College Expansion and the Marginal Returns to Education: Evidence from Russia

December, 2014

Olga Belskaya
UNC-Chapel Hill
belskaya@email.unc.edu

Klara Sabirianova Peter
IZA and CEPR
UNC-Chapel Hill
kpeter@unc.edu

Christian Posso
UNC-Chapel Hill
cpossosu@email.unc.edu

Abstract

This paper 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. Our 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. We use high-quality panel data with multiple wage observations, many birth cohorts, disaggregated location information, and past economic conditions. We find that college expansion attracts individuals with lower returns to college, but the returns for marginal students vary considerably depending on the scale of expansion and the type of location where new campuses are opened. Marginal individuals in smaller cities and locations without college campuses receive the largest benefits from new campuses. The results provide important implications for the design of policies targeting the expansion of higher education.

Keywords: education, returns to college, marginal treatment effect, policy-relevant treatment effect, local instrumental variable, selection bias, college expansion, Russia

JEL Code: J31, I21, P36

Acknowledgement:
We are thankful to Brian McManus, Saraswata Chaudhuri, Donna Gilleskie, and Tiago Pires for useful comments. We also acknowledge helpful comments received from participants of the UNC Applied Microeconomics Workshop and Southern Economic Association meetings in Atlanta, GA.

Web appendix:
http://belskaya.web.unc.edu/files/2014/09/RUS-education.pdf
http://belskaya.web.unc.edu/files/2014/09/MTE-expansion.pdf

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 *et al.* 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 legalized tuition-based programs and branches in public universities (Federal 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 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 Russia's transition to a market economy, but then the premium stabilized.

[1] We provide a brief summary of the Soviet and Russian systems of higher education in the web appendix.
[2] Sources are Regions of Russia (1995, 2004) and authors' calculations.
[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 paper 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 paper.

Our empirical strategy relies on the marginal treatment effect (MTE) method in its normal and semi-parametric versions (Heckman and Vytlacil 2001, 2005, 2007) and a construction of policy treatment parameters (Heckman and Vytlacil 2001; Carneiro *et al.* 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 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

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

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 *et al*. 2011b).[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.

Our paper 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 *et al.* 2010, 2011b; Carneiro and Lee 2009; Heckman *et al.* 2014;

[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, Björklund and Moffitt 1987).

[6] Outside the field of education, the MTE method is only beginning to be applied in a systematic way; e.g., see Basu *et al*. (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.

Heckman and Vytlacil 2001, 2005). MTE estimates outside the U.S. are rare and often limited to a single cross section (Carneiro *et al*. (2011a) 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 *et al.* (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 premium. Carneiro *et al*. (2011b) 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

[7] Carneiro *et al*. (2011a) 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.

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 paper 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 (Andrén *et al.* 2005, Brainerd 1998, Fang *et al.* 2012, Fleisher *et al.* 2005, Gorodnichenko and Sabirianova Peter 2005, Münich *et al.* 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 the later period.

Our paper is structured as follows. In the next section, we present the empirical framework used in this study. Section 3 discusses the data and identifying variables. Section 4 presents

[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 Sabirianova Peter 2014).

estimates of the marginal treatment effect and discusses the identification validity. Section 5 performs policy simulations and Section 6 concludes.

2. Econometric Framework

Model set-up

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

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

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:

$$S_i = \mathbf{1}\left(E_i'\gamma_E + X_i^{S'}\gamma_X + I_i'\gamma_I > \epsilon_i\right) \quad (2)$$

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 $\mathrm{P}(Z)$ denote the probability of selecting into treatment (college) given Z. It is convenient to rewrite the selection equation (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

[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).

the unit interval [0, 1]. Different values of v_i denote different percentiles of ϵ_i. Thus, the selection equation becomes

$$S_i = \mathbf{1}(P(Z_i) > v_i) \tag{3}$$

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

$$Y_{i,s} = \mu_s(X_i) + \varepsilon_{i,s} \text{ for } s = 0, 1 \tag{4}$$

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 individual's 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:

$$\begin{aligned} Y_i &= S_i Y_{\mathrm{i},1} + (1 - S_i) Y_{\mathrm{i},0} \\ &= S_i\left(X_i'\beta_1 + \varepsilon_{\mathrm{i},1}\right) + (1 - S_i)\left(X_i'\beta_0 + \varepsilon_{\mathrm{i},0}\right) \\ &= X_i'\beta_0 + S_i(X_i'\beta_1 - X_i'\beta_0) + S_i\left(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0}\right) + \varepsilon_{\mathrm{i},0} \end{aligned} \tag{5}$$

The gross returns to college education are given by $\Delta_i = Y_{\mathrm{i},1} - Y_{\mathrm{i},0} = (X_i'\beta_1 - X_i'\beta_0) + \left(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0}\right)$ 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:

$$\begin{aligned} \mathrm{E}[Y_i \mid X_i = x, P(Z_i) = p] &= \mathrm{E}[Y_i \mid S_i = 1, x, p]\, p + \mathrm{E}[Y_i \mid S_i = 0, x, p]\,(1 - p) \\ &= x'\beta_0 + (x'\beta_1 - x'\beta_0)p + \int_0^p \mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0}\right) \mid X_i = x, v_i = v\right] dv \end{aligned} \tag{6}$$

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 (ε's) 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 (6) cancels out; (ii) unobserved endowments are heterogeneous but mean

independent of college decisions, i.e., $\mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1}-\varepsilon_{\mathrm{i},0}\right) \mid X_i=x, v_i=v\right]=\mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1}-\varepsilon_{\mathrm{i},0}\right)\right]$; again the last term in equation (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 (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 econometrician's point of view, may make different college decisions; that is, unobserved characteristics ϵ 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 $\mathrm{E}[(X_i'\beta_1 - X_i'\beta_0)|X_i=x] = (x'\beta_1 - x'\beta_0)$ and an individual-specific unobserved component $\mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1}-\varepsilon_{\mathrm{i},0}\right) \mid X_i=x, v_i=v\right]$.

