ESSAYS ON THE EXPANSION OF HIGHER EDUCATION

Volha Belskaya

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2015

Approved by:

Klara S. Peter

Charles Becker

David Guilkey

Tiago Pires

Helen Tauchen

ABSTRACT

VOLHA BELSKAYA: Essays on the Expansion of Higher Education.
(Under the direction of Klara S. Peter)

Over the past twenty years, many developing countries expanded their higher education in order to become more competitive on international markets in future. The largest developing countries, Brazil, India, and Russia, tripled the number of college students per 100,000, and China increased the number of students twelve-fold. This expansion led to the influx of college graduates into the labor market, which had to adjust to the increase in the supply of educated workers. Existing literature shows how the adjustments associated with college expansion happen but many questions remain unanswered.

This dissertation evaluates the expansion of higher education in Russia and the effect of expansion on Russian labor market. The dissertation focuses on two features of the expansion. First, college expansion is usually associated with an increasing participation of women in college education. When the share of educated female workers grows faster than the share of educated male workers, the gender gap in higher education narrows. Between 1990 and 2008, the number of female students in higher education in Russia tripled from 1.4 to 4.3 million and the share of female students rose from 50 to 58 percent. The first chapter estimates education externalities created by the educated men and women in the labor markets and evaluates whether the faster growth of college participation among women affects the gender wage gap through education externalities. Second, during the expansion many new campuses open, providing the access to college to individuals who were previously constrained. The second chapter co-authored with Klara S. Peter and Christian M. Posso evaluates whether the expansion of higher education is economically worthwhile based on a recent surge in the number of campuses and college graduates in Russia. The empirical strategy relies on the marginal treatment effect method in both normal and semi-parametric versions, and estimating policy-influenced treatment parameters for the marginal students who are

directly affected by college expansion. Both of these questions associated with college expansion are heavily understudied in economics and this is where my dissertation contributes to the literature.

ACKNOWLEDGMENTS

I would like to thank my advisor, Klara S. Peter, for her constant support and encouragement. I would also like to express gratitude to my committee members, Charles Becker, David Guilkey, Tiago Pires, and Helen Tauchen, for their expert guidance and help. I am thankful to Donna Gilleskie, Brian McManus and the participants of the Applied Microeconomics Workshop, whose comments and suggestions improved my dissertation.

I would like to thank my parents, Iryna Belskaya and Uladzimir Belski, and my sister, Nadzeya Belskaya, for their love and encouragement throughout my graduate studies. I also thank Victor Standrityuk, Vance Midgett, Christian Posso, Riha Vaidya, Diana Shaw, Didem Pekkurnaz, Nazire Ozkan, and Giang Nguyen for being wonderful friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE GENDER GAP IN HIGHER EDUCATION IN RUSSIA: THE IMPACT ON THE GENDER WAGE GAP THROUGH EDUCATION EXTERNALITIES

1.1 Introduction

Growing participation of women in higher education has become common in many developed and developing countries. Between 1970 and 2011, the number of male students in tertiary education quadrupled from 19 to 90 million while the number of female students rose seven-fold from 13 to 92 million (Bank (2014)).[1] Together with higher female graduation rates, it led the share of women with tertiary education to exceed that of men in many regions of the world.[2] Despite this achievement in female educational attainment, it may not necessarily lead to improvements in women's labor markets outcomes compared to those of men. The goal of this chapter is to estimate the effect of the narrowing of the gender education gap on the wage gap between male and female workers.

The relationship between the narrowing of the gender education gap and the gender wage gap has started to gain an increasing attention in the literature. Gayle and Golan (2012) suggest that over the past decades a decline in the gender education gap might have been one of the sources of the reduced gender earnings gap in the U.S. Autor and Wasserman (2013) show that the reversal of the gender gap in college enrollment in the U.S. coincided with larger gains in earnings among college educated females compared to college educated males, i.e., between 1979 and 2010, real hourly wages of 25-39-year old female college graduates grew by 24 percent compared to 13 percent wage growth among educated male workers. As a result, the gender gap in earnings among

[1] Between 1970 and 2011, male and female population of tertiary age increased by 88 and 86 percent, respectively (Bank (2014)). Therefore, higher growth of female college enrollment was not driven by changes in the relative cohort size.

[2] In 2012, female graduation ratio in tertiary (ISCED 5A) education exceeded male graduation ratio in 52 out of 62 sampled countries (Bank (2014)).

25-39-year old college graduates declined from 32 percentage points in 1979 to 19 percentage points in 2010. Goldin (2014) suggests that over the last three decades a decline in the U.S. gender gap in earnings has largely been due to an increase in the productive human capital of female relative to male workers. Current predictions indicate that the share of female students will continue to grow, which will lead to a further increase in the tertiary education gap favoring females (OECD (2012)).

One channel through which the stock of educated workers affects their wages is education externalities. Positive education externalities arise when growth in the number of educated workers increases aggregate workers' productivity through the sharing of knowledge and skills among workers (Lucas (1988)) or induces a skill-biased technological change (Acemoglu (1998)). Higher aggregate productivity as well as higher demand for labor increase equilibrium wages. When female educational attainment grows faster than the educational attainment of males, wages of female workers may grow faster than wages of male workers for two reasons. First, changes in the aggregate productivity of female workers may exceed those of males, which leads to a higher growth of female wages and a decrease in the gender wage gap. Second, skill-biased technological change favoring female workers leads to a higher increase in female wages (Parro (2012)).[3] This chapter tests the hypothesis that the narrowing of the gender education gap leads to the changes in the gender wage gap because of the existence of education externalities.

This chapter uses a longitudinal survey of Russia, one of the largest developing countries, which experienced a reversal of the gender gap in higher education a decade ago. In Russia, the expansion of higher education was accompanied by a rapidly growing number of female students whose share reached 56 percent in 2011 and led to the reversal of the gender gap in higher education in the early 2000s (Figure 2.1).

One of the main challenges associated with establishing a causal relationship between the share of educated individuals and wages is the endogeneity problem. The share of college graduates in the labor market is likely to be correlated with wages for several reasons. First, workers with higher

[3]Parro (2012) provides some examples of skill-biased technological changes favoring females such as the introduction of computers.

levels of unobserved ability may choose to work in the labor markets with better-educated labor if those markets reward unobserved ability more. In this case, the estimate of the effect of college share on wages will be biased upward if the ability measure is omitted from the wage equation. Second, unobserved characteristics of the labor market may attract more educated workers. For example, a labor market characterized by high productivity of educated workers pays higher wages, which attracts more educated workers to the area.[4] To address the endogeneity problem, this chapter implements an instrumental variable strategy. An instrument for the share of individuals with higher education should affect college education of the majority of workers in a given labor market but should not be correlated with contemporaneous region-specific shocks. The number of college campuses in the past, 20 years before education and wages are recorded, represents such an instrument. This identification strategy parallels the one of Moretti (2004) and Muravyev (2008), who use the supply of higher education institutions in cities in the U.S. and Russia, respectively, to identify education externalities.

The chapter shows that in response to the growing share of educated individuals the wages of educated females grew faster than the wages of educated males and this contributed to the narrowing of the gender wage gap over time. Thus, increasing the access to higher education for women in developing countries generates benefits through education externalities. Male workers also benefit from the expansion of college education but the benefit for them, compared to female workers, is smaller.

The chapter contributes to several strands of the literature. First, it contributes to the literature on human capital externalities. Existing studies primarily ignore the possibility that human capital externalities may differ by gender. Some of the studies estimate education externalities for male workers only (Acemoglu and Angrist (2001); Conley and Tsiang (2003); Iranzo and Peri (2009); Kirby and Riley (2008); Lange and Topel (2006)). Other studies include a dummy variable for gender in the wage equation to account for the gender wage gap but do not interact the share of college

[4]The internet boom in the 1990s drove up the demand for educated workers, increased their wages and attracted more of them to the Silicon Valley in California (Moretti (2004)).

graduates with a gender dummy (Ciccone and Peri (2006); Dalmazzo and de Blasio (2007a), Dalmazzo and de Blasio (2007b); Liu (2007); Muravyev (2008); Rauch (1993); Sand (2013)). Moretti (2004) is the only study that tests for the difference in human capital externalities between male and female workers. The study does not reject the hypothesis of their equality. Despite that, education externalities may differ by gender. For example, a higher share of skilled workers in the labor force implies a larger market size for skill-complementary technologies and encourages faster productivity upgrading of skilled workers (Acemoglu (1998)). An increase in the supply of skills induces skill-biased technological change and increases the skill premium. Therefore, a faster growth in the supply of female skills may lead to a faster growth in the female skill premium. Parro (2012) suggests that the recent skill-biased technological change favors female workers more than males and improves their labor market outcomes. In case of equal education externalities across genders, a faster increase in the share of educated females may lead to a higher growth of female wages and a subsequent decrease in the gender wage gap. This chapter contributes to the literature by estimating education externalities by gender and linking the changes in education externalities to the changes in the gender wage gap. Additionally, this chapter accounts for the nonrandom selection of individuals into employment by means of the inverse propensity weighting method.

Second, the chapter contributes to the literature on the increasing participation of women in tertiary education (Autor and Wasserman (2013); Becker and Murphy (2010); Ganguli (2013); Goldin (2006); Parro (2012)). While most of the literature focuses on the U.S., the analysis of other countries which experience similar changes but differ from the U.S. in their economic conditions, labor markets, and institutions is important. This is the first paper that documents the reversal of the gender gap in higher education in one of the largest developing countries and analyzes its effect on the labor market.

Third, the chapter contributes to the literature on changes in the gender wage gap over time (Blau and Kahn (1997), Blau and Kahn (2000), Blau and Kahn (2006); Goldin (2006), Goldin (2014), among others). Finally, the use of the number of campuses in the past as an instrument for the contemporaneous share of college graduates builds on the applications of the supply-side shifters in identifying education externalities, such as an increase in the supply of higher education in a province due to an educational reform (Bratti and Leombruni (2014)), the presence of a land-grant college in a city (Iranzo and Peri (2009); Moretti (2004)), or the number of universities in a city before the transition to a market economy (Muravyev (2008)).

1.2 The Reversal of the Gender Gap

1.2.1 World Trend in the Gender Gap in Higher Education

Over the last decades, female college participation has grown steadily around the world. While there were seven female students per ten male students in tertiary education in 1970, the ratio had grown to ten female students per ten male students by 2011 (Bank (2014)). Starting from the 1970s, more females than males entered tertiary education in each subsequent cohort and, as a result, the share of educated females grew faster than the share of educated males over time. Panel A of Figure 2.2 depicts the difference between the share of 25-34-year old males and females with complete tertiary education (gender gap in tertiary education). The gap declined from 2.6 percentage points in 1970 to -2.3 percentage points in 2010, i.e., while the share of males with tertiary education exceeded the share of females by 2.6 percentage points in 1970, today the share of females with tertiary education exceeds that of males by 2.3 percentage points.

The observed decline in the gender education gap occurred in many geographic regions, except for South Asia and the Middle East (Panel C), and North Africa (Panel D).[5] The speed of the decline, however, varied by region, with a more rapid change occurring in European and East and Central Asian countries (Panel B of Figure 2.2). In developed countries, the gap changed from being positive in 1970 to negative in 2010 (Figure 2.3). On the other hand, not all developing countries experienced a reversal of the gender gap in tertiary education. Figure 2.4 depicts the evolution of the gap in four largest developing countries, Brazil, Russia, India, and China, which experienced a rapid growth of higher education over the last decades. Brazil and Russia followed the worldwide trend of a declining gender education gap, while India and China had a constant or even increasing gender gap in education (Figure 2.4). Russia experienced the largest decline in the

[5] South Asia and Middle East and North Africa represent regions with high fertility rates, low educational attainment, and limited access of girls to education.

gender education gap.

Existing literature attributes the growing number of women obtaining college education to the emerging differences in the costs and benefits of college education across genders. Studies on gender-specific costs of college education show that girls' preparedness for college and their lower nonpecuniary (effort) costs of college preparation increase their college attainment compared to boys. Specifically, an improvement in girls' high school preparedness (Goldin (2006)), the diverging high school graduation rates between boys and girls (Heckman and LaFontaine (2010)), and the non-cognitive behavioral factors (Jacob (2002)) provide an explanation for the females' advantage in the probability of continuing to college. While in college, higher non-cognitive skills explain higher female college completion rates compared to males and the widening of college completion rates between male and female students over time (Becker and Murphy (2010); Pekkarinen (2012)). Another strand of the literature attributes some part of the increase in female college enrollment to an increase in the labor market benefits from college education, which include higher female returns to education (Charles and Luoh (2003); Dougherty (2005)).

1.2.2 The Reversal of the Gender Education Gap in Russia

Over the last two decades, the largest developing countries experienced an expansion of higher education (Carnoy and Wang (2013)). In Russia, the expansion was initiated in the early 1990s, when public universities received a permission to open tuition-based programs and private colleges and universities to operate in the market of higher education. Over the next decade, the total number of colleges and universities in Russia more than doubled (Figure 2.1). A notable feature of the expansion was a significantly higher participation of women in college. Between 1990 and 2008, the number of female students in higher education tripled from 1.4 to 4.3 million and the share of female students rose from 50 to 58 percent (Panels A and B of Figure 2.1).

Higher female college participation had an impact on the educational composition of population across geographic regions. Between 1989 and 2002, the share of female population with complete higher education grew across all regions while the share of educated male population stagnated in some areas (Panels A and B of Figure 2.5). Between 2002 and 2010, the share of educated male and female population increased universally across regions (Panels C and D of Figure 2.5). The growth in the share of educated individuals occurred together with a growing dispersion of gender-specific human capital across regions. In the 2000s, the share of educated males and females increased more in the regions with the highest share of educated individuals in 2002 (Panels C and D of Figure 1.6). Table C.1 shows that the interquartile range of the share of educated males increased from 2.3 percentage points in 2002 to 3.6 percentage points in 2010 and from 2.2 percentage points in 2002 to 3.7 percentage points in 2010 among females. The difference between the 90th and the 10th percentile of the share of population with higher education also increased more rapidly among females and exceeded that of males in 2010. Thus, female human capital became more dispersed across regions over time.[6]

To graph the relationship between the changes in the educational composition of population and wages over time, Figure 1.7 plots the region-level wage changes against the changes in the fraction of population with higher education in the 1990s and the 2000s. Panels A and B show a negative relationship between the two variables in the 1990s for males and females while Panels C and D show that the relationship became positive in the 2000s. Between 2002 and 2010, regions with the fastest growing share of educated individuals experienced the highest wage increases, and the relationship between the two was stronger for female population.

The faster growth of female college participation resulted in a reversal of the gender gap in higher education in Russia. In 2002, the share of female population with higher education exceeded the share of educated males by 0.8 percentage points (Panel C of Figure 2.1). Between 2002 and 2010, the share of educated females grew faster than that of males, which led to a three-percentage-point gender education gap in favor of females in 2010. The evolution of the gender education gap in the U.S. follows a similar pattern, however, the gap in higher education closed only recently (Panel D of Figure 2.1). Changes in the gender education gap are primarily driven

[6]Selective migration across regions is unlikely to be responsible for the observed changes. Compared to other developed and developing countries, Russia has relatively low migration across regions (Andrienko and Guriev (2004); Guriev and Vakulenko (2014)).

by the younger cohorts. However, while in Russia the share of educated young females was already higher than the share of young males in late 1980s, a small gap in favor of educated young males existed in the U.S. in 1990. In Russia, the gender gap in higher education between young individuals tripled from 4 to 12 percentage points between 1989 and 2010 (Panel A of Figure 1.8). Over time, young females in Russia become much more likely than males to receive higher education. These changes provide motivation for the model which links together changes in the share of male and female population with higher education in a labor market and male and female wages.

1.3 Econometric Framework

1.3.1 Theoretical Model

The goal of the theoretical model is to show how to identify education externalities by gender, i.e., the effect of an increase in the relative supply of educated male and female workers in a labor market on the wages of male and female workers with different levels of education. The model predicts that when the relative supply of educated workers increases, wages of uneducated workers benefit both from imperfect substitution and education externality, while the wages of educated workers decrease because of the increased supply of educated workers but benefit from the education externality.

Assume that a country consists of a number of geographic regions, which represent competitive labor markets.[7] Each labor market produces a single good y and employs educated and uneducated male and female labor and capital. The production is represented by a Cobb-Douglas function[8]:

---

[7]Geographic regions are usually represented by cities(Ciccone and Peri (2006); Conley and Tsiang (2003); Moretti (2004); Muravyev (2008); Rauch (1993); Sand (2013)), provinces (Bratti and Leombruni (2014)), or states (Acemoglu and Angrist (2001); Ciccone and Peri (2006); Iranzo and Peri (2009); Lange and Topel (2006)).

[8]This form of the production function rules out substitutability between male and female workers in the production. This assumption is valid if there is high degree of gender segregation across occupations and/or industries. In Russia, women were highly concentrated in some industries, such as education (82% women, 18% men), healthcare (80% women, 20% men), services (70% women, 30% men) in 2012, and not highly represented in other industries, such as construction (15% women, 85% men) and transportation (27% women, 73% men). Compared to the U.S., Russian occupational segregation across genders is higher, which motivates this form of the production function.

$$y = \sum_{g=m,f} (\theta_0^g L_0^g)^{\alpha_0} (\theta_1^g L_1^g)^{\alpha_1} K^{g1-\alpha_1-\alpha_0} \tag{1.1}$$

where L_0^g is the number of uneducated workers of gender *g* in a labor market; L_1^g is the number of educated workers of gender *g* in a labor market; K^g is capital allocated to workers of gender *g* in a labor market; and the θ's are gender- and education-specific productivity shifters, which depend on the share of educated workers of gender *g* in a market:

$$log(\theta_0^g) = \phi_0^g + \mu^g(\frac{L_1^g}{L_0^g + L_1^g}), g = m, f \tag{1.2}$$

$$log(\theta_1^g) = \phi_1^g + \mu^g(\frac{L_1^g}{L_0^g + L_1^g}), g = m, f \tag{1.3}$$

where ϕ_0^g and ϕ_1^g capture the effect of gender *g* workers' human capital on productivity and $\frac{L_1^g}{L_0^g+L_1^g} < 1$ is the share of gender *g* workers with higher education in a labor market. In a competitive market, wages are equal to the marginal product of labor.[9]

The wages of uneducated workers of gender *g* are equal to

$$logw_0^g = log\alpha_0 + \alpha_0 log(\phi_0^g + \mu^g(\frac{L_1^g}{L_0^g + L_1^g})) + (\alpha_0 - 1)log(1 - \frac{L_1^g}{L_0^g + L_1^g}) \tag{1.4}$$

$$+ \alpha_1 log(\phi_1^g + \mu^g(\frac{L_1^g}{L_0^g + L_1^g}))(\frac{L_1^g}{L_0^g + L_1^g}) + (1 - \alpha_1 - \alpha_0)log(\frac{K^g}{L_0^g + L_1^g})$$

Wages of educated workers of gender *g* are equal to

$$logw_1^g = \alpha_0 log(\phi_0^g + \mu^g \frac{L_1^g}{L_0^g + L_1^g})(1 - \frac{L_1^g}{L_0^g + L_1^g}) + log\alpha_1 + \alpha_1 log(\phi_1^g + \mu^g \frac{L_1^g}{L_0^g + L_1}) \tag{1.5}$$

$$+ (\alpha_1 - 1)log(\frac{L_1^g}{L_0^g + L_1^g}) + (1 - \alpha_1 - \alpha_0)log(\frac{K^g}{L_0^g + L_1^g})$$

[9] Appendix 3 shows how wages are derived.

Assuming the population size, $L_0^g + L_1^g$, stays constant (i.e., no demographic shocks), the effect of changes in the share of educated workers of gender g can be derived by taking the derivative of wages with respect to the share of educated workers of gender g:

$$\frac{\partial log(w_0^g)}{\partial \frac{L_1^g}{L_0^g+L_1^g}} = \frac{\alpha_0 \mu^g}{\phi_0^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} + \frac{1-\alpha_0}{1-\frac{L_1^g}{L_0^g+L_1^g}} + \frac{\alpha_1 \mu^g}{\phi_1^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} + \frac{\alpha_1}{\frac{L_1^g}{L_0^g+L_1^g}} \tag{1.6}$$

$$\frac{\partial log(w_1^g)}{\partial \frac{L_1^g}{L_0^g+L_1^g}} = \frac{\alpha_0 \mu^g}{\phi_0^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} - \frac{\alpha_0}{1-\frac{L_1^g}{L_0^g+L_1^g}} + \frac{\alpha_1 \mu^g}{\phi_1^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} + \frac{\alpha_1 - 1}{\frac{L_1^g}{L_0^g+L_1^g}} \tag{1.7}$$

Equation (1.6) shows that when the share of male/female workers with higher education increases, wages of uneducated male/female workers increase because 1) uneducated workers' productivity increases due to the imperfect substitution between workers with different levels of education

$$\frac{1-\alpha_0}{1-\frac{L_1^g}{L_0^g+L_1^g}} + \frac{\alpha_1}{\frac{L_1^g}{L_0^g+L_1^g}} > 0$$

and 2) there is a positive externality from education

$$\frac{\alpha_0 \mu^g}{\phi_0^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} + \frac{\alpha_1 \mu^g}{\phi_1^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} > 0$$

Equation (1.7) shows that the effect on wages of male/female workers with higher education depends on two effects: 1) a supply effect, which moves the labor market along a downward sloping demand curve

$$\frac{\alpha_1 - 1}{\frac{L_1^g}{L_0^g+L_1^g}} - \frac{\alpha_0}{1-\frac{L_1^g}{L_0^g+L_1^g}} < 0$$

and 2) an education externality effect

$$\frac{\alpha_0 \mu^g}{\phi_0^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} + \frac{\alpha_1 \mu^g}{\phi_1^g + \mu^g(\frac{L_1^g}{L_0^g+L_1^g})} > 0$$

Hence, changes in the wages of educated male/female workers in response to an increase in the share of educated male/female workers depend on which of the two effects (supply effect or education externality effect) is stronger.

1.3.2 Empirical Model

This section describes an empirical specification of the wage equation based on the theoretical model derived above and discusses problems associated with estimating education externality effect. The wage of individual *i* of gender *g* living in region *r* in period *t* is equal to the market equilibrium wage and is determined by an equation of the form

$$log(w^g_{irt}) = \beta^g X^g_{it} + \varphi^g S^g_{rt} + \alpha^g Z_{rt} + \gamma_r + \gamma_t + u^g_{irt}, g = m, f \tag{1.8}$$

where X^g_{it} is a vector of individual characteristics; S^g_{rt} represents the percentage of college educated workers of gender *g* in region *r* in year *t*; Z_{rt} is a vector of characteristics of region *r* at time *t* which may be correlated with S^g_{rt}; γ_r represents region fixed effect; and γ_t is year effect. The error term, u^g_{irt}, is the sum of three components:

$$u^g_{irt} = \nu^g_r \lambda^g_{ir} + \xi^g_{rt} + \eta^g_{irt} \tag{1.9}$$

where λ^g_{ir} is a permanent individual unobservable component (e.g., ability); ν^g_r is a factor loading which represents the return to the unobserved component λ^g_{ir} in region *r*; ξ^g_{rt} represents time-varying shocks to labor demand and supply in region *r* in period *t*; η^g_{irt} is the transitory component of log wages which is assumed to be independently and identically distributed over individuals, regions and time.

When estimating the effect of changes in the share of workers with higher education, S^g_{rt}, on the wages of workers, an OLS estimate of the effect of share on wages, $\hat{\varphi}^g$, is likely to be biased due to the correlation between the share, S^g_{rt}, and the error term, u^g_{irt}, i.e., $corr(S^g_{rt}, u^g_{irt}) \neq 0$. First, the share of educated workers in the region could be correlated with the individual unobserved component λ^g_{ir}, i.e., $corr(S^g_{rt}, \lambda^g_{ir}) \neq 0$. Second, the share of educated workers in the region could be correlated with the time-varying demand or supply shocks, i.e., $corr(S^g_{rt}, \xi^g_{rt}) \neq 0$. In this case, an OLS estimate $\hat{\varphi}^g$ is biased either downward or upward, depending on whether the variation in the relative number of educated workers is driven by the unobserved supply or demand factors (Moretti 2004). If the variation in college share across regions is driven by the supply factors, the unobserved heterogeneity biases the OLS estimate downward. This may occur when some unobserved characteristics of the region attract more educated workers to the area raising the share of educated individuals (for example, geographic location, climate, amenities). To get a consistent estimate of the education externality effect, a researcher needs an instrumental variable that is uncorrelated with the region unobserved characteristics but correlates strongly with college share in the region.

On the other hand, an upward bias in the OLS estimate, $\hat{\varphi}^g$, of the education externality effect arises from the heterogeneity in the demand for educated workers across regions. In this case, the OLS coefficient in a regression of wages of educated workers on the share of educated workers assigns all of the observed correlation between wages and the share of educated workers to education externality and yields an estimate that is upward biased. An instrumental variable uncorrelated with factors that affect the productivity of educated workers and, therefore, the demand for them would generate a consistent estimate of education externality.

1.3.3 Identification

The main challenge associated with estimating education externalities is the endogeneity problem arising due to the correlation between the share of college educated workers in the region and wages. Acemoglu and Angrist (2001) is one of the first studies that addresses the endogeneity problem by using differences in the compulsory schooling and child labor laws across U.S. states as instruments for the average human capital in the labor market. However, the variation in this instrument primarily affects secondary education while sizeable externalities may instead be generated by college education.[10] Therefore, recent studies shifted their focus to higher levels of education such as college education. For example, Moretti (2004) uses the age structure of a

[10] For example, Rosenthal and Strange (2008) show that the benefits of spatial concentration are driven by proximity to college educated workers.

local labor market in the past and the presence of a land-grant college as the instruments for the share of college graduates and finds sizeable human capital externalities of college education. Dias and Tebaldi (2014) provide evidence that sector concentration of highly qualified workers (with at least a college degree) generates knowledge externalities, i.e., workers learn from their peers. To reconcile the mixed evidence on human capital externalities generated by high school and college education, Iranzo and Peri (2009) develop a model that predicts positive externalities from increased college education and negligible external effects from high school education.

