returns to education in the population are similar to that obtained by IV, thus similar to the local average treatment effect of attainment of a higher education degree. This fact is due to the low magnitude of the sorting on the gains effect, which constitutes the bias of the IV estimator for the average returns to education, or in other words, which accounts for the difference between the local average treatment effect and the average treatment effect of a higher education degree. Returns to a higher education degree decrease with the unobservable characteristics, which reduce the probability of obtaining higher education (as shown by the estimation of the marginal returns to education). Those who increase their educational attainment alongside the increasing access to higher education likely have higher values of unobservable characteristics, which decrease the probability of higher education attainment, than those who attain it when access to higher education is tighter. Those who increasingly pursue higher education when access to higher education becomes greater would thus have lower returns to education.

This observed pattern of declining returns to higher education in the population could have several explanations. First, within the theoretical framework of education “enhancing productivity,” the returns to higher education could be lower for people with lower ability levels. Expansion of the higher education system in Russia provided an opportunity to increase educational attainment for a portion of the population. Since selection to universities was competitive, based on individual abilities, the periods with greater access to higher education are thus characterized by a lower average level of ability of university entrants. Second, within the framework of the “signalling” theory, when education is considered as a “signal of ability,” the labour market would reward higher education with a lower payoff when the signal of ability is less precise. When a larger cohort of higher education graduates comes to the labour market, the signal of ability is more vague compared with that of periods of stricter access to higher education. Thus, the expansion of the higher education system makes people both with and without higher education degrees worse in terms of the average level of ability. Therefore, the signal of ability through educational attainment becomes more fuzzy, and the labour market could lower the returns to higher education

for that reason. Finally, the expansion of higher education could have occurred to a different extent for high-quality and low-quality universities, as well as within different majors. Such a changing composition within the educational system could impact the returns to education.¹⁷ However, the current study analyzes neither the changes in the average quality of educational institutions and the resulting influence on the returns to education, nor the changing composition of majors among graduates. Instead, this study focuses on the average returns to a higher education degree, averaged by major and university quality. Nevertheless, analysis of the effects of changing educational quality and the composition of major could be a promising direction for further research.

6 Conclusion

This paper analyzes the effects of a 15-year expansion of the Russian Federation educational system on labour market outcomes. It identifies the returns to education for the population and quantifies the effects of this expansion on employment and wages.

The study reports evidence of significant returns to higher education in terms of both employment and wages. However, these returns are slightly smaller when controlling for parents' educational background. No significant equilibrium effects of this expansion on wages are revealed. The estimation results of the marginal returns to education suggest declining returns to education in the population, based on unobservable characteristics negatively affecting higher education attainment. This study determines a light positive sorting on the gains, meaning that youths self-select to universities based on their expected returns to education. However, this sorting is low in magnitude, and thus similar estimation results are obtained for the returns to education using the IV and MTE methods.

Expansion of the higher education system in Russia during 1990–2005 therefore had positive effects on wages for those who increased their educational attainment. However, their returns to education were lower than those of youths who obtained a higher education degree during the periods of tougher access to universities.

The large expansion of the higher education system in Russia provides an exogenous variation in access to higher education, significantly increasing university enrolment. This study estimates the effects of such expansion specifically on labour market outcomes. However, this natural experiment could potentially be used to analyze the influence of education on other life outcomes: occupational choices, work-life balance, marriages (including a partner's "quality"), childbearing decisions (number and timing of births, time dedicated to children's education), educational outcomes of children of the next generation, health attitudes, risky behaviour, 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 determining the importance of education.

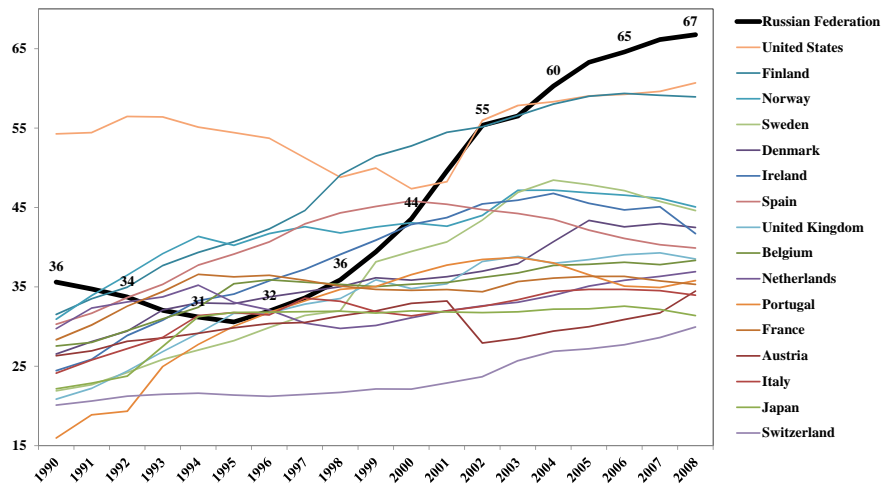
¹⁷For instance, characteristics such as the student-teacher ratio and student-educational surface size ratio could serve as proxies for quality of higher education (being negatively correlated with it). Looking at these ratios over the analyzed period, there was a small increase in these measures in the beginning of expansion followed by a stabilization and decrease. Moreover, the increasing proportion of social science majors (such as economics and law) is also observed during this period.