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]

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., $\left[\left(\varepsilon_{\mathrm{i},1}-\varepsilon_{\mathrm{i},0}\right) \mid X_i=x, v_i=v\right]$ (Heckman and Vytlacil 2005, 2007). Using equation (6), the $MTE(x,v)$ is defined as follows

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

$$MTE(x,v) = \frac{\partial \mathrm{E}[Y \mid X = x, P(z) = p]}{\partial p} = (x'\beta_1 - x'\beta_0) + \mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0}\right) \mid X_i = x, \nu_i = \nu\right] \quad (7)$$

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

$$MTE(x,v) = (x'\beta_1 - x'\beta_0) + \left(\mathrm{Cov}\left(\varepsilon_{\mathrm{i},1}, \epsilon_i\right) - \mathrm{Cov}\left(\varepsilon_{\mathrm{i},0}, \epsilon_i\right)\right) \Phi^{-1}(\nu) \quad (8)$$

Heckman and Vytlacil (2005, 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 $MTE(x,v)$, where weights are given in Appendix Table A3.

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).

To illustrate our parameter of interest, we perform the following experiment using the estimated parameters of selection equation (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 $MTE(x, v)$ for those individuals who switch their decision because of the expansion, such that $\tilde{\epsilon}_i < E_i^{*\prime}\gamma_E + X_i'\gamma_X + I_i'\gamma_I$, 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 $MTE(x, v)$ for switchers and is defined as:

$$PRTE(x, \alpha) = \int MTE(x, v)\, f_{v|X}(\tilde{v}|x, S_i(E_i = e, x, I) = 0, S_i(E_i = e + \alpha, x, I) = 1) d\tilde{v} \qquad (9)$$

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 *et al.* 2010, 2011b; Eisenhauer *et al*. 2015) and a 10 percent reduction in the distance to upper secondary schools in Indonesia (Carneiro *et al.* 2011a). The PRTE approach has not been applied yet with respect to college expansion.

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

$$MPRTE(x) = \lim_{\alpha \to 0} PRTE(x, \alpha). \qquad (10)$$

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.

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 (6) can be re-written as follows:

$$\mathrm{E}[Y_i \mid X_i = x, P(Z_i) = p] = x'\beta_0 + (x'\beta_1 - x'\beta_0)p + K(p) \tag{11}$$

where $K(p) = \mathrm{E}\left[\left(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0}\right) \mid X_i = x, P(Z_i) = p\right]p$ 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 $R = Y - x'\hat{\beta}_0 - \left(x'\hat{\beta}_1 - x'\hat{\beta}_0\right)\hat{p}$. 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 (7):

$$\widehat{MTE}(x, v) = \frac{\partial \mathrm{E}[Y \mid X=x, P(z)=p]}{\partial p} = \left(x'\hat{\beta}_1 - x'\hat{\beta}_0\right) + K'(\hat{p}) \tag{12}$$

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 *et al*. 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.

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.

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.

[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.

To estimate the college decision equation (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 2A 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 and moved to a current residence after completing college) is unknown. We

[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.

impute their location at age 17 based on college location and the type of birthplace such as village, township, and city (see Appendix Table A1 for details of imputation).[14]

Table 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 mother's 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 A1 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 2B). In our

[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 Yearbook of Russia 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.

sample period between 1985 and 2003, however, the fluctuations over time were relatively mild, though significant regional variation remained.

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 Sabirianova Peter 2014). 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). 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 2C). 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 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.

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 *et al*. 2011b, Zamarro 2010, Eisenhauer *et al*. 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 *et al.* 2011b); see Table 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

[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 in Russia, 2008).

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 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 A1 for data sources). Throughout the Soviet period, manual workers were rewarded with high pay, while wages of non-manual workers were artificially compressed. Figure 2D 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.

Summary Statistics for the Estimation Sample

The baseline estimation sample (age $\geq$ 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 mother's education, the missing values in other covariates are trivial (about 0.6 percent).[16] Table A1 in the appendix details how each variable is constructed. The descriptive statistics in Table 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, 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.

4. Marginal Treatment Effect

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)-(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 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, $\widehat{\mathrm{P}(Z)}$, 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

[16] Because mother's 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.

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 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 v) tend to do well in the labor market once they graduate

[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; mother's 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.

($\sigma_{1\epsilon} = -0.150$), but the same individuals would be worse off if they don't go to college ($\sigma_{0\epsilon} = 0.222$).

Following Carneiro *et al*. (2011b) and Heckman *et al*. (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 (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 A2). The test rejects that the returns to college are not correlated with S or that $MTE(x, v)$ is constant in v.

Another test for selection on gains relies on the estimated parameters from the normal switching regression model (Table 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 A2 that does not impose the joint normality assumption.

Using the estimated parameters in Table 4, we compute the MTE according to Equation (8). Figure 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 v (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 v have low MTE. The heterogeneity in the MTE across the distribution of v is substantial: the returns vary from -18.2 percent for the highest v person who would lose from attending college to 32.9 percent for the lowest v person, with the average return of 7.3 percent per one year of college education.

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

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 *et al.* (2011b). 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).

Alternative Instruments

In Table 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 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.

[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.

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 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 instrument is going to be compromised if:

- $\varepsilon's$ in wage equation affect E in college equation; that is, anticipated wage shocks in 2004-2011 influence where new campuses are opened in 1985-2003;
- E and $\varepsilon's$ 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

[20] Interestingly, in a paper on the effect of mother's 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 mother's educational attainment than the opening of private colleges.

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 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 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 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 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 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.

[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.

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.

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, $\overline{X}(\hat{\beta}_1 - \hat{\beta}_0)$, and the covariance component of MTE, $(\sigma_{1\epsilon} - \sigma_{0\epsilon})$. The estimates of additional specifications are presented in Table 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 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

[22] Results are shown in web appendix.

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 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 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 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 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

[23] See Table 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.

future empirical analysis. The policy treatment parameters discussed in the next section should be interpreted from a partial equilibrium perspective.