To identify human capital externalities by gender in Russia, this paper explores the supply of higher education during the Soviet period. Specifically, the number of campuses of higher education institutions in the region of residence twenty years before the wage data is observed is used as an instrument for the current share of individuals with higher education in a region. A number of reasons support the idea that the number of campuses in the past represents a valid instrument for the contemporaneous share of population with higher education in a given labor market. First, a higher number of campuses in a given location increases the local youths' probability of receiving higher education (Frenette (2009); Oppedisano (2011)). College graduates are also more likely to stay and work in the location where they receive education (Bound and Turner (2004); Groen (2004); Winters (2011)). Therefore, the presence and the number of campuses of higher education institutions in the region of residence have a positive effect on the number and the share of individuals with higher education in a given market. Second, the Soviet government regulated the allocation of college campuses across the country based on the needs and political objectives of planned economy. The goal of the central authorities was to allocate higher education institutions over the vast territory of the country and ensure access to them. For example, there was one university/institute of higher education per 110,000 individuals in Russian cities before the collapse of the Soviet Union (Muravyev (2008)). The wages were also equalized across regions through the so-called wage grid and the skill-biased migration was very low. Therefore, the allocation of college campuses across regions during the Soviet period can be regarded as exogenous with respect to the market-based wages prevailing during the transition (the 1990s) and market period (the 2000s). Location of college campuses determined by the Soviet government had little to do with the demand for skilled labor in the emerging market economy of the 1990s and the 2000s.

1.4 Data

The data for this study comes from the Russia Longitudinal Monitoring Survey (RLMS), a longitudinal survey of the population of Russia initiated in the early 1990s to measure the effect of the market reforms on the economic well-being of the households and individuals.[11] The distribution of the initial sample of households by gender, age, and urban-rural location compared well with the corresponding distribution of the Soviet Census 1989. In later years, the sample was replenished to preserve the representativeness of the population of Russia. RLMS collects a rich set of information on demographic characteristics, education, health, labor market outcomes, and community characteristics, among many other. This chapter uses sixteen waves of RLMS spanning the period from 1994 to 2011.[12] The sample includes 22-59-year old men and 22-54-year old women.[13] I drop youths under the age of 22 because some of them may still be completing their education.

Using the Census data, I calculate the share of male and female population with higher education to proxy for the share of workers with higher education, S^g_{rt}.[14] The Census data are disaggregated by gender, geographic region, age (5-year groups), and the type of residence (urban/rural).[15] The share of male/female population with higher education in a region, in a particular age group, and in an urban/rural location is calculated as the number of males/females with complete higher education divided by the total male/female population. This measure is then merged with the RLMS based on the workers' current region of residence, their age, and the type of location where

[11]RLMS is organized by the National Research University Higher School of Economics, Moscow together with the Carolina Population Center at the University of North Carolina at Chapel Hill and the Institute of Sociology at the Russian Academy of Sciences. See http://www.cpc.unc.edu/projects/rlms-hse/project/study for the description of the study.

[12]RLMS was not conducted in 1997 and 1999.

[13]The official retirement age is 55 for women and 60 for men.

[14]The primary reason for this decision is a lack of data on the share of educated workers across regions and over time. I assume linear growth and interpolate the share of population with higher education in years for which Census data is not available.

[15]Age groups are 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, and 55-59. Urban locations include cities and townships. Rural locations include villages.

they live (urban/rural).

Labor market characteristics of regions such as the unemployment rate and the employment by industry and gender come from the Russian Federal Statistical Services. These characteristics are merged with the RLMS based on the respondent's current region of residence and survey year. They vary both across regions in a given year and over time due to the transition to the market economy during the period of study. The transition was characterized by the changing industrial structure and employment in different industries (Bank (2003), Bank (2005)). For example, the share of employment in manufacturing declined from 27 percent in 1994 to 19 percent in 2011. The share of employment in agriculture, construction, and transportation also declined. On the other hand, the share of employment in trade and education increased over time.

The Russian University Database provides the number of campuses of higher education institutions during the Soviet period (Belskaya and Peter (2015)).[16] This database contains detailed information on more than 1,000 institutions of higher education in Russia. Table A.2 shows the distribution of the number of campuses by year. The number of campuses varies considerably over time and across regions. The mean number of campuses per region increased from 19 campuses in 1974 to 22 campuses in 1991. The standard deviation decreased in the 1970s but increased steadily until the 1980s and declined again in the late 1980s.

Table A.3 presents the descriptive statistics of three samples used in the empirical analysis: workers of all levels of education, workers with some higher education or more (college dropouts, college graduates, and postgraduate graduates), and workers with less than higher education (secondary school dropouts, secondary school graduates, and specialized secondary education graduates). There are 40,121 male individuals in the sample and 40,660 female individuals. Columns 1 and 2 of Table A.3 present descriptive statistics of the samples of male and female workers. There is a 0.24 log points difference in hourly wages between male and female workers. Average work experience of male workers exceeds the experience of female workers by one year. On the other hand, female workers have 0.75 more years of education and the share of female population with

[16]This dataset is constructed based on the university websites and the Federal Portal "Russian Education" (www.edu.ru).

higher education in a local labor market exceeds the share of male population by 0.06 percentage points. Columns 3 and 4 present descriptive statistics for the sample of educated workers. There is a 0.15 log points difference in hourly wages, female college graduates have 0.14 more years of education compared to male college graduates, and the share of educated females exceeds the share of educated males by 0.05 percentage points. Columns 5 and 6 present descriptive statistics for the sample of workers without college education. The sample of less educated workers differs from the sample of educated workers in many respects. First, the gender wage gap of 0.36 log points is more than double than the gender wage gap between educated workers. Second, less educated females have 0.48 more years of education than less educated males, which is a three times larger gap compared to the gender gap in education among educated workers.

1.5 Results

1.5.1 Social Returns to Education

This section presents the estimates of the social returns to education for male and female workers as well as the changes in the social returns from mid-1990s to late 2000s. I estimate equation (1.8) by gender and report the results in Table A.4. Individual characteristics X^g_{irt} include years of education, work experience, and work experience squared. To correct for selection into the labor market, I estimate an inverse propensity weight from the probit model of non-missing wages on individual characteristics and dummies for being married and having children under the age of seven. I use the inverse of the estimated weight in all models. Columns 1 and 2 of Table A.4 show that there is a 2.819 percent increase in wages of male workers and a 2.331 percent increase in wages of female workers in response to a one-percentage point increase in the share of male and female population with higher education. The estimates also suggest that private return to one year of education for female workers exceeds that for male workers throughout the period. Another interesting finding is positive and significant returns to work experience for female workers and negative returns for male workers.

Columns 3 and 4 provide the estimates of the model that controls for the regional unemployment rate which proxies for the time-varying labor demand shocks. The estimates of the social returns to education decrease which implies that labor demand shocks matter. In columns 5 and 6, I include regional unemployment rate and an additional control for the labor demand shifts for male versus female labor across industries. Following Katz and Murphy (1992), shifts for male versus female labor across industries are predicted by the nationwide employment growth in industries, weighted by the changes in the region-specific employment share of male and female workers in those industries:

$$shock_{grt} = \sum_{i=1}^{8} \kappa_{irt} \triangle E_{gi} \tag{1.10}$$

where i indexes industry; $shock_{grt}$ represent the predicted employment change for workers of gender g in region r at time t; κ_{irt} is the share of total employment in industry i in region r at time t; $\triangle E_{gi}$ is the change in the log of employment in the same industry nationally between 1994 and t by workers of gender g.[17] The inclusion of the index lowers the estimates to 2.240 percent for male and 1.806 percent for female workers. The fact that Katz-Murphy index changes the estimates of the external returns to education further supports the idea that the demand shocks may introduce the bias. Therefore, it is important to control for them. Finally, to control for the unobserved permanent region-specific characteristics, I estimate equation (1.8) with the region fixed effects in column 7 and 8. Controlling for the heterogeneity across regions lowers the external returns to education to 1.705 percent for male and 0.951 percent for female workers but the estimates remain statistically significant.

Given that the share of educated individuals grew over time, I estimate changes in the social returns to education by year. Column 1 of Table 2.5 suggests that the returns to college share for male workers increased from 3.250 percent in 1994 to 3.974 percent in 2000 and decreased thereafter, reaching a minimum of 1.909 percent in 2011. The returns for female workers followed a similar pattern. Accounting for the heterogeneity across regions in columns 3 and 4 reveals a similar time trend, however, the returns become lower. The finding that the estimates of social returns from using single cross-sections are considerably larger than the estimates that control for region

[17] Classification of industries is presented in Appendix A.

fixed effects suggests that at least part of the relationship between the share of educated individuals and wages is due to omitted labor markets characteristics. A decline in the external returns to education in the 2000s coincided with an increase in the supply of college-educated workers who entered colleges during the expansion in the 1990s and started to join the labor force in the early 2000s. This finding parallels Sand (2013), who finds declining external returns to education in the U.S. between 1980s and 1990s. It is possible that an increase in the education level of population in a country with high level of educational attainment (such as Russia and the U.S.) decreases the external returns to education. Overall, the analysis in this subsection shows the existence of correlations between the main variable of interest, the share of male/female population with higher education, and wages of male and female workers.

1.5.2 External Returns to Education by Education Group

The model presented in Section 3 shows that the external returns to education represent a sum of the imperfect substitution effect and the spillover effect. To identify education externalities, the model needs to be estimated separately for workers with different levels of education. This section presents the estimates the effect of changes in the share of educated males and females on wages of workers with different levels of education.

Panel A of Table 2.6 reports the estimates of equation (1.8) for the sample of male and female workers with higher education. The coefficient on the share of population with higher education is found to be positive for skilled workers, implying that the positive education externality effect exceeds the negative supply effect. There is a 2.255 percent increase in wages of educated male workers and a 1.8 percent increase in wages of educated female workers in response to a one-percentage point increase in the share of male and female population with higher education. Accounting for local demand shifts and regional heterogeneity in column 7 and 8 lowers the estimates to 0.726 percent for educated males and 0.54 percent for educated females but does not alter the sign and statistical significance of the estimates. In Panel B of Table 2.6, I estimate equation (1.8) for a sample of workers with less than higher education. The estimates support the prediction of the theoretical model that the external return associated with an increase in college share is positive for the unskilled workers. Compared to Panel A, the estimates are higher due to the positive supply effect as indicated by the theoretical model.[18]

1.5.3 Instrumental Variable Estimates

The primary concern associated with estimating education externalities is that the share of educated individuals in the labor market is endogenous. To address the endogeneity concern, I estimate the model using the instrumental variable method. I use the number of college campuses in the region of residence in the past to instrument the contemporaneous share of college educated male and female workers. The results are reported in Table 2.7. The first stage results show that the number of college campuses has a strong and significant effect on the share of individuals with higher education. Compared to OLS estimates presented in Table 2.6, IV estimates are somewhat larger which suggests the presence of a downward bias in the OLS estimates due to the unobserved supply factors driving the variation in the share of educated individuals across labor markets. A one percentage point increase in the share of educated male population implies a 3.005 percent increase in male wages (Column 1). A similar increase in educated female population leads to a 3.478 percent increase in female wages. Controlling for the unobserved demand shocks using local unemployment and Katz-Murphy index lowers these effects to 2.045 percent for male workers and 3.647 percent for female workers (Column 5 and 6). Thus, the yearly increase in the share of male population with higher education of 0.76 percentage points observed in Russia between 2002 and 2010 implies an increase of 1.5 percent in male wages. An increase in the share of educated females of 1.06 percentage points per year observed in Russia between 2002 and 2010 implies a 3.86 percent increase in female wages. A higher increase in female wages during the period of study implies that education externalities contributed to the narrowing of the gender wage gap among educated workers in Russia.

---

[18]In Table A3, I estimate equation (1.8) by education level of the workers controlling for the share of male and female workers with higher education together. The estimates suggest that the share of educated male workers has an effect on both male and female workers but the effect on female workers becomes zero when I account for regional heterogeneity.

In columns 7 and 8, I estimate a model that controls for the unobserved heterogeneity across labor markets. The interpretation of the estimates in this specification changes as the main variation in the instrument comes from changes in the number of campuses in a given location over time. Therefore, the estimates of 7 percent for male workers and 6 percent for female workers are interpreted as a Local Average Treatment Effect (Imbens and Angrist, 1994).

Panel B of Table 2.7 reports the estimates for a sample of uneducated workers. The IV estimates of education externalities in a model that controls for the labor demand shocks suggest a 2.380 percent increase in male and a 3.491 percent increase in female wages in response to a one percentage point increase in the share of educated male and female workers, respectively. Given the annual growth of the share of educated male population by 0.76 percentage points and educated female population by 1.06 percentage points, wages of unskilled male workers increased by 1.8 percent and wages of unskilled female workers increased by 3.7 percent. Compared to the wage increase among educated workers due to the growth in the share of educated male and female population, the gender wage gap among unskilled individuals decreased at a slower pace.

1.5.4 Robustness Checks

Table 2.8 presents estimation results of the alternative specifications. In Column 1 and 2, I exclude the period of the transition to a market economy and estimate the model limiting the time period to 2000-2011. The estimates of education externality for both male and female workers increase compared to the baseline (Column 5 and 6 of Table 2.7), which implies that the labor markets become more responsive to changes in the educational composition of the labor force compared to the transition economy. In Column 3 and 4, I report the estimates when two largest cities, Moscow and St.Petersburg, are excluded from the sample. These cities differ from the remaining regions in terms of educational attainment of the labor force, the number of college campuses, and labor market characteristics. The estimates of education externalities decrease quite significantly highlighting the fact that externalities are probably higher in larger labor markets or markets with high level of educational attainment. In Column 5 and 6, I restrict the sample to individuas who are 25-34-year old as they represent the group with the largest gender education gap in favor of females. Finally, in Column 7 and 8, I estimate the model using the sample of individuals older than 35. Compared to the baseline estimate of of education externality of 2.045 percent, younger male workers experience a 2.425 percent increase in their wages when the share of educated male workers increases by one percent. On the other hand, the wages of older male workers would only increase by 1.613 percent in response to such a change. The estimates of education externalities for female workers are also significantly higher for younger females, 4.290 percent compared to 2.952 percent for females older than 35. Overal, Table 2.8 shows that educated female workers younger than 35 working in urban areas receive the largest gains from education externalities.

1.6 Conclusion

This chapter documents the reversal of the gender gap in tertiary education around the world and establishes a causal relationship between the changes in the share of educated male and female population and the gender wage gap through education externalities. In light of the growing female college participation, the analysis of this chapter becomes important for policy makers around the world because an increasing number of educated women will be joining the labor force in the nearest future. Empirical analysis relies on the case of Russia, one of the largest developing countries, which experienced a reversal of the gender gap in higher education over a decade ago. To account for the possibility of the region-wide labor demand shocks that increase wages in a region and attract more educated workers, I estimate these shocks with an index of demand shifts and by the instrumental variable technique. The results suggest that the growth in the share of educated male and female population similar to the one observed between 2002 and 2010 caused the wages of educated male and female workers to grow by 1.5 and 3.86 percent, respectively. Wages of unskilled male and female workers increased by 1.8 and 3.7 percent. Together these changes contributed to the narrowing of the gender wage gap over time

This chapter uncovers the impact of the growing female participation in higher education on the gender wage gap through one channel, namely, education externalities. Other channels can be important as well and evaluation of the impact of other channels as well as the improvements in female health, childbirth, and family formation represent important directions for future empirical research in this area.

Table 1.1: Population with Higher Education by Census Year

	Mean	Variance	P75-P25	P90-P10
Males with Higher Education				
1989	10.2	0.1	2.8	5.8
2002	13.7	0.1	2.3	5.7
2010	18.9	0.2	3.6	7.3
Females with Higher Education				
1989	9.8	0.07	2.1	4.9
2002	14.8	0.09	2.2	5.1
2010	22.7	0.1	3.7	7.7

Note: Sample includes 78 regions and 2 federal cities - Moscow and St. Petersburg.

Table 1.2: Number of Campuses of Higher Education Institutions, 1974-1991

	Men					Women				
	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max	N
1974	19.00	22.54	0	91	2,012	19.65	23.01	0	91	1,774
1975	17.88	21.17	0	92	1,849	18.70	21.84	0	92	1,658
1976	18.09	21.56	0	92	1,771	18.24	21.55	0	92	1,657
1978	17.43	19.79	0	93	1,792	17.35	19.38	0	93	1,835
1980	15.43	16.05	0	93	1,898	16.21	17.03	0	93	1,964
1981	21.43	24.63	0	93	2,101	21.75	24.55	0	93	2,213
1982	22.30	25.97	0	93	2,201	22.89	26.05	0	93	2,358
1983	22.26	25.65	0	93	2,263	21.78	24.95	0	93	2,406
1984	23.07	26.25	0	93	2,328	22.22	25.21	0	93	2,413
1985	22.71	25.83	0	93	2,270	21.27	24.45	0	93	2,302
1986	21.79	25.20	0	94	2,817	21.81	25.37	0	94	2,918
1987	21.45	25.07	0	95	2,869	20.93	24.35	0	95	2,901
1988	21.35	24.84	0	97	2,762	21.43	24.98	0	97	2,867
1989	21.38	24.98	0	96	2,704	20.81	24.30	0	96	2,833
1990	20.77	24.72	0	100	4,230	20.50	24.35	0	100	4,301
1991	22.19	28.40	1	117	4,254	22.53	28.74	1	117	4,260

Note: years 1977 and 1979 are not listed because RLMS was not conducted in 1997 and 1999.

Table 1.3: Summary Statistics, RLMS 1994-2011

	All education levels		Some college or more		Less than college	
	Men	Women	Men	Women	Men	Women
Individual-level variables						
Log hourly wage	3.17	2.93	3.59	3.44	3.03	2.67
	(1.43)	(1.40)	(1.36)	(1.30)	(1.42)	(1.39)
Education	12.37	13.12	15.83	15.97	11.22	11.70
	(2.78)	(2.71)	(2.10)	(1.96)	(1.88)	(1.75)
Experience	20.03	18.94	16.03	14.84	21.36	20.97
	(11.03)	(10.04)	(10.60)	(9.64)	(10.84)	(9.61)
Married	0.72	0.59	0.74	0.59	0.71	0.60
	(0.44)	(0.49)	(0.43)	(0.49)	(0.45)	(0.48)
Children younger than 7	0.37	0.32	0.37	0.32	0.38	0.32
	(0.60)	(0.55)	(0.58)	(0.54)	(0.61)	(0.55)
Region-level variables						
Share of college graduates	0.21	0.27	0.25	0.30	0.20	0.25
	(0.09)	(0.10)	(0.09)	(0.11)	(0.09)	(0.10)
Unemployment rate	7.71	7.73	7.18	7.29	7.89	7.96
	(3.37)	(3.40)	(3.48)	(3.48)	(3.31)	(3.33)
N	40,121	40,660	10,047	13,479	30,074	27,181

Table 1.4: Estimates of the Education Externalities by Gender

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
College share	2.819***	2.331***	2.462***	1.938***	2.240***	1.806***	1.705***	0.951***
	(0.108)	(0.119)	(0.097)	(0.080)	(0.112)	(0.088)	(0.105)	(0.077)
Educ*94	0.041***	0.067***	0.045***	0.068***	-	-	-	-
	(0.015)	(0.019)	(0.015)	(0.020)	-	-	-	-
Educ*95	0.045***	0.084***	0.046***	0.084***	-	-	-	-
	(0.007)	(0.008)	(0.006)	(0.008)	-	-	-	-
Educ*96	0.053***	0.082***	0.053***	0.082***	0.052***	0.081***	0.046***	0.072***
	(0.011)	(0.007)	(0.011)	(0.007)	(0.012)	(0.007)	(0.010)	(0.007)
Educ*98	0.062***	0.092***	0.061***	0.092***	0.065***	0.092***	0.054***	0.086***
	(0.008)	(0.013)	(0.008)	(0.013)	(0.008)	(0.013)	(0.006)	(0.012)
Educ*00	0.068***	0.104***	0.070***	0.106***	0.071***	0.107***	0.063***	0.101***
	(0.009)	(0.011)	(0.009)	(0.011)	(0.022)	(0.011)	(0.007)	(0.009)
Educ*01	0.047***	0.090***	0.046***	0.088***	0.048***	0.090***	0.045***	0.084***
	(0.009)	(0.011)	(0.009)	(0.011)	(0.009)	(0.012)	(0.006)	(0.012)
Educ*02	0.046***	0.081***	0.046***	0.081***	0.046***	0.081***	0.044***	0.081***
	(0.008)	(0.010)	(0.009)	(0.010)	(0.009)	(0.011)	(0.006)	(0.011)
Educ*03	0.050***	0.097***	0.049***	0.097***	0.050***	0.097***	0.047***	0.093***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.005)	(0.006)
Educ*04	0.049***	0.101***	0.050***	0.101***	0.052***	0.101***	0.046***	0.096***
	(0.008)	(0.007)	(0.008)	(0.007)	(0.009)	(0.007)	(0.007)	(0.007)
Educ*05	0.047***	0.087***	0.048***	0.085***	0.050***	0.086***	0.045***	0.081***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.006)	(0.008)
Educ*06	0.052***	0.087***	0.052***	0.087***	0.048***	0.084***	0.049***	0.078***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.006)	(0.006)
Educ*07	0.050***	0.084***	0.051***	0.085***	0.044***	0.083***	0.043***	0.080***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.005)	(0.007)
Educ*08	0.051***	0.070***	0.052***	0.071***	0.051***	0.071***	0.048***	0.065***
	(0.006)	(0.008)	(0.005)	(0.008)	(0.005)	(0.008)	(0.005)	(0.007)

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Educ*09	0.042***	0.074***	0.042***	0.075***	0.043***	0.074***	0.041***	0.069***
	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	(0.005)
Educ*10	0.049***	0.067***	0.050***	0.067***	0.050***	0.066***	0.046***	0.063***
	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Educ*11	0.050***	0.072***	0.052***	0.073***	0.053***	0.074***	0.048***	0.070***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)
Exp*94	-0.005***	0.006**	-0.005***	0.005**	-	-	-	-
	(0.002)	(0.002)	(0.001)	(0.002)	-	-	-	-
Exp*95	-0.001	0.008***	-0.001	0.007***	-	-	-	-
	(0.001)	(0.002)	(0.001)	(0.002)	-	-	-	-
Exp*96	-0.004	0.009***	-0.004	0.008***	-0.003	0.008***	-0.003	0.006***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Exp*98	-0.001	0.010***	-0.001	0.009***	-0.001	0.009***	0.000	0.008***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Exp*00	-0.002	0.013***	-0.002	0.012***	-0.002	0.012***	-0.001	0.012***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Exp*01	-0.006***	0.009***	-0.006***	0.008***	-0.005***	0.008***	-0.004***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*02	-0.008***	0.007***	-0.007***	0.008***	-0.007***	0.008***	-0.005***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*03	-0.009***	0.007***	-0.008***	0.007***	-0.008***	0.007***	-0.006***	0.008***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*04	-0.008***	0.008***	-0.007***	0.008***	-0.007***	0.008***	-0.005***	0.008***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Exp*05	-0.008***	0.007***	-0.007***	0.006***	-0.007***	0.006***	-0.005***	0.006***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Exp*06	-0.006***	0.007***	-0.005***	0.006***	-0.005***	0.006***	-0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*07	-0.004***	0.009***	-0.004***	0.008***	-0.005***	0.007***	-0.004***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*08	-0.003***	0.009***	-0.003***	0.008***	-0.003***	0.008***	-0.002**	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exp*09	-0.001	0.012***	-0.001	0.010***	-0.001	0.010***	-0.001	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*10	-0.001	0.012***	-0.001	0.011***	-0.001	0.010***	-0.000	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exp*11	-0.001	0.011***	-0.001	0.010***	-0.001	0.010***	-0.000	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cons	-0.510***	-1.217***	-0.265	-0.887***	0.370*	-0.095	0.153	-0.424***
	(0.172)	(0.190)	(0.180)	(0.205)	(0.219)	(0.165)	(0.171)	(0.136)
R^2	0.78	0.78	0.78	0.78	0.72	0.72	0.76	0.77
N	40,121	40,660	40,121	40,660	35,831	36,889	35,831	36,889
Unemployment			Yes	Yes	Yes	Yes	Yes	Yes
Katz-Murphy index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

Notes: Dependent variable is log hourly wage. All specifications include year fixed effect. Standard errors, clustered by region and year, are reported in parentheses. The inverse propensity weight from the probit model of non-missing wages is applied in all models. Asterisks denote significance levels: 1 percent (***), 5 percent (**), and 10 percent (*).