References

- Angrist, J.D., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434): 444–455.
- Angrist, J.D. and A.B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106(4): 979–1014.
- Angrist, J.D. and J.S. Pischke. 2009. "Mostly Harmless Econometrics: An Empiricist's Companion." *Princeton University Press* .
- Arcidiacono, P. 2005. "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica* 73(5 (Sept.)): 1477–1524.
- Belokonnaya, L., V. Gimpelson, T. Gorbacheva, O. Gihareva, R. Kapelushnikov, and A. Lukianova. 2007. "Wage formation: analysis in the light of work professions." *Wages in the Russian Federation: Evolution and Differentiation*. Moscow: State University Higher School of Economics.
- Belzil, C. 2006. "The Return to Schooling in Structural Dynamic Models: A Survey." *IZA Discussion Paper No. 2370*.
- Bjorklund, A. and R. Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics* 69.
- Card, D. 1995. "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* 201–222.
- . 1999. "The Causal Effect of Education on Earnings." *Handbook of Labor Economics* 3: 1801–1863.
- Card, D. and T. Lemieux. 2001a. "Can Falling Supply Explain the Rising Return to College for Young Men." *Quarterly Journal of Economics* 116: 705–746.
- . 2001b. "Education, Earnings, and the "Canadian G.I. Bill"." *Canadian Journal of Economics* 34(2): 313–344.
- Carneiro, P., J. Heckman, and E. Vytlacil. 2001. "Estimating the Return to Education When It Varies Among Individuals." *Working Paper, University of Chicago*.
- Cheidvasser, S. and H. Benitez-Silva. 2007. "The Educated Russian's Curse: Returns to Education in the Russian Federation During the 1990s." *Labor* 21(1): 1–41.
- Conneely, K. and R. Uusitalo. 1997. "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model." *Discussion Paper, Industrial Relations Sections, Princeton University*.
- Duflo, E. 2001. "Consequences of School Construction in Indonesia." *The American Economic Review* 91(4): 795–813.
- Froelich, M. and B. Melly. 2008. "Unconditional quantile treatment effects under endogeneity." *IZA Discussion Paper No. 3288*.
- . 2010. "Estimation of quantile treatment effects with STATA." *Stata Journal*.
- Gohberg, L.M., I.U. Zabaturina, N.V. Kovaleva, V.I. Kuznetsova, O.K. Ozerova, and L.A. Rosovetskaya. 2007. *Education in the Russian Federation*. Statistical Yearbook. SU-HSE, Moscow.
- Gregory, P. and J. Kohlhase. 1988. "The Earnings of Soviet Workers: Evidence from The Soviet Interview Project." *The Review of Economics and Statistics* 70(1).
- Heckman, J.J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32(3): 441–462.