5. Policy Treatment Parameters

Conventional Treatment Parameters

Recall the gross returns to college education are given by $\Delta_i = Y_{i,1} - Y_{i,0} = (X_i'\beta_1 - X_i'\beta_0) + (\varepsilon_{i,1} - \varepsilon_{i,0})$, 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:

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

$$= E\left(X_i'\beta_1 + \varepsilon_{i,1}\middle|X_i, S_i = 1\right) - E\left(X_i'\beta_0 + \varepsilon_{i,0}\middle|X_i, S_i = 0\right)$$

$$= E(\Delta_i|X_i) + E\left(\varepsilon_{i,1}\middle|S_i = 1\right) - E\left(\varepsilon_{i,0}\middle|S_i = 0\right)$$

$$= \underset{\text{(ATE)}}{E(\Delta_i|X_i)} + \underset{\text{(Sorting gain)}}{E(\varepsilon_{i,1} - \varepsilon_{i,0}|S_i = 1)} + \underset{\text{(Selection Bias)}}{[E\left(\varepsilon_{i,0}\middle|S_i = 1\right) - E\left(\varepsilon_{i,0}\middle|S_i = 0\right)]}$$

$$= \underset{\text{(TT)}}{E(\Delta_i|X_i, S_i = 1)} + \underset{\text{(Selection Bias)}}{[E\left(\varepsilon_{i,0}\middle|S_i = 1\right) - E\left(\varepsilon_{i,0}\middle|S_i = 0\right)]}, \quad (11)$$

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 $[E\left(\varepsilon_{i,0}\middle|S_i = 1\right) - E\left(\varepsilon_{i,0}\middle|S_i = 0\right)]$ = 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 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 A3 and Figure A1). 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, $\mathrm{E}(\varepsilon_{\mathrm{i},1} - \varepsilon_{\mathrm{i},0} | S_i = 1) > 0$, 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, $\mathrm{E}(\varepsilon_{\mathrm{i},0} | S_i = 1) < \mathrm{E}(\varepsilon_{\mathrm{i},0} | S_i = 0)$. 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 Rosen's (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 ($\varepsilon_{\mathrm{i},1} \neq \varepsilon_{\mathrm{i},0}$) and positive sorting gain. As evident from Appendix Figure A1, the IV estimand weighs a higher valued segment of $MTE(x, v)$ more heavily.

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 *et al*. (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 *et al*. (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+α), $P_\alpha = (1 + \alpha) \cdot P(Z)$. 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 (10) and weights provided in Appendix Table A3 and Figure A1, we calculate marginal policy-relevant treatment effects of three policy regimes from the normal selection model and report them in Table 8, Panel A. Similar to Carneiro *et al*. (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 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 v, and we reject the null hypothesis that the returns to college are not correlated with S or that $MTE(x, v)$ is constant in v based on the test results reported in Appendix Table A2. 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 8 presents the MPRTE parameters from the semi-parametric model 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

[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.

[26] A MPRTE parameter is a weighted average of the MTE values estimated semi-parametrically on the basis of equation (12). The MPRTE weights are reported in Appendix Table A3 and also plotted in Figure A1.

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 (9).[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, $P_{PRTE} = P\big(N_{m,92} + \Delta N_{m,t}\big)$. 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 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 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

[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 α=1.

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 A4). 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 A4 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).

5. Conclusions

This paper 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.

References

Andrén, Daniela, John Earle, and Dana 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.

Basu, Anirban, James J. Heckman, Salvador Navarro-Lozano, and Sergio 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.

Belskaya, Olga, and Klara Sabirianova Peter, 2014. 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.

Björklund, Anders, and Robert Moffitt, 1987. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models, *Review of Economics and Statistics* 69(1): 42-49.

Brainerd, Elizabeth, 1998. Winners and Losers in Russia's Economic Transition, *American Economic Review* 88(5): 1094-1116.

Cameron, Stephen, and James 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, Stephen, and Christopher Taber, 2004. Estimation of Educational Borrowing Constraints Using Returns to Schooling, *Journal of Political Economy* 112(1): 132-182.

Card, David, 1995. Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in *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, 201-222.

Card, David, 2001. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica* 69(5): 1127-1160.

Carneiro, Pedro, and Sokbae 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.

Carneiro, Pedro, James Heckman, and Edward Vytlacil, 2010. Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin, *Econometrica* 78(1): 377-394.

Carneiro, Pedro, James Heckman, and Edward Vytlacil, 2011b. Estimating Marginal Returns to Education, *American Economic Review* 101(4): 2754-2781.

Carneiro, Pedro, Michael Lokshin, Cristobal Ridao-Cano, and Nithin Umapathi, 2011a. Average and Marginal Returns to Upper Secondary Schooling in Indonesia, Policy Research Working Paper 5878.

Carnoy, Martin, Prashant Loyalka, Maria Dobryakova, Rafiq Dossani, Isak Froumin, Katherine Kuhns, Jandhyala B. G. Tilak, and Rong Wang, 2013. *University Expansion in a Changing Global Economy: Triumph of the BRICs?* Stanford, California: Stanford University Press.

Currie, Janet, and Enrico Moretti, 2003. Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings, *Quarterly Journal of Economics* 118(4): 1495-1532.

Demography Yearbook of Russia, 2002. Moscow: Goskomstat.

Duflo, Esther, 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 in Russia, 2008. Moscow: Goskomstat

Eisenhauer, Philipp, James J. Heckman, and Edward Vytlacil, 2015. The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs, *Journal of Political Economy*, 132(2), forthcoming.

Evans, Heather, and Anirban Basu, 2011. Exploring Comparative Effect Heterogeneity with Instrumental Variables: Prehospital Intubation and Mortality, Health, Econometrics and Data Group Working Papers 11/26.

Fang, Hai, Karen Eggleston, John Rizzo, Scott Rozelle, and Richard Zeckhauser, 2012. The Returns to Education in China: Evidence from the 1986 Compulsory Education Law, Working Paper, No. 18189, National Bureau of Economic Research, Cambridge, MA.

Federal Law, 1992. Ob obrazovanii [On Education], July 10, 1992.

Available at http://www.consultant.ru/popular/edu/.