Table 1.5: Cross-Sectional Estimates of External Returns by Gender

	1994				1995			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	3.250***	3.072***	3.260***	2.156**	3.352***	2.821***	3.604***	1.431*
	(0.452)	(0.373)	(0.737)	(0.881)	(0.466)	(0.457)	(0.733)	(0.805)
R^2	0.11	0.08	0.27	0.22	0.14	0.13	0.30	0.27
N	2,012	1,774	2,012	1,774	1,849	1,658	1,849	1,658
	1996				1998			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	2.413***	2.885***	2.231**	2.247***	3.233***	2.612***	3.502***	1.618***
	(0.531)	(0.470)	(0.832)	(0.607)	(0.561)	(0.441)	(0.600)	(0.505)
R^2	0.14	0.14	0.33	0.31	0.18	0.17	0.43	0.34
N	1,771	1,657	1,771	1,657	1,792	1,835	1,792	1,835
	2000				2001			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	3.974***	3.185***	3.134***	2.049***	3.789***	3.303***	2.944***	1.850***
	(0.714)	(0.406)	(0.557)	(0.568)	(0.583)	(0.451)	(0.418)	(0.475)
R^2	0.19	0.18	0.40	0.34	0.19	0.21	0.40	0.38
N	1,898	1,964	1,898	1,964	2,101	2,213	2,101	2,213
	2002				2003			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	3.533***	2.727***	2.311***	0.731*	3.620***	3.071***	2.014***	0.993**
	(0.558)	(0.443)	(0.469)	(0.404)	(0.470)	(0.475)	(0.617)	(0.382)
R^2	0.27	0.19	0.45	0.35	0.24	0.24	0.38	0.36
N	2,201	2,358	2,201	2,358	2,263	2,406	2,263	2,406

	2004				2005			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	3.509***	3.008***	2.183***	1.018**	3.405***	2.880***	2.342***	1.213**
	(0.477)	(0.410)	(0.629)	(0.400)	(0.427)	(0.404)	(0.473)	(0.539)
R^2	0.30	0.26	0.41	0.37	0.24	0.23	0.39	0.35
N	2,328	2,413	2,328	2,413	2,270	2,302	2,270	2,302
	2006				2007			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	2.949***	2.791***	1.882***	1.166***	2.715***	2.615***	1.770***	1.250***
	(0.426)	(0.477)	(0.448)	(0.284)	(0.331)	(0.479)	(0.362)	(0.316)
R^2	0.24	0.23	0.37	0.35	0.24	0.25	0.37	0.38
N	2,817	2,918	2,817	2,918	2,869	2,901	2,869	2,901
	2008				2009			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	2.635***	2.477***	1.355***	0.898**	2.335***	1.751***	0.984***	0.436**
	(0.382)	(0.468)	(0.382)	(0.332)	(0.363)	(0.413)	(0.272)	(0.169)
R^2	0.27	0.22	0.40	0.37	0.24	0.21	0.38	0.37
N	2,762	2,867	2,762	2,867	2,704	2,833	2,704	2,833
	2010				2011			
	OLS		FE		OLS		FE	
	Men	Women	Men	Women	Men	Women	Men	Women
College share	2.101***	1.498***	0.803***	0.239	1.909***	1.557***	0.709***	0.452***
	(0.364)	(0.412)	(0.199)	(0.166)	(0.376)	(0.421)	(0.223)	(0.139)
R^2	0.24	0.19	0.37	0.37	0.23	0.19	0.37	0.36
N	4,230	4,301	4,230	4,301	4,254	4,260	4,254	4,260

Notes: Dependent variable is log hourly wage. Standard errors, clustered by region, are reported in parentheses. The inverse propensity weight from the probit model of non-missing wages is applied in all models. Asterisks denote significance levels: 1 percent (***), 5 percent (**), and 10 percent (*).

Table 1.6: Estimates of Education Externalities by Education Group

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Some college or more				
College share	2.255***	1.800***	1.639***	1.349***	1.382***	1.186***	0.726***	0.540***
	(0.122)	(0.158)	(0.110)	(0.104)	(0.132)	(0.125)	(0.145)	(0.095)
Education	0.025***	0.016***	0.025***	0.015***	0.022***	0.015***	0.014***	0.008**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Experience	0.018***	0.024***	0.021***	0.027***	0.022***	0.028***	0.028***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Experience	-0.054***	-0.039***	-0.060***	-0.051***	-0.065***	-0.057***	-0.076***	-0.073***
squared/100	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.006)	(0.006)
Constant	-0.252***	-0.543***	0.172***	-0.157	1.444***	1.031***	1.003***	0.855***
	(0.079)	(0.086)	(0.095)	(0.109)	(0.122)	(0.121)	(0.117)	(0.094)
R^2	0.78	0.76	0.79	0.76	0.71	0.69	0.75	0.74
N	10,047	13,479	10,047	13,479	9,126	12,511	9,126	12,511
				B. Less than college				
College share	2.903***	2.450***	2.563***	2.049***	2.343***	1.934***	1.569***	0.892***
	(0.117)	(0.119)	(0.112)	(0.086)	(0.128)	(0.093)	(0.126)	(0.098)
Education	0.042***	0.060***	0.042***	0.061***	0.046***	0.060***	0.035***	0.058***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Experience	0.002	0.008***	0.004*	0.010***	0.006***	0.010***	0.009***	0.019***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Experience	-0.019***	0.001	-0.022***	-0.006	-0.027***	-0.006	-0.031***	-0.033***
squared/100	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
Constant	-0.634***	-1.372***	-0.374***	-1.011***	2.504***	0.217*	2.168***	2.334***
	(0.084)	(0.079)	(0.104)	(0.100)	(0.106)	(0.115)	(0.083)	(0.074)
R^2	0.78	0.78	0.78	0.78	0.72	0.72	0.78	0.77
N	30,074	27,181	30,074	27,181	26,705	24,378	26,705	24,378
Unemployment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

Notes: All specifications include year fixed effect. Standard errors, clustered by region and year, are reported in parentheses. The inverse propensity weight from the probit model of non-missing wages is applied in all models. Asterisks denote significance levels: 1 percent (***), 5 percent (**), and 10 percent (*).

Table 1.7: Instrumental Variable Estimates by Gender and Education Group

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Some college or more								
First stage								
Number of	0.002***	0.002***	0.002***	0.001***	0.002***	0.001***	0.001***	0.001***
campuses	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F test	978.47	1207.06	942.89	1163.26	917.40	1158.22	737.95	735.05
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	3.005***	3.478***	2.306***	3.490***	2.045***	3.647***	7.426***	6.338***
	(0.097)	(0.093)	(0.142)	(0.138)	(0.173)	(0.164)	(2.817)	(2.253)
Education	0.016***	0.011***	0.018***	0.011***	0.016***	0.011***	-0.011	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)	(0.004)
Experience	0.015***	0.016***	0.018***	0.016***	0.019***	0.013***	-0.003	-0.007
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.012)	(0.015)
Experience	-0.046***	-0.002	-0.052***	-0.002	-0.057***	0.008	0.003	0.093
squared/100	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.032)	(0.064)
N	10,047	13,479	10,047	13,479	9,126	12,511	9,126	12,511
Unemployment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				B. Less than college				
First stage								
Number of	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***	0.002***	0.001***
campuses	(0.0001)	(0.0000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F test	2026.59	2094.40	1950.04	2022.70	1904.66	1925.15	2195.73	1333.05
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	3.388***	3.842***	2.733***	3.276***	2.380***	3.491***	-1.050	0.695
	(0.059)	(0.069)	(0.080)	(0.089)	(0.091)	(0.108)	(0.679)	(1.203)
Education	0.034***	0.056***	0.038***	0.057***	0.040***	0.055***	0.044***	0.058***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Experience	0.001	0.000	0.003	0.003*	0.006***	0.001	0.023***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.007)
Experience	-0.014***	0.027***	-0.021***	0.016***	-0.025***	0.025***	-0.058***	-0.033
squared/100	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.006)	(0.024)
N	30,074	27,181	30,074	27,181	26,705	24,378	26,705	24,378
Unempoyment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

Notes: All specifications include year fixed effect. The inverse propensity weight from the probit model of non-missing wages is applied in all models. Robust standard errors are reported in parentheses. Asterisks denote significance levels: 1 percent (***), 5 percent (**), and 10 percent (*).

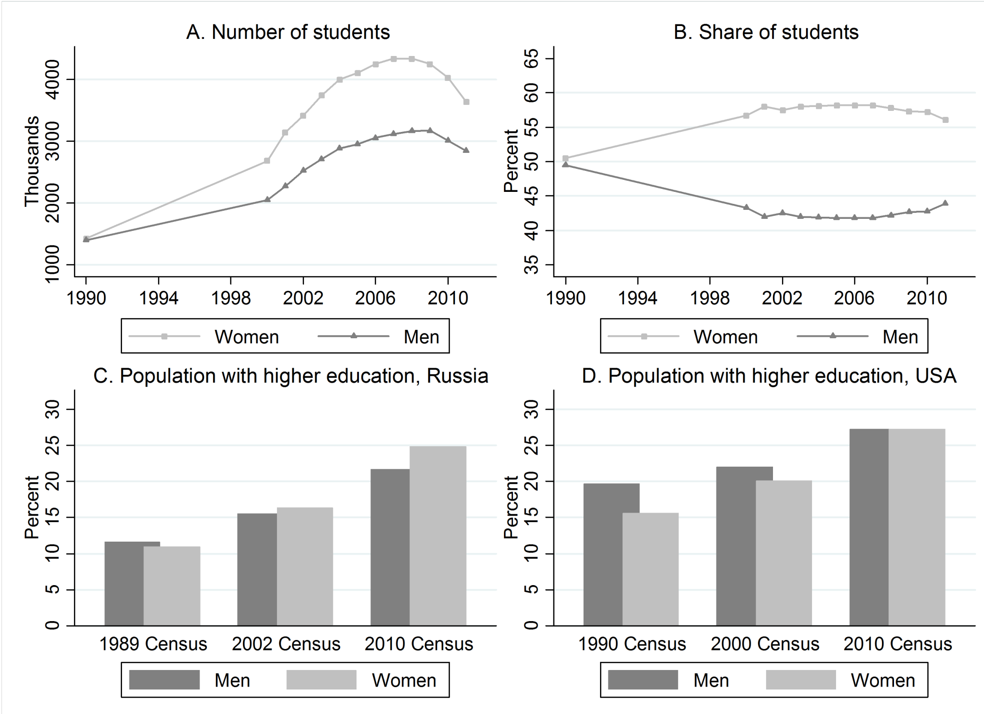
Table 1.8: Robustness Checks for the Baseline IV Specification

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Some college or more								
First stage								
Number of	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***	0.002***	0.002***
campuses	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F test	924.10	1168.82	135.50	351.45	374.34	484.05	1300.10	2513.49
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	2.132***	3.768***	1.066***	1.950***	2.425***	4.290***	1.613***	2.952***
	(0.182)	(0.168)	(0.331)	(0.321)	(0.306)	(0.333)	(0.208)	(0.173)
Education	0.018***	0.013***	0.017***	0.016***	0.015**	0.013*	0.018***	0.025***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.007)	(0.007)	(0.005)	(.004)
Experience	0.020***	0.011***	0.025***	0.027***	0.014	0.017	0.036***	0.014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.013)	(0.013)	(0.008)	(0.008)
Experience	-0.058***	0.016*	-0.071***	-0.043***	-0.048	0.042	-0.094***	-0.009
squared/100	(0.006)	(0.008)	(0.007)	(0.010)	(0.088)	(0.093)	(0.017)	(0.019)
N	8,326	11,599	7,481	10,570	3,316	4,373	5,031	6,894
Unemployment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KM index	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

B. Less than college								
First stage								
Number of	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***	0.002***	0.001***
campuses	(0.0001)	(0.0000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F test	2026.59	2094.40	1950.04	2022.70	1904.66	1925.15	2195.73	1333.05
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	3.388***	3.842***	2.733***	3.276***	2.380***	3.491***	-1.050	0.695
	(0.059)	(0.069)	(0.080)	(0.089)	(0.091)	(0.108)	(0.679)	(1.203)
Education	0.034***	0.056***	0.038***	0.057***	0.040***	0.055***	0.044***	0.058***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Experience	0.001	0.000	0.003	0.003*	0.006***	0.001	0.023***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.007)
Experience	-0.014***	0.027***	-0.021***	0.016***	-0.025***	0.025***	-0.058***	-0.033
squared/100	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.006)	(0.024)
N	30,074	27,181	30,074	27,181	26,705	24,378	26,705	24,378
Unempoyment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KM index	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

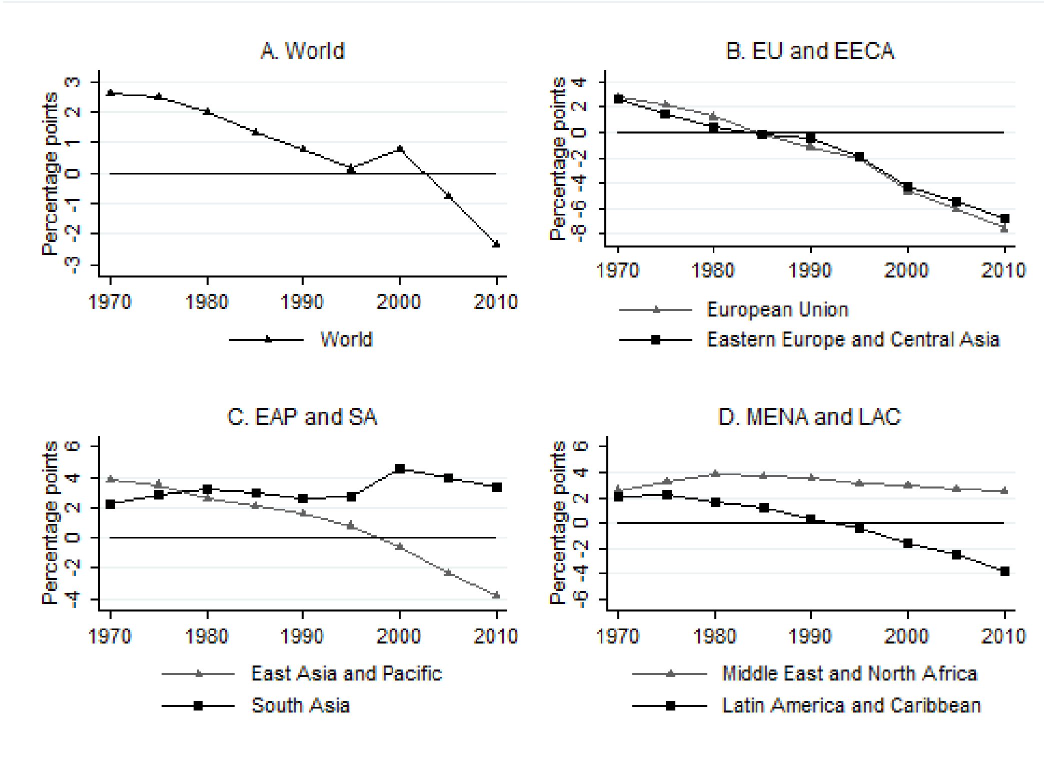
Notes: All specifications include year fixed effect. The inverse propensity weight from the probit model of non-missing wages is applied in all models. Robust standard errors are reported in parentheses. Asterisks denote significance levels: 1 percent (***), 5 percent (**), and 10 percent (*).

Figure 1.1: Female College Participation, Population with Higher Education



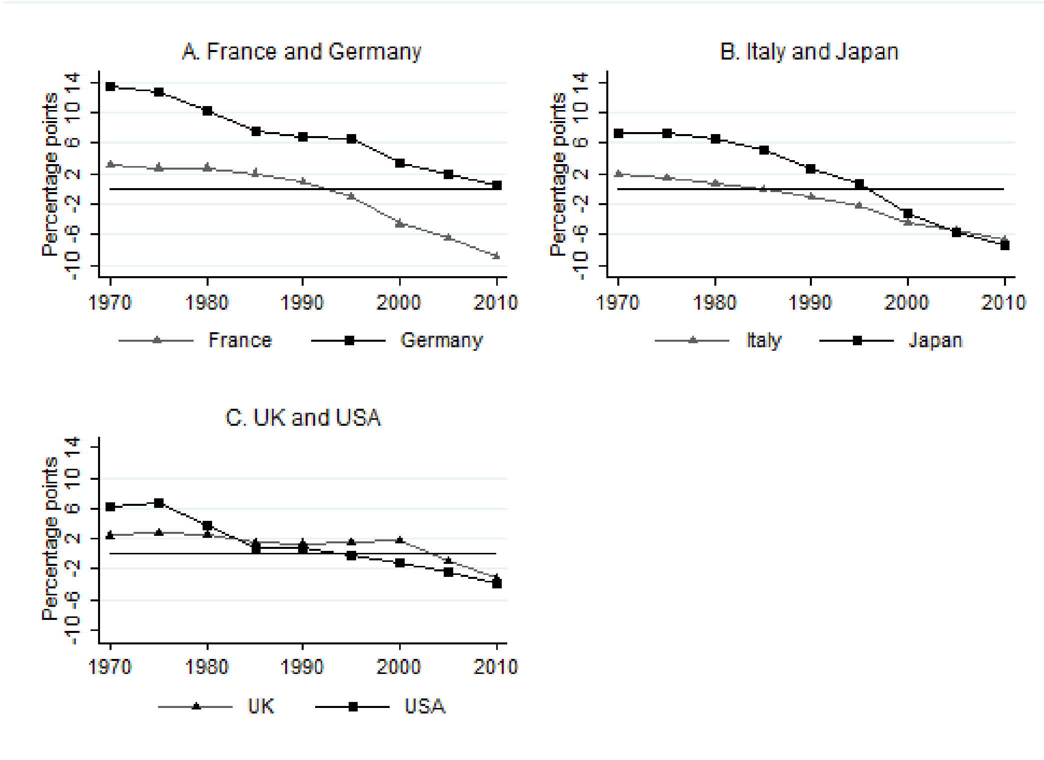
Notes: Panel A depicts the total number of male and female students enrolled in public and private higher education institutions in Russia in a given year. Panel B shows the share of male and female students in the total number of students enrolled in public and private higher education institutions in Russia. Panel C shows the share of Russian male and female population with complete higher education or higher level of education (Postgraduate degree; individuals are 15 years old and over). Panel D shows the share of the U.S. male and female population with a Bachelor's degree or higher (Master's, Professional, or Doctoral degree; individuals are 15 years old and over).

Figure 1.2: Gender Gap in Tertiary Education: World Regions



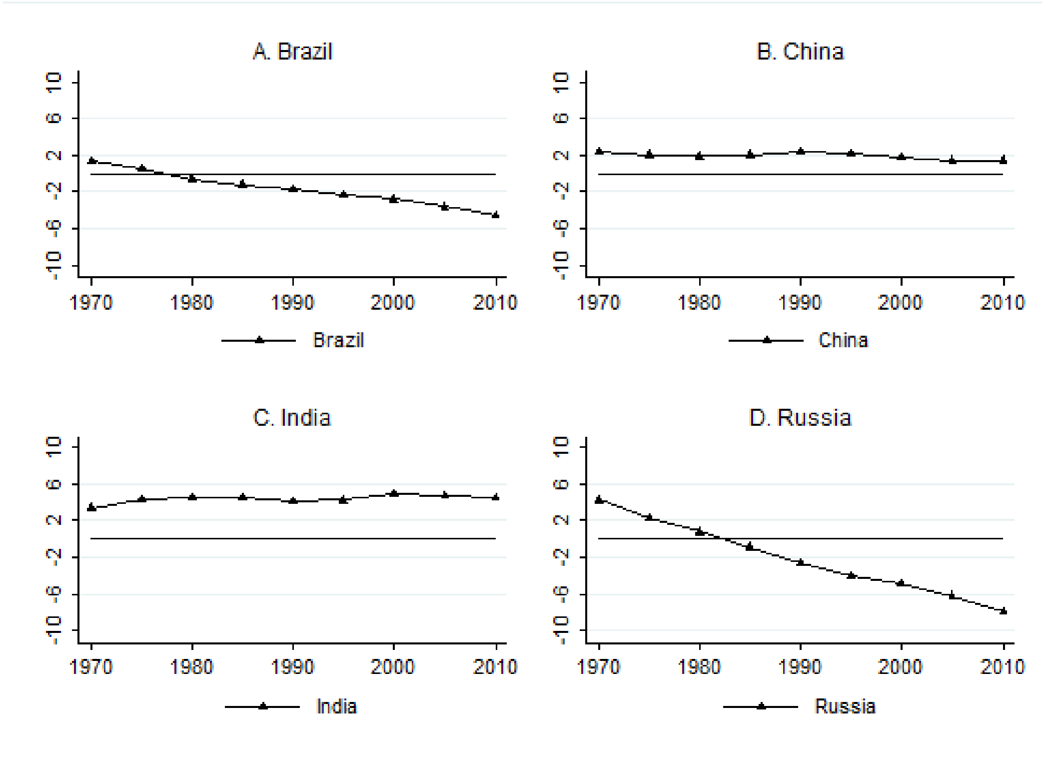
Notes: This figure displays the gender gap in tertiary education (ISCED 5 and 6) for 25-34-year old individuals across the world regions. The gender gap in tertiary education is the difference between the share of men with complete tertiary education and the share of women with complete tertiary education. Panel A shows the world's average trend in the gender gap in tertiary education from 1970 to 2010. Panel B shows the evolution of the gender gap in European Union and Eastern and Central Asia. Panel C shows the evolution of the gender gap in East Asia and Pacific and South Asia. Panel D shows the evolution of the gender gap in Middle East and North Africa and Latin America and Caribbean. Tertiary education includes both higher education institutions such as universities and post-secondary schools such as community colleges in U.S. The sample includes 146 countries. Data are taken from the World Bank Education Statistics, which combine data collected by Barro and Lee (2013) and IIASA/VID (International Institute for Applied Systems Analysis/Vienna Institute of Demography). Barro and Lee (2013) data set provides educational attainment data for 173 countries in 5-year intervals from 1970 to 1995 and 1-year intervals from 1995 to 2010.

Figure 1.3: Gender Gap in Tertiary Education: Developed Countries



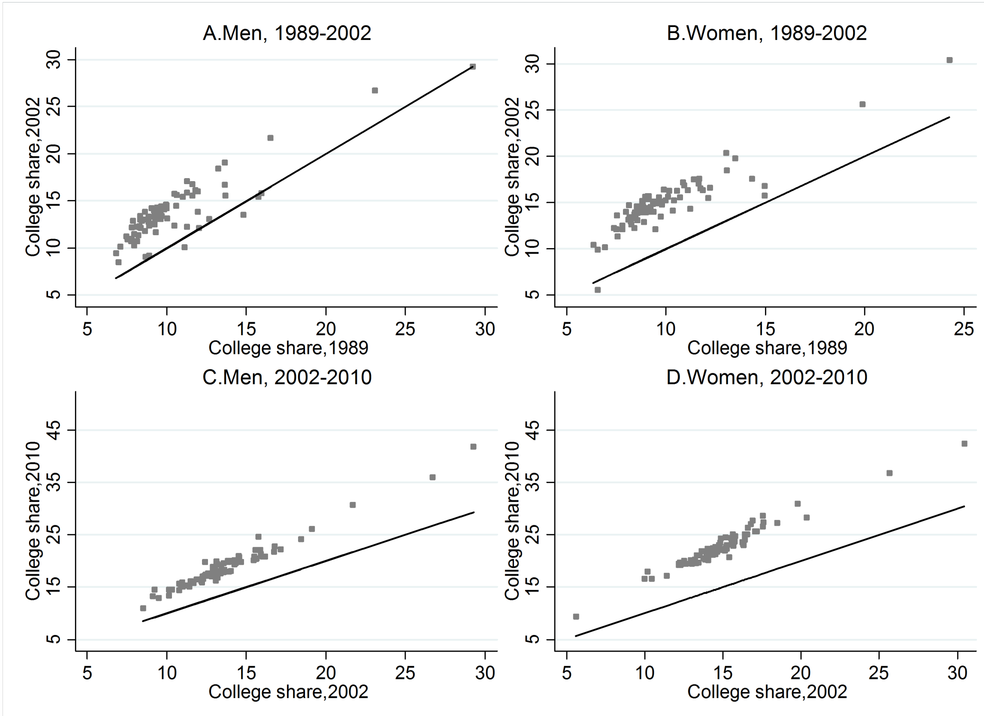
Notes: This figure displays the gender gap in tertiary education (ISCED 5 and 6) for 25-34-year old individuals in selected developed countries. The gender gap in tertiary education is the difference between the share of men with complete tertiary education and the share of women with complete tertiary education. Panel A shows the evolution of the gender gap in France and Germany. Panel B shows the evolution of the gender gap in Italy and Japan. Panel C shows the evolution of the gender gap in the UK and US. Tertiary education includes both higher education institutions such as universities and post-secondary schools such as community colleges in U.S.

Figure 1.4: Gender Gap in Tertiary Education: Developing Countries



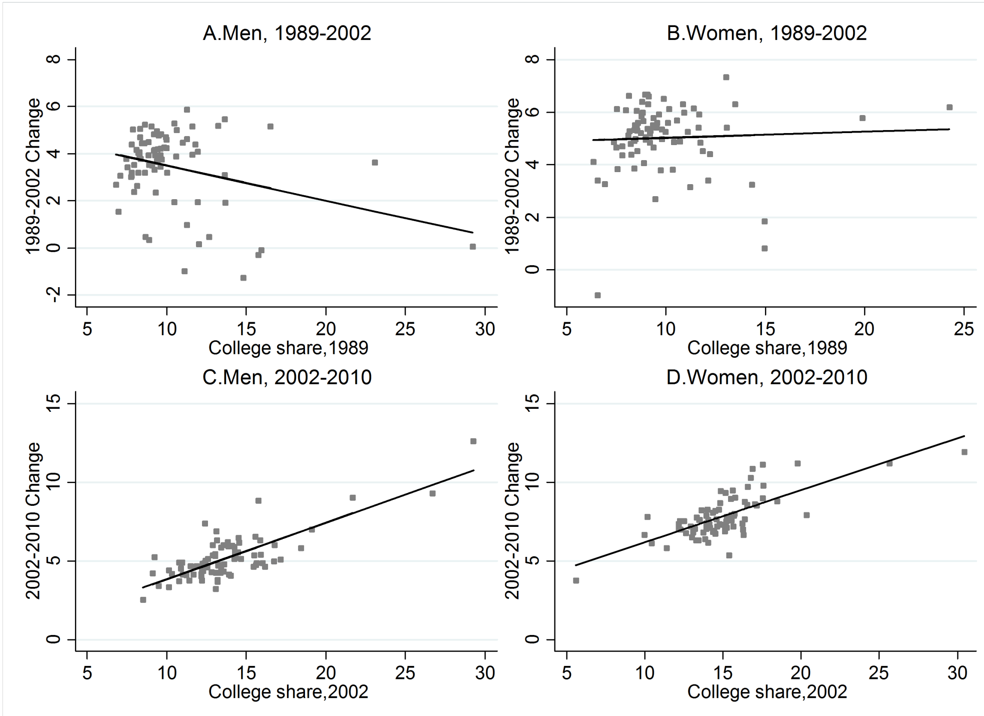
Notes: This figure displays the gender gap in tertiary education (ISCED 5 and 6) for 25-34-year old individuals in four largest developing countries, Brazil, China, India, and Russia (BRIC). The gender gap in tertiary education is the difference between the share of men with complete tertiary education and the share of women with complete tertiary education. Tertiary education includes both higher education institutions such as universities and post-secondary schools such as secondary professional schools (technicums) in Russia.

