NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Tommaso Agasisti, Aleksei Egorov,
Daria Zinchenko, Oleg Leshukov

UNIVERSITIES' EFFICIENCY AND REGIONAL ECONOMIC SHORT-RUN GROWTH: EMPIRICAL EVIDENCE FROM RUSSIA

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 203/EC/2018

Tommaso Agasisti[1], *Aleksei Egorov*[2],
Daria Zinchenko[3], *Oleg Leshukov*[4]

UNIVERSITIES' EFFICIENCY AND REGIONAL ECONOMIC SHORT-RUN GROWTH: EMPIRICAL EVIDENCE FROM RUSSIA[5]

This paper analyses the link between the efficiency of regional higher education systems and the rates of regional economic development between 2012 and 2015 in Russia. The efficiency scores are calculated at the institutional level using a double-bootstrap data envelopment analysis (DEA) procedure, taking into account the different internal characteristics of universities which may affect their production process, and the scores are then aggregated at the regional level. We formulate a regional economic growth model that considers the efficiency of regional higher education systems as one of the explanatory variables. As an econometric method, we employ a robust GMM estimator. The model also includes spatial interactions between regional economies and between regional higher education systems in neighboring regions. The findings highlight a positive, substantial and statistically significant effect of HEI efficiency on the regional economic growth rate. We also found negative spillover effects indicating that efficient regional higher education systems may extract resources from neighboring regions.

JEL Classification: I25, I21, E02.

Keywords: Regional economic development; efficiency in higher education; knowledge spillovers; economic growth; Russia.

---

[1] Politecnico di Milano School of Management, Italy; Department of Management, Economics & Industrial Engineering. Associate Professor; E-mail: agasisti@mip.polimi.it

[2] National Research University Higher School of Economics (Moscow, Russia); Laboratory for University Development of Institute of Education. Research-fellow; E-mail: aegorov@hse.ru

[3] National Research University Higher School of Economics (Moscow, Russia); Laboratory for University Development of Institute of Education. Analyst; E-mail: dzinchenko@hse.ru

[4] National Research University Higher School of Economics (Moscow, Russia); Laboratory for University Development of Institute of Education. Research-fellow; E-mail: oleshukov@hse.ru

[5] Acknowledgements. Previous versions of this paper have been presented at the following academic conferences: 4th LEER Conference on Education Economics, Belgium, 2018; conference "Higher Education in Modern Ecosystems: Efficiency, Society and Policies", Germany, 2018; XIX April International Academic Conference on Economic and Social Development, Russia, 2018; Russian Summer School on Institutional Analysis, Russia, 2018; CHER 31st annual conference, Russia, 2018; 6th international Workshop on "Efficiency in Education, Health and other Public Services", United Kingdom, 2018. We are grateful to the participants for their useful comments and suggestions. All eventual errors are our solely responsibility.

1. Introduction

Higher education institutions (HEI) are often considered to be economic agents and are analyzed in terms of their economic activity. Universities may be engaged in the social and economic development of the territories where they operate and, consequently, they might contribute to economic growth (Belenzon and Schankerman, 2013; Pinheiro et al., 2012; Varga, 1997). There is a large number of empirical studies that show a positive and causal relationship between the development of HEI and rates of economic development (Valero and Van Reenen, 2016). Moreover, most economic growth theories consider human capital to be one of the most important determinants of economic development (Hanushek, 2016).

Human capital is usually measured as the number of years of schooling and this proxy-variable is used in economic growth functions (Barro and Lee, 2013). A positive empirical relationship between the average years of schooling in a country or region and economic growth rates reflects the fact that all levels of education may have a positive impact on economic growth. However, higher education is especially important in this context, since this level of education provides specific sets of skills needed for generating new ideas and innovations (Hanushek, 2016). In such a setting, universities are considered not a burden for state budgets, but an investment in human capital development, which can bring positive returns in the future. Such a perception of higher education has policy implications. The development of regional or national higher education systems is often considered a policy instrument that can lead to positive economic outcomes, both in the short and the long run.

Being treated as economic agents, universities are the subject of debate on the efficiency of their activities (Kosor, 2013). If universities have an economic impact and public funds are invested in them in order to intensify this impact, society can require the effective allocation of these funds. As economic agents, HEI have the goal of maximizing their outputs (teaching, research and knowledge transfer – third mission) using limited resources. The debate on the economic impact of universities and the debate on the efficiency of their activities are very close to each other and this relationship is very important for public policy in the field of higher education.

Russia provides a good case-study for the investigation of these issues. The current federal public policy tries to take into account both the engagement of universities in social and economic development and the efficiency of their activities. In 2012 the "Annual Monitoring of Performance of Higher Education Institutions" was launched by the Ministry of Education and

Science. Using this policy tool, the Ministry aimed to identify the universities that were inefficient and make managerial decisions in such cases, including reorganization. In response to the challenge of the limited links between HEI and regional administrations and enterprises, a special federal program (the Flagship universities program) was launched in 2015 to increase university efficiency in terms of having a positive impact on regional economic development.

Empirical studies that analyze the relationship between the development level of the education system and economic development in Russia (Egorov et al., 2017) give evidence that the scale and quantity of the higher education system matters for economic development. However, the discussion regarding the role of university efficiency is limited. This paper explores the link between the efficiency of regional higher education systems in Russia and the rates of economic growth of the regions where these systems operate. Specifically, we answer the following research question: *does the efficiency of universities operating in a certain region affect the economic growth of that area?*

The analysis is based on economic growth theory and consists of three methodological steps. First, the efficiency scores of particular universities are estimated using a double-bootstrap data envelopment analysis (DEA) following Simar and Wilson (2007). Second, these results are aggregated at the regional level and efficiency scores for regional higher education systems are obtained. Third, a model for regional economic growth is proposed, treating the efficiency of regional higher education systems as one of the explanatory variables. Specifications of the models take into account the structure of regional economies and spatial effects both in gross regional product (GRP) growth rates and the efficiency of regional higher education systems. Such a specification allows us to test the hypothesis about existence of spillover effects. The model is estimated by means of a robust GMM system that handles the endogeneity problem between university efficiency and economic performance. This paper is particularly innovative because it is the first that exploits data from the "Annual Monitoring of Performance of Higher Education Institutions" for investigating regional economic growth.

The paper is organized as follows. Section 2 discusses the background and particularities of regional development in Russia and specific features of the Russian higher education system. Section 3 reviews the literature on the economic impact of universities and details the concept of efficiency in higher education. Section 4 contains the underlying theoretical framework. Section 5 illustrates the methodology of the study and describes the dataset. Section 6 discusses the main results, and Section 7 contains some policy implications and concluding remarks.

2. The background of higher education and regional development in Russia

2.1 Higher education in Russia

The Russian system of higher education has undergone unprecedented reform over the past 30 years and the landscape of the university system is still being transformed. This process was primarily the result of the collapse of the USSR and the transition to a market economy (Froumin et al., 2014).

In Soviet times, universities were part of a unified system of national economy and they were obligated to integrate into national supply chains. The curriculum base, specialization, size and even the territorial location of universities were all regulated by the central government. The higher education system was centralized and subject to rigid control, in adherence to the state's political agenda (Johnson, 2008). The principle of mandatory job placement of university graduates into specific workplaces across the entire country allowed the government to plan the recruitment for certain specialties demanded by the Soviet economy.

With the emergence of a new nation, market mechanisms and new branches of the economy, after 1991, the higher education system was forced to adapt to new social and economic realities. Such changes entailed the following consequences:

- the emergence of a private higher education sector (the opening of private universities and the right of public universities to charge tuition fees);
- the liberalization of education programs and an increase in university autonomy (Froumin and Leshukov, 2015);
- the massification of higher education and a sharp increase in the number of universities. The number of head universities has increased 1.5 times during the past 20 years (see Figure 1). The growth of the higher education sector was ensured by the opening of branches – as of 2017 there were 840 branches of universities in Russia (597 public and 243 private). This led to an increase of higher education coverage – as of 2016, the share of young people aged 17-25 years attending higher education programs was 31.8%.
- the adaptation of universities to the needs of the market, primarily at the local level. The termination of mandatory job placements for graduates forced the universities to adjust to the needs of the regional population.

Most of these changes proceed in a chaotic way, not accompanied by any coordination at the national level. Therefore, since the mid-2000s, a period of rational public reforms in the sphere of higher education has been undertaken. This has determined the basis of the current system. The main tasks of public policy in higher education were maintaining control over universities and improving the quality of their educational and scientific activities.

[Insert Figure 1 here]

However, the issue of the governance of a large and heterogeneous country as the Russian Federation remains particularly urgent. The regions of Russia are significantly heterogeneous in terms of the scale of higher education systems. A third of all head universities are concentrated in two regions – Moscow and the Moscow region (185 universities) and St. Petersburg (66 universities), while in half of other regions the number of universities does not exceed five.

The Soviet legacy has defined the current centralization of university governance – 90% of universities are regulated by the federal government which makes the Russian education governance system one of the most centralized in the world (Morsy et al., 2018). The financial and regulatory possibilities for the participation of regional authorities in the development of universities are substantially restricted (Froumin and Leshukov, 2015). Therefore, assessing the contribution of universities to the economic development of Russian regions in such centralized conditions is especially topical from an academic and institutional viewpoint.

2.2 Regional development in Russia

Russia consists of 85 regions which, in accordance with the Constitution, are equal their in relations with the federal government. There are different legal types of regions in Russia ("oblast", "republic", "krai" etc.), but all of them have their own legislative, executive and judicial branches.

Historically, the main feature of Russian regional development was the high level of regional differentiation, especially in terms of their socio-economic characteristics. The Russian population is distributed unevenly, regions located in the European part of Russia make up only 30% of the total area, while their share of the population is 73% (Sardadvar and Vakulenko, 2016). Three regions (Moscow, St. Petersburg and Tyumen region) generate 35% of the entire gross national product.

These differences became stronger during the transformation period after the collapse of Soviet Union (Carnoy et al., 2018). In order to deal with this challenge, at the beginning of 2000 the federal authorities started a regional development policy to support poorly performing regions. The federal center distributed a substantial amount of resources in favor of regions with budget deficits. In other words, underdeveloped regions began to be supported, via the federal government, by more developed ones.

The current state policy is aimed at supporting regions on a competitive basis. Particularly, special economic zones with tax exemptions have been established in such regions in order to make them more attractive for investors. Another government program has established clusters in particular regions that bring together and intensify the interaction between industry, scientific organizations, universities, etc. This interaction is supposed to strengthen the competitiveness of the territory where it occurs. These reforms force regions to be proactive in their development, while maintaining a high level of the control from the center (Carnoy, 2018).

These policies have had a positive impact, although regional inequality in Russia remains substantial. For instance, in 2015 GRP growth rates ranged between -6.2% and 6.9%; GRP per capita were between 0.09 and 1.1 million rubles; the share of the workforce with higher education was between 22.1% and 47.8%; the share of students in the age cohort was between 14.8% and 46.6%. Despite this differentiation, Russian regions are not closed from an economic prospective, they actively interact with each other in order to find drivers for their development. There are empirical studies that confirm the existence of spillover effects (Demidova and Ivanov, 2016; Egorov et al., 2017).

Current regional development policy pays special attention to higher education as a source of regional growth. Universities are considered to be organizations which can attract resources to a region and support its competitiveness. This paradigm was reflected, particularly, in a special government program to form “flagship universities” to facilitate regional development by their teaching, research activities, knowledge transfer and their active role in society. Regional development in Russia raises the questions of how universities can contribute to regional development and how this contribution can be increased. These questions are addressed by this paper.

3. Literature on the economic impact of universities and the concept of efficiency in higher education

There are different approaches to the description of how universities may be engaged in social and economic development, ranging from the triple helix (Etzkowitz, 1993) to different econometric approaches based on the theoretical macroeconomic (Romer, 1986) and regional development (Capello, 2011) models. This paper considers universities as economic agents and analyzes them in terms of macroeconomic models of economic growth.

Three main frameworks to evaluate universities' economic impact can be found in the literature. The first one is the *traditional economic approach* (Elliott et al., 1998) suggesting that universities make different economic transactions within regional economies – they pay salaries, taxes, purchase goods and services, and provide jobs in the local labor market. These transactions increase the aggregate demand in regional economies and, consequently, transform this demand into GRP with multiplier effects. The second approach is the *skill-based framework* of university contribution (Bluestone, 1993). Within this framework universities are considered generators of human capital for regional labor markets. This line of reasoning considers two main channels. First of all, human capital theory (Becker, 1964) suggests that workers with higher levels of education tend to earn higher salaries (Mincer, 1974) and therefore spend more in regional economies. Secondly, a more educated labor force is usually considered to be more productive and to influence GRP growth rates directly.

The third approach to HEI economic impact evaluation considers universities as *drivers of regional innovation systems* (Huggins and Johnston, 2009), which can cooperate with innovative companies, stimulate entrepreneurship, and contribute to the development of new economic activities.

This paper is based on the macroeconomic theory of economic growth and analyses higher education as one factor in economic development. The theory of economic growth has been developed and extended over a long period of time. Robert Solow (1956; 1957) provided the basic modern framework for economic growth modeling. He elaborated the long-run economic growth model, which included technological progress in addition to the standard determinants of economic growth, i.e. physical capital accumulation and an increase of the labor-force. The technological progress variable was determined exogenously and contained those parts of the economic growth rate which could not be explained by the increase of the labor-force or physical capital. This neoclassical growth model suggested that physical capital accumulation and labor-force increases constitute the foundation of economic growth, but the growth of workforce productivity and capital increased as a result of this technological progress. Therefore, human capital is not a driver of economic growth in the neoclassical paradigm of growth modeling; its increase does not lead to a greater long-run growth rate.