Fleisher, Belton, Klara Sabirianova Peter, and Xiaojun 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, Marc, 2009. Do Universities Benefit Local Youth? Evidence from the Creation of New Universities, *Economics of Education Review* 28(3): 318-328.

Garen, John, 1984. The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable, *Econometrica* 52(5): 1199-1218.

Gorodnichenko, Yuriy, and Klara Sabirianova 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, Jeffrey A., 2004. The Effect of College Location on Migration of College-Educated Labor, *Journal of Econometrics* 121(1-2): 125-142.

Heckman, James J., John Eric Humphries, Gregory Veramendi, and Sergio Urzua, 2014. Education, Health and Wages, Working Paper No. 19971, National Bureau of Economic Research, Cambridge, MA.

Heckman, James and Edward Vytlacil, 2001. Policy-Relevant Treatment Effects, *American Economic Review* 91(2): 107-111.

Heckman, James and Edward Vytlacil, 2005. Structural Equations, Treatment Effects, and Econometric Policy Evaluation, *Econometrica* 73(3): 669–738.

Heckman, James and Edward Vytlacil, 2007. Econometric Evaluation of Social Programs, *Handbook of Econometrics*, Volume 6B: 4877-5143.

Heckman, James and Xuesong Li, 2004. Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000, *Pacific Economic Review* 9(3): 155-171.

Heckman, James, Daniel Schmierer, and Sergio Urzua, 2010. Testing the Correlated Random Coefficient Model, *Journal of Econometrics* 158(2): 177-203.

Heckman, James, Lance Lochner, and Petra Todd, 2006. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, *Handbook of the Economics of Education*, Volume 1, Chapter 7: 307-458.

Ichino, Andrea and Rudolf Winter-Ebmer, 2004. The Long-Run Educational Cost of World War II, *Journal of Political Economy* 22(1): 57-87. Kane, Thomas and Cecilia Rouse, 1995. Labor - Market Returns to Two- and Four-Year College, *American Economic Review* 85(3): 600-614.

Kane, Thomas J. and Cecilia Elena Rouse, 1995. Labor-Market Returns to Two- and Four-Year College, *American Economic Review* 85(3): 600-614.

Katz, Lawrence and Kevin Murphy, 1992. Changes in Relative Wages, 1963-1987: Supply and Demand Factors, *Quarterly Journal of Economics* 107(1): 35-78.

Kyui, Natalia, 2013. Expansion of Higher Education, Employment and Wages: Evidence from the Russian Transition, Bank of Canada Working Paper 2013-45.

Moffitt, Robert, 2008. Estimating Marginal Treatment Effects in Heterogeneous Populations, *Annales d'Economie et de Statistique* 91/92: 239-261.

Moffitt, Robert (2014). Estimating Marginal Treatment Effects of Transfer Programs on Labor Supply, Working Paper.

Münich, Daniel, Jan Svejnar, and Katherine 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.

OECD, 2008. *Education at a Glance: OECD Indicators.* OECD Publishing.

Oppedisano, Veruska, 2011. The (Adverse) Effects of Expanding Higher Education: Evidence from Italy, *Economics of Education Review* 30(5), 997-1008.

Regions of Russia. Moscow: Goskomstat, various years.

Robinson, Peter, 1988. Root-N-Consistent Semiparametric Regression, *Econometrica* 56(4): 931‐54.

Roy, Andrew, 1951. Some Thoughts on the Distribution of Earnings, *Oxford Economic Papers* 3(2): 135-146.

Topel, Robert, 1997. 'Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality, *Journal of Economic Perspectives* 11(2): 55–74.

Willis, Robert and Sherwin Rosen, 1979. Education and Self-Selection, *Journal of Political Economy* 87(5): S7-S36.

Winters, John V., 2011. Why Are Smart Cities Growing? Who Moves and Who Stays? *Journal of Regional Science* 51(2): 253-270.

Yang, Dennis Tao, 2005. Determinants of Schooling Returns during Transition: Evidence from Chinese Cities. *Journal of Comparative Economics*, 33(2): 244-264.

Zamarro, Gema, 2010. Accounting for Heterogeneous Returns in Sequential Schooling Decisions, *Journal of Econometrics* 156(2): 260-276.

Table 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 vector 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: 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 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)	
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 A1. Standard deviations are in parentheses.

Table 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)
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 ≥ 25, age 17 ∈ [1985, 2003], wage ∈ [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 5: College Equation: Alternative Set of Instruments

Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*
N of colleges (main campuses)	0.007***	…	…	…	…
	(0.002)				
N of college 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)
Number 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)	
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: N=17,911. *** 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 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 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 [p-value]	370.6 [0.000]	164.0 [0.000]	170.1 [0.000]	517.3 [0.000]	237.8 [0.000]
χ^2–test for independence of equations [p-value]	39.4 [0.003]	41.2 [0.003]	40.6 [0.001]	32.1 0.000	81.5 [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 4 (age ≥ 25, age 17 ∈ [1985, 2003], wage ∈ [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 ∈ [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 7: Treatment Parameters

Treatment Parameter	*Estimated 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 $MTE(x,v)$ using weighting functions of Heckman and Vytlacil (2005):

$$Parameter(x) = \int_0^1 MTE(x,v)\, h_j(x,v) dv, \text{ where } j = ATE, TT, TUT, IV, OLS.$$