Figure 1.5: Changes in the College Share of Male and Female Population



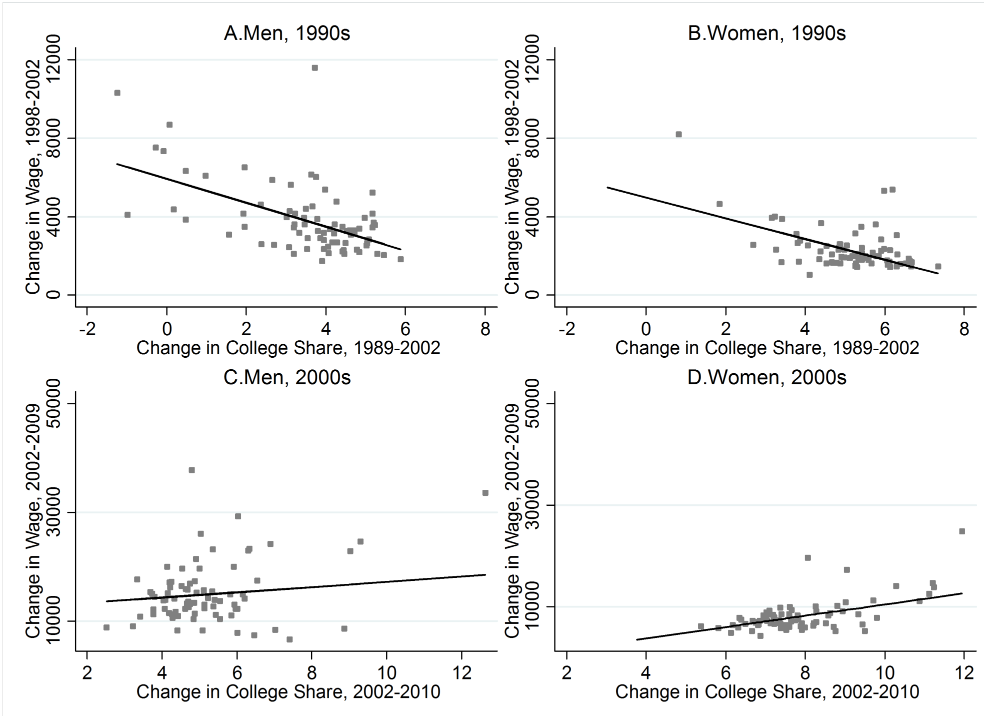
Notes: Panel A plots the relationship between the share of men with higher education in 1989 (on the X-axis) and the share of men with higher education in 2002 (on the Y-axis). Also plotted is the 45-degree line. Panel B plots the relationship between the share of women with higher education in 1989 (on the X-axis) and the share of women with higher education in 2002 (on the Y-axis). Also plotted is the 45-degree line. Panel C plots the relationship between the share of men with higher education in 2002 (on the X-axis) and the share of men with higher education in 2010 (on the Y-axis). Also plotted is the 45-degree line. Panel D plots the relationship between the share of women with higher education in 2002 (on the X-axis) and the share of women with higher education in 2010 (on the Y-axis). Also plotted is the 45-degree line. The share with higher education is calculated as population over 15 with complete higher education (including postsecondary education) over the total population over 15. Sample includes 78 regions and 2 federal cities, Moscow and St. Petersburg.

Figure 1.6: Changes in the College Share of Male and Female Population



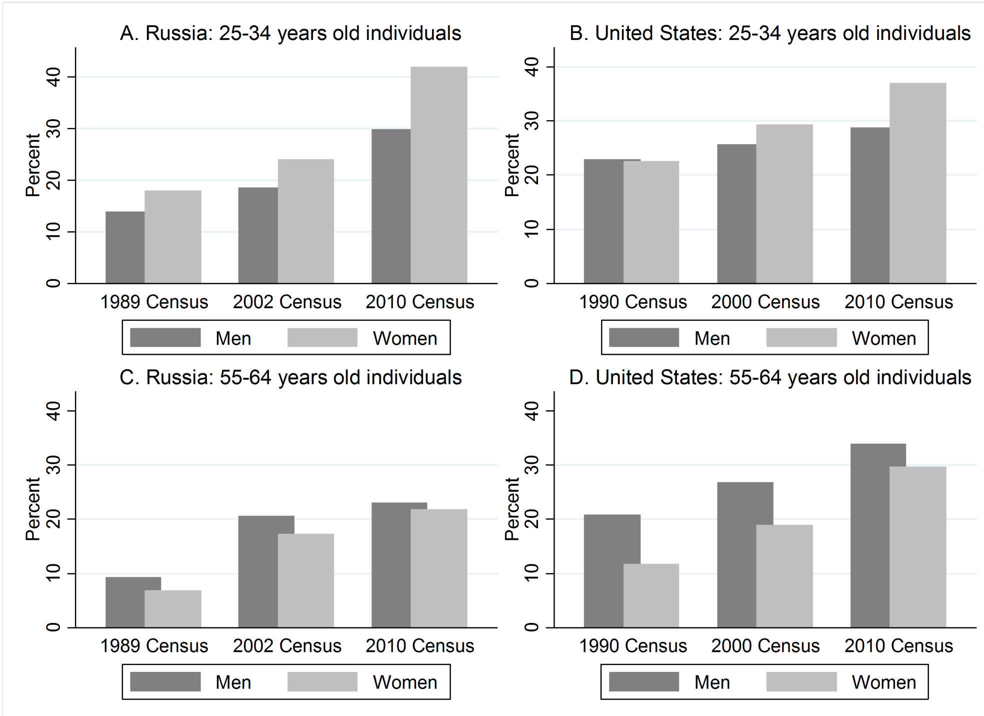
Notes: Panel A plots the relationship between the share of men with higher education in 1989 (on the X-axis) and the change in the share of men with higher education between 1989 and 2002 (on the Y-axis). The slope (standard error) of the OLS fitted line is -0.148(0.051). Panel B plots the relationship between the share of women with higher education in 1989 (on the X-axis) and the change in the share of women with higher education between 1989 and 2002 (on the Y-axis). The slope (standard error) of the OLS fitted line is 0.022(0.054). Panel C plots the relationship between the share of men with higher education in 2002 (on the X-axis) and the change in the share of men with higher education between 2002 and 2010 (on the Y-axis). The slope (standard error) of the OLS fitted line is 0.359(0.032). Panel D plots the relationship between the share of women with higher education in 2002 (on the X-axis) and the change in the share of women with higher education between 2002 and 2010 (on the Y-axis). The slope (standard error) of the OLS fitted line is 0.331(0.034). The share with higher education is calculated as population over 15 with complete higher education (including postsecondary education) over the total population over 15. Sample includes 78 regions and 2 federal cities, Moscow and St. Petersburg.

Figure 1.7: Changes in the Male and Female College Share and Wages



Notes: Panel A plots the relationship between the change in the share of men with higher education between 1989 and 2002 (on the X-axis) and the change in male average wages between 1998 and 2002 (on the Y-axis). The slope (standard error) of the OLS fitted line is -656(109). Panel B plots the relationship between the change in the share of women with higher education between 1989 and 2002 (on the X-axis) and the change in male average wages between 1998 and 2002 (on the Y-axis). The slope (standard error) of the OLS fitted line is -475(95). Panel C plots the relationship between the change in the share of men with higher education between 2002 and 2010 (on the X-axis) and the change in female average wages between 2002 and 2009 (on the Y-axis). The slope (standard error) of the OLS fitted line is 1321(410). Panel D plots the relationship between the change in the share of women with higher education between 2002 and 2010 (on the X-axis) and the change in female average wages between 2002 and 2009 (on the Y-axis). The slope (standard error) of the OLS fitted line is 1562(223). Nominal wages are measured in rubles. Sample includes 76 regions and 2 federal cities Moscow and St. Petersburg. Wage data for Chechen republic and Chelyabinsk region is not available.

Figure 1.8: Russia/U.S. Comparison of Male/Female College Share by Cohort



Notes: Panel A depicts the share of Russian 25-34 years old male and female population with complete higher education or a Postgraduate degree. Panel B shows the share of the US 25-34 years old male and female population with a Bachelor's degree or higher (Master's, Professional, or Doctoral degree). Panel C shows the share of Russian 55-64 years old male and female population with complete higher education or a Postgraduate degree. Panel D shows the share of the US 55-64 years old male and female population with a Bachelor's degree or higher (Master's, Professional, or Doctoral degree).

CHAPTER 2

COLLEGE EXPANSION AND THE MARGINAL RETURNS TO EDUCATION: EVIDENCE FROM RUSSIA (WITH KLARA S. PETER AND CHRISTIAN M. POSSO)

2.1 Introduction

Over the last 20 years, many countries have expanded higher education and increased the number of college graduates. Between 1990 and 2010, the number of students in tertiary education per 100,000 people has tripled in Brazil, India and Russia, and increased twelve-fold in China; the number of colleges has also increased several-fold in the largest developing countries (Carnoy and Wang (2013)). Answering pertinent policy question whether such rapid and massive college expansion is economically worthwhile requires proper analytical tools. To evaluate the effectiveness of college expansion policies, analysts need to estimate the returns to college for the marginal individual who previously would not have had an opportunity to study in college and who is affected by college expansion.

We evaluate the marginal returns to college in response to college openings using the case of Russia.[1] Between 1992 and 2003, the number of students in Russian higher education almost tripled, the number of universities increased two-fold, and the number of municipalities with at least one campus more than doubled.[2] The legal framework for the Russian college expansion was provided by the 1992 Law on Education, which allowed the opening of private universities and

[1] We provide a brief summary of the Soviet and Russian systems of higher education in the web appendix.

[2] Authors calculations.

legalized tuition-based programs and branches in public universities (Law (1992)).[3] The early expansion period was accompanied by a rapid increase in the average college wage premium, as can be seen in Figure 2.1. Following the low-return Soviet period with its compressed centralized wage structure, the OLS-estimated average wage premium per year of college education surged from 2.8 percent in 1990 to about 8.7 percent in 1998 during Russias transition to a market economy, but then the premium stabilized.

Our empirical strategy relies on the marginal treatment effect (MTE) method in its normal and semi-parametric versions (Heckman and Vytlacil (2001), Heckman and Vytlacil (2005), Heckman and Vytlacil (2007)) and a construction of policy treatment parameters (Heckman and Vytlacil (2001); Carneiro and Vytlacil (2010)). Unlike the previous MTE studies that are based on either cross-sectional data or one-cohort panels such as the National Longitudinal Survey of Youth (NLSY), we study multiple cohorts, including ones that made their college decision in the pre-expansion period (1985-1992) as well as cohorts that entered college during the expansion period (1993-2003). Thus, we can evaluate the outcomes of actual, real-world college expansion rather than a hypothetical increase in the probability of college attendance.

The identification of MTE relies on the variation in variables directly affecting the college decision. These variables capture the institutional environment and economic conditions that prevail during individuals late teenage years, the time when individuals make their college decisions.[4] Our key instrument, the number of campuses in the municipality of residence at age 17, measures the extent of college expansion in Russia over time and across locations. This instrument is an improvement over a commonly used binary instrumental variable (IV) - the presence of a college in the U.S. county of residence during teenage years (Card (1995); Cameron and Taber (2004)), as it allows for computing policy-relevant treatment effects separately for the establishment of the first

[3]Russian institutions of higher education are referred to by a variety of names, including Universities, Institutes, Academies, and Higher Schools. For compatibility of terminology with international literature on this topic, this chapter uses the terms Universities and Colleges, even though the latter term is not technically accurate from the Russian language point of view. Thus, such terms as College, University, and Institution of higher education are used interchangeably throughout this chapter.

[4]See Card (2001) for a review of the variables (mainly cost shifters) that affect the schooling decision.

campus in the locality that never had a college as well as for the marginal change in the number of campuses. Since the instrument is numeric, its polynomial functional form is utilized to capture the strong non-linear effects of college availability on college attainment and on the marginal returns. We also provide several arguments and statistical tests to offset potential concerns with the validity of this instrument.

Consistent with the MTE literature, we find that the returns to college are heterogeneous, and that individuals with the highest propensity to go to college enjoy the largest marginal gains from college education. We also find strong evidence of the positive sorting of individuals into college based on unobserved gains. The magnitude of the sorting gain in Russia a 5.7 percent wage increase per year of college falls between the corresponding estimate of 2 percent for China (Heckman and Li (2004)) and 7.6 percent for the U.S. (Carneiro and Vytlacil (2011)).[5]

In addition to conventional treatment parameters, we estimate the returns to college for the individuals who change college participation in response to a marginal increase in the number of campuses. The estimated marginal policy-relevant treatment effect parameter of 9.6 percent wage increase per year of college indicates large gains for the marginal individuals affected by the establishment of new campuses. Our policy simulations show that the opening of a college campus in constrained municipalities - smaller non-capital cities or municipalities that did not have institutions of higher education prior to college expansion - attracts students with higher returns compared to the effect of the same policy in unconstrained municipalities with more college choices. In another policy simulation, we show that if the number of campuses per municipality did not increase or remained at the 1992 level, then a considerable portion of population with high potential gains from college education would not have been able to realize these gains. We also find larger discounted net benefits for students from constrained municipalities by applying the estimated returns in the traditional cost-benefit analysis.

[5]Our results support a comparative advantage model of the labor market rather than a single ability model, which is consistent with the past literature on self-selection that goes back to Roy (1951). A large number of earlier studies established empirical evidence of the non-random sorting of individuals into different levels of education and called for the selectivity correction in the returns to schooling (e.g., Willis and Rosen (1979), Garen (1984), Bjorklund and Moffitt (1987)).

This chapter contributes to several strands of the literature. First, we contribute to the literature on the marginal treatment effect of college education.[6] Most of this literature uses the National Longitudinal Survey of Youth (NLSY) and considers changes in tuition as a policy affecting the college decisions (Carneiro and Vytlacil (2010), Carneiro and Vytlacil (2011); Carneiro and Lee (2009); Heckman and Urzua (2014); Heckman and Vytlacil (2001), Heckman and Vytlacil (2005)). MTE estimates outside the U.S. are rare and often limited to a single cross section (Carneiro and Umapathi (2011)) for Indonesia; Heckman and Li (2004) for China; Kyui (2013) for Russia; Zamarro (2010) for Spain).[7] College expansion in the MTE literature is either not modelled at all or simulated as an additive exogenous change in the probability of college attendance. The key contribution of our study is the estimation of the expansion-related marginal treatment effects. To the best of our knowledge, this is the first study that utilizes data on actual college openings to derive policy-relevant treatment parameters and evaluate counterfactual scenarios of college expansion. We also show that the returns to college for marginal students vary considerably depending on the scale of college expansion and the location of new campuses. Additionally, we extend the MTE literature by using higher quality panel data with multiple wage observations, several birth cohorts, disaggregated location information, and past economic conditions. As shown by Heckman and Todd (2006), cohort-based models fitted on repeated cross sections provide more reliable estimates of the returns to education than the estimates obtained from a single cross section.

Second, our study is directly related to the literature that evaluates the effect of college expansion on the returns to college. A number of earlier papers (Katz and Murphy (1992), Topel (1997), among others) linked the increased supply of college graduates with a lower average college wage

---

[6]Outside the field of education, the MTE method is only beginning to be applied in a systematic way; e.g., see Basu and Urzua (2007) and Evans and Basu (2011) for the MTE applications in health related topics. Moffitt (2014) uses the MTE method to estimate the effect of a transfer program on labor supply.

[7]Carneiro and Umapathi (2011) computes the MTE of post-secondary education in Indonesia in 2000 using the distance from the current village center to the nearest secondary school as an identifying restriction. Heckman and Li (2004) estimate the MTE of college degree for a cross-sectional sample of young Chinese workers in 2000, using limited exclusion restrictions such as parental income and parental education. Kyui (2013) estimates standard college-related treatment parameters using 2006 RLMS and provides the first application of the MTE method to Russia. Zamarro (2010) develops a new method for estimating MTE for multiple schooling levels and applies it to the evaluation of the education reform in Spain.

premium. Carneiro and Vytlacil (2011) and Moffitt (2008) show that an increase in college participation leads to a decline in the marginal returns to college. We also find that college expansion draws individuals with lower marginal returns to college. However, the general equilibrium effects were either offsetting or not strong enough to shift considerably the aggregate equilibrium skill prices and their distribution. In this paper, we do not disentangle varying general equilibrium effects of college expansion and leave this question for future research.[8]

Third, this chapter contributes to the literature evaluating the effect of new universities on college enrollment. We find that the opening of the first campus in Russian municipalities where there were previously no colleges increases the probability of receiving a college degree by 11 percentage points. Studies from other countries also report positive, but lower effects of a new university on college enrollment, e.g., 8 percentage points in Italy (Oppedisano (2011)) and 6.4 percentage points in Canada (Frenette (2009)).

Fourth, our use of the number of campuses in the municipality of residence at age 17 as a supply-side instrument for college education builds upon previous applications of supply-side shifters in identifying the returns to schooling; some earlier examples include construction of elementary schools in Indonesia (Duflo (2001)), a dummy for living near college (Card (1995)), the distance to the nearest college (Kane and Rouse (1995)) and supply disruptions caused by a war (Ichino and Winter-Ebmer (2004)).

Finally, our paper contributes to a large literature on the returns to education during the transition to a market economy (Andren and Sapatoru (2005); Brainerd (1998); Fang and Zeckhauser (2012); Fleisher and Wang (2005); Gorodnichenko and Peter (2005); Munich and Terrell (2005); Yang (2005), among many others). We improve this literature by estimating the distribution of returns, applying a more rigorous identification strategy, and deriving policy-relevant parameters for individuals at the margin of choice. Yet, we also confirm a previously documented pattern of increasing returns to schooling during the early transition period, followed by their levelling out in

[8] In the accompanying paper, we show that the expansion of higher education in Russia worsened the quality of college entrants, reduced resources per student, and increased the market returns to college quality (Belskaya and Peter (2015)).

the later period.

2.2 Econometric Framework

2.2.1 Model set-up

To estimate the heterogeneous returns to college, we follow a semi-structural method developed by Carneiro and Vytlacil (2011). The decision rule of an individual *i* is characterized by a latent variable model of college enrollment:

$$S_i = \mathbf{1}(S_i^* > 0), \quad \text{where} \quad S_i^* = \mu_S(Z_i) - \epsilon_i \tag{2.1}$$

S_i is a binary variable indicating college enrollment; it equals one for college graduates and zero for high school graduates.[9] We approximate $\mu_S(Z_i)$ by a linear form given by $Z_i'\gamma$, where Z_i is a vector of observable characteristics that affect the college decision. Vector Z_i includes three type of variables: (i) variables that affect both the schooling decision and wages, X_i^S; (ii) variables that measure the extent of college expansion at the time when individuals make their college decisions, E_i; and (iii) other instrumental variables/exclusion restrictions that also shift the latent variable S_i^*, I_i:

$$S_i = \mathbf{1}(E_i'\gamma_E + X_i^{S\prime}\gamma_X + I_i'\gamma_I > \epsilon_i) \tag{2.2}$$

We assume that ϵ_i is an unobserved to the econometrician error term that is statistically independent of Z_i. It captures the marginal cost of obtaining a college education for individual *i*.

Let P(Z) denote the probability of selecting into treatment (college) given Z. It is convenient to rewrite the selection equation (8.2) using the following innocuous transformation. Define $\nu_i = F_\epsilon(\epsilon_i)$, where F_ϵ is a cumulative distribution function, and ν_i is distributed uniformly in the unit interval [0, 1]. Different values of ν_i denote different percentiles of ϵ_i. Thus, the selection equation

[9] Following the tradition of the literature, the college drop-out decision is not modelled in this study. Russia has a relatively high completion rate of 79 percent of college entrants completing a tertiary type A program compared to the OECD average of 69 percent and the U.S. college completion rate of 56 percent (OECD (2008)).

becomes

$$S_i = \mathbf{1}(P(Z_i) > \nu_i) \tag{2.3}$$

Let the outcome of interest be Y_i =ln(w_i), where w_i is an hourly wage. The potential outcomes are defined according to the level of education achieved such that

$$Y_{i,s} = \mu_S(X_i) + \varepsilon_{i,s} \text{ for } s = 0, 1 \tag{2.4}$$

where $X_i = (X_i^w, X_i^S)$ is a vector of observable characteristics that affect hourly wages, and X_i^w include exogenous wage determinants that are excluded from the college choice equation (e.g., an individuals age at time *t* and unanticipated transitory shocks to local labor markets). We assume that $\mu_s(X_i)$ is linear, $\mu_s(X_i) = X_i'\beta_s$.

The observed outcome Y_i can be written in a switching regression form:

$$Y_i = S_iY_{i,1} + (1 - S_i)Y_{i,0} \tag{2.5}$$

$$= S_i(X_i'\beta_1 + \varepsilon_{i,1}) + (1 - S_i)(X_i'\beta_0 + \varepsilon_{i,0})$$

$$= X_i'\beta_0 + S_i(X_i'\beta_1 - X_i'\beta_0) + S_i(\varepsilon_{i,1} - \varepsilon_{i,0}) + \varepsilon_{i,0}$$

The gross returns to college education are given by $\Delta_i = Y_{i,1} - Y_{i,0} = (X_i'\beta_1 - X_i'\beta_0) + (\varepsilon_{i,1} - \varepsilon_{i,0})$ and the model assumes that agents know their returns.

The mean outcome Y_i, conditional on $(P(Z_i) = p, X_i = x)$, is a sum of mean outcomes for each level of education and weighted by the probability of being at each level of education:

$$E[Y_i|X_i = x, P(Z_i) = p] = E[Y_i|S_i = 1, x, p]p + E[Y_i|S_i = 0, x, p](1 - p)$$

$$= x'\beta_0 + (x'\beta_1 - x'\beta_0)p + \int_0^p E[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu]d\nu \tag{2.6}$$

The vector (Z, X) is observed by both the agent and econometrician, while $(\varepsilon_1, \varepsilon_0, \epsilon)$ is known by the agent but unobserved by the econometrician. To understand the possible effects of unobserved endowments in the wage equation (εs) on the returns, we consider three cases: (i) unobserved endowments are homogeneous, i.e. $\varepsilon_{i,0} = \varepsilon_{i,1} = \bar{\varepsilon}$ for all individuals, such that the last term in equation (2.6) cancels out; (ii) unobserved endowments are heterogeneous but mean independent of college decisions, i.e., $E[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu] = E[(\varepsilon_{i,1} - \varepsilon_{i,0})]$; again the last term in equation (2.6) cancels out; and (iii) unobserved endowments are heterogeneous, $\varepsilon_{i,0} \neq \varepsilon_{i,1}$, and correlated with the unobserved characteristics from the college decision equation (ϵ_i), in which case the last term in equation (2.6) cannot be ignored.

In the first two cases, the standard IV approach might be sufficient to identify the average returns to college education. In the third case, which is more realistic, individuals, who are observationally identical from an econometricians point of view, may make different college decisions; that is, unobserved characteristics influencing the college decision are correlated with unobserved endowments ε's in the wage equation. As a result, the returns to college, for observationally identical individuals, will depend upon a conditional mean component E$[(X_i'\beta_1 - X_i'\beta_0)|X_i = x] = (x'\beta_1 - x'\beta_0)$ and an individual-specific unobserved component E$[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu]$.

In the third case, the standard IV approach may not be appropriate for estimating the parameters of interest. It is widely recognized that an IV estimator identifies the local average treatment effect (LATE), which is the expected gain from the treatment (e.g., college graduation) of those individuals who switch from no treatment to treatment when an instrument changes. In the case of several instruments, each of the IVs identifies a different margin of the return to college that varies across individuals. In general, there is no simple interpretation of the IV estimator in the presence of several instruments that induce individuals to go to college.[10]

---

[10] Heckman and Vytlacil (2005), Heckman and Vytlacil (2007) present a limited number of cases where it is possible to determine the exact average treatment parameter identified by each of the instruments.

2.2.2 Marginal Treatment Effect from the Normal Selection Model

The marginal treatment effect (MTE) approach allows researchers to obtain the entire distribution of the individual-specific returns (i.e., $[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu]$ (Heckman and Vytlacil (2005), Heckman and Vytlacil (2007)). Using equation (8.6), the $MTE(x, \nu)$ is defined as follows

$$MTE(x, \nu) = \frac{\partial E[Y|X = x, P(z) = p]}{\partial p}$$

$$= (x'\beta_1 - x'\beta_0) + E[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu] \tag{2.7}$$

We assume that Z_i includes at least one exclusion restriction, and Z_i and X_i are statistically independent of $(\varepsilon_{i,0}, \varepsilon_{i,1}, \epsilon_i)$. We first estimate $MTE(x, \nu)$ from the parametric selection model that assumes a multivariate normal distribution of errors, $(\varepsilon_{i,0}, \varepsilon_{i,1}, \epsilon_i) \sim N(0, \Omega)$. Under the above assumptions, E$[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, \nu_i = \nu]$ =(Cov$(\varepsilon_{i,1}, \epsilon_i)$-Cov$(\varepsilon_{i,0}, \epsilon_i))\Phi^{-1}(\nu)$, where $\Phi^{-1}()$ is the inverse of the standard normal cumulative distribution function, and $MTE(x, \nu)$ is given by

$$MTE(x, \nu) = (x'\beta_1 - x'\beta_0) + (Cov(\varepsilon_{i,1}, \epsilon_i) - Cov(\varepsilon_{i,0}, \epsilon_i))\Phi^{-1}(\nu) \tag{2.8}$$

Heckman and Vytlacil (2005), Heckman and Vytlacil (2007) show how, based on the MTE distribution, to recover several standard treatment parameters such as the average treatment effect in the population $ATE(x)$, the average treatment effect on the treated $TT(x)$, the average treatment effect on the untreated $TUT(x)$, the ordinary least squared estimator $OLS(x)$, and the IV estimator $IV(x)$. In short, each treatment parameter can be obtained as a weighted average of $MTE(x, \nu)$, where weights are given in Appendix Table C3.