The new theories of economic growth proposed since the late 1960s, on the contrary, confirm that human capital has a positive impact on the long-term development of economic growth rates. This is the endogenous growth model, discussed in the seminal contribution of Romer (1986). A more detailed analysis of human capital as an economic growth factor in the context of endogenous growth theory was provided in the models of Lucas (1988) and was continued in the study of Mankiw, Romer and Weil (1992). The most recent models indicate greater importance of tertiary education particularly for countries near the technological frontier (i.e. the most economically-developed countries), where growth requires new innovations (Vandenbussche et al., 2006; Aghion et al., 2009).

In the literature, there are several approaches to measuring human capital, such as school attainment and the results of standardized achievement tests of students (Hanushek, 2016). Quantifying human capital in terms of school attainment was early suggested by Mincer (1974). This approach has some limitations. The first is the assumption that one year of schooling correlates with the same amount of learning in all countries. The second limitation is that schooling is considered as the only source of human capital (Hanushek, 2016). A new approach to measuring human capital via the assessment of skill levels was first applied in Hanushek and Kimko (2000) and was extended in Hanushek and Woessmann (2007, 2015). Such a position measures human capital directly and reflects a net indicator of knowledge capital. Hanushek (2016) demonstrated that while measures of human capital are included simultaneously in the model, school attainment was not significantly related to growth. This can be attributed to the direct relationship between skills and years of schooling.

The endogenous economic growth model is constantly being reviewed and extended. In Romer (1986), technological progress is determined by knowledge accumulation, while Lucas (1988) considers education and human capital accumulation as drivers of productivity growth. In the latter approach to endogenous economic growth models, human capital stock and the development level HEIs can determine technological progress. The fundamental idea is that the development and expansion of higher education contributes to economic growth. Thus, a new branch of research on the economic impact of universities emerged. A variety of proxy-variables for the development level of HEIs have been proposed. Usually, the development level of HEIs are measured simply by student population, the number of universities and the share of student in the appropriate age-cohort (Valero and Van Reenen, 2016). In this paper, we develop this a step further and show the role of university efficiency in explaining regional economic growth, supplementing the findings of Barra and Zotti (2017), as described in our theoretical framework.

4. Theoretical framework

Based on the literature review we formulated the following theoretical framework describing how the efficiency of a higher education system may be related with rate of regional economic development.

We assume that a regional economy can be described by the production function (1):

$$Y = F(K, H, A) \tag{1}$$

where Y is GRP, K is the stock of physical capital in the regional economy, H is the stock of human capital in the regional economy, A is the technological progress (total factor productivity).

We also assume for simplicity that the function F has the form (2):

$$Y = AK^{\alpha}H^{\beta} \tag{2}$$

Human capital can be represented as the product of the total labor resource in the regional economy (L) and average amount of human capital per unit of the labor force (h) (3):

$$H = hL \tag{3}$$

Taking logs from both parts of equation (2), differentiating with respect to time and using the fact that $\frac{dlog(x)}{dt} = g_x$ we obtain (4):

$$g_y = g_A + \alpha g_K + \beta g_H \tag{4}$$

where g_y is GRP growth rate, g_A is the total factor productivity gross rate, g_K is the physical capital stock growth rate, and g_H is the human capital stock growth rate.

Since we assumed relationship (3), we can split the growth rate of the human capital stock into two components: labor force (L) growth rate and the average amount of human capital per unit of the labor force (h) growth rate (5):

$$g_H = \alpha g_h + \gamma g_L \quad (5)$$

where α and γ are some constant weights.

The growth rate of human capital per unit of the labor force is determined by the dynamics of the share of the time spent by individuals on the accumulation of human capital (u) and efficiency of human capital accumulation (ϕ). In other words, we assume that $g_h = f(u, \phi)$ and consider the exponential form of the function $f(f(u, \phi) = e^{\phi u})$. In such a setting, the total stock of human capital in the regional economy can be represented by the formula (6):

$$H = f(u, \phi)L = e^{\varphi u}L \quad (6)$$

In the model represented by (1)–(6) the total human capital stock in the regional economy is determined by the three factors simultaneously. The first one is the size of the labor force, which, in practice, can be measured by the total number of the population employed. The second factor is the average time spent by individuals on the accumulation of human capital, which can be approximated by the share of the workforce with a particular level of education. Finally, human capital stock is determined by how efficiently it is accumulated. This efficiency can be associated with the productivity of the educational system. The productivity of regional higher education systems is based on concept of efficiency proposed by Farrell (1957) and is calculated using DEA methodology. This approach gives a measure which can be considered a proxy variable for the efficiency of the human capital accumulation process. The time spent by individuals on human capital accumulation is measured by the share of workforce with higher education.

Our theoretical framework suggests that the efficiency of HEI is related to GRP growth rates through human capital stock, since relatively productive universities have a greater positive impact on human capital accumulation. However, there are different indirect effects of university efficiency on economic growth rates. As shown in Agasisti et al. (2017) (i) a relatively efficient university may employ fewer staff to achieve the same results, and this additional labor force may find alternative employment in the regional economy and generate additional GRP; (ii) if the universities are perceived by society as efficient organizations, they may be more successful in building relationships and collaborations with other organizations in the region; (iii) efficient universities create incentives for the institutions with which they interact to become more efficient; (iv) more efficient universities can produce more with the same resources.

In the empirical validation of the theoretical model, described in detail below, we use a broader list of control variables which control for the structure of the regional economy, the convergence effect, and the scale of the regional higher education system. One more important consideration in our theoretical setting is spatial interaction. There is an argument that the borders of regional economies do not always correspond to the administrative borders of real regional economies in Russia. It is important, therefore, to analyze spatial effects which account for this particularity of regional economic development in Russia. The same argument may be formulated for regional higher education systems, so the model should also include spatial effects for their efficiency levels. In other words, we expect that one regional higher education system may contribute to the human capital accumulation and, consequently, to the economic growth of neighboring regions.

Based on this theoretical framework we formulate the following research hypotheses:

(H1) The efficiency of a regional higher education system is an important determinant of regional economic growth;

(H2) Positive spillovers exist: the rates of regional economic growth are determined by the efficiency of higher education systems in neighboring regions.

5. Methodology and data selection

We employ a three-step methodology to analyze the link between the efficiency of regional higher education systems and the rates of regional economic growth, and to test hypotheses H1 and H2.

5.1 Efficiency estimation on the institutional level

We employ a two-stage DEA and the bootstrap procedures suggested by Simar and Wilson (2007). Simar and Wilson (1999) demonstrated that DEA scores obtained in the first stage are biased, and the environmental variables from the second stage are correlated to the output and input variables, therefore a bootstrap procedure is recommended. The two-stage DEA assumes that environmental variables might affect university outputs and proposes a re-estimation of the DEA model with adjusted outputs for environmental variables through the bootstrap procedure.

DEA involves the selection of the orientation (input or output) and the type of returns to scale. We consider an output-oriented model with the assumption of constant returns to scale (CRS). CRS was considered suitable, as universities cannot increase scale effects rapidly (Nazarko and Saparauskas, 2014). An output-oriented model evaluates how much outputs can be increased holding inputs fixed (Agasisti and Perez-Esparrells, 2010). This seems to be a reasonable assumption in the context of the efficiency in higher education due to the existing regulations that fix education production costs.

The linear programming model, assuming an output-oriented framework and CRS, must be solved for the $k-th$ Decision Making Unit (DMU). In our case, DMUs are universities, which transform an input vector $X_k = (x_{1k} \dots x_{jk}) \epsilon R_+^j$ into an output vector $Y_k = (y_{1k} \dots y_{sk}) \epsilon R_+^s$:

$$\max \theta_k \tag{7}$$

$$\theta_k y_{ks} \leq \sum_{i=1}^{N} \lambda_i y_{is}, s = 1..S; S = \#\{outputs\}$$

$$x_{kj} \geq \sum_{i=1}^{N} \lambda_i x_{ij}, j = 1..J;\ J = \#\{inputs\}$$

$$\lambda_i \geq 0$$

Here θ_k is the value of the efficiency of the $k-th$ DMU and satisfies $\theta_k \geq 1$. θ_k measures the efficiency of the $k-th$ unit as the distance to the frontier, which is the linear combination of the best practice units. $\theta_k \geq 1$ means that the DMU is below the best practice frontier (inefficient), while $\theta_k = 1$ means that the DMU lies on the estimated frontier (efficient). λ is a $(k \times 1)$ vector of intensity variables (Simar and Wilson, 2007).

Following Simar and Wilson's (2007) approach, the second stage model is constructed as a censored regression:

$$\widehat{\theta_k} = \alpha + \beta Z_k + \varepsilon_k, \qquad k = 1, \dots, N \tag{8}$$

where $\widehat{\theta_k}$ is the efficiency score obtained by solving equation (7); Z_k is the vector of environmental variables that can influence efficiency through the vector of parameters $\hat{\beta}$, and ε_k indicates statistical noise. Equation (8) is estimated only for $\widehat{\theta_k} > 1$. A more detailed description of the algorithm used, and the results are presented in Annex 2.

In order to formulate the efficiency model, it is necessary to make some assumptions regarding production processes in universities and about the input and output sets. At this step, we have to consider universities as multi-product organizations (Baumol at al., 1982) which utilize different inputs in order to produce different outputs. Following the literature, we assume university production technology with three inputs. The first input is the financial resource of the university measured by income from all sources at constant prices (normalized by the number of academic staff). This variable is common in research concerning university production functions (Agasisti and Johnes, 2009; Agasisti and Perez-Esparrells, 2010). This indicator is also a good proxy of the quality of input (i.e. the availability of academic staff, facilities, etc.). The second input is the relative weight of academic staff with advanced degrees (Candidate of Sciences, i.e. the Russian analogue to a PhD or Doctor of Sciences) in the total number of academic staff (excluding part-time staff and independent contracts). This variable measures the human resources available for HEI in order to carry-out teaching and research activities. Research usually uses the total number of academic staff as the input that measures the university's human capital resources (Agasisti and Johnes, 2009; Agasisti and Perez-Esparrells, 2010; Agasisti and Pohl, 2012; Wolszczak-Derlacz and Parteka, 2011). However, by using the share of academic staff with advanced degrees, we want to capture not the quantity, but the quality of human resources available. The third input is the average unified state exam (the entrance exam for Russian universities) score which reflects the quality of entrants. This variable is important for universities, however, often it is not clear if this indicates an output or an input in the production process (Abankina et al., 2013). If we treat this variable as an input, we assume that more prepared students are an important resource for the university. If we consider it as an output, the underlying assumption is that this variable reflects the ability of university to attract the most talented students, indicating the reputation of the university. We consider the average entrance exam score as an input variable, following (Johnes, 2006; Barra and Zotti, 2017; Agasisti et al., 2017).

The set of outputs consists of three variables reflecting three different activities of higher education institutions – research, collaborations with industrial partners as a proxy for knowledge transfer, and teaching. The first output is the total number of publications indexed in Web of Science, Scopus and the Russian science citation index (normalized by the total number of academic staff). This variable reflects the scientific productivity of university academic staff (Parteka and Wolszczak-Derlacz, 2013; St. Aubyn et al., 2009). The second output is the total income from grants obtained for applied research carried out by the university. This variable reflects the engagement of the university in collaborations with industrial partners and partially measures the money spent by companies on applied researches conducted by the university. In the Russian context, this factor is a good proxy for knowledge transfer which reflects cooperation between universities and industry. Finally, the third output is the total number of students per academic staff member. The most widespread indicator of university teaching activity is the number of graduates (Agasisti and Johnes, 2009; Bonaccorsi et al., 2007; Agasisti and Pohl, 2012). Due to data availability constraints, we cannot use the same indicator. However, since the dropout ratio in Russia is very close to zero, the correlation between the total number of students in the university and the total number of graduates is very high, so in Russia, we can replace the total number of graduates with the total number of students without losing much information.

In order to take into account the different internal characteristics of HEI which may potentially affect the production process inside the university, we used a set of exogenous variables to determine the efficiency scores correctly. Different papers examining efficiency in the higher education sector stress that there are different factors besides inputs and outputs which may affect the production process of the university. For instance, some of the environmental variables are the population size of the city where the university is located, the percentage of students receiving need-based financial aid (Nazarko and Saparauskas, 2014); the year of the university's foundation (Agasisti et al., 2017; Wolszczak-Derlacz and Parteka, 2011); a dummy variable reflecting the presence of a medical school in the university (Agasisti et al., 2017; Wolszczak-Derlacz and Parteka, 2011); real GDP per capita in the region of the university, the number of different faculties, the share of core funding revenues in total revenues, the share of women in the academic staff (Wolszczak-Derlacz and Parteka, 2011); the number of years since a technology transfer office opened at the university, the percentage of dropouts, and funding received from central government (Agasisti et al., 2017).