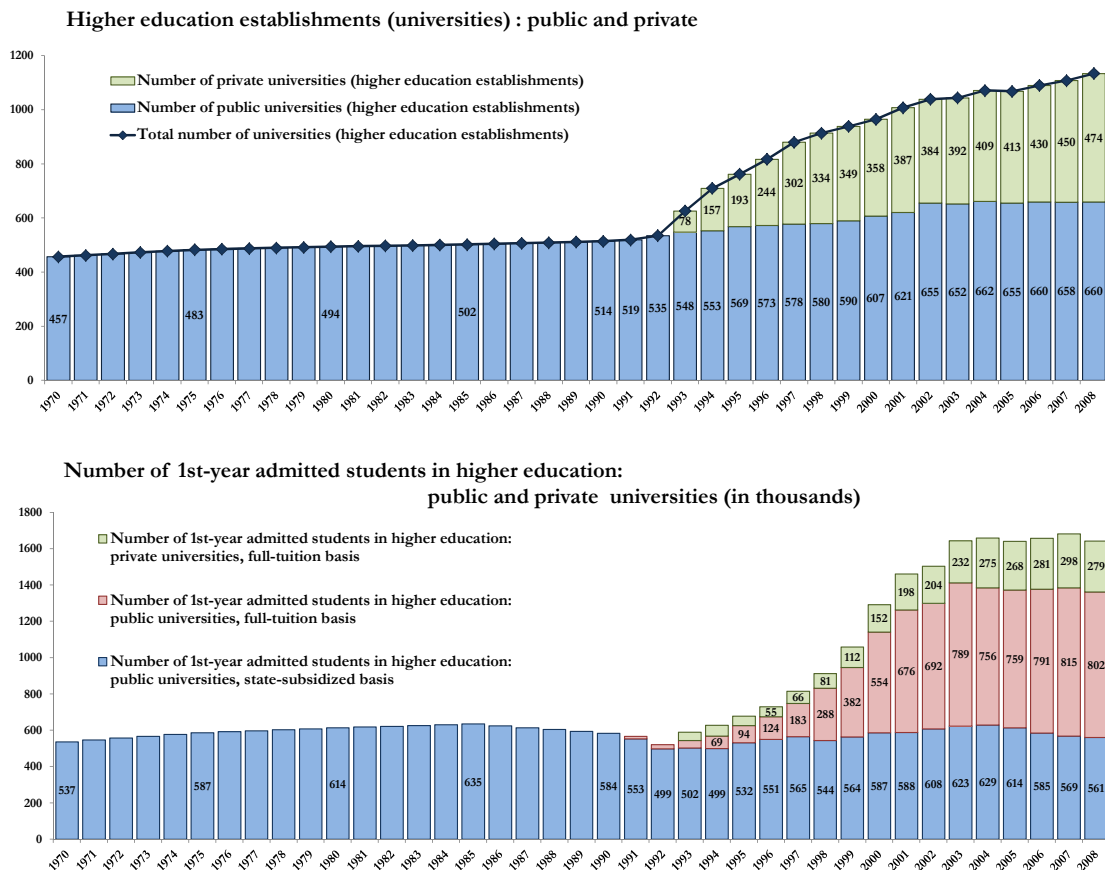
- Heckman, J.J. and X. Li. 2003. "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000." *IFAU Working Paper* No. 17.
- Heckman, J.J., L.J. Lochner, and P.E. Todd. 2003. "Fifty Years of Mincer Earnings Regressions." *NBER Working Paper* No. 9732.
- Heckman, J.J., S. Urzua, and E. Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88(3).
- Heckman, J.J. and E. Vytlacil. 1999. "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences*, No. 96.
- . 2007. "Econometric Evaluation of Social Programs." *Handbook of Econometrics*.
- ISCED. 2006. "International Standard Classification of Education (ISCED) 1997." *UNESCO Institute for Statistics*.
- Kane, T.J. and C.E. Rouse. 1993. "Labor Market Returns to Two- and Four-Year Colleges: is a Credit a Credit and Do Degrees Matter?" *NBER Working Paper* No. 4268.
- Katz, K. 1999. "Were there no returns to education in the USSR? Estimates from Soviet-period household data." *Labor Economics* (6).
- Katz, L. and K. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107: 35–78.
- Kyui, N. 2011. "Microeconomic Analysis of Tertiary Education Demand and Returns to Education." *PhD Dissertation, Paris School of Economics - University Paris-1 Pantheon Sorbonne*.
- Lukov, V.A. 2005. "Educational Reforms." *Encyclopaedia of the Humanities* (3): 217–219.
- OECD. 2013. "Education at a Glance 2013: OECD Indicators." *OECD Publishing* (<http://dx.doi.org/10.1787/eag-2013-en>).
- Ofer, G. and A. Vinokur. 1992. *The Soviet Household Under the Old Regime. Economic Conditions and Behaviour in the 1970s*. CUP, Cambridge.
- Oreopoulos, P. 2006a. "The Compelling Effects of Compulsory Schooling: Evidence from Canada." *Canadian Journal of Economics* 39(1): 22–52.
- . 2006b. "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter." *The American Economic Review* 96(1): 152–175.
- Pischke, S. and T. von Wachter. 2008. "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation." *Review of Economics and Statistics* 90(3): 592–598.
- Robinson, P. 1988. "Root-N consistent semiparametric regression." *Econometrica* (56): 931–954.
- Swafford, M., M. Kosolapov, and P. Kozyreva. 1999a. "International standards for evaluating survey research." *Universe of Russia* 1-2: 281–302.
- . 1999b. "The Russian Longitudinal Monitoring Survey: Measuring the Well-Being of Russians during the 1990s." *Universe of Russia* 3: 153–172.
- Walker, I. and Y. Zhu. 2008. "The College Wage Premium and the Expansion of Higher Education in the UK." *Scandinavian Journal of Economics* 110(4): 695–709.
- Wang, X., B.M. Fleisher, H. Li, and S. Li. 2007. "Access to Higher Education and Inequality: The Chinese Experiment." *IZA Working Paper* No. 2823.

Figure 1: Number of Tertiary Education Students in Russia and OECD countries per 1,000 inhabitants (in %)