2.2.3 Policy Parameters

In addition to the standard treatment effect parameters, we also calculate the returns to college for those individuals who went to college due to the college expansion policy. The objective here is to calculate the gains of the marginal graduates who would have had limited access to higher education if the college expansion did not occur, but they chose to go to college in response to the opening of more colleges in the place of their residence. In other words, we are interested in the returns of the marginal graduates whose college decision was influenced by an increase in the instrument E_i in equation (2.2).

To illustrate our parameter of interest, we perform the following experiment using the estimated parameters of selection equation (2.2). Suppose that an individual *i* gets a draw given by $\tilde{\epsilon_i}$ such that $\tilde{\epsilon_i} > E_i'\gamma_E + X_i'\gamma_X + I_i'\gamma_I$, which implies that $S_i = 0$. Further, suppose that the total number of college campuses available to an individual *i* in his region of residence at the age of 17 increases such that the new number of campuses is given by $E_i^* = E_i + \alpha$. Our parameter of interest is the average of $MTE(x, \nu)$ for those individuals who switch their decision because of the expansion, such that $\tilde{\epsilon_i} < E_i^{*'}\gamma_E + X_i'\gamma_X + I_i'\gamma_I$, which implies that $S_i' = 1$.

Heckman and Vytlacil (2001) develop a policy relevant treatment effect $(PRTE(x))$ parameter that measures the average returns to college education for individuals induced to change their schooling decisions in response to a specific policy (for instance, $E_i^* = E_i + \alpha$). Essentially, PRTE(x) is the average $MTE(x, \nu)$ for switchers and is defined as:

$$PRTE(x, \alpha) = \int MTE(x, \nu) f_{\nu|X}(\tilde{\nu}|x, S_i(E_i = e, x, I) = 0, S_i(E_i = e + \alpha, x, I) = 1) d\tilde{\nu} \tag{2.9}$$

To identify $PRTE(x, \alpha)$, the support condition needs to be met, that is, the range of $P(E_i^*, X_i, I_i)$ must be contained in the range of $P(E_i, X_i, I_i)$. PRTE is generally applied to a fixed discrete policy change. Examples of specific policies evaluated with this method in the past include a tuition change in the US (Heckman and Vytlacil (2001); Carneiro and Vytlacil (2010), Carneiro and Vytlacil (2011); Eisenhauer and Vytlacil (2015)) and a 10 percent reduction in the distance to upper secondary schools in Indonesia (Carneiro and Umapathi (2011)). The PRTE approach has not been applied yet with respect to college expansion.

Alternatively, Carneiro and Vytlacil (2010) proposed the marginal policy relevant treatment effect ($MPRTE(x)$), which corresponds to a marginal change in policy α. That is,

$$MPRTE(x) = lim_{\alpha \to 0} PRTE(x, \alpha) \tag{2.10}$$

The MPRTE method essentially estimates the mean benefits of college for the marginal individuals who are indifferent between participating and not participating in college. The estimator requires the availability of a continuous instrument, assuming an infinitesimal change in α.

2.2.4 Marginal Policy Effects Using Local IV

An alternative approach to estimating MTE is based on the semi-parametric method of local instrumental variables (Heckman and Vytlacil (2005)). Under the assumption that Z_i and X_i are statistically independent of $(\varepsilon_{i,0}, \varepsilon_{i,1}, \epsilon_i)$, equation (2.6) can be re-written as follows:

$$E[Y_i|X_i = x, P(Z_i) = p] = x'\beta_0 + (x'\beta_1 - x'\beta_0)p + K(p) \tag{2.11}$$

where $K(p) = E[(\varepsilon_{i,1} - \varepsilon_{i,0})|X_i = x, P(Z_i) = p]p$ for $p \in P(Z_i)$

The MTE estimation process follows Carneiro and Lee (2009) and consists of three steps:

1) In the first step, we obtain the propensity score, $\hat{p}$, from a probit regression of S_i on $Z_i = \{X_i^S, I_i, E_i\}$.

2) In the second step, we estimate $\{\hat{\beta}_0, \hat{\beta}_1\}$ using a partially linear regression estimator of Robinson (1988) and then compute $R = Y - x'\hat{\beta}_0 - (x'\hat{\beta}_1 - x'\hat{\beta}_0)\hat{p}$. The estimates of $\{\hat{\beta}_0, \hat{\beta}_1\}$ are provided in the web appendix.

3) In the third step, we estimate a locally quadratic regression of R on $\hat{p}$ and calculate the derivatives of the conditional mean estimate $K'(\hat{p})$.

After completing these steps, the MTE estimates are obtained according to equation (2.7):

$$M\hat{T}E(x, \nu) = \frac{\partial E[Y|X = x, P(z) = p]}{\partial p} = (x'\hat{\beta}_1 - x'\hat{\beta}_0) + K'(\hat{p}) \tag{2.12}$$

Compared to the normal selection model, the local IV method for estimating MTE is more flexible as it does not assume the normal structure of the error term. But it is also more restrictive in the sense that the MTE can be estimated with this method only over the common support of $P(Z_i)$, which rarely takes the entire full unit interval [0, 1]. Since it is practically impossible to estimate the MTE over the full unit interval, the standard treatment parameters such as ATE, TT, TUT, as well as PRTE cannot be identified using the semi-parametric approach. Nonetheless, it is possible to identify the MPRTE parameters because only the marginal support of $P(Z_i)$ is required in this case (Carneiro and Vytlacil (2010)). In our empirical work, we apply the local IV method to recover the marginal returns to college for the individuals responding to a marginal change in the probability of college participation and in the number of campuses.

The following section discusses how each of the model variables (E_i, X_i, I_i) is measured in the data.

2.3 Data and Identification Variables

The primary data source for this study is the 1995-2011 Russia Longitudinal Monitoring Survey - Higher School of Economics (RLMS-HSE), which is a nationally representative stratified sample of households of the Russian Federation.[11] We limit the sample to high school graduates, age 25 and older, assuming that the majority of people complete their formal education by age 25. We also restrict our sample to people born between 1968 and 1986. Thus, considering that the college decision in Russia is typically made at age 17, our sample includes individuals who made their college decision in the pre-expansion period (1985-1992) as well as those who did it during the expansion period (1993-2003). Compared to the previous studies that are based on either cross-sectional data or one-cohort panels such as NLSY, our multi-cohort panel not only has multiple wage observations over the individual life cycle, but also permits the analysis of schooling decisions for several cohorts, thus making the study of college expansion feasible.

[11] RLMS-HSE is organized by the National Research University Higher School of Economics, Moscow together with the Carolina Population Center at the University of North Carolina at Chapel Hill and the Institute of Sociology at the Russian Academy of Sciences. The panel started in 1994. The RLMS-HSE surveyed individuals in 32 out of 83 regions and all seven federal districts of the Russian Federation (according to the official classification of regions as of January 1, 2010). We drop the year 1994, because the education questionnaire in that year is incompatible with the one in subsequent survey rounds.

The wage variable is the log of deflated hourly wage rate at primary job. As mentioned above, we restrict the period of college decision making to years 1985-2003. Our preferred specification employs individual wage data from the post college decision period, 2004-2011. However, in the robustness checks, we consider other intervals of wage data.

To estimate the college decision equation (8.2), we define the treatment group as college graduates and above (four years or more of tertiary education with a diploma) whereas the control group as secondary school graduates. The latter category includes graduates of both general and professional secondary schools, but excludes college drop-outs.[12] In some of the robustness checks, the definition of the treatment group includes college drop-outs with three or more years of college education. Figure 2.2 shows that the share of college graduates increased from 23 percent in 1995 to 39 percent in 2011 in the 25-44 age group of the RLMS sample.

We linked RLMS respondents with the local characteristics at age 17, which are available either at the level of municipality (such as the number of campuses by category) or at a more aggregate regional level (such as the size of cohort, earnings, and unemployment rate). The location of respondents at age 17 is not directly reported by the RLMS respondents, but it can be inferred in the majority of cases from the survey questions related to migration history, completion of high school, and college location. For example, location at 17 is the same as the place of current residence in three instances: (i) for individuals who permanently moved to their present location before they turned 18 years old (including those who never moved); (ii) for individuals who completed high school at present location; and (iii) for individuals who are born in the place of current residence and who temporarily moved to another location after age 17. College location is taken as location at 17 for those respondents who reported to reside in the same community before going to college. Thus, location at age 17 is known with high confidence for 83 percent of our sample.[13] The location of the remaining 17 percent of our sample at age 17 (e.g., those born in a different municipality

[12]Professional secondary schools include vocational schools (PTU), technical schools (technikums) and the specialized schools that train associate professionals in various fields (such as medicine, education, business, etc.). They offer 3-4-year programs after 9 years of secondary schools or 2-3-year programs after 11 years of complete general secondary education.

[13]Our sample excludes respondents who were born and studied outside Russia.

and moved to a current residence after completing college) is unknown. We impute their location at age 17 based on college location and the type of birthplace such as village, township, and city (see Appendix Table C1 for details of imputation).[14]

Table C.1 categorizes the variables used in the wage and schooling equations. Individual characteristics such as gender (=1 if female), nationality (=1 if ethnically Russian), and mothers education (=1 if mother has a college degree) enter both the wage and schooling equations. Since current urban residence is likely to be anticipated at the time of college decision, it theoretically explains both the wage and schooling outcomes. Year dummies, age, and age squared are controlled for in the wage equation, while four birth cohort dummies with 5-year intervals are included in the college equation. Dummies for the Moscow city and seven federal districts are determined at age 17 in the college equation and at the current time in wage equations.

Another variable that is expected to affect both the college decision and wage is the regional cohort size, which is the log of regional population of age 17 at the time when an individual was 17. Individuals from larger cohorts are likely to face more intense competition for a limited number of university slots in their region and thus have a lower probability of getting into college. The wage effect of larger cohorts is ambiguous due to a downward pressure from the labor supply side and an upward pressure from a bigger market size and potentially higher demand for labor. The data source for this variable is the Russian Census and other official demographic statistics (see Appendix Table C1 for details). Historically, the cohort size varied significantly not only between regions but also within regions over time due to the large demographic changes associated with the long-lasting effects of WWII and low birth rates in the 1990s in Russia (see Figure 2.2). In our sample period between 1985 and 2003, however, the fluctuations over time were relatively mild,

[14]In the 1990s, more than half of the internal migration flows were registered within a given region, e.g., from rural areas to cities of the same region (Demography (2002)). Furthermore, in our RLMS sample, 62 percent of college students with the known place of birth studied in the same municipality as their birthplace, 22 percent studied in a different municipality of the same region, and only 16 percent studied in a different region. Given these facts, we can randomly assign the missing location at age 17 based on the type of birthplace (village, township, and city), college location for students, and current residence for non-students. For example, if a student was born in a township, then an arbitrary township would be chosen within the region of his college as a place of his college decision. For non-students, we choose a random municipality of the same type as their birthplace within the region of their current residence. Note that such imputations affect only one variable of interest, which is the number of campuses per municipality, since all other local variables are constructed at the regional level.

though significant regional variation remained.

2.3.1 Number of Campuses per Municipality

The key identification variable through which we perform policy simulations is the number of campuses available in a municipality at age 17. We discuss its identification validity in Section 4. This variable is constructed based on the opening/closing dates of campuses from the Russian university database (Belskaya and Peter (2015)). The number of campuses in each municipality measures the local availability of college education at the time when individuals make their college enrollment decisions, and it serves as a proxy for E_i in equation (2.2). It is also broken down by categories of public-private and main campus-branch.

The descriptive statistics for this variable confirm a rapid expansion of college availability in Russia following the 1992 Law on Education. Between 1985 and 2003, the number of campuses in the Russian university database surged from 812 to 2245, and the number of municipalities with at least one campus increased from 198 to 442. Excluding Moscow city, the average number of campuses per municipality in the RLMS sample more than doubled over the sample period, with mean = 6.9 and std. dev. = 13.9 (see Figure 2.2). The upsurge of new campuses was also particularly striking in the Moscow region where their number quadrupled.

Despite seemingly large changes in the number of campuses over time, the share of within-municipality variance is only about 13 percent of the total variance. The variance decomposition reported in Table A.2 indicates that most variation in the total number of campuses is between municipalities (87 percent). However, the composition of variance differs considerably by the type of campus. For instance, the fact that the number of public universities barely changed since the Soviet times is reflected in the near to zero within-municipality variance (less than 1 percent of the total variance); practically all the variation in this variable comes from the differences between municipalities. At the same time, the within-municipality variation over time is substantial for the number of branches (both public and private) and for the number of private universities (32 to 55 percent each). These features of the main variable will be utilized in the validity tests of our instrument.

2.3.2 Local Labor Market Conditions

Besides the number of campuses per municipality at age 17, we also use labor market conditions in the location of residence at the time of college decisions as additional exclusion restrictions. Local earnings and unemployment rate are frequently used as instruments in the MTE literature (e.g., Cameron and Heckman (1998), Cameron and Taber (2004), Carneiro and Vytlacil (2011), Zamarro 2010, Eisenhauer and Vytlacil (2015), among others).[15] In our case, regional earnings at 17 and regional unemployment rate at 17 capture the opportunity cost of going to college.

To control for the potential correlation between local wage/unemployment at age 17 and current individual wage, we include (i) the contemporaneous (transitory) component of regional variables in the wage equations and (ii) the permanent (predicted) component of regional variables in both the schooling and wage equations (similar to Cameron and Heckman (1998); Carneiro and Vytlacil (2011)); see Table C.1.

The transitory shocks to regional labor markets are proxied by the deviation of contemporaneous regional earnings and unemployment rate from the trend-adjusted regional average. Specifically, regional transitory earnings are defined as $\hat{\epsilon_{rt}}$ in a regression of the log of regional monthly earnings in region *r* and year *t* on the set of regional dummies μ_r and a quadratic time trend, as in $\ln(w_{rt}) = \mu_r + \varphi(t) + \epsilon_{rt}$. The regional transitory unemployment rate is defined similarly.

The permanent regional variables are predicted from the regressions of regional monthly earnings and unemployment rate on region dummies and a quadratic time trend over the period 1995-2011. Essentially, these are trend-adjusted average characteristics of the local labor market. As an alternative measure of permanent earnings, we compute average regional monthly earnings over the first 10 years following the college decision. To account for the possibility that individuals making a college decision under the Soviet centralized wage structure may not have anticipated the future market earnings due to a large structural break in the economy, we also compute permanent

[15] Local tuition as another commonly used IV is not applicable to our case because college tuition was not charged during the Soviet period and only 28 percent of all public students paid a fee for their education during our sample period of college expansion, 1993-2003 (Education (2008)).

regional earnings separately for the Soviet period (1980-1991) and the market period (1992-2011).

In essence, with the above structure of regional variables, we assume that individuals making their college decision are not only influenced by the observed contemporaneous characteristics of their location at age 17, but also have an imperfect foresight at age 17 of future local labor market conditions.

In addition to the regional monthly earnings and unemployment rate at age 17, the opportunity cost of going to college is captured by the country-level skill wage ratio at age 17, which is the ratio of average wages of manual workers to average wages of non-manual workers in industry (see Appendix Table C1 for data sources). Throughout the Soviet period, manual workers were rewarded with high pay, while wages of non-manual workers were artificially compressed. Figure 2.2 shows that the wage ratio plummeted from 0.94 to 0.55 during our sample period (1985-2003), reflecting a historic trend of the rising wage premium for skilled workers.

2.3.3 Summary Statistics for the Estimation Sample

The baseline estimation sample (age ≥ 25, age 17 $\in$ [1985, 2003], wage $\in$ [2004, 2011], and non-missing variables) amounts to 17,911 person-year observations. Conditional on the exogenous sample constraints based on age, survey year, and birth cohorts and for a given definition of the college variable, wage is missing for about 20 percent of respondents either due to non-employment or non-reporting. In robustness checks, we apply the inverse propensity score re-weighting to deal with the selection into wages (i.e., employment or reporting of wages). With the exception of mothers education, the missing values in other covariates are trivial (about 0.6 percent).[16] Table C1 in the appendix details how each variable is constructed. The descriptive statistics in Table A.3 is reported separately for college graduates and secondary school graduates. As expected, college graduates are more likely to be female, reside in urban areas, have a mother with college degree, live in municipalities with more campuses at age 17, and earn a higher wage rate. Interestingly,

[16]Because mothers education is unknown for the 12 percent of our estimation sample, we include a binary indicator for missing values of this variable to prevent sample loss. We drop observations with missing values in other covariates.

when college graduates were 17, they resided in areas with lower earnings and higher unemployment rates, but their regions of residence at the time of the survey had better permanent labor market characteristics.

2.4 Marginal Treatment Effect

2.4.1 Baseline Estimates

In this section, we report the MTE estimates of the marginal benefits of college education and discuss the validity of exclusion restrictions. We begin by estimating the components of the switching regression model given in Equations (2.2)-(2.4). The random variables ($\varepsilon_{i,0}, \varepsilon_{i,1}, \epsilon_i$) are assumed to have a multivariate normal distribution, and the model is estimated using maximum likelihood. Table A.4 presents the estimates of the wage equation for secondary school graduates (column 1), the wage equation for college graduates (column 2), the college equation (column 3), and the marginal effects of variables in the college equation (column 4). The probability of having a college degree, $\hat{P}(Z)$, is significantly higher among females, ethnically non-Russians, urban residents, and those whose mother went to college. As expected, individuals from smaller regional cohorts at age 17 tend to have a higher rate of college attainment, as they face less competition for a given number of college openings in their region. More people go to college in the regions with higher permanent earnings, but college decisions do not appear to be responsive to regional unemployment rates and to regional earnings at age 17. Predictably, better salaries of manual workers relative to non-manual workers - observed at age 17 - deter people from pursuing a college degree.

The instruments are jointly strong predictors of college decision (p-value=0.000). The key instrument measuring the college availability at the time when an individual was 17 is the total number of college campuses per municipality. This variable serves as a cost shifter of college decisions. In order to capture potential non-linear effects of college availability and to account for the fact that some of the larger cities had a wide choice of colleges even before the college expansion (e.g., Moscow had 100 colleges in 1990), we add a quadratic term on the number of campuses. Because the distribution of the number of campuses is clumping at zero, we also introduce a binary variable to indicate municipalities with no campuses (mean=0.386). The latter variable is analogous to a commonly used binary IV, such as the presence of a college in the county of residence during teenage years (Card (1995); Cameron and Taber (2004)). Results in Table A.4 suggest that the opening of the first campus in a municipality without a college increases the probability of receiving a college degree by 0.114, i.e., by 11 percentage points, on average, *ceteris paribus*. The establishment of additional campuses also improves the college attainment, though at a diminishing rate per new campus. Thus, the effect of college availability on college attainment appears to be much stronger for constrained municipalities with fewer campuses.

The estimates of the wage equations for the two groups are standard.[17] The results suggest that each group has a comparative advantage in the labor market, that is, the individuals with a higher propensity to go to college (low ν) tend to do well in the labor market once they graduate ($\sigma_{1\epsilon} = -0.150$), but the same individuals would be worse off if they dont go to college ($\sigma_{0\epsilon} = 0.222$).

Following Carneiro and Vytlacil (2011) and Heckman and Urzua (2010), we perform a test for selection on gains (i.e., whether returns to college are correlated with S). A simple test involves estimating equation (2.6) where the last term is approximated with a polynomial in $\hat{p}$ (obtained from the probit of college equation) and testing whether the coefficients on higher order polynomial terms are jointly statistically significant. The results of this test support the hypothesis that individuals in our sample select on college gains (see Appendix Table C2). The test rejects that the returns to college are not correlated with S or that $MTE(x, \nu)$ is constant in ν.

Another test for selection on gains relies on the estimated parameters from the normal switching regression model (Table A.4). The null hypothesis that the slope of the MTE is zero, i.e., H_0 : $\sigma_{1\epsilon} - \sigma_{0\epsilon} = 0$, is rejected at the 1 percent level. We estimate that $\sigma_{1\epsilon} - \sigma_{0\epsilon}$=-0.372 with a standard error of 0.065. This finding supports the conclusion of selection on gains in Table C2 that does not impose the joint normality assumption.

[17] The results worth mentioning are: a very large gender wage gap in the control group (0.45 log points or 56 percent difference favoring males) compared to 27 percent in the treatment group; mothers education affects wages of college graduates, but not wages of secondary school graduates; the individual wage rate in both groups is positively influenced by higher transitory and permanent local earnings; only in the control group, the wage rate responds positively to local unemployment (both transitory and permanent); the cohorts with larger regional population at 17 receive a higher wage premium today.

Using the estimated parameters in Table A.4, we compute the MTE according to Equation (2.8). Figure 2.3 plots the estimated MTE for a grid of values v between zero and one,[18] with 90-percent confidence bands, evaluated at mean values of *X*. We obtain annualized estimates of the returns to college by dividing the MTE by 4.5, which is the average difference in years of schooling between the treatment and control groups in our sample. The negative slope of MTE implies that individuals with low values of ν (those who are more likely to go to college) have the largest marginal returns to one year of college education. Conversely, individuals with high values of ν have low MTE. The heterogeneity in the MTE across the distribution of ν is substantial: the returns vary from -18.2 percent for the highest ν person who would lose from attending college to 32.9 percent for the lowest ν person, with the average return of 7.3 percent per one year of college education.

Given the vast differences in labor market institutions and data characteristics between Russia and the U.S., our MTE estimates for Russia turned out to be surprisingly close to the findings of Carneiro and Vytlacil (2011). In a sample of 28 to 34-year old white males from NLSY, they find that the college returns in the U.S. vary from -15.6 percent to 28.8 percent per year of college with the mean of 6.7 percent and the MTE slope $\sigma_{1\epsilon} - \sigma_{0\epsilon} = -0.239$. At the same time, our MTE estimates for Russia are lower at the mean and have a much larger variance than the corresponding estimates obtained by Heckman and Li (2004) in a cross-sectional sample of Chinese workers (their MTE ranges from 5 to 15 percent per year of college, mean=10.8).

2.4.2 Alternative Instruments

In Table 2.5, our main instrument - number of campuses per municipality - is broken down by categories of main campus-branch (column 1), public-private [19] (column 2), Moscow vs. other cities (column 3). Overall, these results are not different from the baseline specification. We find a

[18]In practice, we restrict the grid of values of ν to be between 0.001 and 0.999, with 999 equally spaced points.

[19]We use a dummy indicator for private campuses instead of the number of campuses due to considerable clustering of observations at values of zero and one: only 10 percent of all municipalities in the RLMS sample had more than one private campus.

negative MTE slope of similar magnitude, average returns of 7 to 8 percent per year of college, and a clear concave relationship between college availability and college attainment for all indicators. But the college probability function is estimated to be considerably more concave with respect to the number of branches. This result could be partly explained by branches rather than main campuses being opened in less-populated areas where the first few local branches may have a substantial impact on local college attainment (hence, large positive linear term), and where the establishment of further branches may also quickly overcrowd a local market for higher education services (hence, large negative quadratic term). Lower tuition fees and laxer admission criteria in branches could also contribute to a high response of the probability of going to college to the opening of the first few branches in locality.

Another noteworthy finding is that the establishment of a private campus in a municipality with existing public campuses does not have a statistically significant effect on the probability of college attainment.[20] This result could be due to the fact that private education institutions in Russia are charging higher tuition, are often small in size and tend to open in larger cities where individuals have other options of pursuing a college degree. In the third column of Table 2.5, we try to isolate the effect of college-crowded Moscow by the interaction of the number of campuses per municipality at 17 with Moscow residence at 17 in a linear specification. The estimates suggest that the college attainment in Moscow is not affected by a further increase in the number of campuses, but the opening of an additional campus in other municipalities with at least one campus increases the probability of receiving a college degree by 0.4 percentage points. All parameters of interest remain akin to the baseline specification.

Starting with the paper of Card (1995)), the literature raised two major issues with the validity of a binary indicator for college presence as an instrument, including non-random college construction and the Tiebout-type geographic sorting of individuals in response to college presence. These are legitimate concerns with regard to the number of campuses as well. The validity of the

---

[20]Interestingly, in a paper on the effect of mothers education on birth outcomes in the U.S., Currie and Moretti (2003) also found that the opening of public colleges has a larger effect on mothers educational attainment than the opening of private colleges.

instrument is going to be compromised if:

- ϵs in wage equation affect E_i in college equation; that is, anticipated wage shocks in 2004-2011 influence where new campuses are opened in 1985-2003;
- E_i and ϵs are jointly determined; for example, individuals with, let say, positive wage draws choose to reside in the location of new campuses.

To isolate the potential correlation between future expected earnings and campus openings, both wage and schooling equations control for the predicted regional earnings and unemployment rate in the place of residence at 17. Adding district fixed effects and a dummy for the type of location serves the same purpose. These controls partly address the above concerns, including non-random college openings in response to higher regional income and the sorting of individuals with higher wage draws into the locations with better labor market conditions, which also happened to be places with more campuses.

The variance decomposition in Table A.2 gives us useful hints to check whether our baseline results are driven by the time-series or cross-sectional variation in the number of campuses. A large within-municipality variation in the number of campuses over time provides more room for the endogenous decisions to open campuses in response to unobserved (by the econometrician) future wage shocks. At the same time, the cross-sectional/geographic variation that is created in a different economic system long before the college expansion started does not allow for such strategic behavior. Table A.2 shows that most of the variation in the number of campuses is geographic rather than over time (87 percent vs. 13 percent). Furthermore, we re-estimate our baseline specification with the number of public colleges only. This number remained practically the same since the Soviet times; the within-municipality variance is less than 1 percent of the total variance (Table A.2). We can reasonably assume that the college construction decision under the centrally-planned industrial structure is not correlated with wage innovations in the market economy. The alternative specification with the number of public colleges does not change the value and the slope of the marginal treatment effect (column 4 of Table 2.5).

We also ran several placebo tests by adding future college openings in the college equation.[21] The goal here is to check whether college decisions are influenced by the future college expansion. One of such tests is shown in column 5 of Table 2.5 and suggests that newly opened campuses in the same municipality between the ages of 25 and 30 do not have any significant impact on the likelihood of going to college at age 17.