Given the inputs and outputs of the efficiency model, we employ five exogenous variables. The first two variables reflect the structure of the student body: the share of master's students in the total number of students; and the share of full-time students. Such indicators influence the university's strategy and the structure of the production process. For instance, if most students at the university are part-time, the university utilizes a fundamentally different educational model with a different structure of costs and resources. The third exogenous variable is a dummy variable for the university being located in the capital city of the region. The underlying assumption here is that a university located in the capital city is usually oriented towards students from the whole region. Capital cities are usually more attractive for living, so compared to universities in other cities, these institutions may be more attractive for students from other cities and regions, and this heterogeneous level of attractiveness might affect efficiency. The level of competition in the regional higher education market is also an important factor which may determine the level of efficiency (Leshukov et al., 2015). Universities that operate in a highly competitive environment tend to consolidate their resources and perform better (see also the conceptual discussion in Agasisti, 2009). The general measure of competition, used as an exogenous variable in the efficiency model, is the share of students in the total number of students in the region. Finally, we use a dummy variable which indicates the presence of a medical faculty in the university. A field of study such as medicine may have a strong influence on the technology used by university. For example, universities with a medical faculty usually have their own hospitals and the structure of the education process differs from the study programs in other fields. These particularities may also affect the production technology and the associated efficiency.

Table 1 presents the inputs, outputs and environmental variables used for the efficiency model.

[Insert Table 1 here]

5.2 The efficiency of regional higher education systems

In addition to considering environmental variables for the assessment of the efficiency scores of HEIs one more issue here is to measure the efficiency at the level of the whole regional higher education system. Due to the particularities of regional structures in Russia and data availability constraints, we cannot associate one university with a particular sub-regional territory (c.f. Agasisti et al., 2017), so we aggregate the efficiency scores obtained at the institutional level to the regional level. Efficiencies calculated at the second step report weighted averages of university performances by the total number of students of local universities in certain regions according to formula (9):

$$eff_i = \frac{\sum_{j=1}^{n} eff_j \times students_j}{\sum_{j=1}^{n} students_j}, \tag{9}$$

where i is number of regions; j is number of universities in region i; n is quantity of universities in region i; eff is the efficiency score; $students$ is the total number of students in university.

At the end of this aggregation, we obtain one synthetic indicator that measures the average level of university efficiency for any given region, and the subsequent analysis of economic growth is conducted at the regional level.

5.3 Modeling regional economic growth

The third and final step of our methodology is to estimate an economic growth model with the regional higher education system efficiency score among the explanatory variables. Within the theoretical model, we expect education, in general, and the efficiency of higher education systems, in particular, to cause economic growth (Hanushek, 2016), as discussed in sections 2 and 3.

The regional economic growth model is represented by the formula (10):

$$\begin{aligned}\Delta GRP_{t,j} = \alpha_0 + \alpha_1 \Delta GRP_{j,t-1} + \alpha_2 log\left(GRP_{j,t-1}\right) + \alpha_3 \Delta INV_{j,t} + \alpha_4 \Delta POP_{j,t} + \\ \alpha_5 PSEC_{j,t} + \alpha_6 CME_{j,t} + \alpha_7 IND_{j,t} + \alpha_8 EMPHE_{j,t} + GRAD_{j,t} + \alpha_9 EFF_{j,t} + \\ \alpha_{10}(EFF \times W)_{j,t} + \alpha_{11}(\Delta GRP \times W)_{j,t} + \mu_{j,t} + \tau_t + \varepsilon_{j,t}\ ,\end{aligned} \quad (10)$$

where $\Delta GRP_{t,j}$ is GRP growth rate; $log(GRP_{j,t-1})$ is the log of GRP in the previous period; $\Delta INV_{j,t}$ is the investment growth rate; $\Delta POP_{j,t}$ is the population growth rate; $PSEC_{j,t}$ is the share of the public sector in GRP; $CME_{j,t}$ is the share of commercial mineral extraction in GRP; $IND_{j,t}$ is the share of industries in GRP; $EMPHE_{j,t}$ is the share of the employed population with higher education; $GRAD_{j,t}$ is the total number of university graduates (bachelor + masters); $EFF_{j,t}$ is the efficiency level of the regional higher education system; $EFF \times W_{j,t}$ is the efficiency spatial lag; $\Delta GRP \times W_{j,t}$ is the spatial lag of GRP growth rates; $\mu_{j,t}$ are individual region-specific effects; τ_t are time effects; $\varepsilon_{j,t}$ are errors.

The regional economic growth model contains standard variables which correspond to our theoretical framework and which are used in most research devoted to economic growth modeling. These variables are the investment growth rate as a proxy for the physical capital stock growth rate; employed population growth rates; the share of employed population with higher education as a measure of average time spent by individual on human capital accumulation; the log of GRP in the previous period, which captures the convergence effect

(Sala-i-Martin, 1994). As suggested by the theoretical framework, the economic growth model also contains the efficiency measure of regional higher education systems reflecting the efficiency of human capital accumulation.

As the human capital of one region could be accumulated by higher education institutions of other regions (due to educational mobility which is high in Russia (Kashnitsky et al., 2016)) we include the total number of university graduates in the region, which controls for the scale of the regional higher education system. We also employ a set of variables to capture the structure of regional economies. These variables are the share of the public sector (education, public administration, healthcare) in GRP, the share of commercial mineral extraction in GRP and the share of industries in GRP. There are alternative ways of taking into account the regional economy's structure in economic growth modeling. For instance, Kufenko (2015) used commercial mineral extraction per capita. Our model accounts not only for commercial mineral extraction, but also for the public sector and industries, so the optimal way to take into account the economy's structure is to consider the shares of these sectors in GRP.

A third set of the variables included in the model is needed to resolve the second research hypotheses of this study. These variables are the spatial interaction of the efficiency which accounts for spillover effects (Agasisti et al., 2017); the growth spatial interaction which accounts for the positive spatial correlation in regional growth rates in Russia (Demidova, 2015).

In order to construct spatial interactions, we used a simple inverse distance matrix. This choice is based on evidence that the estimates and inferences in spatial regression models are not sensitive to the choice of the spatial weight matrix (LeSage and Pace, 2014).

5.4 Data and descriptive analysis

The source of the data for the efficiency evaluation is the "Annual Monitoring of Performance of Higher Education Institutions"[8] conducted by the Russian Ministry of Education and Science. This monitoring was launched in 2012, so data for the period from 2012 to 2016 are available. However, we used only data for the period from 2012 to 2015 because of missing data in the last year. Given the limited number of years for which data are available, we must interpret the results as only the effect for short-run region economic growth.

Only public and head universities of Russia were included in the analysis. Such a limitation is imposed on the sample to reduce the level of university heterogeneity in terms of

[8] www.miccedu.ru

their production functions. The limitation does not reduce its representativeness, since non-public universities account for 18% of the 5-year average of the total student population[9]. Given all the constraints, the sample contains 449 universities located in 77 regions and has data for each year within the period from 2012 to 2015.

Since the outlier problem and missing values should be taken into account in the efficiency analysis, preliminary data processing was implemented. In order to deal with missing values, we use an imputation procedure based on the classification and regression trees algorithm. In order to eliminate outliers, we used capping correction – upper outliers were replaced by the values that correspond to the quantile 0.975; the lower outliers were replaced by the values that correspond to the quantile 0.025. A detailed description of the preliminary data processing is presented in Annex 4.

Our starting point is a descriptive analysis of the institutional characteristics needed to assess university efficiency levels. Table 2 presents the key descriptive statistics on the institutions in our sample. The first variable represents the total income from all sources per faculty member. The value of this variable increased during the period 2012–14 (9% growth in 2014 compared to 2012 with the largest increment of 7% in 2014 compared to 2013). In 2015 the income per academic staff member fell by 17% in comparison with 2014. This reduction was largely due to the financial crisis in Russia 2014–15 and the related budget cuts in the educational sector. The relative weight of academic staff with advanced degrees is characterized by stable growth – it increased from 64% in 2012 to 70% in 2015. The third input variable is the average entrance exam score, which has stable values.

The first output indicator is the number of publications in journals indexed in Russian Science Citation Index, Web of Science and Scopus, per academic staff member. This indicator increased significantly during the period under review: the indicator was 58% higher in 2015 in comparison with 2012. This dramatic rise can be attributed to the Performance Based Funding scheme in 2015. The general principle of this scheme is that the number of publicly funded student slots available for the university is determined according to a formula, and scientific productivity is one of the components in this formula. Another explanation for this increase is that during the period several large government projects were launched to increase the scientific productivity of Russian universities (the excellence initiative "5-100 project", the establishment of national research universities, the establishment of flagship universities). The next output variable is the total quantity of R&D per academic staff member. The value of this indicator

[9] These calculations are based on the data from Russian Federation Federal State Statistics (www.gks.ru).

decreased for the period 2012–14. The largest reduction of this variable value (18%) was between 2014 and 2015. This reduction was determined by the negative macroeconomic shock in Russian economy and consequently by a decrease of private funds for R&D implementation. The last output indicator is the student to academic staff ratio. This indicator was stable and the value averages is about 10 students to each member of academic staff.

The share of masters' students, which is the first environmental variable, rose by 2.9%: and reached 7.3% of the total number of students in 2015. This growth was primarily due to the increase of number of state-funded places in master's programs. The share of full-time students in the total number of students was virtually unchanged for the period. The absence of a clear dynamic was related to the stable proportion of budget slots for full-time students. The average market share of the university increased during the period from 10.6% in 2012 to 12.3% in 2015 which was related to the policy of consolidation in the national higher education system. This included university mergers and the closing down of low-performing universities. The share of universities located in regional capital cities was stable at the level 91%, as was the share of universities with medical faculties (11%).

Overall, we can observe multidirectional dynamics of the institutional characteristics, so we can expect ambiguous dynamics of university efficiency during the period.

[Insert Table 2 here]

The source of the data for the regional economic growth model estimation is the Russian Federal State Statistics Service[10]. The sample includes the 77 regions that have efficiency scores for their regional higher education systems. We used data for most of the variables for the period from 2012 to 2015, but the variable for GRP growth rate covers the period from 2011 to 2015. The extension of the period for this variable is needed to use more lags for the dependent variable in the model identification using the sys-GMM approach (see section 5.5). In order to deal with the outlier problem we used the same capping procedure outline above (see Annex 4).

The descriptive statistics for the variables used for the regional economic growth model estimation are presented in Table 3. The average GRP growth rate declined 2012–15. Due to the economic crisis, in 2015 Russian regions on average experienced economic contraction (-0.2% in 2015 compared to the previous year). This decrease in GRP growth rate was followed by a decline in the investment growth rate. The investment growth rate decreased by 12.5 percentage

[10] http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/

points – from 2.6% in 2012 to -9.89% in 2015. The most rapid decline (4.3%) was between 2014 and 2015. The total number of graduates steadily declined by 9.1% – from 18,400 in 2012 to 16,700 in 2015. This reduction is primarily due to demographic factors. The peak of student numbers (7.5 million) was observed in 2008. For demographic reasons from that time the student body began to decrease and in 2014 it was 5.2 million. The dynamics of the share of the employed population with higher education corresponds to the non-standard mechanism according to which the Russian labor market adapts to macroeconomic shocks. The distinctiveness of this mechanism lies in the dominance of cost adaptation over quantitative adaptation. In other words, the Russian labor market reduces wages rather than the number of staff. Wage flexibility mitigates negative shocks by protecting employment and stemming unemployment growth (Gimpelson and Kapeliushnikov, 2015). Another variable that grew, in spite of negative economic shocks, was the efficiency level of regional higher education systems. This indicator was 7% higher in 2015 compared to 2012. The other variables used in economic growth model estimation had stable values.

[Insert Table 3 here]

5.5 Dealing with endogeneity and other econometric details

In the economic growth model represented by equation (10) we acknowledge the problem of endogeneity i.e. the correlation between some regressors and the error term. Ignoring this problem, we obtain biased and inconsistent parameter estimates that lead to an incorrect interpretation of modeling results. Particularly, we may assume here that the efficiency level of a regional higher education system does not influence GRP growth rate, but relatively efficient regional higher education systems tend to be formed in the regions with high rates of economic development. The most widespread way of overcoming this problem is using instrumental variables. In order to deal with this problem we employ a GMM dynamic panel data estimator (sys-GMM) (Arellano and Bover, 1995). The sys-GMM estimator acts like an instrumental variables approach and instruments endogenous variables by their lags. Using this technique gives us evidence that we are exploring causal relationship between the efficiency of regional higher education system and GRP growth rates. The basic argument here is that efficiency in year t affects the efficiency of year $t+n$ but not the economic growth in the same year $t+n$ directly. In order to check the reliability of the model, the Hansen-Sargan test for the overidentifying restriction, and a second-order autocorrelation test are used. The estimation strategy can be summarized as follows. At the first step the model is estimated by taking the first differences (Wooldridge, 2002). Second, lagged levels of the dependent variables are added in order to estimate the two-step sys-GMM model (Roodman, 2006). In order to obtain robust results, we use corrected standard errors (Windmeijer, 2005). We use all available lags as instruments in the sys-GMM regression.

6. Results of econometric analysis

6.1 Estimation of HEI's efficiency

In order to obtain efficiency scores on the institutional level, taking into account exogenous factors, a double-bootstrap data envelopment procedure was implemented. A detailed description of the double-bootstrap procedure, as well as the intermediate calculation results are presented in Annex 2. The double-bootstrap procedure uses Farrell's efficiency concept (Farrell, 1957) to analyze the efficiency level of HEI. According to this concept, it fulfills the condition of $\theta_k \geq 1$, with a value of one indicating that a university belongs to the production frontier and is identified as efficient, while values between one and infinity correspond to inefficient universities located below the best practice frontier. Within the economic growth model estimation, we prefer to work with Shephard measures which are simply the inverse of the Farrell ones (Bogetoft and Otto, 2010). The Shephard concept satisfies the condition $\theta_k \leq 1$, with unity representing an efficient observation, and values between zero and one indicate an inefficient observation. The distributions of the DEA efficiency scores obtained at the institutional level are presented in Figure 3.