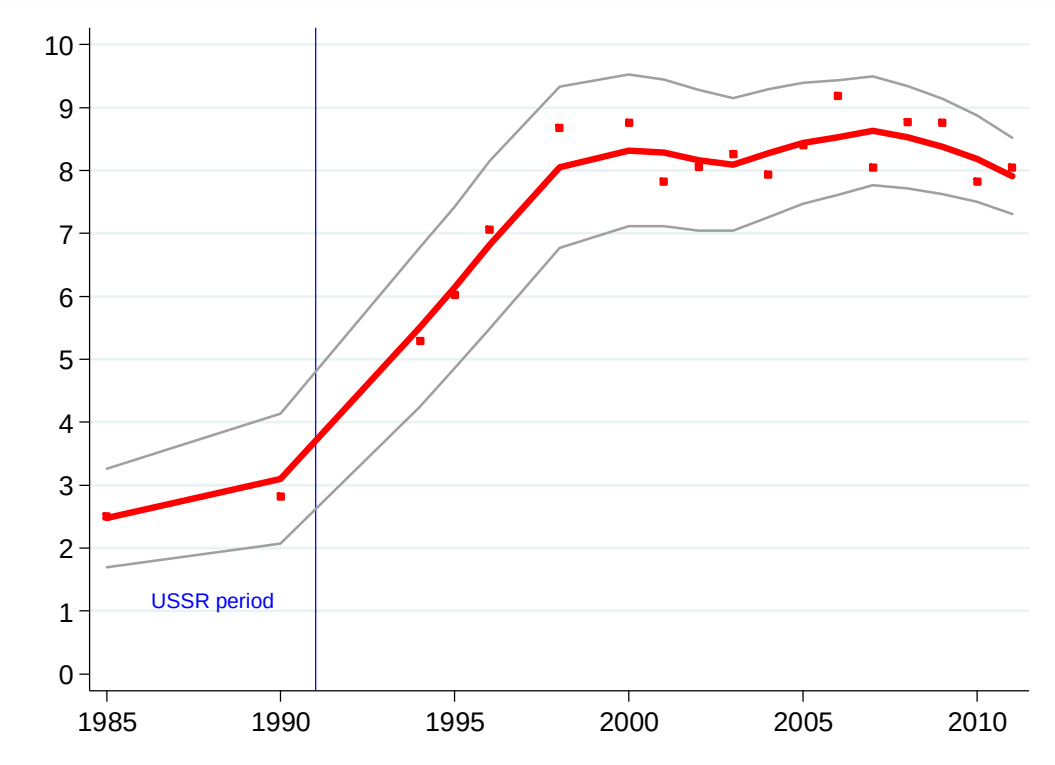
The weighting functions $h_j(x,v)$ are defined in Appendix Table A3. Linear IV estimates use $P(Z)$ as an instrument. Bootstrapped standard errors are in parentheses (250 replications). The reported returns are annualized by dividing the estimates by 4.5, which is the difference in average years of schooling between treatment and control groups. Table reports the results under the baseline specification (Table 4).

Table 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)
$P_\alpha = (1+\alpha) \cdot P(Z)$	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)
$P_\alpha = (1+\alpha) \cdot P(Z)$	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)
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)

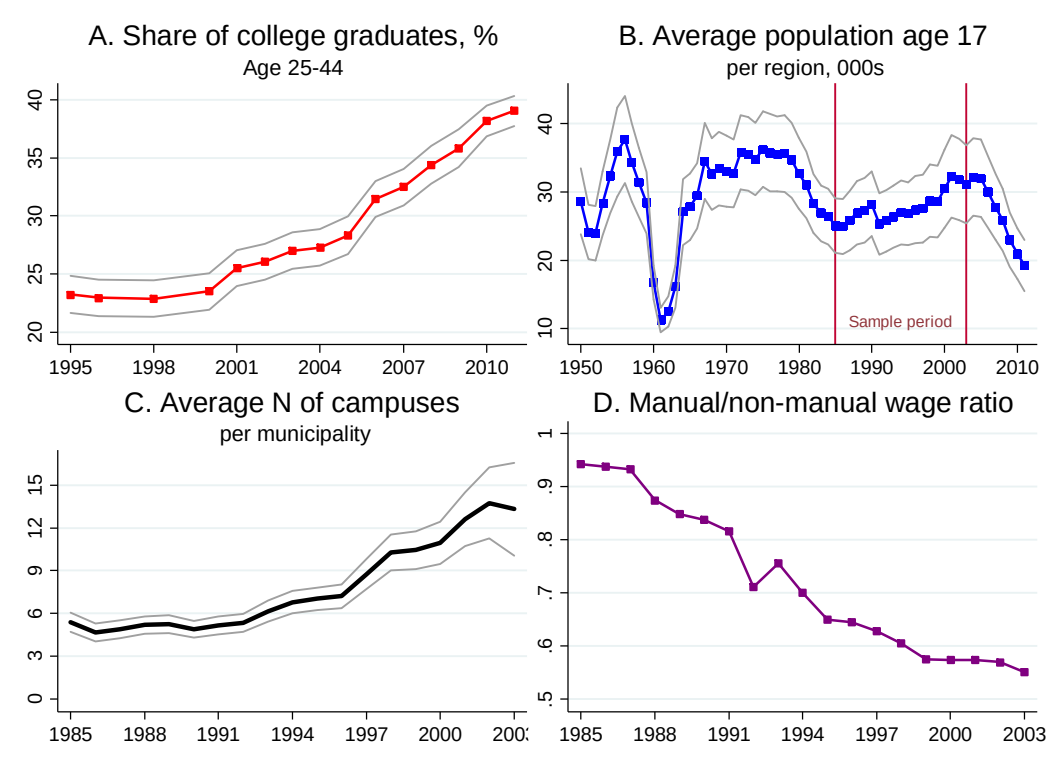
Notes: Table reports marginal policy-relevant treatment effect (MPRTE) for the normal selection and semi-parametric models as well as policy-relevant treatment effect (PRTE) from policy experiments. The reported returns are per year of college. $N_{m,t}$ is the number of campuses in the municipality 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 (8) in Panels A and C; equation (12) in Panel B. The MPRTE are obtained by integrating the $MTE(x,v)$ using the weighting functions reported in Appendix Table A3. 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 1: Average Returns to College Education in Russia