Therefore, provided that the anticipated local labor market conditions are sufficiently controlled for, the number of campuses at age 17 can be used as a sensible instrument for the marginal treatment effect estimation and subsequent policy simulations.

2.4.3 Alternative Specifications

Before calculating the treatment parameters, we check whether our results are sensitive to changes in specifications. Specifically, we focus on the sensitivity of the mean component of MTE, $\bar{X}(\hat{\beta}_1 - \hat{\beta}_0)$, and the covariance component of MTE, $(\sigma_{1\epsilon} - \sigma_{0\epsilon})$. The estimates of additional specifications are presented in Table 2.6. In all of these specifications, the instruments are jointly strong predictors of college decisions, and the null hypothesis that the MTE slope is zero, $H_0 : \sigma_{1\epsilon} - \sigma_{0\epsilon} = 0$, is rejected at the one percent level. In column 1, the model is estimated using an alternative definition of the treatment group that includes college dropouts with three or more years of higher education. The rationale for changing the dependent variable is that college dropouts with some years of schooling might also have been treated by studying in college.

In column 2, we use an alternative definition of regional permanent earnings calculated as average regional earnings over the first ten years since college decision at age 17. Specification in column 3 applies the same definition as in column 2 but replaces permanent earnings with the Soviet period average regional earnings (1980-1991) for individuals who turned 17 before market reforms. Our motivation here is that teenagers raised under central planning may not have foreknown future earnings in the market economy. As we introduce Soviet earnings, the association

[21] For example, we controlled for the number of campuses established in the same municipality between the ages of 30 and 35; then we tried adding new campuses opened between 2007 and 2011. None of these variables have any significant impact on the likelihood of going to college at age 17.

between individual wages and regional permanent earnings becomes weaker but remains statistically significant.[22] Overall, modifications in the first three columns do not cause any significant deviation of the estimated MTE from a baseline specification.

To check for potential bias due to non-random missing wages, specification in column 4 applies the inverse propensity weighting (IPW) to our baseline specification, where the weight is the inverse of the predicted probability of having non-missing wages. The propensity score is constructed using the model covariates X^S and X^W, which are available for all respondents, including those with missing wages.[23] The MTE estimates are not affected by the IPW correction.

We also estimate the model without assuming the joint normality between the errors of the wage and college equations. Figure 2.4 shows the MTE using the semiparametric procedure we described in section 2. Overall, the shape of the MTE and policy parameters we discuss in the next section are consistent with the fully parametric normal model.

Next, we test whether our results change if we use wage data for different survey years. Figure 2.5 plots the average treatment effect and the MTE slope along with the 95-percent confidence interval for different periods of the wage data; the point estimates from the 1995-2011 sample period is also shown in column 5 of Table 2.6. Although the average returns to college appear to decline over time, there is no statistically significant (at the 5 percent level) difference in the average returns to college between different survey periods. The MTE slope is also constant over different periods. It seems that the general equilibrium effects of college expansion are either canceling each other out or they are not sufficiently strong to shift the equilibrium skill prices and their distribution. By attracting marginal students with lower returns, college expansion may alter the aggregate composition of college graduates and thus put a downward pressure on the average returns to college. However, the composition effect in our estimates is intertwined with the price effect that partly could be demand-driven. That is, skill-biased demand shocks along with the positive productivity

[22]Results are shown in web appendix.

[23]See Table C.1 for the list of covariates. The MTE estimates are not sensitive if the IPW correction is performed separately for missing wages among the employed due to non-reporting (15 percent of all missing wages) and for missing wages due to non-employment. The probit models and MTE estimates with IPW for different sources of missing data are shown in web appendix.

spillovers from the increased stock of human capital may compensate for the supply-side effects keeping the equilibrium college wage premium constant over time. Disentangling these varying general equilibrium effects of college expansion remains an area for future empirical analysis. The policy treatment parameters discussed in the next section should be interpreted from a partial equilibrium perspective.

2.5 Policy Treatment Parameters

2.5.1 Conventional Treatment Parameters

Recall the gross returns to college education are given by $\Delta_i = Y_{i,1} - Y_{i,0} = (X_i'\beta_1 - X_i'\beta_0) + (\epsilon_{i,1} - \epsilon_{i,0})$, with E$(\Delta_i X_i) =$ E$[(\beta_1 - \beta_0)|X_i]$. Following Heckman and Li (2004), the probability limit of the OLS estimator can be written as:

$$plim(\hat{\Delta}_{OLS}) = E(Y_i|X_i, S_i = 1) - E(Y_i|X_i, S_i = 0) \tag{2.13}$$

$$= E(X_i'\beta_1 + \epsilon_{i,1}|X_i, S_i = 1) - E(X_i'\beta_0 + \epsilon_{i,0}|X_i, S_i = 0) \tag{2.14}$$

$$= E(\Delta_i|X_i) + E(\varepsilon_{i,1}|S_i = 1) - E(\varepsilon_{i,0}|S_i = 0) \tag{2.15}$$

$$= \underbrace{E(\Delta_i|X_i)}_{"ATE"} + \underbrace{E(\varepsilon_{i,1} - \varepsilon_{i,0}|S_i = 1)}_{"Sortinggain"} + \underbrace{[E(\varepsilon_{i,0}|S_i = 1) - E(\varepsilon_{i,0}|S_i = 0)]}_{"SelectionBias"} \tag{2.16}$$

$$= \underbrace{E(\Delta_i|X_i, S_i = 1)}_{"TT"} + \underbrace{[E(\varepsilon_{i,0}|S_i = 1) - E(\varepsilon_{i,0}|S_i = 0)]}_{"SelectionBias"}$$

where ATE= E$(\Delta_i|X_i)$ is the average treatment effect of college education for a randomly chosen individual; TT= E$(\Delta_i|X_i, S_i = 1)$ is the treatment effect on the treated; the sorting gain, E$(\varepsilon_{i,1} - \varepsilon_{i,0}|S_i = 1)$=TT-ATE, is the mean gain of the unobservables for people who select college, and the selection bias [E$(\varepsilon_{i,0}|S_i = 1) - E(\varepsilon_{i,0}|S_i = 0)$] = OLS - TT is the mean difference in unobservables between secondary school graduates and college graduates if the latter would not go to college.

In Table 2.7, we report the above treatment parameters, which are constructed by integrating MTE with the appropriate weights developed by Heckman and Vytlacil (2005) (Appendix Table C3). Standard errors are bootstrapped, and all parameters are annualized. For a randomly chosen individual, the average treatment effect of one year of college education is about 7.3 percent. The treatment effect on the treated is a 13 percent wage increase for college graduates compared with what they would receive without college degree. At the same time, the treatment effect on the untreated (TUT) is only 1.9 percent wage increase for secondary school graduates if they would go to college. The OLS estimate is 7.7 percent. We estimate the sorting gain of 5.7 percent and the selection bias of 5.3 percent. Positive sorting gain, E$(\varepsilon_{i,1} - \varepsilon_{i,0}|S_i = 1) > 0$, implies that individuals sort into college on the basis of unobserved gains. Negative selection bias means that if college graduates did not complete college education, they would be worse off in terms of the unobserved wage component in comparison with secondary school graduates, E$(\varepsilon_{i,0}|S_i = 1) <$ E$(\varepsilon_{i,0}|S_i = 0)$. In other words, both college and secondary school graduates have a comparative advantage in the labor market, which is consistent with the analysis of Willis and Rosens (1979) seminal paper. An IV estimate of the returns to college (with the propensity score $\hat{p}$ used as IV) is 16.1 percent, and it is predictably upward biased compared to ATE due to heterogeneity ($\varepsilon_{i,1} \neq \varepsilon_{i,0}$) and positive sorting gain. As evident from Appendix Figure C1, the IV estimand weighs a higher valued segment of $MTE(x, \nu)$ more heavily.

2.5.2 Policy Effects

In addition to conventional treatment parameters, we also estimate the treatment effect for individuals at the margin of indifference between going and not going to college ($\mu_S(Z_i) = \epsilon_i$). As shown by Carneiro and Vytlacil (2010), the marginal treatment effect at the indifference point is equivalent to the effect of the marginal policy change, which is college expansion in our case. Following Carneiro and Vytlacil (2010), we assess the marginal returns to college for three alternative policy regimes. The first policy exogenously increases the probability of graduating from college by an infinitesimal amount α, so that $P_\alpha = P(Z) + \alpha$. An alternative policy changes the probability of college by a tiny proportion $(1 + \alpha)$, $P_\alpha = (1 + \alpha)P(Z)$. The third policy may involve a small change in one of the continuous components of Z. In our Z vector, we have a direct measure of college expansion, $N_{m,t}$, which is the number of campuses in the municipality of residence *m* in year *t* (corresponding to age 17), such that $P_\alpha = P(N_{m,t}+\alpha)$.[24] Using equation (8.10) and weights provided in Appendix Table C3 and Figure C1, we calculate marginal policy-relevant treatment effects of three policy regimes from the normal selection model and report them in Table 2.8, Panel A. Similar to Carneiro and Vytlacil (2010), we find that the MPRTE for a marginal additive change in P is estimated to be higher (9.9 percent) than the MPRTE for a marginal proportional change in P (7.4 percent), but the estimates for Russia are greater in magnitude in both cases. The third policy regime, which is more explicit and not yet reported in the literature, has returns of 9.6 percent per one year of college for a marginal person who is indifferent between going or not going to college and who would change college participation in response to a marginal increase in the number of campuses.

Alternatively, we report MPRTE parameters from the MTE distribution estimated using a semi-parametric method of local IV. The estimation process is described in Section 2, and the MTE estimates, evaluated at mean values of X, are plotted with 90-percent confidence bands in Figure 2.4. We find that the semi-parametric method produces MTE with the same shape as the parametric one, but with somewhat larger standard errors. Similar to the normal model, the MTE is declining in ν, and we reject the null hypothesis that the returns to college are not correlated with S or that $MTE(x, \nu)$ is constant in ν based on the test results reported in Appendix Table C2. The common support of the $P(Z)$ estimated from our sample ranges from a minimum of 0.070 to a maximum of 0.938.[25] Panel B of Table 2.8 presents the MPRTE parameters from the semi-parametric model

[24]In our sample, the number of campuses per municipality varies from 0 to 299 and is treated as a continuous variable. An infinitesimal change in $N_{m,t}$ can be interpreted as a new classroom or an additional student slot in a locality.

[25]Common support is defined as the intersection of the support of $P(Z)$ given $S = 1$ and the support of $P(Z)$ given $S = 0$. 76 observations, or 0.4 percent of our sample, fall outside the common support.

for the three alternative policy regimes described above.[26] In particular, the MPRTE for a marginal change in the number of campuses is estimated to be 10.7 percent wage increase for one year of college. These estimates are only slightly higher (by about one percentage point) than the MPRTE estimates from the normal model, so we adhere to our earlier conclusions with regard to the marginal returns to college. Regardless of the method, we find that a marginal student at the indifference point enjoys relatively high returns to college in Russia, even though his returns are, as expected, lower than those earned by current college graduates (TT).

Next, we perform a few policy experiments for a discrete change in the number of campuses. The experiments are evaluated using the policy-relevant treatment effect (PRTE) estimator proposed by Heckman and Vytlacil (2001) and given in equation (8.8).[27] In our case, PRTE captures the average MTE for the individuals who changed college participation in response to a fixed increase in the number of campuses from the 1992 pre-expansion level, $P_{PRTE} = P(N_{m,92}+\Delta N_{m,t})$. In the first simulation, we add one campus in each municipality to the 1992 level in 1993-2003 and find the PRTE returns to be 9.8 percent wage increase per year of college. In the second set of simulations, we add one campus per municipality in different locations. Results in Table 2.8, Panel C show that the returns to college vary depending on the place where campuses are established. The returns are estimated to be larger for constrained municipalities, including smaller non-capital cities, rural districts, and municipalities that did not have institutions of higher education in 1992 (9.9 percent wage increase in all three cases). At the same time, the opening of a campus in localities with previously existing campuses or in the largest regional cities-capitals attracts individuals with lower returns to college (7.7 and 7.5 percent, respectively).

In the final set of simulations, we ask what would the returns to college be for the affected individuals (those who shifted to treatment) if the number of campuses increased by only a half

[26]A MPRTE parameter is a weighted average of the MTE values estimated semi-parametrically on the basis of equation (8.12). The MPRTE weights are reported in Appendix Table C3 and also plotted in Figure C1.

[27]The PRTE requires that the empirical support of the distribution of $P(Z)$ to be within the unit interval, and that the empirical support of P_α has to be contained in the support of $P(Z)$. These conditions are satisfied in our setting. Given the multivariate normal structure of the errors, the support of $P(Z)$ is the unit interval by construction. Additionally, the empirical support of $P(Z)$ is given by (0.044, 0.949), while the empirical support of P_α is (0.045, 0.949) for $\alpha = 1$.

of the actual increase, $N_{m,t}* = N_{m,92} + \frac{1}{2}(N_{m,t} - N_{m,92})$, or by 50 percent more of the actual increase, $N_{m,t}* = N_{m,92} + \frac{3}{2}(N_{m,t} - N_{m,92})$. Compared to earlier simulations in Panel C where we add uniformly one campus in selected municipalities, here we simulate changes which are large in magnitude and varying in size depending on the scale of actual college expansion in each municipality. On average, four additional campuses per municipality were established between 1992 and 2003. Predictably, more sizeable college expansions attract individuals with lower returns, though the scale effect is not linear. PRTE is estimated to be 9.8 percent for a one-campus establishment policy, 8 percent for half expansion, 6 percent for full expansion, and 6.2 percent for 1.5 of actual expansion.

Our PRTE estimates of the returns to college imply that the present value of additional earnings stream for a marginal student over a 33-year working life (age 22-55) is equal to about \$41,657 gains in non-capital cities vs. \$36,261 in capital cities (see details of calculations in Appendix Table C4). The difference in net benefits between the two types of cities is even greater, since students in capital cities, on average, pay more tuition and forego higher earnings while in school compared to students in smaller cities. A rough cost-benefit calculation in Table C4 suggests that the net present benefits for a marginal student amount to \$15,357 in non-capital cities and \$7,334 in capital cities after subtracting the present value of foregone earnings, average tuition and other college-related expenses. Considering that relocation to a large city imposes additional costs (e.g., transportation, greater living expenses, etc.), new students in constrained municipalities clearly benefited from college openings in the place of their residence. The stock of human capital in the remote labor markets is also likely to increase, as college graduates tend to stay in the area where they receive their education (Groen (2004); Winters (2011)).

2.6 Conclusion

This chapter estimates marginal returns to college education in Russia. Despite the vast differences in labor market institutions and data characteristics between Russia and the U.S., our results are consistent with the previous U.S. literature on MTE in that we also find (i) a large degree of heterogeneity in returns to college, varying from -18 to 33 percent increase in lifetime wages for a year of college; (ii) a negatively-sloped MTE showing greater marginal benefits from college among individuals with the highest propensity to go to college; (iii) the positive sorting of individuals into college based on observed and unobserved market gains associated with college; (iv) lower returns for a marginal student than for an average student (10 vs. 13 percent per year of college); (v) a larger IV estimate of the returns to college compared to an OLS estimate; 16 and 8 percent, respectively. Implicitly these results support the MTE approach and the need for the precise characterization of different treatment parameters.

The main focus of this study is on the evaluation of the large-scale college expansion that occurred in Russia and resulted in mass openings of new colleges, both public and private, and their branches in many localities where college education was not previously offered. Specifically, we are interested in the returns of the marginal individual who switched into treatment as a result of college expansion. Unlike previous MTE studies where college expansion is characterized by an exogenous hypothetical shift in the probability of attending college by some random amount or proportion, we evaluate actual, real-world college expansion using the number of campuses (and their types) at a highly disaggregated level of municipality.

We establish that individuals with lower returns enter colleges as more campuses open. However, for the marginal individuals who are affected by the establishment of an additional campus at the time of making a college decision, the overall gains from attending college are large and positive; we estimate a 10 percent wage gain for these individuals. Furthermore, we find that the opening of a new campus in constrained municipalities - smaller non-capital cities or municipalities that did not have institutions of higher education before college expansion - attracts students with higher returns compared to the effect of the same policy in unconstrained municipalities with at least one college existing before the expansion. Our policy estimate also indicates that if the number of campuses per municipality did not increase, then a considerable share of population with high potential labor market gains from college would not have been able to realize these gains in the market.

Other results highlight the important distinction between public and private colleges and between main campuses and branches and show that public campuses and local branches are estimated to be more effective in influencing local college participation. We also find the college probability function to be concave with respect to the number of campuses, with a very large kink point for the first campus opened in a locality. The effect of each additional campus on the local college attainment diminishes and eventually vanishes in college-rich locations such as Moscow city.

Overall, our findings show the direction for future policies targeted at expanding college education in developing countries and identifying locations for the future college construction that would attract individuals with the highest potential gains from additional education.

Table 2.1: Variables in Wage and College Equations

Variable	**Time frame**	**Level of aggregation**	**Common** X^S	**Wage eq** X^W	**College eq E and I**
Female	Fixed	individual	x		
Urban residence	Fixed	individual	x		
Ethnically Russian	Fixed	individual	x		
Mother's education	Fixed	individual	x		
Age, age squared	Current year	individual		x	
Birth cohorts	Fixed	individual			x
Federal districts	Current year	individual		x	
Federal districts at 17	Age 17	individual			x
Survey year dummies	Current year	individual		x	
Regional cohort size at 17	Age 17	region	x		
Permanent regional earnings	Fixed	region	x		
Permanent unemployment rate	Fixed	region	x		
Transitory regional earnings	Current year	region		x	
Transitory unemployment rate	Current year	region		x	
Regional earnings at 17	Age 17	region			x
Unemployment rate at 17	Age 17	region			x
Skill wage ratio at 17	Age 17	country			x
Number of campuses at 17	Age 17	municipality			x

Notes: The X^W includes wage determinants; E is the number of campuses in the municipality of residence at age 17; the I vector includes other exclusion restrictions in the college decision equation; and the common vector X^S belongs to both equations.

Table 2.2: Variance Decomposition for the Number of Campuses

Variables	**Mean**	**Total Variance**	**Within-municipality %**	**Between-municipality %**
Total number of campuses	1.43	77.3	86.8	13.3
Number of colleges - public	0.77	22.3	99.4	0.6
Number of colleges - private	0.27	17.5	59.6	40.4
Number of branches - public	0.33	1.0	68.3	31.8
Number of branches - private	0.07	0.2	44.9	55.1

Notes: The number of campuses, colleges, and branches is calculated for every municipality in the RLMS estimation sample. The panel of municipalities used in the variance decomposition is balanced with 733 municipalities and 19 time periods from 1985 to 2003. A campus refers to all buildings of the same college in one municipality. Branch is a campus located outside the municipality of the main campus. Total number of campuses is the sum of the number of main campuses (which is equivalent to the number of colleges) and the number of branches.

Table 2.3: Sample Statistics

	S = 0 (N = 10,962)	**S = 1 (N = 6,949)**	**Mean comparison t-test (p-value)**
Hourly wage rate (log)	3.831	4.286	0.000
	(0.751)	(0.721)	
Age	32.855	31.919	0.000
	(4.808)	(4.708)	
Regional transitory earnings (log)	0.020	0.013	0.001
	(0.139)	(0.146)	
Regional transitory unemployment rate, %	0.111	0.239	0.000
	(1.687)	(1.641)	
Female	0.492	0.606	0.000
	(0.500)	(0.489)	
Urban residence (binary)	0.754	0.873	0.000
	(0.431)	(0.333)	
Ethnically Russian (binary)	0.907	0.911	0.319
	(0.291)	(0.284)	
Mother's education (binary)	0.067	0.302	0.000
	(0.249)	(0.459)	
Mother's education missing (binary)	0.117	0.115	0.648
	(0.322)	(0.319)	
Regional cohort size at 17 (log)	10.456	10.551	0.000
	(0.610)	(0.674)	

	S = 0 (N = 10,962)	S = 1 (N = 6,949)	Mean comparison t-test (p-value)
Regional permanent earnings (log)	8.519	8.626	0.000
	(0.387)	(0.396)	
Regional permanent unemployment rate, %	6.841	6.201	0.000
	(2.615)	(2.886)	
N of campuses per municipality	11.879	26.432	0.000
	(36.070)	(57.131)	
Municipality with no campuses (binary)	0.465	0.260	0.000
	(0.499)	(0.439)	
Skill wage ratio at 17	0.760	0.725	0.000
	(0.128)	(0.126)	
Regional earnings at 17 (log)	7.875	7.844	0.000
	(0.463)	(0.460)	
Regional unemployment rate at 17, %	4.817	5.808	0.000
	(5.209)	(5.164)	

Notes: Descriptive statistics are provided for the baseline estimation sample (age $\geq$ 25, age 17 $\in$ [1985, 2003], wage $\in$ [2004, 2011], and non-missing variables). The t-test compares means of variables between college graduates (S=1) and graduates of secondary schools (S=0). The definition of all variables is described in Appendix Table C1. Standard deviations are in parentheses.

Table 2.4: Maximum Likelihood Estimates of the Normal Switching Regression Model

Variables	**Wage equations**		**College equation**	
	S=0	**S=1**	**Coefficients**	**ME**
Age	0.013	0.055***	...	...
	(0.015)	(0.019)		
Age squared	-0.001	-0.060**	...	...
	(0.023)	(0.028)		
Regional transitory earnings (log)	0.733***	0.780***	...	...
	(0.095)	(0.119)		
Regional transitory unemployment rate, %	0.010**	-0.001	...	...
	(0.005)	(0.006)		
Female	-0.465***	-0.239***	0.374***	0.142***
	(0.013)	(0.017)	(0.020)	(0.008)
Urban residence (binary)	0.212***	0.308***	0.183***	0.069***
	(0.015)	(0.024)	(0.031)	(0.011)
Ethnically Russian (binary)	-0.040**	-0.015	-0.065*	-0.027*
	(0.019)	(0.024)	(0.038)	(0.015)
Mother's education (binary)	-0.041	0.167***	1.053***	0.403***
	(0.034)	(0.036)	(0.029)	(0.010)
Mother's education missing (binary)	-0.021	0.110***	0.156***	0.059***
	(0.017)	(0.023)	(0.031)	(0.012)
Regional cohort size at 17 (log)	0.191***	0.064***	-0.181***	-0.066***
	(0.017)	(0.018)	(0.026)	(0.010)
Regional permanent earnings (log)	0.618***	0.796***	0.281***	0.112***
	(0.040)	(0.051)	(0.051)	(0.019)
Regional permanent unemployment rate, %	0.020***	0.003	-0.000	0.002
	(0.005)	(0.006)	(0.009)	(0.003)

Variables	Wage equations		College equation	
	S=0	**S=1**	**Coefficients**	**ME**
Instruments				
N of campuses per municipality	...	...	0.011***	0.005***
			(0.001)	(0.000)
N of campuses per municipality squared/100	...	...	-0.003***	-0.001***
			(0.000)	(0.000)
Municipality with no campuses (binary)	...	...	-0.312***	-0.110***
			(0.026)	(0.010)
Skill wage ratio at 17	...	...	-1.313***	-0.527***
			(0.368)	(0.145)
Regional earnings at 17 (log)	...	...	0.040	0.016
			(0.043)	(0.017)
Regional unemployment rate at 17, %	...	...	-0.007	-0.003
			(0.006)	(0.002)
$\hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon} = -0.372^{***}(0.065)$				
$\bar{X}(\hat{\beta}_1 - \hat{\beta}_0) = 0.330^{***}(0.057)$; annualized=7.3				
χ^2 - test for joint significance of instruments = 360.9***				
χ^2 - test for independence of equations = 40.6***				

Notes: *** Significant at 1%. ** Significant at 5 %. * Significant at 10 %. N = 17,911 (age $\geq$ 25, age 17 $\in$ [1985, 2003], wage $\in$ [2004, 2011], and non-missing variables). This table shows the maximum likelihood estimates of wage and college equations for the normal switching regression model. Robust standard errors are in parentheses. The standard error of $(\sigma_{1\epsilon} - \sigma_{0\epsilon})$ is computed using the Delta method. The college equation also includes dummies for four cohorts, Moscow residence at age 17, and seven federal districts at age 17; wage equations include dummies for survey years, current Moscow residence, and seven federal districts at the time of the survey.

Table 2.5: College Equation: Alternative Set of Instruments

Variables	(1)	(2)	(3)	(4)	(5)
N of colleges (main campuses)	0.007***	...	...	...	...
	(0.002)				
N of colleges squared/100	-0.002***	...	...	...	...
	(0.000)				
N of branches	0.065***	...	...	...	...
	(0.013)				
N of branches squared/100	-0.287***	...	...	...	...
	(0.100)				
Municipality with no campuses (dummy)	-0.258***	-0.250***	-0.324***	-0.276***	-0.312***
	(0.030)	(0.029)	(0.026)	(0.027)	(0.026)
N of public campuses	...	0.024***	...	...	...
	(0.002)				
N of public campuses squared/100	...	-0.018***	...	...	...
	(0.003)				
Municipality with a private campus (dummy)	...	-0.008	...	...	...
		(0.031)			
N of campuses	...	...	0.009***	...	0.011***
			(0.001)		(0.001)
x Moscow residency at 17	...	...	-0.009***	...	...
			(0.001)		
N of public colleges	...	...	...	0.025***	...
				(0.002)	
N of public colleges squared/100	...	...	...	-0.020***	...
				(0.003)	

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
N of campuses squared/100	...	...	...	...	-0.003***
					(0.000)
Change in N of campuses between ages 25-30	...	...	...	...	-0.001
					(0.001)
$\sigma_{1\epsilon} - \sigma_{0\epsilon}$	-0.353***	-0.352***	-0.387***	-0.373***	-0.370***
	(0.066)	(0.066)	(0.063)	(0.064)	(0.066)
$\bar{X}(\hat{\beta}_1 - \hat{\beta}_0)$	0.374***	0.369***	0.318***	0.347***	0.338***
	(0.061)	(0.060)	(0.054)	(0.058)	(0.059)
χ^2 - test for joint significance of instruments	388.4	412.2	352.3	394.4	362.1
p -value	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
χ^2 - test for independence of equations	49.6	49.2	44.0	50.3	41.6
p -value	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log-likelihood	-24,466	-24,451	-24,480	-24,460	-24,476

Notes: *** Significant at 1%. ** Significant at 5 %. * Significant at 10 %. N=17,911. Table shows the maximum likelihood estimates of the coefficients of the N of campuses per municipality in the college equation using alternative definitions. The number of campuses, colleges, and branches is computed at the level of municipality. All specifications use the same set of variables and the same sample constraints as in the baseline specification reported in Table A.4 (age $\geq$ 25, age 17 $\in$ [1985, 2003], wage $\in$ [2004, 2011], and non-missing variables). The standard error of $(\sigma_{1\epsilon} - \sigma_{0\epsilon})$ is computed using the Delta method. Robust standard errors are in parentheses.