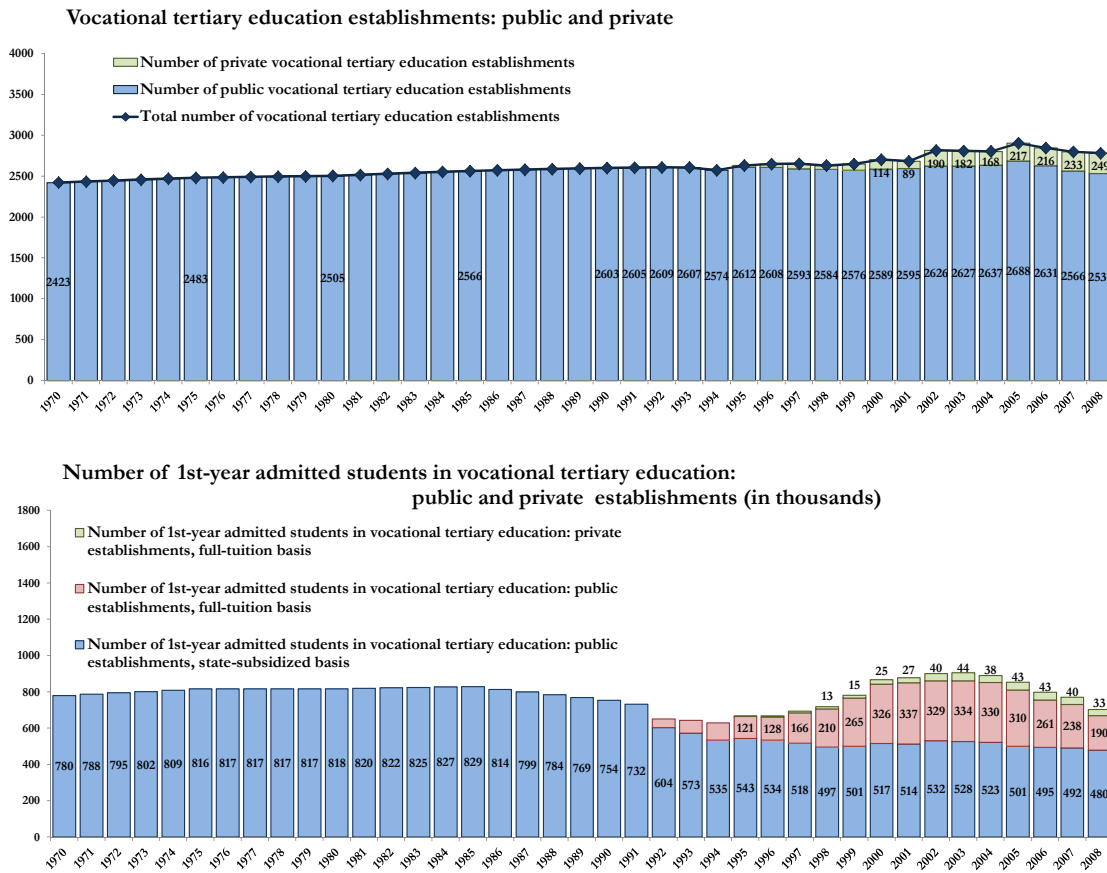
Source: World Bank databank (<http://databank.worldbank.org/data/home.aspx>).

Figure 2: Higher Education System in Russia: 1970-2008



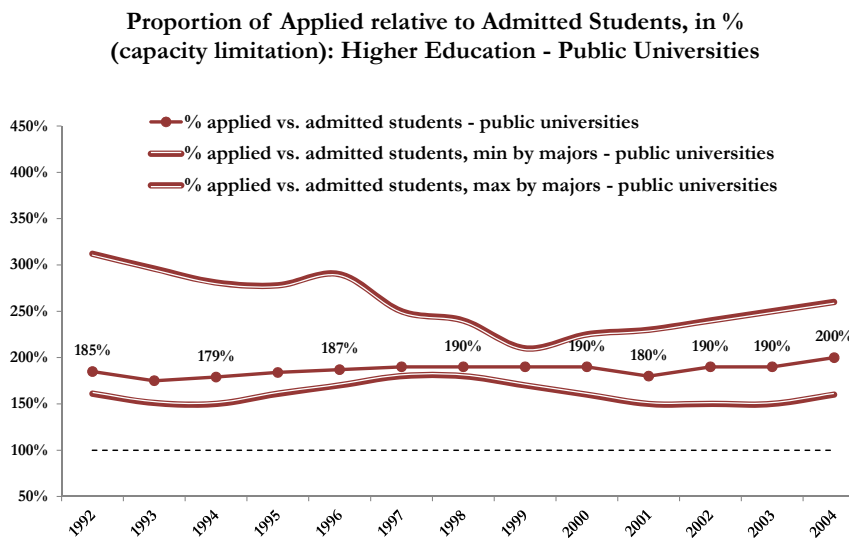
Sources: Gohberg et al. (2007), Russian Federal Committee for Statistics (<http://www.gks.ru>)