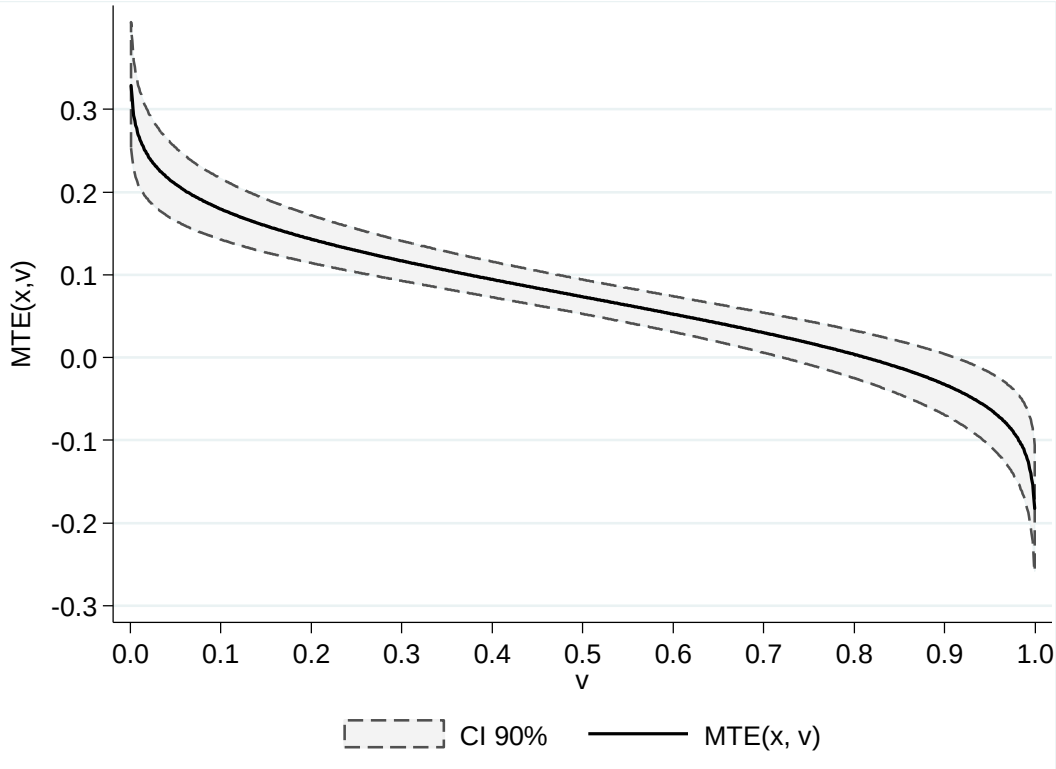
Notes: Returns to college education are calculated from the OLS regression of the log of monthly earnings at primary job on college degree, gender, age, age squared, urban place of birth, a dummy for Russian nationality, and seven federal districts. The comparison group includes graduates of general 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% confidence interval computed using robust standard errors. The vertical line demarks 1991 as the breakup of the USSR.

Figure 2: Trends in Key Variables



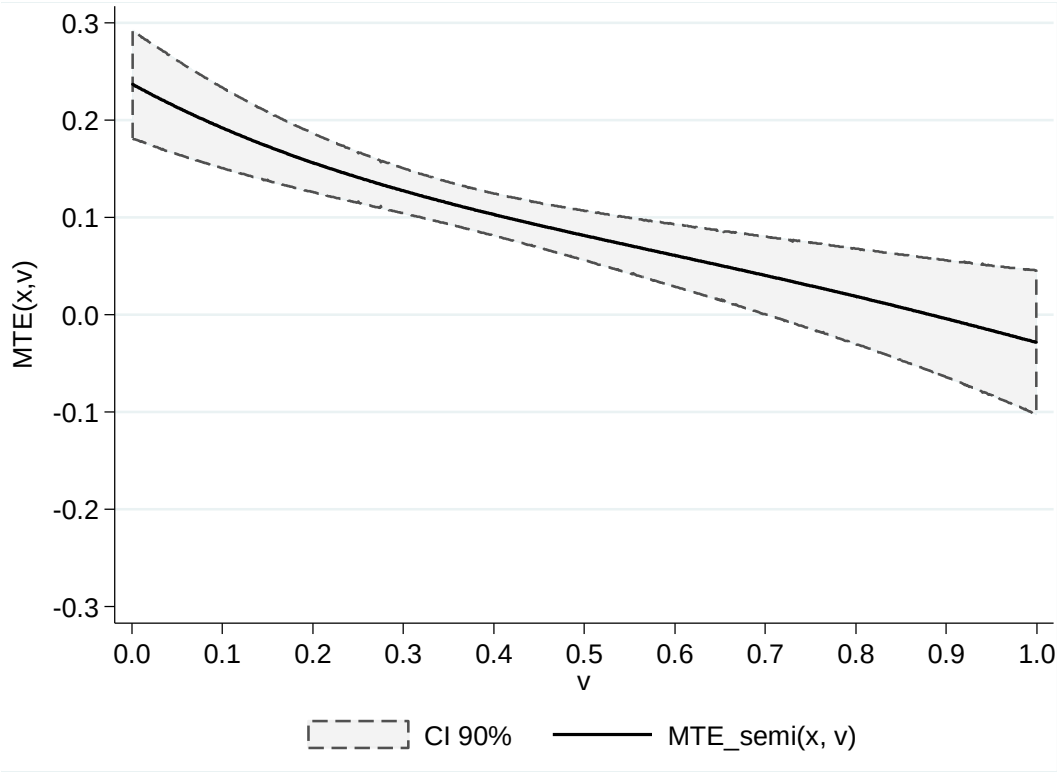
Notes: Panel A shows the percent share of college graduates in RLMS among 25-44-year old individuals. The comparison group includes graduates of general and professional secondary schools, but excludes college drop-outs. Panel B depicts the average size of 17-year-old population in thousands people across all Russian regions. Vertical lines define the sample period – from 1985 to 2003. Also shown is the 90% confidence interval. Panel C shows the average number of campuses per municipality at age 17 in the RLMS sample, excluding Moscow city. Panel D displays the ratio of wages of manual workers to the wages of non-manual workers in the industrial sector during the sample period 1985-2003.

