INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 27 (2020) 1911–1929
DOI: 10.1111/itor.12742

The Russian Excellence Initiative for higher education: a nonparametric evaluation of short-term results

Tommaso Agasisti[a], Ekaterina Shibanova[b,*], Daria Platonova[b] and Mikhail Lisyutkin[b]

[a] *Department of Management, Economics & Industrial Engineering, Politecnico di Milano School of Management Milan, 20156, Milan, Italy*

[b] *National Research University Higher School of Economics, Moscow, 101000, Russia*

E-mail: tommaso.agasisti@polimi.it [Agasisti]; eshibanova@hse.ru [Shibanova]; dplatonova@hse.ru [Platonova]; mlisyutkin@hse.ru [Lisyutkin]

Received 24 December 2018; received in revised form 4 October 2019; accepted 8 October 2019

Abstract

This research studies the short-term effects of the Russian Excellence Initiative Project 5–100 on participating universities. To trace the effect, we develop a quasi-experimental methodology. A control group of universities comparable to the Project 5–100 universities at the starting point of the program's implementation was singled out using propensity score matching. Data envelopment analysis was conducted, and the Malmquist productivity index was calculated to trace how and why the efficiency of the "participants" and "nonparticipants" of the Project 5–100 has changed. We find statistically significant positive effects of the policy both on the productivity and on the efficiency of the participating universities.

Keywords: efficiency in higher education; excellence initiative; management of universities; data envelopment analysis; Malmquist index

1. Introduction

Global competition in higher education has had a major impact on the priorities of national governments in recent years (Chirikov, 2016). The global ranking of universities has become a powerful tool, influencing the perception of success and excellence in higher education at national and institutional levels (Hazelkorn, 2014, 2015; Altbach and Hazelkorn, 2017). Many countries have launched programs to develop a group of the so-called "world-class" universities (Altbach and Salmi, 2011). Such policies are known as Excellence Initiatives (ExIn), and are aimed at pushing particular higher education institutions (HEIs) to compete successfully in international education and research markets. Since 2000, more than 40 excellence-driven initiatives have been launched in more than 20 countries. More than US$60 billion has been invested in these initiatives (Salmi,

*Corresponding author.

2016). Inspired mostly by the success of the Chinese ExIn, Russia initiated Project 5–100 in 2012. The basic aim of the project is for the least five Russian universities to be ranked among the world's top 100 according to key world university rankings by 2020. Fifteen universities were selected on a competitive basis in 2013 and have received additional funding each year.

ExIns in higher education are mostly aimed at boosting the production of a single key output: the rate at which universities produce internationally recognized research (Salmi, 2009; Froumin and Lisyutkin, 2015). A number of academic papers have evaluated changes in publication productivity for Chinese universities (Zhang et al., 2013; Yang and You, 2017; Ou, 2017; Zong and Zhang, 2019), German universities (Möller et al., 2016; Menter et al., 2018), Russian universities (Turko et al., 2016; Poldin et al., 2017), and Korean universities (Seong et al., 2008; Shin, 2009). The design of recent studies is quasi-experimental. Zong and Zhang (2019), Ou (2017), and Poldin et al. (2017) compare the results of participants and nonparticipants of ExIn. Zong and Zhang (2019) and Menter et al. (2018) use difference-in-differences models. Ou (2017) develops a propensity score matching (PSM) model, specifically nearest neighbor matching, to evaluate the effects of China's Project-211.

ExIns not only fund programs to achieve higher output in terms of research productivity; they also aim to bring organizational transformation to institutions and institutional environments (see Chirikov, 2018). These include changes in the internal activities and the efficiency of universities. To the best of our knowledge, only two studies assess the effects of ExIn by evaluating university efficiency. Gawellek and Sunder (2016) examine effects of the German ExIn on the efficiency of universities using data envelopment analysis (DEA) as well as the Malmquist index and its decomposition. Yaisawarng and Ng (2014) also use DEA and compute the annual efficiency scores to test if Chinese Project-211 universities perform better than non-211 universities. They also compute the Malmquist index to examine whether productivity changes and technological advancement took place over a three-year period.

The main objective of this study is to estimate the changes in the activities of Project 5–100 universities. The key research question is:

Have there been any changes in the efficiency of participating universities since the implementation of the project?

A striking inadequacy among studies evaluating the effects of ExIns on university activities is the lack of evaluation of performance and efficiency score changes between participant and nonparticipant groups, which would help solve the attribution challenge using PSM. This study addresses this gap by deploying an innovative quasi-experimental design to study the causal effects of ExIns. We employ an empirical analysis based on three steps.

Step 1: Single out HEIs that are comparable to excellence-driven universities at the start of the program, which are not part of the ExIn, using PSM to create a control group; check whether after the launch of Project 5–100 there is a significant difference between the control and treatment groups in their key performance indicator—publication activity.