[Insert Figure 3 here]

The key descriptive statistics of efficiency estimates at institutional and regional levels are presented in Table 4. Figure 3 and Table 4 show that efficiency values for each considered year are above 0.67 and fluctuate around 0.74. Distributions of efficiency scores have a normal distribution but are characterized by higher standard deviations; therefore, we note that DEA efficiency scores discriminate universities in the sample well. The average efficiency scores for different regions are also graphically depicted in Figure 4.

[Insert Table 4 and Figure 4 here]

The standard deviations presented in Table 4 show that after aggregation at the regional level, the standard deviations of our efficiency scores become lower. DEA efficiency scores, however, still discriminate universities well in terms of their efficiency. At both the regional and institutional levels, efficiency of HEI in Russia increased in the period from 2012 to 2015. The average annual growth rate of efficiency at the institutional level was 3% and at the regional level was 1%.

Table 4 also demonstrates that the standard deviations of efficiency scores, characterizing the level of heterogeneity of regional higher education system in terms of efficiency, are not stable over time. We assume here that the heterogeneity or homogeneity of regional higher education systems in terms of the efficiency of particular universities may also affect the rates of regional economic development, so we include also the standard deviation of within-region efficiency scores into the model for explaining economic growth, in addition to the average regional efficiency.

6.2. Regional economic growth model estimation

The economic growth model was estimated using Stata 13 software and the package xtabond2 (Roodman, 2006). Four different specifications were considered, and the results are reported in Table 5.

The baseline model (Model 1) includes the standard variables used in economic growth modeling (Kufenko, 2015) as pointed out in section 5.3. These variables are the rate of investment growth; the rate of population growth; the share of the employed population with higher education as a measure of accumulated human capital in the region; the total number of university graduates, which controls for the scale of regional higher education system; the log of GRP in the previous period in order to capture convergence effects; and variables that reflect the structure of the regional economy: the share of the public sector in GRP, the share of commercial mineral extraction in GRP and the share of industries in GRP. The parameters of this model are statistically significant and the signs of the estimated parameters correspond to the underlying theoretical assumptions: GRP growth rate is positively related to the employed population growth rate, the growth rate in previous period, the share of the employed population with higher education, and the total number of universities graduates. As expected, GRP growth rate negatively relates to the total GRP in the previous period, confirming the existence of convergence in growth rates: poor regions tend to grow faster than the rich ones.

Model 2 contains an additional explanatory variable: the DEA average efficiency score of regional higher education systems. This model confirms the main hypothesis of this study and demonstrates that the efficiency of regional higher education systems is an important determinant of regional economic development growth rates even if we control for the number of graduates. The positive and statistically significant relationship between GRP growth and regional higher education system efficiency is stable and can be observed in all subsequent specifications of the model. We also implemented a special robustness check in order to obtain additional evidence that this positive relationship exists. We used the stochastic frontier analysis (SFA) efficiency score instead of the DEA efficiency score in the economic growth model. The detailed description of SFA efficiency estimation and the results of the regional economic growth model with SFA efficiency scores are presented in Annex 3.

[Insert Table 5 here]

Model 3 additionally contains variables reflecting the standard deviation of university efficiency scores. The elasticity of GRP growth rates by this variable is positive and statistically significant, thus it is better for regional development to have heterogeneous efficiency in regional higher education systems. A possible explanation for this finding is that universities with different levels of efficiency contribute to the development of different areas of regional economies. Particularly, strong and efficient higher education institutions may create and adopt innovations and new ideas, while weaker and less efficient universities may concentrate on the production of workers for regional labor market.

Finally, Model 4 checks the hypothesis about spillover effects, showing that the parameter of efficiency spatial interaction is statistically significant and has a negative sign. This means that we have negative spillover effects. Such a finding may be explained by the fact that regional higher education systems in Russia tend to compete with each other. Each university tries to attract the brightest students and the best academic staff from neighboring regions. If, in one region regional, the higher education system is very strong and efficient and in the neighboring regions is not, this efficient system will extract resources, first of all human resources, from the other regions. That is why, if a region is located near regions with efficient higher education systems, its competitiveness in terms of competition with other regions will be relatively lower. The parameter of economic growth spatial interaction, as expected, is positive and statistically significant. As mentioned, existing evidence suggests that spatial correlation in regional economic growth rates in Russia is positive, so our results are in line with the findings of previous studies.

7. Discussion and concluding remarks

Universities are multi-product organizations (Baumol et al., 1982) with multiple impact channels. Usually three different types of university contributions to economic development are highlighted – a general economic approach, suggesting that universities as economic agents generate additional aggregate demand in regional economy (Elliot et al., 1988); a skill-based approach, which analyses the contribution of higher education in terms of human capital reproduction (Bluestone, 1993); and an innovation approach, which considers universities as integrators of regional innovation ecosystems.

There is evidence that the economic impact of universities in Russia in terms of these approaches is positively related to the scale of higher education systems (Egorov et al., 2017). In other words, the quantity of higher education matters for regional economic development. This study is a first attempt to shed the light on the question of whether the efficiency level of regional higher education systems is positively related to regional economic growth rates in Russia. Using a framework which considers the efficiency level as a good instrument to capture the impact of universities on the community (Agasisti et al., 2017; Barra and Zotti, 2016) we estimate higher education efficiency at the institutional level, then aggregate these estimates in order to obtain an aggregate measure of regional higher education system efficiency and, finally, construct a regional economic growth model which treats regional higher education system efficiency as one of the explanatory variables. In order to evaluate the efficiency, we assume that there are different exogenous factors that are out of university management control and employ a 2-stage DEA procedure (Simar and Wilson, 2007). For the causal inference we employ the sys-GMM approach for the identification of economic growth models. We also employ spatial econometrics techniques in order to analyze spillover effects (the positive economic impact of universities on neighboring regions).

The estimated economic growth models shows that DEA efficiency scores (corrected for exogenous factors) are statistically significantly related to GRP growth rates. Moreover, we find statistically significant and negative spillover effects. The explanation behind this finding is that strong and efficient regional higher education systems may extract resources (predominantly human resources) from neighboring regions. Particularly, these regional higher education systems are more attractive for students and scholars from other regions. Such findings suggest that universities can ensure the competitiveness of the regions where they are located in relation to other regions. In such a highly centralized higher education system such as the Russian one (approximately 90% of all state-owned universities are governed by the federal authorities), this fact can be considered a significant incentive for regional authorities to collaborate with the higher education sector more actively.

The main policy implication of this study is that public policy in higher education has to be concentrated not only on the quantity of higher education and its availability, but on its efficiency level as well. Public policy should take into account the other side of this mechanism since efficient higher education systems extract resources from the neighbors, the regions that have weak higher education systems may be damaged. In order to overcome this constraint, it is necessary to develop networks of strong regional universities. Currently, the Russian higher education system is characterized by a geographical concentration of strong HEI. The first steps towards a more even distribution of high-quality universities were made within the framework of the “Flagship universities program” developed by the Russian Government in 2016; this should be continued with more intensity.

References

Abankina, I., Aleskerov, F., Belousova, V., Zinkovsky, K., & Petrushchenko, S. (2013). Evaluating Performance of Universities Using Data Envelopment Analysis. *Educational studies*, (2), 15-48.

Agasisti, T. (2009). Market forces and competition in university systems: theoretical reflections and empirical evidence from Italy. *International Review of Applied Economics*, 23(4), 463-483.

Agasisti, T., & Johnes, G. (2009). Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. *Education Economics*, *17*(1), 59-79.

Agasisti, T., & Perez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. *Higher Education*, *59*(1), 85-103.

Agasisti, T., & Pohl, C. (2012). Comparing German and Italian public universities: Convergence or divergence in the higher education landscape? *Managerial and Decision Economics*, *33*(2), 71-85.

Agasisti, T., Barra, C., & Zotti, R. (2017). *Research, knowledge transfer and innovation: the effect of Italian universities' efficiency on the local economic development 2006-2012*.

Aghion, P., Blundell, R., Griffith, R., Howitt, P., & Prantl, S. (2009). The effects of entry on incumbent innovation and productivity. *The Review of Economics and Statistics*, *91*(1), 20-32.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, *6*(1), 21-37.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, *68*(1), 29-51.

Aubyn, M. S., Garcia, F., & Pais, J. (2009). *Study on the efficiency and effectiveness of public spending on tertiary education* (No. 390). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Barra, C., & Zotti, R. (2017). Investigating the Human Capital Development–growth Nexus: Does the Efficiency of Universities Matter? *International Regional Science Review*, *40*(6), 638-678.

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

Battese, G. E., & Corra, G. S. (1977). Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. *Australian Journal of Agricultural and Resource Economics*, *21*(3), 169-179.

Baumol, W., Panzar, J., & Willig, R. (1982). Contestable markets and the theory of market structure. *Nueva York, Harcourt Brace Javanovich, Inc*.

Becker, G. (1964). human capital, NY. *GS Becker*.

Belenzon, S., & Schankerman, M. (2013). Spreading the word: Geography, policy, and knowledge spillovers. *Review of Economics and Statistics*, *95*(3), 884-903.

Bluestone, B. (1993). *An Economic Impact Analysis*. UMASS/Boston

Bogetoft, P., & Otto, L. (2010). *Benchmarking with Dea, Sfa, and R* (Vol. 157). Springer Science & Business Media.

Bonaccorsi, A., Daraio, C., Lepori, B., & Slipersaeter, S. (2007). Indicators on individual higher education institutions: addressing data problems and comparability issues. *Research Evaluation*, *16*(2), 66-78.

Burgette, L. F., & Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American journal of epidemiology*, *172*(9), 1070-1076.

Capello, R. (2011). Location, Regional Growth and Local Development Theories. *Aestimum*, *58*(1), 1-25.

Chirikos, T. N., & Sear, A. M. (2000). Measuring hospital efficiency: a comparison of two approaches. *Health Services Research*, *34*(6), 1389.

Demidova, O. (2015). Spatial effects for the eastern and western regions of Russia: a comparative analysis. *International Journal of Economic Policy in Emerging Economies*, *8*(2), 153-168.

Demidova, O., & Ivanov, D. (2016). Models of Economic Growth with Heterogenous Spatial Effects: The Case of Russian Regions. *HSE Economic Journal*, *20*(1), 52-75.

Egorov, A., Leshukov, O., & Gromov, A. (2017). The Role of Universities in Economic Development of Russian Regions.

Elliott, D. S., Levin, S. L., & Meisel, J. B. (1988). Measuring the economic impact of institutions of higher education. *Research in Higher Education*, *28*(1), 17-33.

Etzkowitz, H. (1993). Enterprises from science: The origins of science-based regional economic development. *Minerva*, *31*(3), 326-360.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3), 253-290.

Ferrier, G. D., & Lovell, C. K. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of econometrics*, *46*(1-2), 229-245.

Froumin, I., & Leshukov, O. (2015). National-Regional Relationships in Federal Higher Education Systems: The case of Russian Federation. In *Higher education forum: a COE publication* (Vol. 12, pp. 77-94).

Froumin, I., Kouzminov, Y., & Semyonov, D. (2014). Institutional diversity in Russian higher education: revolutions and evolution. *European Journal of Higher Education*, *4*(3), 209-234.

Carnoy, M., Froumin, I., Leshukov, O. & Marginson., S. (2018) *Higher education in federal countries.* SAGE Publishing.

Gimpelson, V., Kapeliushnikov, R., (2015) The Russian Labour Market Model: Trial by Recession. *Journal of the New Economic Association*, *26*(2), 249-254.

Hanushek, E. A. (2016). Will more higher education improve economic growth? *Oxford Review of Economic Policy*, *32*(4), 538-552.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, *90*(5), 1184-1208.

Hanushek, E. A., & Woessmann, L. (2007). *The role of school improvement in economic development* (No. w12832). National Bureau of Economic Research.

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT Press.

Huggins, R., & Johnston, A. (2009). The economic and innovation contribution of universities: a regional perspective. *Environment and Planning C: Government and Policy*, *27*(6), 1088-1106.

Johnson, M. S. (2008). Historical legacies of Soviet higher education and the transformation of higher education systems in post-Soviet Russia and Eurasia. In *The worldwide transformation of higher education* (pp. 159-176). Emerald Group Publishing Limited.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. Economics of Education Review, 25(3), 273-288.

Kashnitsky, I., Mkrtchyan, N., & Leshukov, O. (2016). Interregional youth migration in Russia: a comprehensive analysis of demographic statistical data. Educational Studies, 3, 169-203.

Kosor, M. M. (2013). Efficiency measurement in higher education: Concepts, methods and perspective. *Procedia-Social and Behavioral Sciences*, *106*, 1031-1038.

Kufenko, V. (2015). The resource curse and human capital. In *Economic Growth and Inequality* (pp. 42-61). Springer Gabler, Wiesbaden.

LeSage, J. P., & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2*(4), 217-249.

Leshukov, O., Platonova, D., & Semyonov, D. (2015). Does Competition Matter? The Efficiency of Regional Higher Education Systems and Competition: The Case of Russia.

Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery*, 1(1), 14-23.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, *22*(1), 3-42.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, *107*(2), 407-437.

McMillan, M. L., & Chan, W. H. (2006). University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. *Education economics*, *14*(1), 1-30.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, 435-444.

Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings* (pp. 41-63). NBER.

Morsy, L., Khavenson, T., & Carnoy, M. (2018). How international tests fail to inform policy: The unsolved mystery of Australia's steady decline in PISA scores. *International Journal of Educational Development*, *60*, 60-79.