Figure 3: Vocational Tertiary Education System in Russia: 1970-2008



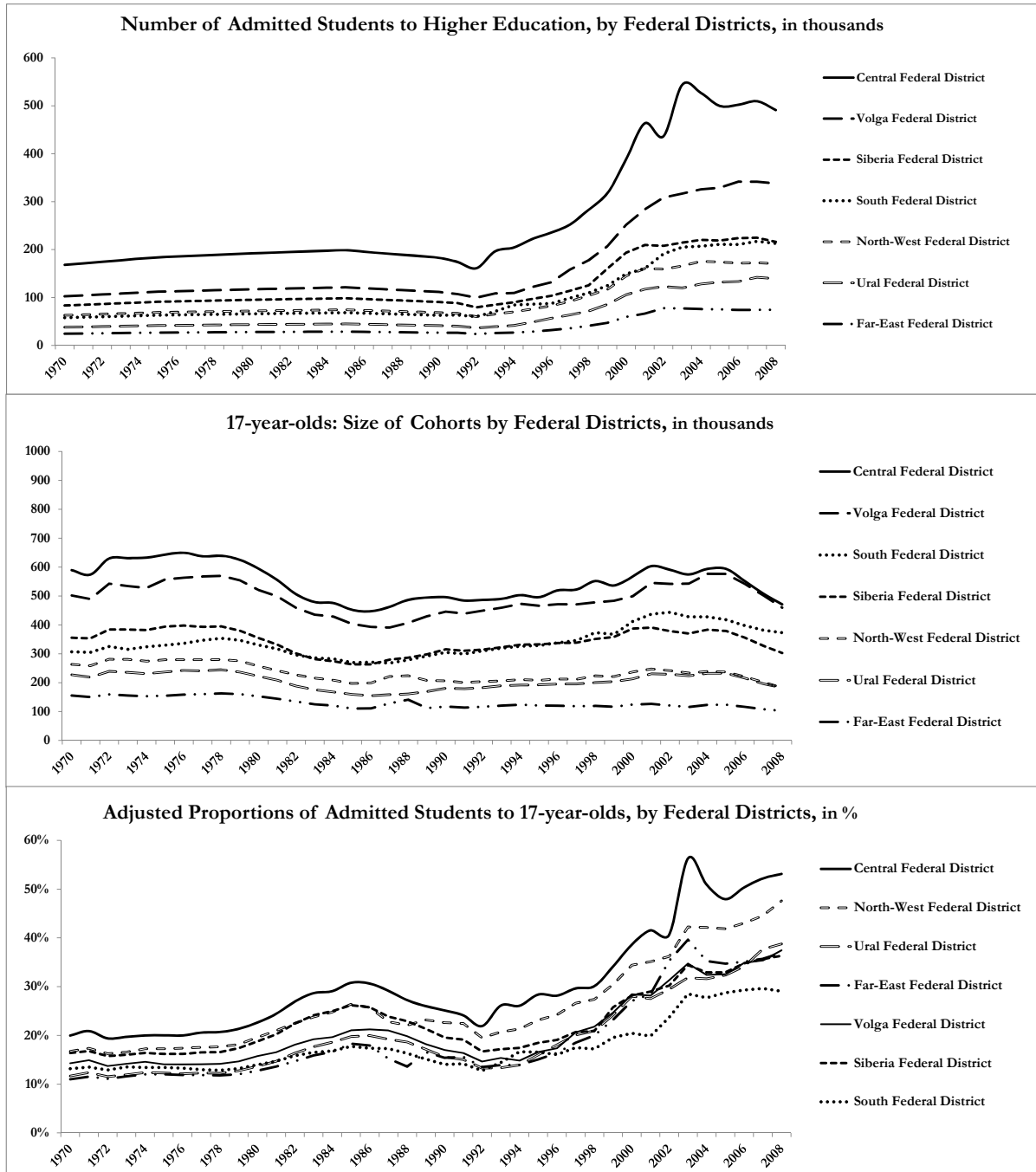
Sources: Gohberg et al. (2007), Russian Federal Committee for Statistics (<http://www.gks.ru>)

Figure 4: The Ratio of Admitted to Applying Students: Public Universities



Source: Russian Federal Committee for Statistics (<http://www.gks.ru>)

Figure 5: Regional Expansion of the Higher Education System in Russia: 1970-2008



Sources: Education in the Russian Federation (Gohberg et al. (2007)), Russian Federal Committee for Statistics (<http://www.gks.ru>)

Figure 6: Estimated Propensity Scores of Higher Education Attainment, 24- to 52-year-olds, 2006