Step 2: Estimate the efficiency of universities participating in the ExIn and their control group over a six-year period (2012–2017) using DEA.

Step 3: Estimate efficiency changes using the Malmquist productivity index, and its decomposition into (a) efficiency change due to internal change (getting closer to or further from the frontier) and (b) technical change due to the overall shift of the efficiency frontier.

The paper is organized as follows. We discuss ExIn as a global phenomenon and provide information about Project 5–100 in Section 2. In Section 3 we describe the methodology of the three-step analysis: PSM, DEA, and the Malmquist productivity index. The key results of the research are presented in Section 4. In Section 5 we develop a discussion of the empirical findings, as well as the limitations of the study.

2. Russian Excellence Initiative: the design of Project 5–100

In 2012, Russia joined the race for global competitiveness in higher education by launching Project 5–100. The basic idea of the project was that at least five Russian universities would enter the world's top 100 universities (according to the international rankings Academic Ranking of World Universities, The Times Higher Education World University Rankings, and Quacquarelli Symonds Rankings) by 2020. The initiative offers both financial and managerial support from the government. In 2013, the first year of the project's deployment, 15 universities were selected on a competitive basis and have been receiving additional funding since then. The group of 5–100 universities was enlarged to 21 at the second stage of the project in 2015. Universities aspiring to become part of Project 5–100 prepared detailed roadmaps for their development by 2020 based on the criteria established by the government. The roadmaps anticipated the achievement of specific performance indicators. An International Council was established by the government to make roadmap assessment more objective and to evaluate the roadmaps according to international standards. In general, the approaches and design of the project is very similar to the Chinese Project-985 (Chirikov, 2018).

During the five years in which the project has been in effect, the 5–100 universities have received more than 50 billion rubles (about US$850 million) from the federal budget. In relative numbers, the annual project subsidy is only 2% of the federal budget for higher education. The subsidy given to universities differs according to their achievements. In 2017, the subsidy for high achievers was twice that for low achievers (among the 15 universities in the first wave). The 5–100 universities vary dramatically in scale. In terms of budget, the largest is eight times the size of the smallest. They differ in size and structure, for example, in the mix of disciplines and departments. However, almost all the 5–100 universities operate in STEM (science, technology, engineering, and mathematics) fields.

There have been a few studies on the effects of Project 5–100. An analysis by Turko et al. (2016) identified that Project 5–100 had a positive impact on publication productivity and enhanced the growth of global competitiveness as expressed in the rankings. Applying mixed method models, Poldin et al. (2017) confirmed the relationship between participation in ExIn and the number of publications (both in general and in high-impact journals), and was positive and significant for two years after the program began. However, as the authors mention, for more sustainable results a longer period should be taken into consideration. Besides, one should expect that a policy aimed at research productivity increase might have a spillover effect on teaching activity as well (De Witte et al., 2013).

3. Methodology and data

This study develops a three-step analysis to assess the effects of ExIn and efficiency gains at participating universities. We use performance monitoring data for Russian HEIs, which was gathered by the Ministry of Education and Science of the Russian Federation (for a detailed description of the monitoring, see Sokolov and Tsivinskaya, 2018a, 2018b). The data cover a six-year period, for the 2012/2013–2017/2018 academic years.

Initially, there were 908 public and private HEIs in total. Taking into account the official requirements for potential participants in the ExIn and data availability, we applied the following limitations to our data sample:

- there must be publicly funded students within the university's educational programs;
- the minimal overall number of students enrolled in the educational programs is 4000;
- the minimal unified state exam (USE) (standardized national entry exam) grade must be equal to or greater than 64.[1]

We also excluded all universities that had been reorganized during the period (2012–2017), to avoid any bias caused by the structural differences of the units. We also excluded universities participating in the ExIn at its second stage. After applying the limitations, we had a sample of 152 universities, including the 15 original participants of the ExIn program. The total sample used for PSM is 152 universities. After performing the PSM we have a total of 30 universities, 15 participants, and 15 in the control group. Only these 30 universities are analyzed at further stages (DEA and Malmquist index).

We note the importance of the stable unit treatment value assumption (SUTVA) (Rubin, 1974), which imposes the absence of direct interaction among the units of our analysis. University activities are open to the market, but we consider the treatment effect to be pure because the scope of the treatment was strictly limited to the participating universities. Even if we assume that the policy was more widespread and occasionally affected the control group, we follow Imbens and Wooldridge (2009) and consider the indirect effects much smaller than the direct treatment; we assume that the SUTVA condition is not violated in our research.