Nazarko, J., & Saparauskas, J. (2014). Application of DEA method in efficiency evaluation of public higher education institutions. *Technological and Economic Development of Economy*, *20*(1), 25-44.

Parteka, A., & Wolszczak-Derlacz, J. (2013). Dynamics of productivity in higher education: Cross-European evidence based on bootstrapped Malmquist indices. *Journal of Productivity Analysis*, *40*(1), 67-82.

Pinheiro, R., Benneworth, P., & Jones, G. A. (Eds.). (2012). *Universities and regional development: A critical assessment of tensions and contradictions*. Routledge.

Raghunathan, T. E., Solenberger, P. W., & Van Hoewyk, J. (2002). IVEware: Imputation and variance estimation software. Ann Arbor, *MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan.*

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy*, *94*(5), 1002-1037.

Roodman, D. (2006). How to do xtabond2: An introduction to difference and system GMM in Stata.

Sala-i-Martin, X. (1994). Cross-sectional regressions and the empirics of economic growth. *European Economic Review*, *38*(3-4), 739-747.

Sardadvar, S., & Vakulenko, E. (2016). Interregional migration within Russia and its East-West divide: evidence from spatial panel regression. *Review of Urban & Regional Development Studies*, 28(2), 123-141.

Shephard, R.W., (1953) *Cost and production functions*. Princeton University Press, Princeton.

Simar, L., & Wilson, P. W. (1999). Of course we can bootstrap DEA scores! But does it mean anything? Logic trumps wishful thinking. *Journal of Productivity Analysis*, *11*(1), 93-97.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, *136*(1), 31-64.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, *70*(1), 65-94.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312-320.

Valero, A., & Van Reenen, J. (2016). *The economic impact of universities: Evidence from across the globe* (No. w22501). National Bureau of Economic Research.

Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of economic growth*, *11*(2), 97-127.

Varga, A. (1997). Regional economic effects of university research: a survey. *Unpublished manuscript, West Virginia University, Regional Research Institute, Morgantown, WV*.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of econometrics*, *126*(1), 25-51.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: a two-stage multicountry approach. *Scientometrics*, *89*(3), 887-917.

Wooldridge, J. (2002). Econometric Analysis of Cross-Section and Panel Data. *MIT Press*, pp. 130, 279, 420-449.

Annex 1. Figures and tables

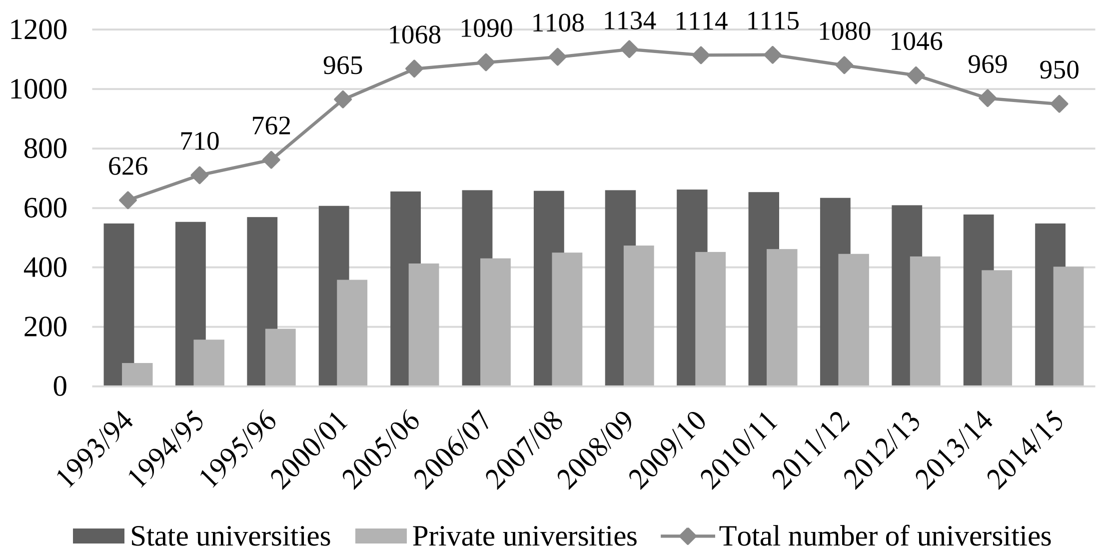

Figure 1. Number of organization of Higher Education in Russian Federation.

Source: Russian Federation Federal State Statistics (http://www.gks.ru)

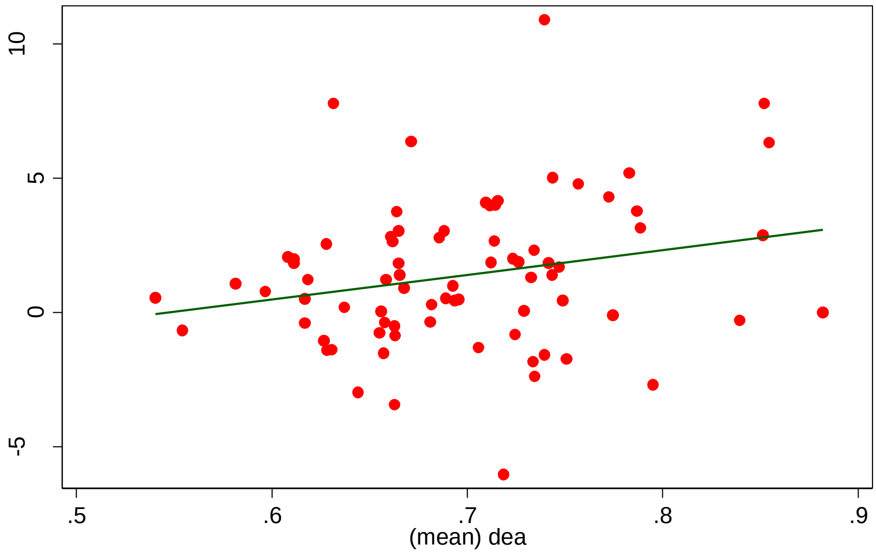

Figure 2. Scatter plot of DEA efficiency scores and GRP growth rates (average values for the period 2012–2015).

Notes: Plot presents the averaged values of efficiency scores of regional HE systems and GRP growth rates over 2012–2015. Regional HE system is a set of universities located within the administrative borders of the region.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» and Russian Federation Federal State Statistics.

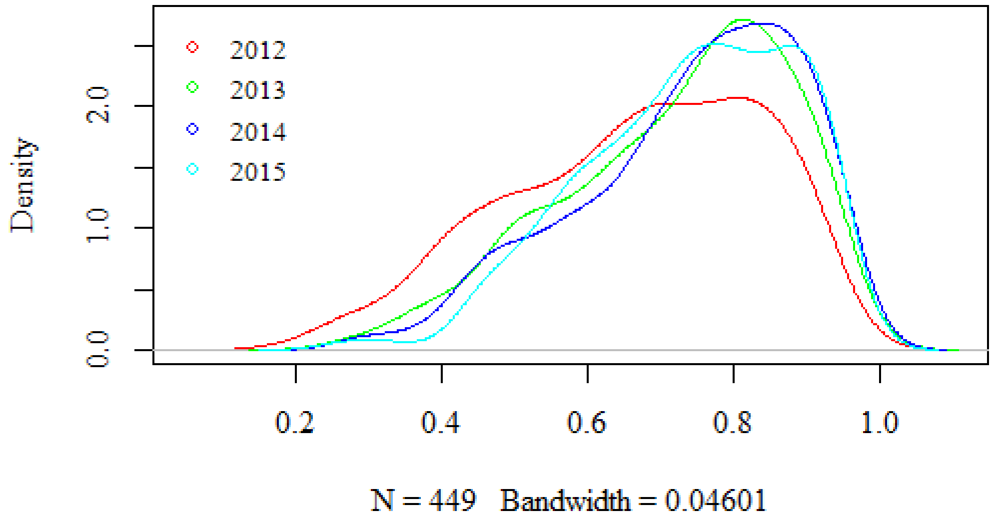

Figure 3*.* DEA efficiency distribution on institutional level.

Notes: Efficiency scores are distributed between 0 and 1. A value of 1 indicates that a university is efficient and lies on the best practice frontier.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions».

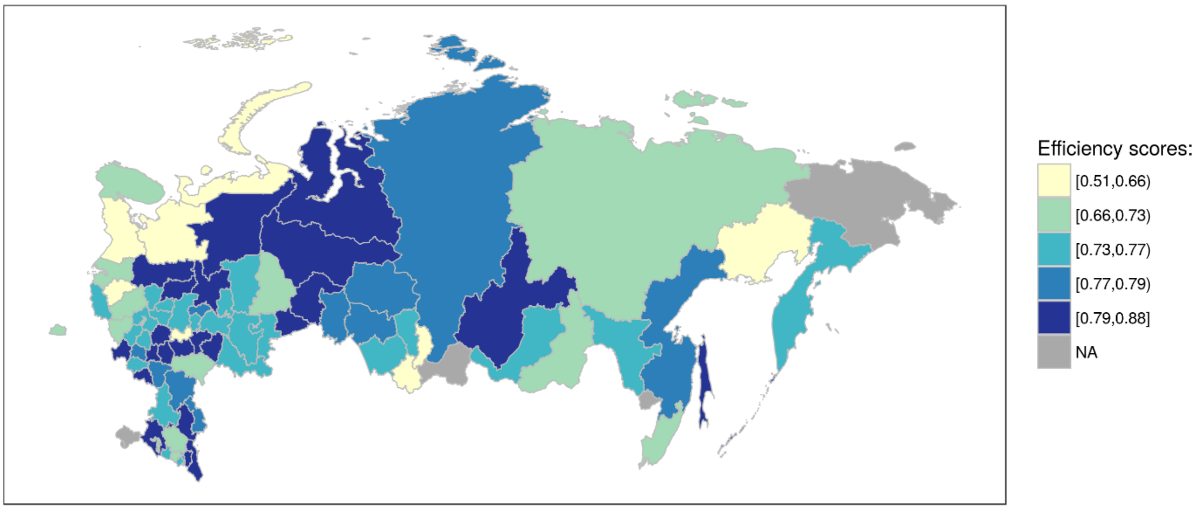

Figure 4. DEA efficiency scores of Regional Higher Education (HE) systems.

Notes: Plot presents the averaged efficiency scores of Regional Higher Education (HE) systems over 2012–2015. Regional HE system is a set of universities located within the administrative borders of the region. A value NA indicates the lack of the data for this region. The sample includes 77 regions.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Table 1. Variables description.

Variable name	Description	Unit of measurement
	Inputs	
Income	The income of educational organization from all sources per number of Faculty members	Thousand rubles
Academic staff	The relative weight of academic staff with advanced degrees	%
Exam scores	The average entrance exam score of entering students	Score
	Outputs	
Publications	The number of publications in science journals indexed in RSCI, Web of Science and Scopus, per capita of academic staff	Unit
R&D	The total quantity of R&D per one employee of academic staff, thousand rubles	Thousand rubles
Students	Students to academic staff ratio	Persons
	Environmental variables	
Masters' students	Share of masters' students in total number of students	%
Full-time students	Share of full-time students in total number of students	%
Capital city	Location of university in the capital city of the region (dummy)	-
Market share	Market share of university – share of students in the university in total number of students in the region	%
Medical faculty	Presence of medical faculty (dummy)	-

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Table 2. Descriptive statistics of the variables used for efficiency evaluation.

Variable name	2012 (N=449)		2013 (N=449)		2014 (N=449)		2015 (N=449)	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Inputs								
Income	2832	1448	2892	1333	3097	1395	2577	1180
Academic staff with	64	16	66	15	69	15	70	15
Exam scores	65	9	68	10	65	10	64	8
Outputs								
Publications	78	49	84	51	128	77	184	119
R&D	236	327	238	333	214	281	176	228
Students	9	3	10	3	10	3	10	2
Environmental variables								
Masters' students	4.4	8.2	5.0	8.2	5.3	7.5	7.3	7.9
Full-time students	62.1	18.1	62.6	18.4	63.9	18.3	61.0	22.7
Capital city	91	-	91	-	91	-	91	-
Market share	10.6	13.2	11.0	14.0	14.2	11.9	12.3	15.3
Medical faculty	11	-	11	-	11	-	11	-

Notes: Information on incomes is adjusted to the level of December 2015 by using the annual national CPI. Data in the case of missing values an imputation procedure based on classification and regression trees algorithm was employed (Loh at al.,2011).

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Table 3. Descriptive statistics of the variables used for economic growth model.

Variable name	2012 (N=77)		2013 (N=77)		2014 (N=77)		2015 (N=77)	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
GRP growth rates (year to year), %	3.41	5.77	2.07	5.68	1.95	4.74	-0.20	2.52
Efficiency measured by DEA	0.71	0.08	0.75	0.07	0.76	0.08	0.76	0.07
Population growth rates, %	- 0.09	0.01	- 0.08	0.01	- 0.01	0.01	- 0.07	0.01
Investment growth rates, %	2.59	13.32	- 3.01	14.27	- 1.88	17.73	- 9.89	13.72
Gross regional product in previous period, bln rub.	910.8	190.1	927.3	196.3	949.7	198.9	827.9	169.4
Total number of universities' graduates thousand people	17.69	29.90	16.345	26.89	15.13	24.32	16.05	26.75
Share of employed with higher education,%	27.85	4.89	29.14	5.01	29.71	4.78	30.65	4.61
Share of public sector in GRP,%	17.87	7.46	18.99	7.68	18.53	7.39	17.12	7.19
Share of commercial minerals extraction in GRP, %	8.13	12.91	7.95	12.82	7.75	13.03	8.36	13.43
Share of industries in GRP, %	17.89	9.94	17.43	9.34	17.48	9.92	18.02	10.21

Notes: Information on incomes is adjusted to the level of December 2015 by using the annual national CPI.