Table 2.6: Switching Regression Model Parameters from Alternative Specifications

	Includes dropouts	**10-year earnings average**	**Soviet earnings**	**IPW**	**Wage 1995-2011**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
$\hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon}$	-0.363	-0.353	-0.365	-0.369	-0.355
	(0.064)	(0.070)	(0.064)	(0.053)	(0.061)
$\bar{X}(\hat{\beta}_1 - \hat{\beta}_0)$	0.345	0.367	0.355	0.397	0.434
	(0.054)	(0.077)	(0.064)	(0.047)	(0.057)
χ^2 - test for joint significance of instruments	370.6	164.0	170.1	517.3	237.8
p -value	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
χ^2 - test for independence of equations	39.4	41.2	40.6	32.1	81.5
p -value	[0.003]	[0.003]	[0.001]	[0.000]	[0.000]
Log-likelihood	-26,100	-24,680	-24,750	-30,589	-31,592
Number of observations	18,823	17,911	17,911	17,911	22,584

Notes: Table reports statistics for alternative specifications of the normal switching regression model. Unless noted otherwise, all specifications use the same set of variables and the same sample constraints as in the baseline specification reported in Table A.4 (age $\geq$ 25, age 17 $\in$ [1985, 2003], wage $\in$ [2004, 2011], and non-missing variables). In column 1, the treatment group, S=1, includes college dropouts with three or more years of college. In column 2, we use an alternative definition of regional permanent earnings calculated as average regional earnings over the first ten years since college decision at age 17. The specification in column 3 applies the same definition as in column 2 but replaces permanent earnings with the Soviet period average regional earnings (1980-1991) for individuals who turned 17 before market reforms. In column 4, we apply the inverse propensity weight from the probit model of non-missing wages on covariates X^S and X^W. In column 5, wage $\in$ [1995, 2011]. The standard error of ($\hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon}$) is computed using the Delta method. Robust standard errors are in parentheses, and p-values are in brackets.

Table 2.7: Treatment Parameters

Treatment Parameter	**Estimates Return**
Average Treatment Effect (ATE)	0.073
	(0.026)
Treatment on the Treated (TT)	0.130
	(0.041)
Treatment on the Untreated (TUT)	0.019
	(0.020)
Instrumental Variables (IV)	0.161
	(0.051)
Ordinary Least Squares (OLS)	0.077
	(0.024)
Sorting gain (TT - ATE)	0.057
Selection bias (OLS - TT)	-0.053

Notes: The standard treatment parameters are obtained by integrating the $MTE(x, \nu)$ using weighting functions of Heckman and Vytlacil (2005):

$$Parameter(x) = \int_0^1 MTE(x, \nu) h_j(x, \nu) d\nu, wherej = ATE, TT, TUT, IV, OLS$$

The weighting functions $h_j(x, \nu)$ are defined in Appendix Table C3. Linear IV estimates use $P(Z)$ as an instrument. Bootstrapped standard errors are in parentheses (250 replications). The reported returns are annualized by dividing the estimates by 4.5, which is the difference in average years of schooling between treatment and control groups. Table reports the results under baseline specification (Table A.4).

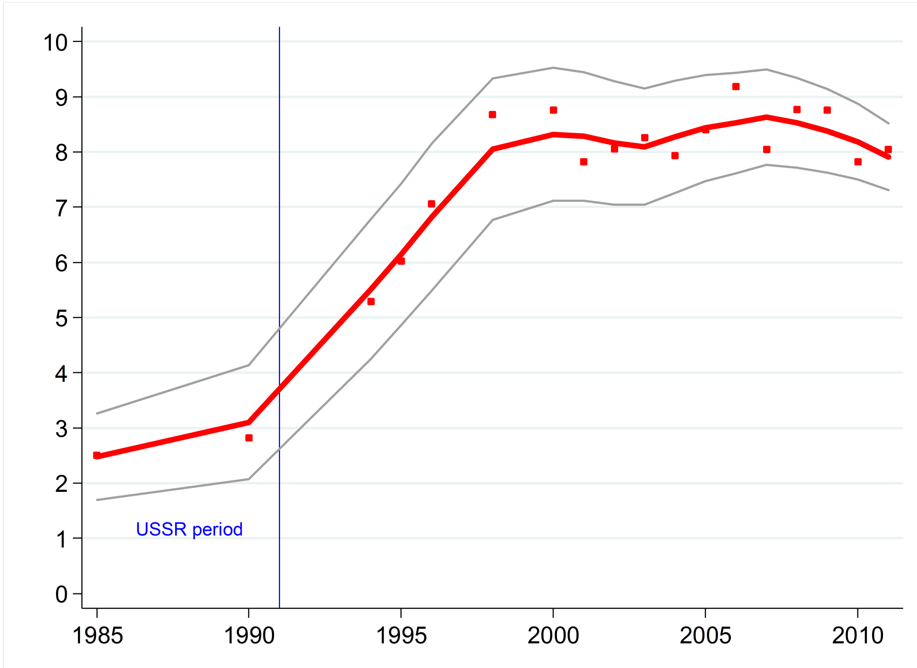
Table 2.8: Policy Parameters

Policy Experiment	**Policy Parameter**
Panel A: MPRTE from the normal selection model (for infinitesimal α)	
$P_\alpha = P(Z) + \alpha$	0.099
	(0.011)
$P_\alpha = (1+\alpha)P(Z)$	0.074
	(0.013)
$P_\alpha = P(N_{m,t} + \alpha)$	0.096
	(0.011)
Panel B: MPRTE from the semi-parametric model (for infinitesimal α)	
$P_\alpha = P(Z) + \alpha$	0.110
	(0.012)
$P_\alpha = (1+\alpha)P(Z)$	0.086
	(0.014)
$P_\alpha = P(N_{m,t} + \alpha)$	0.107
	(0.012)
Panel C: PRTE from policy simulations	
Adding a campus in each municipality	0.098
$N^*_{m,t} = N_{m,92} + 1$	(0.011)
Opening a first campus in 1993-2003	0.099
$N^*_{m,t} = 1$ if $N_{m,92} = 0, t \in [1993, 2003]$	(0.011)
Opening an additional campus in 1993-2003	0.077
$N^*_{m,t} = N_{m,92} + 1$ if $N_{m,92} > 0, t \in [1993, 2003]$	(0.012)
Adding a campus in capital cities in 1993-2003	0.075
$N^*_{m,t} = N_{m,92} + 1, t \in [1993, 2003]$	(0.012)

Policy Experiment	Policy Parameter
Adding a campus in non-capital cities in 1993-2003	0.099
$N^*_{m,t} = N_{m,92} + 1, t \in [1993, 2003]$	(0.010)
Adding a campus in rural districts in 1993-2003	0.099
$N^*_{m,t} = N_{m,92} + 1, t \in [1993, 2003]$	(0.011)
50% of actual expansion	0.080
$N^*_{m,t} = N_{m,92} + \frac{1}{2}(N_{m,t} - N_{m,92})$	(0.012)
Actual expansion ($N_{m,t} - N_{m,92}$)	0.060
$N^*_{m,t} = N_{m,t}$	(0.013)
150% of actual expansion	0.062
$N^*_{m,t} = N_{m,92} + \frac{3}{2}(N_{m,t} - N_{m,92})$	(0.013)

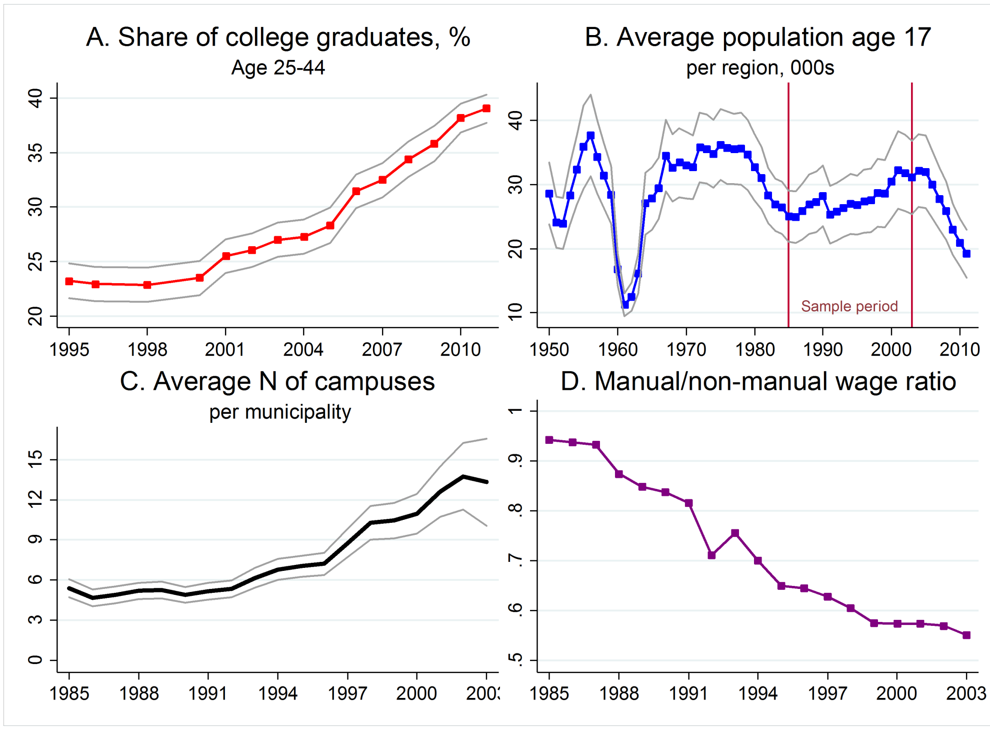
Notes: Table reports marginal policy-relevant treatment effect (MPRTE) for the normal selection and semi-parametric models as well as policy-relevant treatment effect (PRTE) from policy experiments. The reported returns are per year of college. $N_{m,t}$ is the number of campuses in the municipality of residence m at age 17 in year t; $N^*_{m,t}$ is the simulated number of campuses. The MTE estimates are based on equation (2.8) in Panels A and C; equation (2.12) in Panel B. The MPRTE are obtained by integrating the $MTE(x, \nu)$ using the weighting functions reported in Appendix Table C3. The reported PRTE estimates are computed using Quasi-Monte Carlo simulation with Halton sequences; these estimates are essentially identical to the PRTE estimates based on the PRTE weighting function. Bootstrapped standard errors are in parentheses (250 replications).

Figure 2.1: Average Returns to College Education in Russia



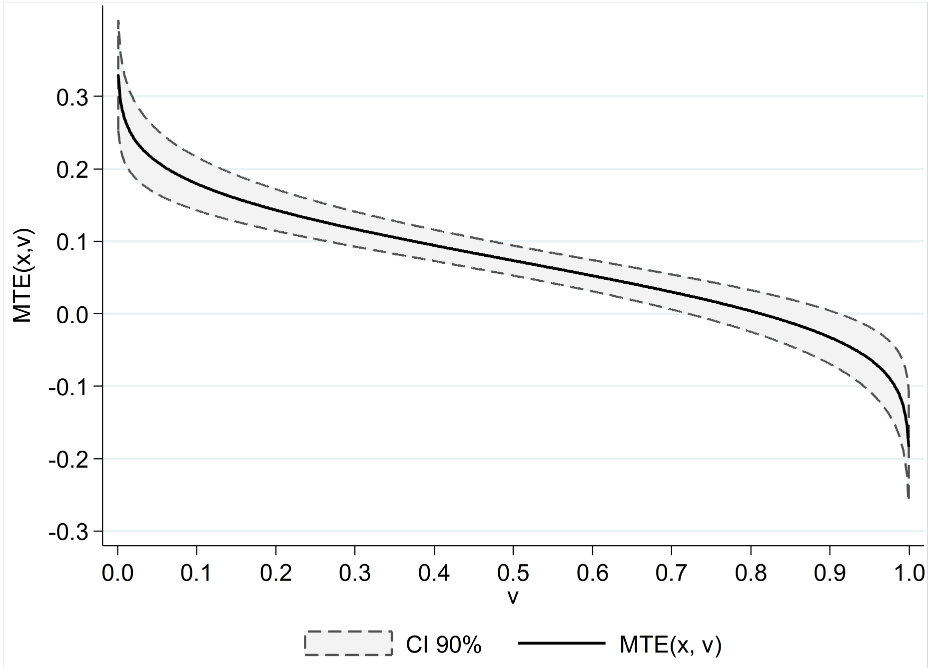
Notes: Returns to college education are calculated from the OLS regression of the log of monthly earnings at primary job on college degree, gender, age, age squared, urban place of birth, a dummy for Russian nationality, and seven federal districts. The comparison group includes graduates of general and professional secondary schools, but excludes college drop-outs. Earnings for 1985 and 1990 are reported retrospectively. Estimation is performed for each year separately using the sample of 25- to 55-year olds. Reported are the estimated coefficients on college degree and the overall trend fitted using non-parametric smoothing (lowess; bandwidth=0.4). Returns are in percent and per year of college (divided by 4.5). Also shown is the 90

Figure 2.2: Trends in Key Variables



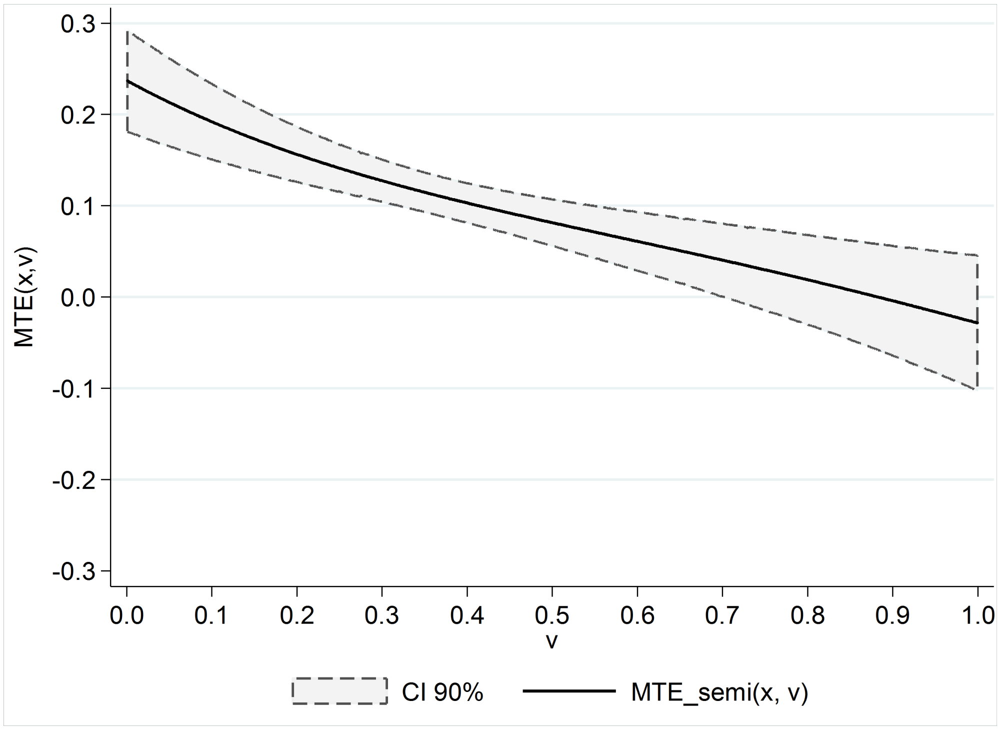
Notes: Panel A shows the percent share of college graduates in RLMS among 25-44-year old individuals. The comparison group includes graduates of general and professional secondary schools, but excludes college drop-outs. Panel B depicts the average size of 17-year-old population in thousands people across all Russian regions. Vertical lines define the sample period from 1985 to 2003. Also shown is the 90

Figure 2.3: Marginal Treatment Effect Estimated from a Normal Selection Model



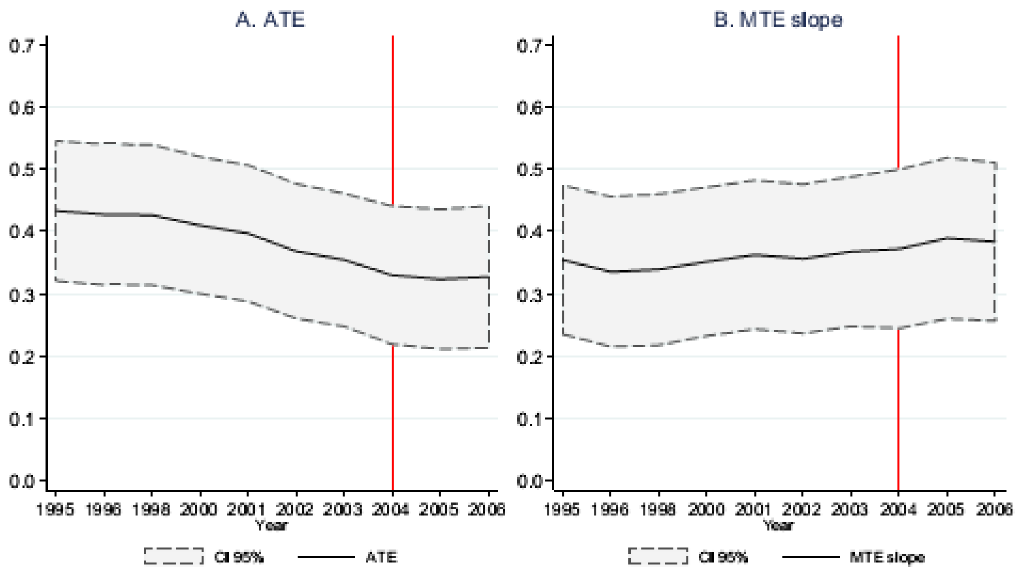
Notes: The figure plots the MTE estimates for a grid of values v between zero and one, with a 90-percent confidence interval, evaluated at mean values of X. We estimate a parametric normal selection model given in equations (2.1)-(2.4) by maximum likelihood. The figure is computed using equation (2.8) and the estimated parameters in Table A.4. The fixed covariance component of $MTE(x,\nu)$ is given by $\hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon}$=-0.372 with a standard error of 0.065. The reported returns are per year of college. The standard errors are computed using the Delta method.

Figure 2.4: Marginal Treatment Effect Estimated from a Semi-Parametric Model with Local IV



Notes: The figure plots the MTE estimates for a grid of values v between zero and one, with a 90-percent confidence interval, evaluated at mean values of X. Estimation steps are described in Section 2. The estimation is performed using a Robinson's (1988) double residual estimator to obtain the non-linear relation between the log of wages and $\hat{p}$, where $\hat{p}$ is the predicted probability of graduating from college. Then, we use a Kernel quadratic local polynomial regression to evaluate the derivatives of equation (8.11), $K'(\hat{p})$, with a bandwidth of 0.32. The reported returns are per year of college. 90 percent standard error bands are calculated using the bootstrap (250 replications).

Figure 2.5: Average Treatment Effect Estimates for Different Sample Periods



Notes: $ATE = \bar{X}(\hat{\beta}_1 - \hat{\beta}_0)$; $MTEslope = \hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon}$. Figure reports the average returns to college (ATE) and the MTE slope along with the 95-percent confidence interval under the baseline specification by adding/subtracting one additional survey year to preserve the panel features of wage series. The last year of each survey period is always 2011 to keep the college decision period (1985-2003) constant for comparability purposes. Year indicates the beginning of the survey period. The red line corresponds to the baseline estimates for the 2004-2011 survey period (Table A.4).

APPENDIX A

DESCRIPTION OF THE VARIABLES I

Table A.1: Description of the Variables

Variable name	**Definition**
	A. Individual-Level Variables
Actual years of schooling	The sum of the years of schooling at all levels of education, including multiple degrees at the same level
Potential experience	(Age-actual years of schooling - 6)
Hourly wage rate	Usual monthly earnings divided by usual monthly hours of work at the primary job. Adjusted for inflation using yearly CPI (year 2000 is the base year).
Region of residence	33 subjects of the Russian Federation or regions, Moscow and Leningrad regions include Moscow city and St. Petersburg, respectively.
	B. Region-Level Variables
Share of men with higher education	Share of male population older than 15 with complete higher education. The share varies by region, type of location (urban vs. rural), and age. Sources: Russian Census 2002, 2010; Demoscope Weekly (www.demoscope.ru/); values in 1995-2001 are imputed based on the shares in 1989 and 2002 for a given cohort, values in 2003-2009, 2011 are imputed based on the share in 2010 for a given cohort.
Share of women with higher education	Share of female population older than 15 with complete higher education. The share varies by region, type of location (urban vs. rural), and age. Sources: Russian Census 2002, 2010; Demoscope Weekly (www.demoscope.ru/); values in 1995-2001 are imputed based on the shares in 1989 and 2002 for a given cohort, values in 2003-2009, 2011 are imputed based on the share in 2010 for a given cohort.
Unemployment rate	Regional level of unemployment, percent. Sources: Labor and Employment (various years); Population Survey on Employment (various years)
Number of campuses	Total number of campuses of higher education institutions located in the region of current residence twenty years ago. Sources: Russian University Database

Table A.2: Classification of Industries

Economic activity	1994	1995-2002	2003-2004	2005-2010	2011
1. Manufacturing	Manufacturing	Manufacturing	Manufacturing	Mining, quarrying, oil and gas extraction	Mining, quarrying, oil and gas extraction
2. Agriculture	Agriculture, forestry	Agriculture, forestry	Agriculture	Agriculture	Agriculture
3. Construction	Construction	Construction	Construction	Construction	Construction
4. Transportation	Transportation, communications	Transportation, communications	Transportation, utilities	Transportation, utilities	Transportation, utilities
5. Wholesale and retail trade	Trade, catering, logistics, distribution and supply and supply	Wholesale, retail trade, catering	Wholesale, retail trade	Wholesale, retail trade, repair of motor vehicles, motorcycles, personal and household goods for personal use	Wholesale, retail trade, repair of motor vehicles, motorcycles, personal and household goods for personal use
6. Health care	Healthcare, physical culture, social security	Healthcare, physical culture, social security	Healthcare, physical culture, social security	Healthcare, social services,	Healthcare, social services,
7. Education	Education, culture, art, science and scientific services	Education, culture, art, science and scientific services	Education,	Education,	Education,
8. Other	Information, computer sciences; general commercial activities to support the market and real estate transactions	Information, computer sciences; general commercial activities to support the market and real estate transactions	Housing, communal services; personal services of population; culture and arts; science and scientific services	Hotels, restairants; financial activity; real estate, renting and business activities; public administration	Hotels, restairants; real estate, renting and business activities, social and personal services

Table A.3: Estimates of Education Externalities by Education Group

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Some college or more				
College share,	0.023***	0.027***	0.014***	0.019***	0.012***	0.019***	0.015***	-0.000
men	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
College share,	-0.000	-0.006	0.002	-0.003	0.001	-0.004	-0.007*	0.006*
women	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Education	0.025***	0.009**	0.026***	0.011***	0.022***	0.011**	0.012***	0.008**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Experience	0.018***	0.023***	0.021***	0.026***	0.022***	0.027***	0.027***	0.032***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Experience	-0.054***	-0.059***	-0.058***	-0.063***	-0.064***	-0.069***	-0.081***	-0.073***
squared/100	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Constant	-0.245***	-0.364***	0.154	-0.089	1.438***	1.072***	1.082***	0.849***
	(0.084)	(0.086)	(0.122)	(0.112)	(0.125)	(0.121)	(0.116)	(0.096)
R^2	0.78	0.76	0.79	0.76	0.71	0.69	0.75	0.74
N	10,047	13,479	10,047	13,479	9,126	12,511	9,126	12,511
				B. Less than college				
College share,	0.023***	0.025***	0.018***	0.018***	0.017***	0.020***	0.029***	0.015***
men	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
College share,	0.005	0.001	0.007*	0.004	0.005	0.001	-0.011***	-0.004
women	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
Education	0.044***	0.054***	0.044***	0.057***	0.047***	0.056***	0.031***	0.054***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Experience	0.003	0.008***	0.004**	0.010***	0.006***	0.010***	0.008***	0.018***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
Experience	-0.016***	-0.014***	-0.018***	-0.017***	-0.024***	-0.018***	-0.036***	-0.040***
squared/100	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Constant	-0.678***	-1.234***	-0.427***	-0.939***	0.653***	0.280**		2.480***
	(0.086)	(0.085)	(0.104)	(0.104)	(0.140)	(0.118)	()	(0.081)
R^2	0.78	0.78	0.78	0.78	0.72	0.72	0.78	0.77
N	30,074	27,181	30,074	27,181	26,705	24,378	26,705	24,378
Unemployment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

Table A.4: Instrumental Variable Estimates - Endogenous Schooling

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Some college or more				
First stage								
Number of	0.226***	0.215***	0.208***	0.195***	0.190***	0.178***	0.112***	0.124***
campuses	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.030)	(0.027)
Military	0.186***	0.203***	0.182***	0.203***	0.170***	0.159***	0.188***	0.129***
service	(0.018)	(0.022)	(0.018)	(0.022)	(0.020)	(0.025)	(0.016)	(0.020)
F test	937.17	1224.99	901.86	1179.30	878.64	1172.57	728.70	739.81
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	0.031***	0.039***	0.024***	0.040***	0.021***	0.045***	0.062**	0.124
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.004)	(0.025)	(0.087)
Education	-0.073***	-.342***	-0.061**	-0.348***	-0.104***	-0.462***	-0.197***	-0.775
	(0.025)	(0.091)	(0.025)	(0.093)	(0.029)	(0.159)	(0.066)	(0.531)
Experience	0.010***	-0.000	0.013***	-0.000	0.011***	-0.008	-0.008	-0.072
	(0.002)	(0.005)	(0.002)	(0.005)	(0.003)	(0.008)	(0.014)	(0.074)
Experience	-0.038***	-0.000	-0.045***	0.002	-0.045***	0.010	0.004	0.223
squared/100	(0.007)	(0.009)	(0.006)	(0.010)	(0.007)	(0.013)	(0.033)	(0.224)
N	10,047	13,479	10,047	13,479	9,126	12,511	9,126	12,511
Unemployment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