Figure 3: Marginal Treatment Effect Estimated from a Normal Selection Model



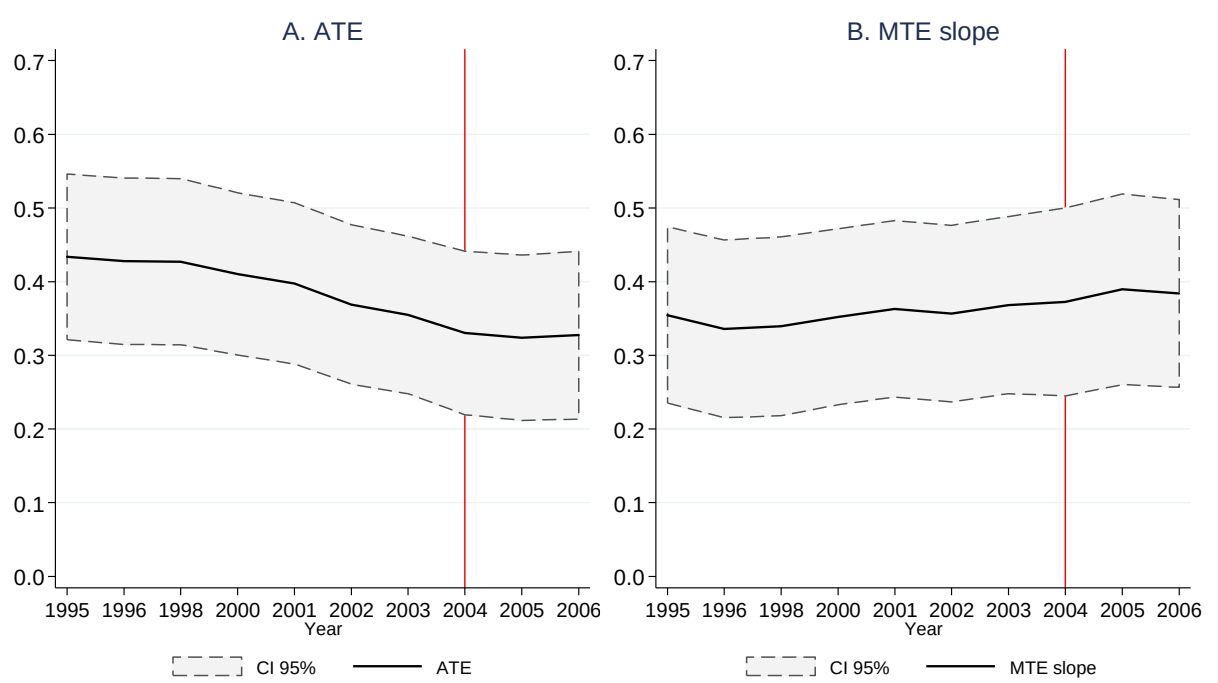
Notes: The figure plots the MTE estimates for a grid of values v between zero and one, with a 90-percent confidence interval, evaluated at mean values of X. We estimate a parametric normal selection model given in equations (1)-(4) by maximum likelihood. The figure is computed using equation (8) and the estimated parameters in Table 4. The fixed covariance component of $MTE(x, v)$ 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 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 (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 5: Average Treatment Effect Estimates for Different Sample Periods



Notes: $ATE = \bar{X}(\hat{\beta}_1 - \hat{\beta}_0)$; $MTE\ slope = \hat{\sigma}_{1\epsilon} - \hat{\sigma}_{0\epsilon}$. 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 4).

Appendix

Table A1: 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 X^S	
Female	= 1 if female
Urban residence (binary)	= 1 if resides in urban location at the time of the survey
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 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 Sabirianova Peter, 2014).*
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, 1989–1993, 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 A2: Tests for Selection on Gains

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

Degree of Polynomial	***2***	***3***	***4***	***5***
p-value of joint test of 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 v for LATE		***Difference in LATEs***	***p-value***	***Joint p-value***
j	***j+1***			
(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, $X \cdot \hat{P}(Z)$ 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 4 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 *v*, 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 $T^j = \left|LATE^j - LATE^{j+1}\right|$ and $T_b^j = \left|\left(LATE_b^j - LATE_b^{j+1}\right) - \left(LATE^j - LATE^{j+1}\right)\right|$ for $b = 1, \ldots 250$. 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 $C_b = \sum_j \left[\left(LATE_b^J - LATE_b^{j+1}\right) - \left(LATE^j - LATE^{j+1}\right)\right]^2$ for $b = 1, \ldots 250$ and all *j* and *j*+1 intervals.

Table A3: Weights for Treatment Effects

$$h_{ATE}(x,v) = 1$$

$$h_{TT}(x,v) = \left[\int_v^1 f(p|X=x)dp\right]\frac{1}{E(\mathrm{P}|X=x)}$$

$$h_{TUT}(x,v) = \left[\int_0^v f(p|X=x)dp\right]\frac{1}{E((1-\mathrm{P})|X=x)}$$

$$h_{PRTE}(x,v) = \left[\frac{F_{P^*,X}(v) - F_{P,X}(v)}{\Delta P}\right], \Delta P = \mathrm{P}^*(Z+\alpha) - \mathrm{P}(Z)$$

$$h_{MPRTE}(x,v) = \left[\frac{f_{p|X}(v) - f_{\epsilon|X}(F^{-1}_{\epsilon|X}(v))}{\mathrm{E}\left[f_{\epsilon|X}\left(F^{-1}_{\epsilon|X}\mu_S(Z_i)\right)|X\right]}\right]$$

$$h_{IV}(x,v) = \left[\int_0^1 \big(p - E(P|X=x)\big)f(p|X=x)dp\right]\frac{1}{Var(P|X=x)} \quad \text{(for } P(Z) \text{ as the instrument)}$$

$$h_{OLS} = \frac{E(U_1|X=x,v)h_1(x,u_D) - E(U_0|X=x,v)h_0(x,v)}{MTE(x,v)} \quad (for\ MTE(x,v) \neq 0)$$

$$h_1(x,v) = \left[\int_v^1 f(p|X=x)dp\right]\frac{1}{E(P|X=x)}$$

$$h_0(x,v) = \left[\int_0^v f(p|X=x)dp\right]\frac{1}{E((1-P)|X=x)}$$

Source: Heckman and Vytlacil (2001, 2005), Carneiro *et al*. (2010)