Source: Authors' calculations from Russian Federation Federal State Statistics Service.

Table 4. Descriptive statistics of DEA efficiency scores over 2012–2015.

Statistics	2012 (N=449)	2013 (N=449)	2014 (N=449)	2015 (N=449)
Institutional level				
Mean	0.665	0.722	0.739	0.741
Median	0.685	0.756	0.767	0.756
Std deviation	0.173	0.156	0.152	0.143
Minimum	0.185	0.259	0.275	0.262
Maximum	0.967	0.981	0.970	0.966
Statistics	**2012** (N=77)	**2013** (N=77)	**2014** (N=77)	**2015** (N=77)
Regional level				
Mean	0.713	0.751	0.763	0.759
Median	0.725	0.758	0.771	0.762
Std deviation	0.077	0.070	0.078	0.070
Minimum	0.435	0.475	0.489	0.583
Maximum	0.897	0.900	0.912	0.882

Notes: Regional efficiency scores conform to estimates of regional HE systems. Regional HE system is a set of universities located within the administrative borders of the region.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Table 5. Results of regional growth model estimation (standard errors are presented in the brackets)

Variable name	Model1	Model2	Model3	Model4
Growth rate in previous period	0.860***	0.867***	0.866***	0.874***
	(0.017)	(0.018)	(0.020)	(0.020)
Investment growth rate	3.873***	4.294***	4.681***	4.382***
	(0.513)	(0.447)	(0.481)	(0.478)
Employed population growth rate	0.329***	0.308***	0.296***	0.297***
	(0.026)	(0.020)	(0.024)	(0.025)
Gross regional product in previous period (log)	-5.203***	-4.491***	-4.599***	-4.804***
	(0.470)	(0.369)	(0.421)	(0.622)
Share of employed with HE	0.171**	0.098*	0.111**	0.140**
	(0.051)	(0.039)	(0.047)	(0.046)
Total number of universities' graduates	0.069***	0.051***	0.052***	0.065***
	(0.011)	(0.008)	(0.010)	(0.012)
Share of commercial minerals extraction in GRP	-0.069**	-0.078***	-0.061**	-0.076**
	(0.023)	(0.021)	(0.021)	(0.025)
Share of industries in GRP	0.065**	0.068**	0.047*	0.038
	(0.021)	(0.021)	(0.023)	(0.027)
Share of public sector in GRP	-0.732***	-0.690***	-0.664***	-0.704***
	(0.047)	(0.050)	(0.058)	(0.059)
Efficiency DEA		1.070*	0.947*	3.111*
		(0.503)	(0.397)	(1.254)
Efficiency (DEA)standard deviation			9.230***	8.923***
			(1.834)	(2.070)
Efficiency spatial interaction				-1.426*
				(0.613)
Growth spatial interaction				7.513**
				(2.648)
Constant	82.011***	71.530***	70.828***	74.718**
	(5.876)	(5.674)	(6.185)	(11.342)
Hansen-Sargan	0.208	0.852	0.395	0.444
AR(2)	0.893	0. 588	0.658	0.747
# of observations	308	308	308	308
Significance levels: *** p-value<0.001; ** p-value<0.01; * p-value<0.05; . p-value<0.1				

Notes: All equations are estimated through a two-step system generalized method moment estimator with Windmeijer (2005) corrected standard error (in brackets).

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» and Russian Federation Federal State Statistics.

Annex 2. Data envelopment analysis calculations details

For DEA efficiency evaluation we used double-bootstrap procedure proposed in (Simar & Wilson, 2007).

The idea of the algorithm is as follows:

1. Compute $\hat{\theta}_k$ for all universities from the sample by solving equation (7).
2. Estimate equation (8) by using the truncated regression model of $\hat{\theta}_k$ on Z_k to obtain regression parameters $\hat{\beta}$ and residuals variance $\hat{\sigma}_\varepsilon$.
3. For each $k = 1, \dots, N$ repeat the following four steps (3.1–3.4) L_1 times to obtain N sets of bootstrap estimates $B_k = \{\hat{\theta}^*_{kb}\}_{b=1}^{L_1}$:

   3.1. For each $k = 1, \dots, N$ draw ε_k from the truncated normal distribution $N(0, \hat{\sigma}_\varepsilon^2)$ with left-truncation at $(1 - Z_k\hat{\beta})$.

   3.2. For each $k = 1, \dots, N$ compute $\theta^*_k = Z_k\hat{\beta} + \varepsilon_k$.

   3.3. For each $k = 1, \dots, N$ correct the outputs vector for the ration of efficiency score without environmental variables $\hat{\theta}_k$ to efficiency score with environmental variables θ^*_k like the follows: $y^*_k = y_k \cdot \frac{\hat{\theta}_k}{\theta^*_k}$.

   3.4. Using the corrected outputs vector y^*_k for each $k = 1, \dots, N$ estimate new efficiency scores $\hat{\theta}^*_k$ by solving equation (7).

4. For each $k = 1, \dots, N$ compute the bias-corrected efficiency scores $\hat{\hat{\theta}}_k$ using the first bootstrap estimates $\hat{\theta}^*_{kb}$ in B_k and the efficiency scores without environmental variables $\hat{\theta}_k$ as follows: $\hat{\hat{\theta}}_k = \hat{\theta}_k - \widehat{BIAS}(\hat{\theta}_k)$, where $\widehat{BIAS}\left(\hat{\theta}_k\right) = \overline{\hat{\theta}^*_k} - \hat{\theta}_k$; $\overline{\hat{\theta}^*_k} = \frac{1}{L_1}\sum_{b=1}^{b=L_1} \hat{\theta}^*_{bk}$.
5. Use the method of maximum likelihood to estimate the truncated regression model of $\hat{\hat{\theta}}_k$ on Z_k to obtain regression parameters $\hat{\hat{\beta}}$ and residuals variance $\hat{\hat{\sigma}}_\varepsilon$.
6. For each $k = 1, \dots, N$ repeat the next three steps (6.1–6.3) L_2 times to yield N sets of bootstrap estimates $\left\{\left(\hat{\beta}^*, \hat{\sigma}^*_\varepsilon\right)_b\right\}_{b=1}^{L_2}$:

   6.1. For each $k = 1, \dots, N$ draw ε_k from the truncated normal distribution $N(0, \hat{\hat{\sigma}}_\varepsilon^2)$ with left-truncation at $(1 - Z_k\hat{\hat{\beta}})$.

   6.2. For each $k = 1, \dots, N$ compute $\theta^{**}_k = Z_k\hat{\hat{\beta}} + \varepsilon_k$.

6.3. Use the maximum likelihood method to estimate the truncated regression of θ_k^{**} on Z_k obtaining estimates $\hat{\hat{\beta}}^*$ and $\hat{\hat{\sigma}}_\varepsilon^*$.

7. Use the bootstrap values and the estimates $\hat{\hat{\beta}}$, $\hat{\hat{\sigma}}_\varepsilon$ from the 5 step to construct estimated confidence intervals for each element of β and for σ_ε as described below.

Full description of the algorithm and underlying assumption can be found in original paper (Simar & Wilson, 2007).

The results of the second-stage regression are presented in the Table 6. The confidence intervals for most environmental variables do not contain zero, so this factors can be considered as statistically significant related to universities' efficiency scores. However, the significance of effect of most determinants has decreased over time, and even has changed the sign for some parameters. Such results can be explained by the economic crises that occurred in Russia during the reporting years. Economic turbulences neutralize some dependencies that existed in 2012.

Table 7 contains intermediate results of the Simar and Wilson algorithm and final efficiency scores. All these results, including confidence intervals values, were aggregated to the regional level.

Table 6. Results of the second-stage truncated regression in double-bootstrap DEA procedure.

Variable name	2012 (N=449)		2013 (N=449)		2014 (N=449)		2015 (N=449)	
	$\hat{\hat{\beta}}^*$	95% Conf. In.	$\hat{\hat{\beta}}^*$	95% Conf. In	$\hat{\hat{\beta}}^*$	95% Conf. In	$\hat{\hat{\beta}}^*$	95% Conf. In
Constant	-0.447	[-1.70, 0.53]	-0.101	[-0.92, 0.54]	-0.035	[-0.84, 0.59]	0.654	[0.13, 1.08]
Masters' students	-0.083	[-0.14, -0.04]	-0.014	[-0.04, 0.01]	-0.012	[-0.04, 0.01]	-0.015	[-0.03, 0.00]
Full-time students	2.055	[1.05, 3.13]	1.3	[0.69, 2.02]	1.368	[0.73, 2.07]	0.788	[0.28, 1.33]
Capital city	0.583	[-0.03, 1.37]	0.375	[-0.03, 0.90]	0.079	[-0.26, 0.53]	-0.037	[-0.34, 0.27]
Market share	-6.162	[-9.39, -3.76]	-2.408	[-3.81, -1.16]	-1.134	[-2.32, -0.13]	-1.1	[-2.04, -0.30]
Medical faculty	0.603	[0.15, 1.05]	0.778	[0.47, 1.12]	0.838	[0.54, 1.18]	0.654	[0.39, 0.92]

Notes: Confidence intervals are estimated by using bootstrap procedure (N = 2000).

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Table 7. Intermediate results of double-bootstrap DEA procedure and final efficiency scores on the regional level.

	$\hat{\theta}_k$				$\hat{\hat{\theta}}_k$				Lower bound				Upper bound				Mean over 2012-15	
id_region	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015	$\hat{\theta}_k$	$\hat{\hat{\theta}}_k$
1	1.43	1.32	1.24	1.25	1.49	1.37	1.28	1.29	1.45	1.33	1.25	1.26	1.53	1.41	1.31	1.31	1.31	1.36
2	1.47	1.26	1.32	1.28	1.51	1.30	1.37	1.32	1.48	1.27	1.33	1.29	1.54	1.33	1.41	1.35	1.33	1.38
3	2.18	2.00	1.97	1.47	2.31	2.11	2.04	1.54	2.23	2.03	1.99	1.49	2.42	2.19	2.10	1.60	1.90	2.00
4	1.26	1.27	1.28	1.21	1.33	1.32	1.32	1.25	1.27	1.29	1.29	1.21	1.39	1.36	1.35	1.28	1.25	1.30
5	1.07	1.11	1.09	1.07	1.14	1.16	1.17	1.14	1.08	1.12	1.10	1.09	1.19	1.21	1.24	1.20	1.08	1.15
6	1.19	1.15	1.07	1.20	1.24	1.20	1.11	1.23	1.20	1.16	1.07	1.20	1.28	1.24	1.13	1.26	1.15	1.20
7	1.29	1.25	1.24	1.29	1.34	1.29	1.30	1.34	1.30	1.26	1.26	1.31	1.39	1.33	1.35	1.39	1.27	1.32
8	1.23	1.24	1.25	1.45	1.34	1.33	1.32	1.52	1.25	1.26	1.26	1.46	1.42	1.40	1.38	1.58	1.29	1.38
9	1.26	1.11	1.19	1.23	1.29	1.14	1.22	1.27	1.26	1.12	1.20	1.24	1.30	1.17	1.24	1.29	1.20	1.23
10	1.31	1.21	1.23	1.26	1.38	1.29	1.29	1.32	1.32	1.23	1.25	1.27	1.42	1.35	1.35	1.36	1.25	1.32
11	1.57	1.43	1.42	1.35	1.68	1.53	1.51	1.42	1.59	1.46	1.44	1.36	1.77	1.61	1.58	1.48	1.44	1.54
12	1.68	1.50	1.51	1.50	1.80	1.62	1.61	1.58	1.72	1.54	1.53	1.52	1.89	1.71	1.68	1.64	1.55	1.65
14	1.38	1.34	1.41	1.63	1.41	1.39	1.45	1.70	1.38	1.36	1.42	1.65	1.44	1.42	1.48	1.76	1.44	1.49
15	1.41	1.31	1.30	1.31	1.50	1.40	1.37	1.37	1.44	1.33	1.31	1.32	1.56	1.47	1.42	1.42	1.33	1.41
16	1.15	1.23	1.32	1.41	1.21	1.26	1.36	1.46	1.15	1.23	1.32	1.42	1.27	1.28	1.40	1.49	1.28	1.32
17	1.23	1.20	1.21	1.17	1.32	1.28	1.28	1.23	1.26	1.23	1.23	1.19	1.38	1.33	1.34	1.27	1.21	1.28
18	1.59	1.42	1.35	1.43	1.66	1.48	1.41	1.50	1.60	1.43	1.36	1.45	1.72	1.53	1.47	1.55	1.45	1.51
19	1.44	1.41	1.45	1.49	1.52	1.48	1.50	1.54	1.46	1.43	1.47	1.50	1.60	1.53	1.54	1.58	1.45	1.51
20	1.52	1.38	1.19	1.12	1.57	1.42	1.21	1.14	1.53	1.39	1.19	1.12	1.61	1.45	1.23	1.16	1.30	1.34
21	1.37	1.44	1.34	1.46	1.45	1.51	1.40	1.51	1.39	1.46	1.36	1.48	1.52	1.56	1.46	1.54	1.40	1.47
22	1.50	1.39	1.25	1.17	1.67	1.42	1.28	1.19	1.47	1.39	1.25	1.16	1.82	1.44	1.30	1.20	1.33	1.39
23	1.46	1.33	1.33	1.31	1.52	1.39	1.39	1.36	1.48	1.34	1.35	1.32	1.56	1.43	1.43	1.39	1.36	1.41
24	1.47	1.22	1.18	1.14	1.51	1.26	1.22	1.18	1.48	1.23	1.19	1.15	1.55	1.28	1.25	1.22	1.25	1.29
25	1.24	1.16	1.20	1.35	1.28	1.20	1.24	1.38	1.25	1.17	1.20	1.35	1.32	1.23	1.28	1.41	1.24	1.28
26	1.24	1.17	1.13	1.14	1.29	1.22	1.18	1.18	1.25	1.19	1.15	1.15	1.33	1.26	1.22	1.21	1.17	1.22