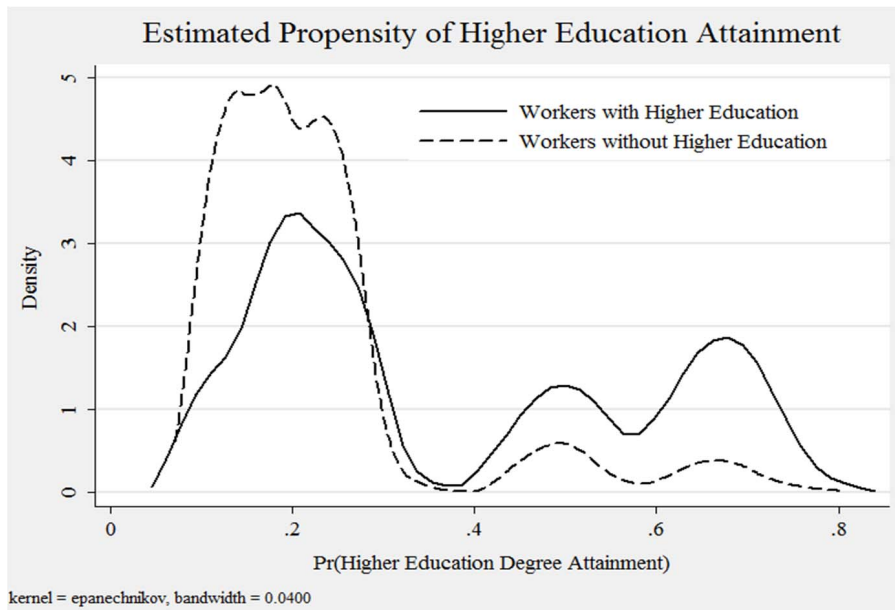
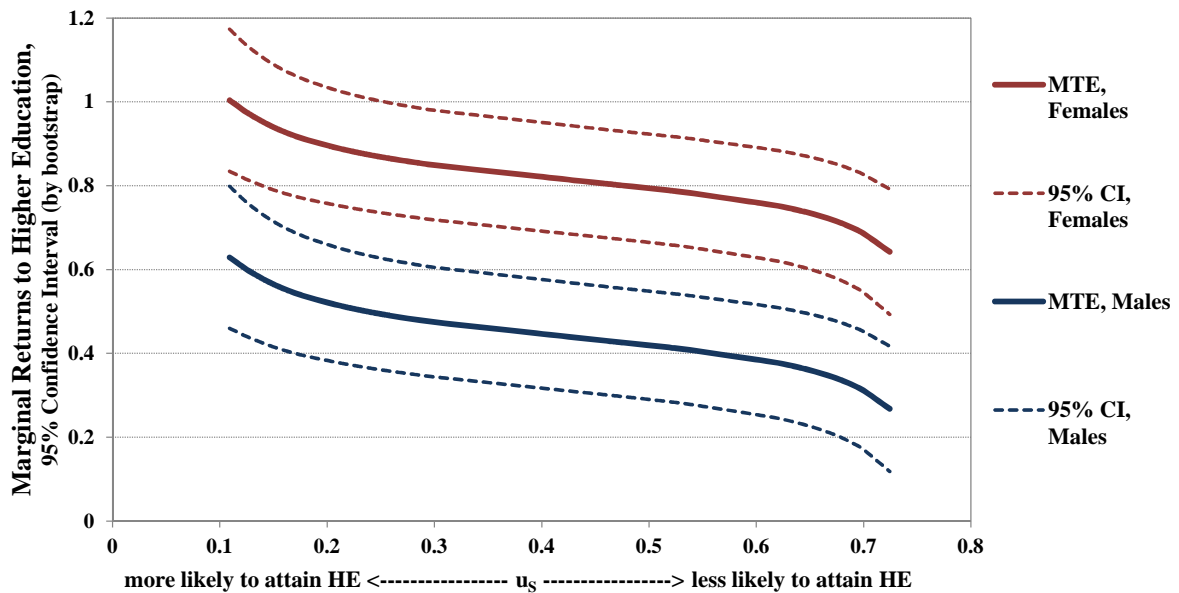


Figure 7: Estimated Marginal Returns to Higher Education (MTE), Males and Females 24–52 years old, 2006



Note: confidence intervals are calculated by bootstrap with 1,000 iterations.

A Technical Appendix

This section discusses the methods of estimation of econometric models used in the paper. The joint model of educational choice and wages (section 1), and joint model of educational choice, employment and wages (section 2) are presented. Section 3 of this appendix describes the estimation procedure for the marginal returns to education.

1. Educational Choice and Wages

Higher Education Degree Attainment and Wages

The empirical model of wages (W_i) with endogenous education, which is described by the binary variable higher education degree attainment (Ed_i), is as follows:

$$Ed_i = I(Y_{2TE,i}^* \geq 0),$$

where $Y_{2TE,i}^* = X_{ed,i}\beta_{ed} + \varepsilon_{1,i};$

$$W_i = X_{w,i}\beta_w + \varepsilon_{2,i}; \tag{16}$$

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

The following normalization is done for identification reasons: $\sigma_{\varepsilon_1}^2 = 1$.

All parameters of the model are estimated, as: $\theta = \{\beta_{ed}; \beta_w; \sigma_{\varepsilon_1\varepsilon_2}; \sigma_{\varepsilon_2}^2\}$.

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

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

The likelihood function for this model can be written as follows:

$$L(\theta, X) = \prod_{i=1}^N \{ (l_{Ed_i=1}(\theta, X_i))^{Ed_i} (l_{Ed_i=0}(\theta, X_i))^{1-Ed_i} \}, \tag{17}$$

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

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

and

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

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

2. Educational Choice, Employment and Wages

Higher Education Degree Attainment, Employment and Wages

This section describes 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:

$$\begin{aligned}
Ed_i &= I(Y_{2TE,i}^* \geq 0), \\
\text{where } Y_{2TE,i}^* &= X_{ed,i}\beta_{ed} + \varepsilon_{1,i}; \\
Empl_i &= I(Empl_i^* \geq 0), \\
\text{where } Empl_i^* &= X_{em,i}\beta_{em} + \varepsilon_{2,i}; \tag{18}
\end{aligned}$$

$$W_i = W_i^* \cdot (Empl_i),$$

where $W_i^* = X_{w,i}\beta_w + \varepsilon_{3,i};$

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

The following normalization is performed for identification reasons: $\sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1.$

All parameters of the model are estimated, as

$$\theta = \{\beta_{ed}; \beta_{em}; \beta_w; \sigma_{\varepsilon_3}^2; \sigma_{\varepsilon_1\varepsilon_2}; \sigma_{\varepsilon_1\varepsilon_3}; \sigma_{\varepsilon_2\varepsilon_3}\}.$$