	Men	Women	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				B. Less than college				
First stage								
Number of	0.258***	0.250***	0.244***	0.237***	0.233***	0.222***	0.261***	0.185***
campuses	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.022)	(0.022)
Military	0.212***	0.145***	0.214***	0.147***	0.206***	0.135***	0.170***	0.122***
service	(0.012)	(0.014)	(0.012)	(0.014)	(0.013)	(0.017)	(0.010)	(0.014)
F test	1942.73	2114.96	1867.36	2043.50	1815.83	1942.38	2216.76	1355.00
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Second stage								
College share	0.034***	0.038***	0.027***	0.033***	0.023***	0.035***	-0.11	0.008
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.007)	(0.011)
Education	-0.004	-0.011	0.022	0.008	0.027	0.023	0.131***	0.105***
	(0.021)	(0.028)	(0.021)	(0.028)	(0.023)	(0.033)	(0.032)	(0.038)
Experience	-0.000	-0.005*	0.003*	-0.000	0.005**	-0.002	0.028***	0.022**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.005)	(0.009)
Experience	-0.15***	0.031***	-0.021***	0.020***	-0.024***	0.028***	-0.061***	-0.033
squared/100	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)	(0.008)	(0.025)
N	30,074	27,181	30,074	27,181	26,705	24,378	26,705	24,378
Unempoyment			Yes	Yes	Yes	Yes	Yes	Yes
KM index					Yes	Yes	Yes	Yes
Region FE							Yes	Yes

APPENDIX B

MATHEMATICAL APPENDIX

$$w_0^g = \frac{\partial y}{\partial L_0^g}$$

$$= \alpha_0 \theta_0^g (\theta_0^g L_0^g)^{\alpha_0 - 1} (\theta_1^g L_1^g)^{\alpha_1} K^{g 1-\alpha_1-\alpha_0}$$

$$= \alpha_0 \theta_0^{g \alpha_0} \left((1 - \frac{L_1^g}{L_0^g + L_1^g})(L_0^g + L_1^g)\right)^{\alpha_0 - 1} \left(\theta_1^g (\frac{L_1^g}{L_0^g + L_1^g})(L_0^g + L_1^g)\right)^{\alpha_1} K^{g 1-\alpha_1-\alpha_0}$$

$$= \alpha_0 \theta_0^{g \alpha_0} \left((1 - \frac{L_1^g}{L_0^g + L_1^g})\right)^{\alpha_0 - 1} \left(\theta_1^g (\frac{L_1^g}{L_0^g + L_1^g})\right)^{\alpha_1} \frac{K^g}{L_0^g + L_1^g}^{1-\alpha_1-\alpha_0}$$

$$log w_0^g = log \alpha_0 + \alpha_0 log \theta_0^g + (\alpha_0 - 1) log(1 - \frac{L_1^g}{L_0^g + L_1^g}) + \alpha_1 log \theta_1^g (\frac{L_1^g}{L_0^g + L_1^g})$$

$$+ (1 - \alpha_1 - \alpha_0) log(\frac{K^g}{L_0^g + L_1^g})$$

$$= log \alpha_0 + \alpha_0 log(\phi_0^g + \mu^g (\frac{L_1^g}{L_0^g + L_1^g})) + (\alpha_0 - 1) log(1 - \frac{L_1^g}{L_0^g + L_1^g})$$

$$+ \alpha_1 log(\phi_1^g + \mu^g (\frac{L_1^g}{L_0^g + L_1^g}))(\frac{L_1^g}{L_0^g + L_1^g}) + (1 - \alpha_1 - \alpha_0) log(\frac{K^g}{L_0^g + L_1^g})$$

$$w_1^g = \frac{\partial y}{\partial L_1^g}$$

$$= \alpha_0 (\theta_0^g L_0^g)^{\alpha_0} \theta_1^g (\theta_1^g L_1^g)^{\alpha_1 - 1} K^{g 1-\alpha_1-\alpha_0}$$

$$= \left(\theta_0^g (1 - \frac{L_1^g}{L_0^g + L_1^g})(L_0^g + L_1^g)\right)^{\alpha_0} \alpha_1 \theta_1^{g \alpha_1} \left((\frac{L_1^g}{L_0^g + L_1^g})(L_0^g + L_1^g)\right)^{\alpha_1 - 1} K^{g 1-\alpha_1-\alpha_0}$$

$$= \left(\theta_0^g (1 - \frac{L_1^g}{L_0^g + L_1^g})\right)^{\alpha_0} \alpha_1 \theta_1^{g \alpha_1} \left((\frac{L_1^g}{L_0^g + L_1^g})\right)^{\alpha_1 - 1} \left(\frac{K^g}{L_0^g + L_1^g}\right)^{1-\alpha_1-\alpha_0}$$

$$log w_1^g = \alpha_0 log \theta_0^g (1 - \frac{L_1^g}{L_0^g + L_1^g}) + log \alpha_1 + \alpha_1 log \theta_1^g + (\alpha_1 - 1) log(\frac{L_1^g}{L_0^g + L_1^g})$$

$$+ (1 - \alpha_1 - \alpha_0) log(\frac{K^g}{L_0^g + L_1^g})$$

$$= \alpha_0 log(\phi_0^g + \mu^g \frac{L_1^g}{L_0^g + L_1^g})(1 - \frac{L_1^g}{L_0^g + L_1^g}) + log\alpha_1 + \alpha_1 log(\phi_1^g + \mu^g \frac{L_1^g}{L_0^g + L_1^g})$$

$$+ (\alpha_1 - 1) log(\frac{L_1^g}{L_0^g + L_1^g}) + (1 - \alpha_1 - \alpha_0) log(\frac{K^g}{L_0^g + L_1^g})$$

APPENDIX C

DESCRIPTION OF THE VARIABLES II

Table C.1: Description of Variables

General notes:

1. The source for all individual-level variables is RLMS.

2. Estimation sample covers 1995-1996, 1998, 2000-2011 time periods; variables are available for all years, unless noted otherwise.

3. The sample is restricted to individuals born between 1968 and 1986 and older than 25 at the time of the survey.

4. Russian regions include two federal cities (Moscow city and St. Petersburg) and 81 territories, which are aggregated into seven federal districts.

5. Due to multiple changes in the administrative-territorial structure of Russia, all past regional data are collected based on the classification of regions according to the 2008 amendment to the Constitution of the Russian Federation.

6. At the time of the survey, respondents resided in 32 regions, but they graduated from universities located in 73 regions.

7. Municipality in RLMS is defined according to the 5-digit municipality code taken from the National Classification of Administrative and Territorial Division (OKATO in Russian). OKATO code consists of two digits for a region and three digits for either city or district (county-equivalent) within a region.

8. Municipality at age 17 is the same as current municipality for individuals who moved to current municipality before 17, finished secondary school in current location, or currently reside in the place of their birth. Municipality at age 17 is the same as college location if individuals resided in the same place before college. Municipality at age 17 is imputed by randomly selecting a municipality for a given type of birthplace (city, township, and village) within the region of current residence for individuals who did not attend college and moved

to a new location after age 17. Municipality at age 17 is imputed by randomly selecting a municipality for a given type of birthplace (city, township, and village) within the region of college for individuals who attended college and moved to a new location after age 17.	
Dependent Variables	
Hourly wage rate (log)	= Labor earnings per month at primary job / Hours of work per month at primary job Labor earnings per month are defined as: - monthly average (over the last 12 months) after-tax labor earnings for an employee [1998-2011]; - total accumulated wage debt divided by the number of months of overdue wages for employees with wage arrears [1995-1996]; monetary portion of last month earnings for employees with no wage arrears [1995-1996]; - monetary portion of last month earnings for self-employed; self-employed include individuals reporting place of work other than an organization as well as those involved in regular individual economic activities [all years]. Hours of work per month are defined as follows: - usual hours of work in a typical week times four [1998-2011]; - actual hours of work last month [1995-1996]; - unusually high hours are top coded at 480 hours per month (16 hours per day times 30 days).
College degree (binary)	Two alternative definitions: (i) = 1 if an individual has a college degree or higher, = 0 if an individual graduated from a general and/or professional secondary school with credentials;

	(ii) = 1 if an individual has three or more years of college education with or without college degree, = 0 if an individual graduated from a general and/or professional secondary school.
Common vector $\mathbf{X}^S$	
Female	= 1 if female
Ethnically Russian (binary)	= 1 if ethnicity is Russian
Mother's education (binary)	= 1 if mother has a college degree *Note: Available in 2006 and 2011 surveys; extrapolated to other years based on individual panel id for cases of consistent reporting of mother's education.*
Mother's education is missing	= 1 if mother's education is missing
Regional cohort size at 17 (log)	Log of 17-year old population in the region of residence at the age of 17, 1985-2003. *Sources: Census 1979, 1989, 2002; Demoscope Weekly (www.demoscope.ru/); Goskomstat Central Statistical Database.*
Regional permanent earnings (log)	Three alternative definitions: (i) Predicted earnings = predicted value from the regression of the log of regional earnings on 83 regional dummies and a quadratic trend, 1994-2011; (ii) 10-year average earnings = the log of the average of regional earnings over the first ten years since college decision at age 17; (iii) 10-year average with a structural break = 10-year average earnings as in (ii) for 17-year olds in the market economy after 1991;

	= the log of the average of regional earnings over the 1980-1991 period for 17-year olds in the Soviet period before 1992. *Note: Regional earnings are the average monthly earnings in a given year deflated in 2000 prices.* *Sources: Goskomstat Central Statistical Database; Labor and Employment (various years).*
Regional permanent unemployment rate, %	Predicted value from the regression of regional unemployment rate on 83 regional dummies and a quadratic trend, 1994-2011. *Note: Alternative definitions were not used because of the lack of variation in the unemployment rate during the Soviet period, when unemployment was considered to be non-existent.* *Sources: Goskomstat Central Statistical Database.*
Wage equation only $\mathbf{X}^W$	
Age, age squared	Year of survey minus year of birth; the mode of birth year is computed in cases of inconsistencies across rounds.
Regional transitory earnings (log)	Residual from the regression of the log of regional earnings on 83 regional dummies and a quadratic trend, 1994-2011. *Sources: Goskomstat Central Statistical Database; Labor and Employment (various years).*
Regional transitory unemployment rate, %	Residual from the regression of regional unemployment rate on 83 regional dummies and a quadratic trend, 1994-2011. *Sources: Goskomstat Central Statistical Database*

Federal districts (dummies)	Set of dummies for residing in Moscow and seven federal districts at the time of the interview.
	College equation, E and I
N of campuses per municipality	Total number of college campuses in the municipality of residence at age 17, 1985-2003. A campus refers to all buildings of the same college in one municipality. Each college has a main campus; some colleges may also have branches with campuses located in other municipalities. Subcategories of campuses include main campuses; branches; public campuses; private campuses. *Sources: Russian University Database (Belskaya and S. Peter, 2015).*
Skill wage ratio at 17	Ratio of average wages of manual workers to average wages of non-manual workers in the industrial sector (manufacturing + mining + electricity + selected industrial services) at age 17, 1985-2003. *Sources: Russian yearbooks (annual issues from 1985 to 2003)*
Regional earnings at 17 (log)	Log of regional earnings in the region of residence at age 17, 1985-2003. *Note: Regional earnings are the average monthly earnings in a given year deflated in 2000 prices.* *Sources: Goskomstat Central Statistical Database; Labor and Employment (various years).*

Regional unemployment at 17, %	Unemployment rate in the region of residence at age 17, 1985-2003. *Note: Available for 1992-2003; assumed zero during the Soviet period, when unemployment was considered to be non-existent.* *Sources: Goskomstat Central Statistical Database.*
Birth cohorts (dummies)	Set of dummies for individuals born in 1985-1988, 1994-1998, and 1999-2003.
Federal districts at 17 (dummies)	Set of dummies for living in Moscow and seven federal districts at age 17.

Table C.2: Tests for Selection on Gains

Panel A. Test of linearity of $E[Y_i|X_i = x, P(Z_i) = p]$ using polynomials in P

Degree of Polynomial	**2**	**3**	**4**	**5**
p-value of joint test on nonlinear terms	0.000	0.000	0.000	0.000

Panel B. Test of equality of LATEs $(H0 : LATE_j - LATE_{j+1} = 0)$

Ranges of ν for LATE **j**	**j+1**	**Difference in LATEs**	**p-value**	**Joint p-value**
(0.00, 0.04)	(0.08, 0.12)	0.035	0.000	0.0000
(0.08, 0.12)	(0.16, 0.20)	0.029	0.000	
(0.16, 0.20)	(0.24, 0.28)	0.025	0.000	
(0.24, 0.28)	(0.32, 0.36)	0.021	0.000	
(0.32, 0.36)	(0.40, 0.44)	0.019	0.000	
(0.40, 0.44)	(0.48, 0.52)	0.017	0.000	
(0.48, 0.52)	(0.56, 0.60)	0.016	0.000	
(0.56, 0.60)	(0.64, 0.68)	0.016	0.000	
(0.64, 0.68)	(0.72, 0.76)	0.017	0.000	
(0.72, 0.76)	(0.80, 0.84)	0.018	0.020	
(0.80, 0.84)	(0.88, 0.92)	0.019	0.044	
(0.88, 0.92)	(0.96, 1.00)	0.019	0.060	

Notes to Panel A: Standard errors are bootstrapped (500 replications) to account for the fact that *P* is estimated. The computation of the test includes two steps. First, we estimate a probit model of college equation and predict the propensity score $\hat{P}(Z)$. Second, we estimate a linear regression of *Y* on *X*, *X(Z)* and $K_j(\hat{P})$, where $K_j(\hat{P})$ is a polynomial of degree $j = 2, 3, 4, 5$. Each column presents the *p*-value of the test associated with the null hypothesis that neither nonlinear term has an explanatory power for each polynomial.

Notes to Panel B: We compute the semi-parametric MTE as described in Section 2 and presented in Figure 12 using 250 bootstrap replications. LATE is the local average treatment effect, and it is defined as the average of MTE in each of the equally-spaced intervals of ν, with the total of 13 non-overlapping intervals *j* separated by a distance of 0.04. The null hypothesis is that the LATEs of two contiguous intervals are equal. The p-value for this null hypothesis is shown in column 4. The p-value is the proportion of bootstrap *b* for which $T_b^j > T^j$ in interval *j*, where $T^j = |LATE^j - LATE^{j+1}|$ and $T_b^j = |(LATE_b^j - LATE_b^{j+1}) - (LATE^j - LATE^{j+1})|$ for $b = 1, 250$. The last column reports the *p*-value associated with the null hypothesis that the differences across all adjacent LATEs are different from zero. The joint *p*-value is the proportion of bootstrap *b* for which $C_b > C$, where

$C = \sum_j (LATE^j - LATE^{j+1})^2$ for all *j* intervals and

$C_b = \sum_j [(LATE_b^J - LATE_b^{j+1}) - (LATE^j - LATE^{j+1})]^2$ for $b = 1, 250$ and all *j* and *j+1* intervals.

REFERENCES

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics 113*(4), 1055–1089.

Acemoglu, D. and J. Angrist (2001). How large are human capital externalities? Evidence from compulsory schooling laws. Ben S. Bernanke and Kenneth Rogoff (eds.). National Bureau of Economics Macroeconomics Annual 2000. MIT Press. Cambridge.

Andren, Daniela, J. E. and D. Sapatoru (2005). The wage effects of schooling under socialism and in transition: Evidence from romania, 1950-2000. *Journal of Comparative Economics 33*(2), 300–323.

Andrienko, Y. and S. Guriev (2004). Determinants of interregional mobility in russia: Evidence from panel data. *Economics of Transition 12*(1), 1–27.

Autor, D. and M. Wasserman (2013). Wayward sons: The emerging gender gap in labor markets and education. *Third Way Report*.

Bank (2003). The russian labor market: Moving from crisis to recovery. *Washington: The World Bank*.

Bank (2005). Enhancing job opportunities in Eastern Europe and the former Soviet Union. *Washington: The World Bank*.

Bank (2014). World Bank Education Statistics.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics 104*, 184–198.

Basu, Anirban, J. J. H. S. N.-L. and S. Urzua (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients. *Health Economics 16*(11), 1133–1157.

Becker, Gary S., W. H. J. H. and K. M. Murphy (2010). Explaining the worldwide boom in higher education of women. *Journal of Human Capital 4*(3), 203–241.

Belskaya, V. and K. S. Peter (2015). How does the expansion of higher education change the returns to college quality? Insights from sixty years of russian history. *UNC Chapel Hill Working Paper*.

Bjorklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics 69*(1), 42–49.

Blau, F. D. and L. M. Kahn (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics 15*(1), 1–42.

Blau, F. D. and L. M. Kahn (2000). Gender differences in pay. *Journal of Economic Perspectives 14*(4), 75–99.

Blau, F. D. and L. M. Kahn (2006). The U.S. gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review 60*(1), 45–66.

Bound, John, J. G. G. K. and S. Turner (2004). Trade in university training: Cross-state variation in the production and stock of college-educated labor. *Journal of Econometrics 121*(1-2), 143–175.

Brainerd, E. (1998). Winners and losers in Russia's economic transition. *American Economic Review 88*(5), 1094–1116.

Bratti, M. and R. Leombruni (2014). Local human capital externalities and wages at the firm level: Evidence from Italian manufacturing. *Economics of Education Review 41*, 161–175.

Cameron, S. and J. J. Heckman (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy 106*(2), 262–333.

Cameron, S. and C. Taber (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy 112*(1), 132–182.

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, ed. by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press*.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Carneiro, Pedro, J. H. and E. Vytlacil (2010). Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica 78*(1), 377–394.

Carneiro, Pedro, J. H. and E. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review*.

Carneiro, Pedro, M. L. C. R.-C. and N. Umapathi (2011). Average and marginal returns to upper secondary schooling in Indonesia. *Policy Research Working Paper*.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Carnoy, Martin, P. L. M. D.-R. D. I. F. K. K. J. B. G. T. and R. Wang (2013). University expansion in a changing global economy: Triumph of the BRICs? *Stanford, California: Stanford University Press*.

Charles, K. K. and M.-C. Luoh (2003). Gender differences in completed schooling. *Review of Economics and Statistics 85*(3), 359–577.

Ciccone, A. and G. Peri (2006). Identifying human-capital externalities: Theory with applications. *Review of Economics and Statistics 73*(2), 381–412.

Conley, Timothy G., F. F. and G. R. Tsiang (2003). Spillovers from local market human capital and the spatial distribution of productivity in Malaysia. *Advances in Economic Analysis Policy 3*(1), 1–45.

Currie, J. and E. Moretti (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics 118*(4), 1495–1532.

Dalmazzo, A. and G. de Blasio (2007a). Production and consumption externalities of human capital: An empirical study for Italy. *Journal of Population Economics*.

Dalmazzo, A. and G. de Blasio (2007b). Social returns to education in Italian local labor markets. *Annals of Regional Science*.

Demography (2002). Demography yearbook of Russia. *Moscow: Goskomstat*.

Dias, Joilson, F. S. and E. Tebaldi (2014). Geographic and sector externalities from highly qualified human capital: The importance of the business service sector. *Applied Economics Letters 21*(5), 329–334.

Dougherty, C. (2005). Why are the returns to schooling higher for women than for men? *Journal of Human Resources 40*(4), 969–988.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(4), 795–813.

Education (2008). Education in Russia. *Moscow: Goskomstat*.

Eisenhauer, Philipp, J. J. H. and E. Vytlacil (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

Evans, H. and A. Basu (2011). Exploring comparative effect heterogeneity with instrumental variables: Prehospital intubation and mortality. *Health, Econometrics and Data Group Working Papers 11/26*.

Fang, Hai, K. E. J. R.-S. R. and R. Zeckhauser (2012). The returns to education in china: Evidence from the 1986 compulsory education law. *Working Paper 18189, National Bureau of Economic Research, Cambridge, MA*.

Fleisher, Belton, K. S. P. and X. Wang (2005). Returns to skills and the speed of reforms: Evidence from Central and Eastern Europe, China, and Russia. *Journal of Comparative Economics 33*(2), 351–370.

Frenette, M. (2009). Do universities benefit local youth? Evidence from the creation of new universities. *Economics of Education Review 28*(3), 318–328.

Ganguli, I. (2013). Fact or fiction? the reversal of the gender education gap across the world and the former soviet union. *FREE Policy Brief Series*, 1–6.

Garen, J. (1984). The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica 52*(5), 1199–1218.

Gayle, G.-L. and L. Golan (2012). Estimating a dynamic adverse-selection model: Labour-force experience and the changing gender earnings gap. *Review of Economic Studies 79*, 227–267.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American Economic Review 96*(2), 1–21.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104*(4), 1091–1119.

Gorodnichenko, Y. and K. S. Peter (2005). Returns to schooling in Russia and Ukraine: A semiparametric approach to cross-country comparative analysis. *Journal of Comparative Economics 33*(2), 324–350.

Groen, J. A. (2004). The effect of college location on migration of college-educated labor. *Journal of Econometrics 121*(1-2), 125–142.

Guriev, S. and E. Vakulenko (2014). Breaking out of poverty traps: Internal migration and interregional convergence in Russia. *SSR Working Paper*.

Heckman, James, D. S. and S. Urzua (2010). Testing the correlated random coefficient model. *Journal of Econometrics 158*(2), 177–203.

Heckman, J. and X. Li (2004). Selection bias, comparative advantage and heterogeneous returns to education: Evidence from China in 2000. *Pacific Economic Review 9*(3), 155–171.

Heckman, J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

Heckman, J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. and E. Vytlacil (2007). Econometric evaluation of social programs. *Handbook of Econometrics 6B*, 4877–5143.

Heckman, James J., J. E. H.-G. V. and S. Urzua (2014). Education, health and wages. *Working Paper No. 19971, National Bureau of Economic Research, Cambridge, MA*.

Heckman, J. J. and P. A. LaFontaine (2010). The american high school graduation rate: Trends and levels. *Review of Economics and Statistics 92*(2), 244–262.

Heckman, James, L. L. and P. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*(7), 307–458.

Ichino, A. and R. Winter-Ebmer (2004). The long-run educational cost of World War II. *Journal* *of Political Economy 22*(1), 57–87.

Iranzo, S. and G. Peri (2009). Schooling externalities, technology, and productivity: Theory and evidence from U.S. states. *Review of Economics and Statistics 91*(2), 420–431.

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review 21*(6), 589–598.

Kane, T. and C. Rouse (1995). Labor - market returns to two- and four-year college. *American Economic Review 85*(3), 600–614.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics 107*(1), 35–78.

Kirby, S. and R. Riley (2008). The external returns to education: UK evidence using repeated cross-sections. *Labour Economics 15*(4), 619–630.

Kyui, N. (2013). Expansion of higher education, employment and wages: Evidence from the russian transition. *Bank of Canada Working Paper 2013-45*.

Lange, F. and R. Topel (2006). The social value of education and human capital. *Handbook of the Economics of Education, ed. Eric A. Hanushek and Finis Welch. Amsterdam and Oxford: Elsevier, North-Holland 1*, 459–509.

Law (1992). Ob obrazovanii [On education], July 10, 1992. *Available at http://www.consultant.ru/popular/edu/*.

Liu, Z. (2007). The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics 61*(3), 542–564.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3–42.

Moffitt, R. (2008). Estimating marginal treatment effects in heterogeneous populations. *Annales dEconomie et de Statistique 91/92*, 239–261.

Moffitt, R. (2014). Estimating marginal treatment effects of transfer programs on labor supply. *Working Paper*.

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics 121*(1), 175–212.

Munich, Daniel, J. S. and K. Terrell (2005). Returns to human capital under the communism wage grid and during the transition to a market economy. *Review of Economics and Statistics 83*(1), 100–123.

Muravyev, A. (2008). Human capital externalities: Evidence from the transition economy of Russia. *Economics of Transition 16*(3), 415–443.

OECD (2008). Education at a glance: OECD indicators. *OECD Publishing*.

OECD (2012). Closing the gender gap: Act now. *OECD Publishing*.

Oppedisano, V. (2011). The (adverse) effects of expanding higher education: Evidence from italy. *Economics of Education Review 30*(5), 997–1008.

Parro, F. (2012). International evidence on the gender gap in education over the past six decades: A puzzle and an answer to it. *Journal of Human Capital 6*(2), 150–185.

Pekkarinen, T. (2012). Gender differences in education. *IZA Discussion Paper No. 6390*.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics 34*(3), 380–400.

Robinson, P. (1988). Root-N-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rosenthal, S. S. and W. C. Strange (2008). The attenuation of human capital spillovers. *Journal of Urban Economics 64*(2), 373–389.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Sand, B. M. (2013). A re-examination of the social returns to education: Evidence from U.S. cities. *Labour Economics 24*, 97–106.

Topel, R. (1997). Factor proportions and relative wages: The supply-side determinants of wage inequality. *Journal of Economic Perspectives 11*(2), 5574.

Willis, R. and S. Rosen (1979). Education and self-selection. *Journal of Political Economy 87*(5), S7–S36.

Winters, J. V. (2011). Why are smart cities growing? Who moves and who stays? *Journal of Regional Science 51*(2), 253–270.

Yang, D. T. (2005). Determinants of schooling returns during transition: Evidence from Chinese cities. *Journal of Comparative Economics 33*(2), 244–264.

Zamarro, G. (2010). Accounting for heterogeneous returns in sequential schooling decisions. *Journal of Econometrics 156*(2), 260–276.