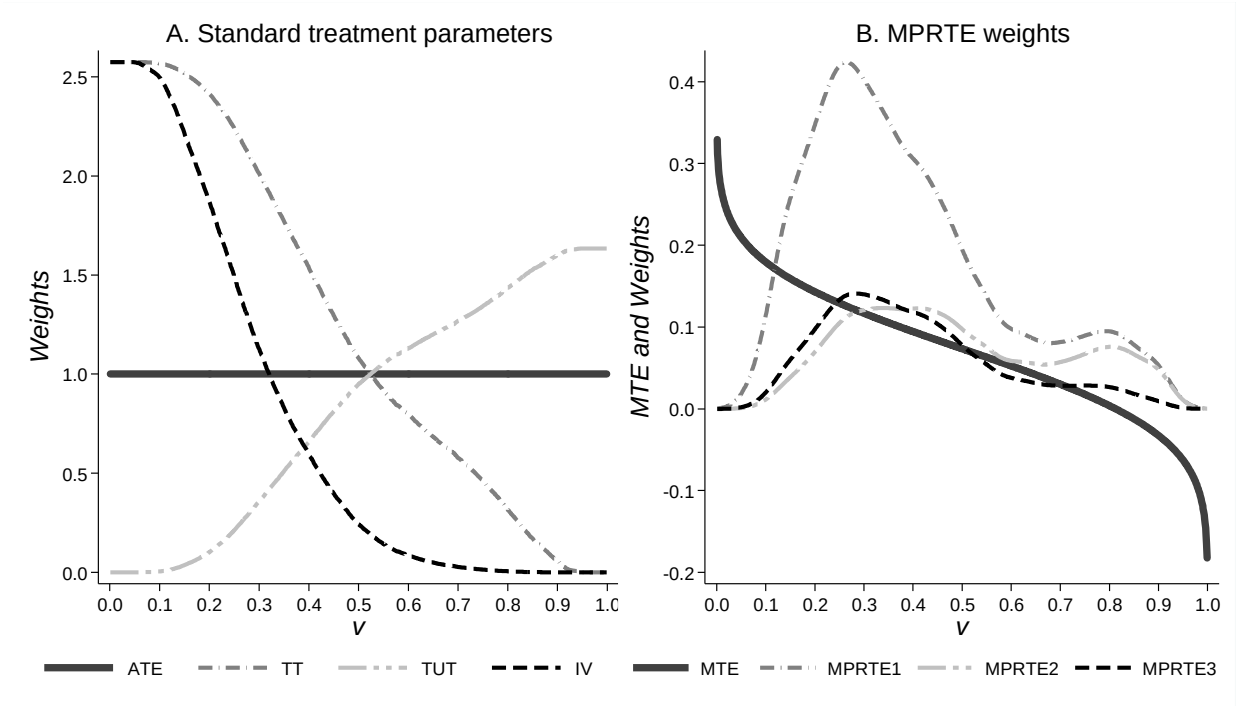
Notes: ATE = average treatment effect; TT = treatment effect on the treated; TUT = treatment effect on the untreated; PRTE = policy relevant treatment effect; MPRTE = marginal policy relevant treatment effect; IV = instrumental variables; OLS = ordinary least square.

Table A4: Cost-Benefit Analysis

	Regional capitals	*Non-capital cities*
COSTS PER COLLEGE STUDENT		
A. Direct costs of college attendance		
Tuition expenditures for fee-paying students per year	$1,376	$1,059
Share of students paying tuition	0.576	0.680
Additional direct costs (books, supplies, tutoring, etc.) per year	$72	$72
Total direct costs of college attendance per year	$865	$792
B. Foregone labor market earnings		
Average annual net earnings of high school graduates, age 18-30	$5,763	$4,706
Probability of being employed of high school graduates, age 18-30	0.831	0.854
Total costs of college attendance per year	$5,656	$4,809
Present value of college costs	$30,927	$26,300
BENEFITS PER COLLEGE STUDENT		
C. Improved labor market income		
Rate of returns to college (%)	7.5	9.9
Probability of being employed of college graduates, age 22-55	0.920	0.929
Additional annual net earnings of college graduate, age 22-55	$1,789	$1,948
Present value of additional earnings stream, 32-year work life	$38,261	$41,657
Net present benefits of college education	$7,334	$15,357

Notes: Regional capitals are most populated cities in the region and have the highest concentration of colleges. All monetary values are deflated in 2010 prices using national CPI and then converted to US dollars ($1 = 30.6 rubles in December 2010). Tuition expenditures and the share of students paying tuition are calculated from the RLMS questions available in 2006-2008 surveys. Additional direct costs are from Goskomstat (*Social Conditions and Welfare of the Russian Population*, 2011) and assumed the same for both types of cities. Real earnings are net of tax, averaged over the 2004-2011 period, and adjusted for the probability of being employed. The rates of returns are taken from table 8 (PRTE from adding a campus in capital and non-capital cities). Additional annual net earnings of college graduate, age 22-55 = (average annual net earnings of high school graduates, age 18-30) x (percent rate of returns to college/100) x (4.5 years of college) x (probability of being employed of college graduates, age 22-55). The present value is calculated at age 22 over 4.5 past years of college and 33 future years of working life at 3 percent real interest per annum. Calculations of net benefits do not account for earnings during school years, stipends and financial support, additional living expenses if a student decides to move to a regional capital, and college benefits besides earnings.

Figure A1: MTE Weights



Notes: Panel A plots the MTE weights used in computing ATE, TT, TUT, and IV with estimates provided in Table 7; see Appendix Table A3 for formulas. Panel B plots the weights on MTE for three different versions of MPRTE: 1) $P_\alpha = P(Z) + \alpha$; 2) $P_\alpha = (1+\alpha) \cdot P(Z)$; and 3) $P_\alpha = P\left(N_{m,t} + \alpha\right)$. The MPRTE estimates are reported in Table 8. The scale of the *y*-axis in Panel B is the scale of the MTE, not the scale of the weights, which are scaled to fit the picture.