Note $X_i = \{X_{ed,i}; X_{em,i}; X_{w,i}\}$ - the set of all variables, which are used in the model. Note also that Ed_i is included in $X_{em,i}$ and $X_{w,i}$, and a 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 follows:

$$L(\theta, X) = \prod_{i=1}^N \{ (l_{Ed_i=1, Empl_i=1}(\theta, X_i))^{Empl_i \cdot Ed_i} \cdot (l_{Ed_i=0, Empl_i=0}(\theta, X_i))^{(1-Empl_i)(1-Ed_i)} \cdot (l_{Ed_i=0, Empl_i=1}(\theta, X_i))^{Empl_i \cdot (1-Ed_i)} (l_{Ed_i=1, Empl_i=0}(\theta, X_i))^{(1-Empl_i)(Ed_i)} \}. \quad (19)$$

The parts of the likelihood function for the unemployed population can be expressed as

For unemployed people without a higher education degree:

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

For unemployed people with a higher education degree:

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

Further, the parts of the likelihood function for the employed population can be expressed as

For employed people without a higher education degree:

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

For employed people with a higher education degree:

$$\begin{aligned}
l_{Ed_i=1, Empl_i=1}(\theta, X_i) &= P(-\varepsilon_1 \leq X_{ed,i}\beta_{ed}; -\varepsilon_2 < X_{em,i}\beta_{em}; W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\
&= f(W_i) \cdot P(-\varepsilon_1 \leq X_{ed,i}\beta_{ed}; -\varepsilon_2 < X_{em,i}\beta_{em} | W_i = X_{w,i}\beta_w + \varepsilon_{3,i}) = \\
&= \frac{1}{\sigma_{\varepsilon_3}} \phi \left(\frac{W_i - X_{w,i}\beta_w}{\sigma_{\varepsilon_3}} \right) \cdot \Phi_2 \left(\frac{X_{ed,i}\beta_{ed} + \frac{\sigma_{\varepsilon_1\varepsilon_3}}{\sigma_{\varepsilon_3}^2} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma_{\varepsilon_1}^2 - \frac{\sigma_{\varepsilon_1\varepsilon_3}^2}{\sigma_{\varepsilon_3}^2}}}; \frac{X_{em,i}\beta_{em} + \frac{\sigma_{\varepsilon_2\varepsilon_3}}{\sigma_{\varepsilon_3}^2} \cdot (W_i - X_{w,i}\beta_w)}{\sqrt{\sigma_{\varepsilon_2}^2 - \frac{\sigma_{\varepsilon_2\varepsilon_3}^2}{\sigma_{\varepsilon_3}^2}}}; \rho_{(1,2|3)} \right), \\
\text{where: } \rho_{(1,2|3)} &= \frac{\sigma_{\varepsilon_1\varepsilon_2} - \frac{\sigma_{\varepsilon_1\varepsilon_3} \cdot \sigma_{\varepsilon_2\varepsilon_3}}{\sigma_{\varepsilon_3}^2}}{\sqrt{(\sigma_{\varepsilon_1}^2 - \frac{\sigma_{\varepsilon_1\varepsilon_3}^2}{\sigma_{\varepsilon_3}^2})(\sigma_{\varepsilon_2}^2 - \frac{\sigma_{\varepsilon_2\varepsilon_3}^2}{\sigma_{\varepsilon_3}^2})}}.
\end{aligned}$$

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

3. Marginal Treatment Effects: Estimation Method

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

1. The propensity scores (predicted probabilities for the higher education degree attainment) for the educational-choice equation (3) are estimated by probit: p . Next, the common support for the propensity scores for workers with and without a higher education degree is constructed: $p_s = \{max(p_{min}^0, p_{min}^1); min(p_{max}^0, p_{max}^1)\}$. p_{min}^1 and p_{min}^0 are the minimal values of propensity scores for the population with ($S = 1$) and without ($S = 0$) a higher education degree, p_{max}^1 and p_{max}^0 are the maximal values of propensity scores for the same groups of workers.
2. Then a semi-parametric two-step procedure is used in order to estimate the marginal treatment effect (Carneiro et al. (2001), Heckman et al. (2006)).
 - (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):

$$\ln W = \alpha_0 + \beta_0 \cdot X + [(\beta_1 - \beta_0) \cdot X] \cdot p + H(p) + \varepsilon_w \quad (20)$$

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

In order to simplify the non-linear component, this equation can be rewritten as follows:

$$\begin{aligned} E(\ln W|p) &= \beta_0 \cdot E(X|p) + (\beta_1 - \beta_0) \cdot E(X \cdot p|p) + H(p) \quad (21) \\ \ln W - E(\ln W|p) &= \beta_0 \cdot (X - E(X|p)) + (\beta_1 - \beta_0) \cdot (X \cdot p - E(X \cdot p|p)) + \varepsilon_w. \end{aligned}$$

This part corresponds to the estimation steps 1-5 of Heckman et al. (2006). $E(\ln W|p)$, $E(X|p)$, and $E(X \cdot p|p)$ can be estimated by the local linear regressions of the $\ln W$, X , and $X \cdot p$ on p . Then, the residuals of these regressions can be calculated: $\ln W - \widehat{(\ln W|p)}$, $X - \widehat{(X|p)}$, $X \cdot p - \widehat{(X \cdot p|p)}$. Finally, by regressing (OLS) the first residuals on the second and third ones, the estimated coefficients for β_0 and $(\beta_1 - \beta_0)$ can be derived.