27	1.34	1.19	1.21	1.38	1.45	1.27	1.28	1.45	1.37	1.21	1.22	1.39	1.53	1.33	1.33	1.51	1.28	1.36
28	1.22	1.19	1.24	1.33	1.24	1.21	1.27	1.36	1.22	1.19	1.24	1.33	1.26	1.23	1.29	1.38	1.25	1.27
29	1.33	1.36	1.28	1.19	1.42	1.46	1.34	1.23	1.35	1.40	1.30	1.19	1.50	1.55	1.38	1.26	1.29	1.36
30	1.30	1.22	1.11	1.16	1.36	1.27	1.16	1.22	1.31	1.22	1.11	1.16	1.40	1.31	1.19	1.27	1.20	1.25
31	1.38	1.41	1.57	1.65	1.45	1.46	1.63	1.69	1.39	1.42	1.58	1.66	1.51	1.50	1.67	1.72	1.50	1.56
32	1.48	1.26	1.26	1.26	1.57	1.35	1.34	1.33	1.49	1.28	1.28	1.28	1.64	1.43	1.40	1.39	1.32	1.40
33	1.61	1.56	1.40	1.31	1.67	1.63	1.45	1.36	1.62	1.59	1.41	1.33	1.71	1.69	1.49	1.39	1.47	1.53
34	1.35	1.33	1.34	1.31	1.43	1.40	1.43	1.37	1.36	1.34	1.36	1.32	1.49	1.45	1.50	1.42	1.33	1.41
35	1.59	1.56	1.50	1.49	1.66	1.63	1.57	1.54	1.61	1.58	1.52	1.49	1.72	1.69	1.62	1.58	1.54	1.60
36	1.37	1.26	1.24	1.22	1.45	1.33	1.31	1.28	1.39	1.28	1.26	1.23	1.51	1.39	1.38	1.33	1.27	1.34
37	1.37	1.26	1.22	1.24	1.44	1.31	1.28	1.30	1.39	1.27	1.24	1.26	1.49	1.35	1.33	1.35	1.27	1.33
38	1.44	1.34	1.30	1.29	1.50	1.39	1.36	1.34	1.45	1.35	1.32	1.29	1.54	1.42	1.39	1.38	1.34	1.39
39	1.28	1.30	1.19	1.26	1.34	1.36	1.25	1.31	1.29	1.32	1.20	1.26	1.39	1.42	1.31	1.35	1.26	1.32
40	1.29	1.19	1.06	1.10	1.35	1.24	1.13	1.17	1.30	1.19	1.07	1.10	1.40	1.28	1.20	1.23	1.16	1.22
41	1.52	1.40	1.30	1.33	1.60	1.47	1.36	1.38	1.53	1.41	1.31	1.34	1.65	1.54	1.41	1.41	1.39	1.45
42	1.55	1.49	1.40	1.32	1.64	1.59	1.49	1.39	1.56	1.50	1.42	1.34	1.72	1.68	1.57	1.45	1.44	1.53
43	1.31	1.28	1.20	1.28	1.35	1.34	1.26	1.33	1.32	1.29	1.21	1.28	1.38	1.39	1.30	1.37	1.27	1.32
44	1.43	1.43	1.40	1.36	1.51	1.48	1.46	1.43	1.44	1.44	1.40	1.37	1.58	1.52	1.51	1.49	1.40	1.47
45	1.44	1.45	1.68	1.66	1.49	1.50	1.75	1.71	1.45	1.46	1.69	1.67	1.54	1.53	1.80	1.76	1.56	1.61
46	1.34	1.30	1.25	1.33	1.41	1.36	1.32	1.38	1.36	1.32	1.27	1.34	1.46	1.40	1.37	1.43	1.30	1.37
47	1.35	1.23	1.32	1.32	1.41	1.27	1.36	1.37	1.36	1.24	1.33	1.33	1.45	1.30	1.39	1.40	1.31	1.35
48	1.27	1.17	1.16	1.22	1.38	1.26	1.23	1.26	1.30	1.19	1.18	1.22	1.46	1.34	1.30	1.29	1.21	1.28
50	1.00	1.22	1.29	1.25	1.11	1.25	1.32	1.29	1.00	1.22	1.30	1.26	1.22	1.26	1.35	1.31	1.19	1.24
51	1.79	1.62	1.64	1.64	1.89	1.71	1.74	1.71	1.81	1.65	1.66	1.65	1.98	1.78	1.82	1.77	1.67	1.76
52	1.15	1.24	1.24	1.20	1.21	1.31	1.29	1.25	1.14	1.24	1.24	1.21	1.27	1.37	1.33	1.29	1.21	1.26
53	1.43	1.31	1.20	1.13	1.49	1.34	1.24	1.17	1.44	1.32	1.21	1.13	1.55	1.38	1.27	1.21	1.26	1.31
54	1.72	1.46	1.55	1.42	1.83	1.54	1.60	1.47	1.75	1.48	1.57	1.42	1.92	1.61	1.63	1.51	1.54	1.61
56	1.52	1.42	1.44	1.44	1.62	1.52	1.53	1.51	1.54	1.44	1.46	1.46	1.70	1.61	1.62	1.58	1.45	1.55
57	1.45	1.42	1.30	1.25	1.51	1.48	1.35	1.31	1.45	1.44	1.32	1.26	1.56	1.54	1.39	1.35	1.36	1.42

58	1.45	1.32	1.22	1.23	1.57	1.41	1.29	1.29	1.47	1.34	1.23	1.24	1.67	1.48	1.34	1.34	1.30	1.39
59	1.28	1.31	1.36	1.35	1.31	1.35	1.42	1.39	1.28	1.32	1.38	1.36	1.33	1.39	1.46	1.42	1.33	1.37
60	1.56	1.43	1.58	1.37	1.63	1.49	1.64	1.40	1.56	1.44	1.59	1.37	1.69	1.53	1.68	1.42	1.48	1.54
61	1.41	1.27	1.37	1.28	1.49	1.34	1.45	1.34	1.43	1.28	1.39	1.29	1.55	1.40	1.52	1.39	1.33	1.40
62	1.31	1.18	1.14	1.14	1.39	1.24	1.21	1.20	1.33	1.19	1.15	1.15	1.46	1.28	1.28	1.25	1.19	1.26
63	1.34	1.27	1.16	1.25	1.43	1.34	1.23	1.32	1.36	1.29	1.17	1.27	1.49	1.40	1.29	1.38	1.26	1.33
64	1.42	1.38	1.40	1.39	1.48	1.44	1.46	1.46	1.43	1.39	1.41	1.40	1.53	1.49	1.51	1.52	1.40	1.46
65	1.27	1.25	1.09	1.22	1.34	1.32	1.14	1.27	1.27	1.25	1.10	1.23	1.40	1.38	1.18	1.32	1.21	1.27
66	1.65	1.42	1.43	1.34	1.74	1.50	1.48	1.41	1.68	1.45	1.44	1.36	1.82	1.57	1.53	1.46	1.46	1.54
67	1.33	1.41	1.32	1.44	1.44	1.48	1.38	1.47	1.36	1.43	1.34	1.44	1.51	1.54	1.43	1.49	1.38	1.44
68	1.58	1.43	1.28	1.20	1.67	1.51	1.36	1.27	1.59	1.45	1.30	1.21	1.74	1.56	1.42	1.33	1.37	1.45
69	1.18	1.12	1.15	1.20	1.25	1.19	1.21	1.26	1.19	1.14	1.16	1.21	1.31	1.25	1.26	1.32	1.16	1.23
70	1.53	1.37	1.33	1.44	1.61	1.42	1.38	1.51	1.55	1.38	1.34	1.45	1.68	1.46	1.42	1.57	1.42	1.48
71	1.30	1.27	1.28	1.21	1.39	1.36	1.37	1.28	1.31	1.28	1.29	1.22	1.46	1.44	1.44	1.35	1.27	1.35
72	1.37	1.20	1.23	1.30	1.43	1.24	1.27	1.35	1.39	1.21	1.25	1.31	1.48	1.26	1.31	1.40	1.27	1.32
73	1.30	1.23	1.20	1.16	1.37	1.29	1.26	1.24	1.32	1.24	1.21	1.19	1.43	1.34	1.30	1.31	1.22	1.29
74	1.37	1.28	1.29	1.30	1.43	1.34	1.34	1.34	1.39	1.29	1.30	1.30	1.48	1.38	1.39	1.37	1.31	1.36
75	1.23	1.18	1.13	1.15	1.30	1.26	1.21	1.23	1.25	1.20	1.15	1.17	1.35	1.33	1.27	1.30	1.17	1.25
76	1.43	1.26	1.22	1.24	1.49	1.33	1.26	1.28	1.44	1.29	1.23	1.25	1.54	1.38	1.30	1.31	1.29	1.34
78	1.43	1.35	1.33	1.36	1.51	1.42	1.39	1.42	1.46	1.37	1.35	1.38	1.56	1.47	1.43	1.47	1.37	1.43
79	1.11	1.06	1.05	1.12	1.19	1.12	1.10	1.17	1.12	1.08	1.06	1.13	1.26	1.17	1.14	1.21	1.08	1.14
80	1.34	1.25	1.16	1.18	1.40	1.31	1.20	1.21	1.35	1.26	1.17	1.18	1.46	1.35	1.24	1.23	1.23	1.28
81	1.35	1.25	1.34	1.21	1.43	1.34	1.43	1.28	1.38	1.28	1.36	1.22	1.50	1.40	1.50	1.35	1.29	1.37

Notes: Efficiency scores and its' confidence intervals from the institutional level were aggregated to the regional level. These aggregated values are weighted averages of universities efficiency scores and its' confidence intervals by the total number of students of local universities in certain regions (for more detailed information see the main body of research).

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Annex 3. Robustness check with Stochastic frontier analysis

Stochastic frontier analysis (SFA) was introduced by Aigner et al. (1977), Meeusen and Van den Broeck (1977), Battese and Corra (1977). This technique is parametric alternative to non-parametric data envelopment analysis. We employ this technique as a robustness check of the results obtained through data envelopment analysis. Using both DEA and SFA techniques we can get reliable information about efficiency ranking (McMillan and Chan, 2006)

In order to implement SFA technique in multi-output case we use input distance function of the following form (Shephard distance function (Shephard, 1953)):

$$D_k(X,Y) = max\{D > 0|(\frac{X}{D}, Y) \in T\} = \frac{1}{E(X,Y)},\ k = 1, \dots, N \quad (11)$$

where $E(X,Y) = min\{E > 0|\ (EX, Y) \in T\}$ −Farrell input [Farrell, 1957]; T −technology set; X −input vector; Y −output vector.

We used the variable of income of educational organization from all sources per number of Faculty members as normalizing input – dependent variable in stochastic frontier regression. All inputs and outputs are the same as in case of data envelopment analysis (see Table 1). In order to estimate stochastic frontier model Cobb-Douglas type production function in logarithmic form was assumed. The results of the SFA regression estimation are presented in the Table 8.

Table 8. Results of the efficiency estimation on institutional level (standard errors are presented in the brackets)

Dependent variable – income	2012 (N=449)	2013 (N=449)	2014 (N=449)	2015 (N=449)
Academic staff	-6.452*** (0.836)	-9.034*** (0.763)	-11.976*** (0.761)	-8.102*** (0.094)
Exam scores	-27.72*** (1.028)	-25.95*** (0.847)	-28.78*** (0.899)	-25.66*** (1.062)
Publications	0.039*** (0.011)	0.048*** (0.011)	0.041*** (0.010)	-0.016** (0.006)
R&D	0.022** (0.008)	0.019** (0.006)	0.010 (0.006)	0.007 (0.006)
Students	-0.045 (0.038)	-0.127*** (0.030)	-0.122*** (0.319)	-0.136*** (0.034)
Masters' students	0.035*** (0.010)	0.034*** (0.009)	0.044*** (0.008)	0.052*** (0.010)
Full-time students	0.365*** (0.092)	0.371*** (0.074)	0.389*** (0.066)	0.418*** (0.068)
Capital city	0.049 (0.034)	0.043. (0.026)	0.009 (0.028)	0.022 (0.031)
Market share	0.222* (0.097)	0.073 (0.075)	0.064 (0.069)	0.076 (0.070)
Medical faculty	0.309*** (0.043)	0.386*** (0.036)	0.348*** (0.041)	0.268*** (0.047)
Lambda	4.414	4.952	3.965	3.412
Log likelihood	224.02	315.73	339.35	326.24
Significance levels: *** p-value<0.001; ** p-value<0.01; * p-value<0.05; . p-value<0.1				

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

Kernel density plot for the efficiency scores obtained using stochastic frontier analysis is presented on the Figure 5. Distributions of SFA efficiency scores have negative asymmetry and higher kurtosis than for DEA efficiency scores (see Figure 3).

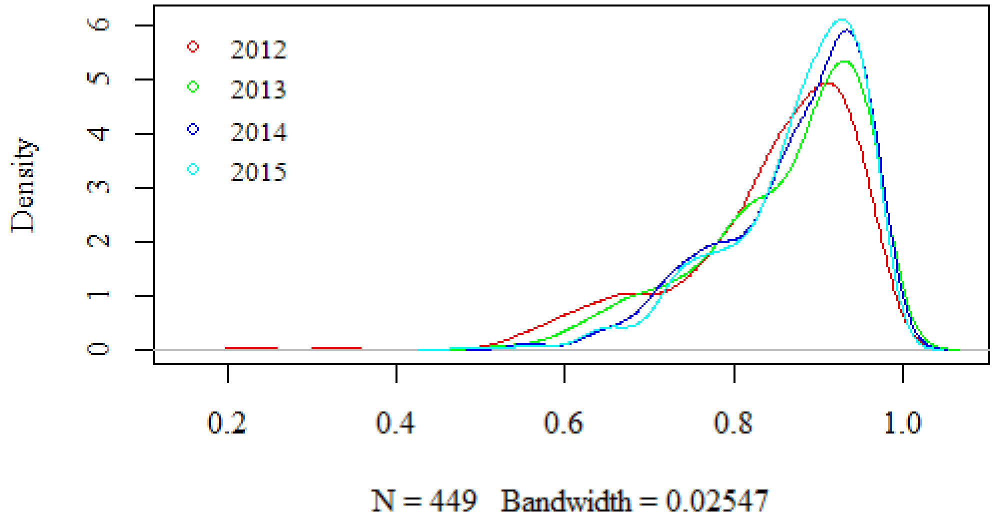

Figure 5. SFA efficiency distribution on institutional level.

Notes: Efficiency scores are distributed between 0 and 1. A value of 1 indicates that a university is efficient and lies on the best practice frontier.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions».

The mean SFA efficiency scores have increased by in 4%: from 0.84 in 2012 to 0.87 in 2015. This dynamic of growth correspond to tendency with DEA efficiency estimates. The values of SFA efficiency scores at the institutional level are on average 15 per cent lower than DEA efficiency scores. The SFA efficiency estimates discriminate universities in the sample worse than DEA scores: the standard deviation of SFA scores are on average 40% lower than standard deviation for DEA estimates. The descriptive statistics of SFA efficiency scores at the institutional level are presented in the Table 9.

SFA efficiency scores were also aggregated to the regional level using weighted average formula with total number of students in the university as a weight. The above mentioned conclusions for institutional efficiency scores are relevant too for regional efficiency values, because the last ones are based on the first. The descriptive statistics for SFA efficiency scores at the regional level are shown in Table 9.

Table 9. Descriptive statistics of SFA efficiency scores over 2012–2015.

Statistics	**2012** (N=449)	**2013** (N=449)	**2014** (N=449)	**2015** (N=449)
Institutional level				
Mean	0.839	0.859	0.869	0.871
Median	0.864	0.889	0.893	0.893
Std deviation	0.109	0.097	0.086	0.084
Minimum	0.226	0.513	0.552	0.490
Maximum	0.982	0.989	0.983	0.984
Statistics	**2012** (N=79)	**2013** (N=79)	**2014** (N=79)	**2015** (N=79)
Regional level				
Mean	0.841	0.862	0.875	0.877
Median	0.855	0.883	0.885	0.884
Std deviation	0.071	0.067	0.056	0.056
Minimum	0.591	0.641	0.669	0.649
Maximum	0.932	0.965	0.961	0.952

Notes: Regional efficiency scores conform to estimates of regional HE systems. Regional HE system is a set of universities located within the administrative borders of the region.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

The distribution of Regional Higher Education systems based on SFA efficiency scores is represented by Figure 6. This distribution corresponds to the analogous regional distribution of DEA efficiency scores (see Figure 4). The most efficient Regional HE systems are concentrated in the central part of Russia and less at the western regions. While the most efficient Regional HE systems according to DEA scores are located more at the western part of Russia and less at the central regions.

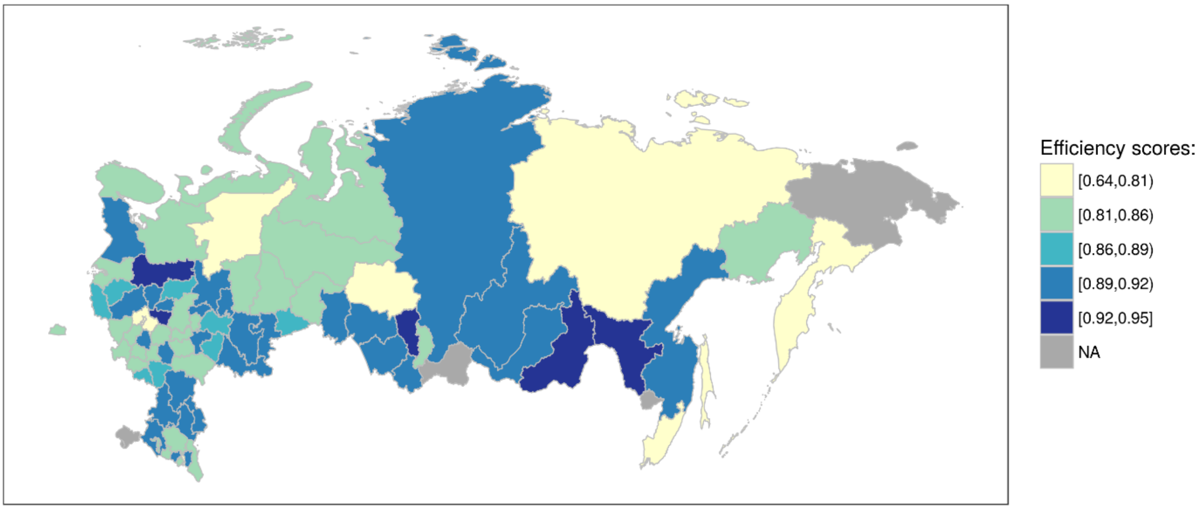

Figure 6. SFA efficiency scores of Regional Higher Education (HE) systems.

Notes: Plot presents the averaged efficiency scores of Regional Higher Education (HE) systems over 2012–2015. Regional HE system is a set of universities located within the administrative borders of the region. A value NA indicates the lack of the data for this region. The sample includes 77 regions.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» data.

The Pearson correlation coefficient between DEA and SFA regional efficiency scores amounts on average 0.15 that corresponds to a low strength of relationship. This correlation is statistically significant ($p<0.05$). The result about low strength of relationship between scores estimated by DEA and SFA was achieved in more early literature (for example, Chirikos and Sear (2000), Ferrier and Lovell (1990)). There are numerous causes for the variance of DEA and SFA results: different functional form of production function and true/false assumption on it, presence of statistical noise, different distributions of inefficiency term and true/false assumption on it, correlation of inefficiency scores with explanatory variables, omission of relevant/inclusion of irrelevant variables etc. The analysis of listed causes goes beyond the scope of the present paper. But since the DEA efficiency scores discriminate universities in the sample better than SFA estimates, the economic growth model estimation involves mainly DEA efficiency values and use SFA scores to check the robustness of the results. The estimation results of economic growth with SFA efficiency scores are presented in Table 10.

Table 10. Results of regional growth model estimation with SFA efficiency scores (standard errors are presented in the brackets)

Variable name	Estimated parameter
Growth rate in previous period	0.902***
	(0.018)
Investment growth rate	4.192***
	(0.408)
Employed population growth rate	0.243***
	(0.032)
Gross regional product in previous period	-3.039***
	(0.543)
Share of employed with HE	0.106*
	(0.052)
Total number of universities' graduates	0.021*
	(0.009)
Share of commercial minerals extraction in GRP	-0.070*
	(0.023)
Share of industries in GRP	0.028
	(0.025)
Share of public sector in GRP	-0.535***
	(0.069)
Efficiency SFA	1.130**
	(0.423)
SFA efficiency standard deviation	1.488***
	(0.240)
Efficiency spatial interaction	-0.784***
	(0.136)
Growth spatial interaction	2.044***
	(0.342)
Constant	36.525***
	(9.019)
Hansen-Sargan	0.391
AR(2)	0.217
# of observations	308
Significance levels: *** p-value<0.001; ** p-value<0.01; * p-value<0.05; . p-value<0.1	

Notes: Equation is estimated through a two-step system generalized method moment estimator with Windmeijer (2005) corrected standard error (in brackets).

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» and Russian Federation Federal State Statistics.

All relationships that were revealed based on the model with DEA scores (see Table 5) are preserved in the model with SFA efficiency scores as one of the explanatory variables. This robustness check confirms the conclusions about positive influence of regional higher education system efficiency on economic growth rates, negative spatial effects of efficiency and positive relationship between standard deviation of universities' efficiency within the region and economic growth rates.

The scatter plot of SFA efficiency and GRP growth rates averaged for the period under review is presented on the Figure 7. As in the case of DEA (see Figure 2), scatter plot shows positive relationship between regional rates of economic growth and regional HE systems efficiency scores. However, the two-dimensional regression line has a smaller slope than in case of DEA.

In overall, we can conclude that results of the paper are robust to the change of efficiency measure.

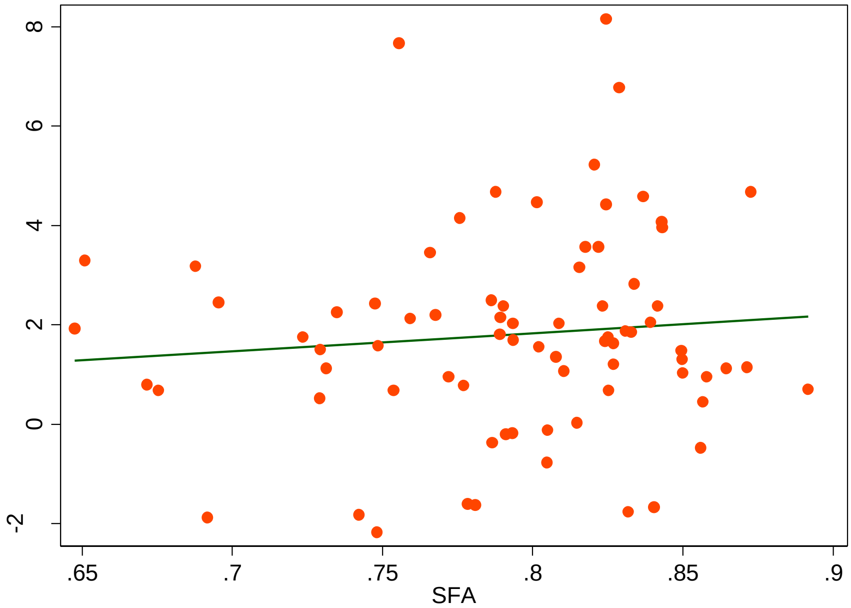

Figure 7. Scatter plot of SFA efficiency scores and GRP growth rates (average values for the period 2012–2015).

Notes: Plot presents the averaged values of efficiency scores of regional HE systems and GRP growth rates over 2012–2015. Regional HE system is a set of universities located within the administrative borders of the region.

Source: Authors' calculations from «Annual Monitoring of Efficiency of Higher Education Institutions» and Federal State Statistics Service.

Annex 4. Handling missing data and outliers

MICE with CART algorithm for missing data imputation

The basic idea of the multiple imputations by chained equations (MICE) (Raghunathan et al, 2002) is as follows. Suppose we have a sample of variables $X_1, \dots, X_n$ that contains missing values. These variables are considered separately starting from X_1. In order to predict missing values in this variable we built a regression for X_1 using $X_2, \dots, X_n$ as predictors:

$$\hat{X}_1 = \beta_1 + \sum_{i=1}^{n} \beta_i X_i \, , i = 2, \dots, n \tag{12}$$

The value of $\hat{X}_1$ predicted by regression (12) is used as a proxy for missing value. This procedure should be repeated for each variable in the sample.

However, sometimes relationships between variables in the sample are interactive and non-linear, thus simple linear regression may give predictions with rather high errors. Moreover, sometimes variables may have difficult distributions that are not captured by linear models (Burgette and Reiter, 2010). That is why it is reasonable to implement alternative models for prediction of missing values, for instance, Classification and Regression Trees algorithm (CART).

In order to implement MICE based on CART algorithm we have to change equation (12) by the regression tree model that is formulated as follows. Suppose we have outcome variable Y and matrix X containing explanatory variables. Regression tree suggests that the space of predictors should be divided into regions A_k in such way that minimize the sum of squared errors and the fitted value in each A_k. More detailed description of CART algorithm can be found in (Loh, 2011).

Capping procedure for outliers

The capping procedure suggests the following algorithm. Suppose we have variable X in our dataset. At the first step we identify outliers X_o according to the formula (13).

$$(X_0 = \{X: X > Q_3 + 1.5 \cdot IQR \; or \; X < Q_1 - 1.5 \cdot IQR\}, \tag{13}$$

where $IQR = Q_3 - Q_1$; $Q_i = X: F(x_i) \leq Q \; and \; F(x_i + 0) \geq Q$; $F(X) = P\{\epsilon < X\} -$ comulative distribution function.

When outliers are identified their values are substituted using the following rule (14):

$$X_{upper\ outlier} = Q_{0.975};\ X_{lower\ outlier} = Q_{0.025} \tag{14}$$

So this procedure replaces the outliers by the tail distribution quantiles, for example by quantiles of the levels 0.975 and 0.025.

Aleksei Egorov
National Research University Higher School of Economics (Moscow, Russia), Laboratory for University Development of Institute of Education. Analyst;
E-mail: aegorov@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.