- (b) The non-parametric term $H(p)$ is determined from the residual \tilde{W} of the equation (20):

$$\tilde{W} = \ln W - \hat{\beta}_0 \cdot X - [\widehat{(\beta_1 - \beta_0)} \cdot X] \cdot p = H(p) + \alpha_0 + \varepsilon_w \quad (22)$$

$$H(p) + \alpha_0 = E(\tilde{W}|p). \quad (23)$$

$\frac{\partial H(p)}{\partial p}$ can then be estimated by local linear regression of \tilde{W} on p .

In the current study, this part is estimated using the *semipar* package in STATA (semi-parametric regression estimator of Robinson (1988)). This non-parametric estimator uses a Gaussian kernel weighted local polynomial fit.

Marginal returns to education are determined as in Heckman et al. (2006), steps 6-7:

$$\begin{aligned} MTE(X = x, u_s) &= \frac{\partial E(\ln W | X, P(Z) = p)}{\partial p} = (\widehat{\beta_1 - \beta_0}) \cdot x + \frac{\partial H(p)}{\partial p} \\ &= (\widehat{\beta_1 - \beta_0}) \cdot x + MTE^U(X = x, u_s). \end{aligned} \quad (24)$$

The cross-validation method is used to choose the smoothing parameter among the values $0.1 \div 0.9$ (package *locreg* in STATA: Froelich and Melly (2008), Froelich and Melly (2010)). To evaluate the marginal returns to higher education, the evenly spaced points of the set of values p_s : $1 \div 99$ centiles of p_s are used.

3. The ATE, TT and TUT effects are calculated using formulas (13) and (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, gender) allows for accounting of 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:

$$\begin{aligned} ATE(x_i) &= \int_0^1 MTE(x_i, u_s) \cdot h_{ATE}(x_i, u_s) du_s = \\ &= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{ATE}(x_i, u_s) du_s = \\ &= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{ATE}(x_i, u_s) du_s, \\ TT(x_i) &= \int_0^1 MTE(x_i, u_s) \cdot h_{TT}(x_i, u_s) du_s = \\ &= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{TT}(x_i, u_s) du_s = \\ &= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{TT}(x_i, u_s) du_s, \quad (25) \\ TUT(x_i) &= \int_0^1 MTE(u_s) \cdot h_{TUT}(x_i, u_s) du_s = \\ &= \int_0^1 ((\widehat{\beta_1 - \beta_0}) \cdot x_i + MTE^U(x_i, u_s)) \cdot h_{TUT}(x_i, u_s) du_s = \\ &= (\widehat{\beta_1 - \beta_0}) \cdot x_i + \int_0^1 MTE^U(x_i, u_s) \cdot h_{TUT}(x_i, u_s) du_s, \end{aligned}$$

where the weights are

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

$$\begin{aligned}
ATE &= \int_X ATE(x_i) dF_X(x) = \\
&= \int_X (\widehat{\beta_1} - \widehat{\beta_0}) \cdot x_i dF_X(x) + \int_X \int_0^1 MTE^U(x_i, u_s) \cdot h_{ATE}(x_i, u_s) du_s dF_X(x) = \\
&= (\widehat{\beta_1} - \widehat{\beta_0}) \cdot E(X) + E(MTE^U), \tag{26}
\end{aligned}$$

$$\begin{aligned}
TT &= \int_{X|S=1} TT(x_i) dF_{X|S=1}(x) = \\
&= \int_{X|S=1} (\widehat{\beta_1} - \widehat{\beta_0}) \cdot x_i dF_{X|S=1}(x) + \\
&\quad + \int_{X|S=1} \int_0^1 MTE^U(x_i, u_s) \cdot h_{TT}(x_i, u_s) du_s dF_{X|S=1}(x) = \\
&= (\widehat{\beta_1} - \widehat{\beta_0}) \cdot E(X|S=1) + E(MTE^U \cdot h_{TT}|S=1), \tag{27}
\end{aligned}$$

$$\begin{aligned}
TUT &= \int_{X|S=0} TUT(x_i) dF_{X|S=0}(x) = \\
&= \int_{X|S=0} (\widehat{\beta_1} - \widehat{\beta_0}) \cdot x_i dF_{X|S=0}(x) + \\
&\quad + \int_{X|S=0} \int_0^1 MTE^U(x_i, u_s) \cdot h_{TUT}(x_i, u_s) du_s dF_{X|S=0}(x) = \\
&= (\widehat{\beta_1} - \widehat{\beta_0}) \cdot E(X|S=0) + E(MTE^U \cdot h_{TUT}|S=0). \tag{28}
\end{aligned}$$

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