3.1. Step 1. Propensity score matching: control-group selection

We use PSM for verifiable control-group selection, which is a common approach for government intervention estimations (Rosenbaum and Rubin, 1985). This is defined as the probability of a treatment that is conditional on a set of observable variables:

$$\text{Propensity score}_i = E(X_i) = \text{probability}(D_i = 1 | X_i = x_i), \quad (1)$$

where D_i is the status of the treatment (0 if not present and 1 if present) and X_i is a set of observed covariates. Matching entities from the treatment group with entities from the control group is needed to construct the control group. As our study is an observable one, the true propensity score is unknown, but we can calculate it through a logistic model, 0 or 1 being the outcome of treatment assignment. The propensity score, which describes the probability of being in the treatment group for entities, must be calculated according to their characteristics, which were used to assign the

[1] We included one of the participating universities in the sample even though it did not fit the minimal USE criteria.

Table 1
Descriptive statistics for PSM variables before matching

	Variables	Matching sample			Treatment group		
		N	Mean	SD	*N*	Mean	SD
Matching variables	Unified state exam	137	71.30	6.991	15	72.73	7.686
	Share of PhD students	137	4.943	2.138	15	5.665	1.778
	Number of publications (Web of Science/Scopus) per 100 staff capita	137	5.875	7.600	15	30.14	20.53
	Research and development income per staff capita (thousand rubles)	137	224.6	280.5	15	779.7	458.6
	Share of foreign students	137	2.251	4.388	15	1.513	1.323
	Total number of students	137	9597	4850	15	16,653	10,062
	Share of foreign research staff	137	0.902	1.984	15	1.447	2.949
Outcome variables	Publications (maximum of Web of Science/ Scopus), 2017/2018	137	41.71	58.83	15	393.26	193.98

Source: Authors' calculations.

treatment. A correctly conducted matching leads to a balanced distribution of indicators between the groups. PSM consists of the following steps:

- the selection of variables potentially influencing the probability of entering the ExIn;
- calculations of propensity scores for each university within the sample;
- using additional variables to balance control between the matched pairs;
- an evaluation of the average effect on treated units based on the matched samples.

We chose a 1:1 matching due to our data limitations, as we wanted to trace potential effects of the ExIn on real universities and their existent counterparts. More specifically, we conducted a nearest neighbor matching, and checked the biases by a test of means in the variables used for matching, as well as additional covariates that could affect the treatment assignment but were not included in the official call for participation in the ExIn (see Tables 1 and 2). We also provided a robustness check and compared the average treatment effect on the treated (ATT) results between the results of the nearest neighbor and radius (caliper) matching procedures (Table 3).

Additional assumptions must also be taken into account. The conditional independence assumption (CIA) implies that selection is solely based on observable characteristics and that all variables influencing the treatment assignment and potential outcomes are simultaneously observed by the researcher. To satisfy the CIA, the values of the characteristics before treatment should be used so that we can be sure all the universities had an equal chance of being chosen. Common support ensures that entities with the same characteristic value have a positive probability of being both participants and nonparticipants. The common support problem means that there can be entities

Table 2
PSM variables and additional covariates after matching and testing for differences in means

		Control group			Treatment group		
	Variables	*N*	Mean	SD	*N*	Mean	SD
Matching variables	Unified state exam	15	69.58***	5.280	15	72.73***	7.686
	Share of PhD students	15	5.243***	1.534	15	5.665***	1.778
	Number of publications (Web of Science/Scopus) per 100 staff capita	15	14.07*	15.01	15	30.14*	20.53
	Research and development income per staff capita (thousand rubles)	15	568.3***	482.8	15	779.7***	458.6
	Share of foreign students	15	1.292***	1.158	15	1.513***	1.323
	Share of foreign research staff	15	0.423***	0.597	15	1.447***	2.949
Additional covariates	Total income per staff capita (thousand rubles)	15	2331*	820.1	15	3190*	1276
	Share of new resource base	15	51.95***	18.76	15	50.71***	26.76
	Total number of students	15	16,250***	6422	15	16,653***	10,062
	Total number of teaching staff	15	1073***	541.6	15	1318***	729.1
	Share of STEM programs	15	58.97***	24.10	15	58.97***	25.89

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Source: Authors' calculations.

Table 3
PSM and ATT results

Outcome variable	Matching method	Sample	Treated	Controls	Difference	SE	*T*-stat
Publications (maximum of Web of Science/Scopus), 2017/2018		Unmatched	393.257	41.714	351.453	22.178	15.85
	Nearest neighbor	ATT	393.257	130.730	262.529	57.267	4.58
	Radius caliper matching	ATT	393.257	207.222	186.035	54.060	3.44

without a matching pair. If the number of such entities is low, they can be dropped. We do not anticipate finding evidence of problems in the validity of the CIA.

Our methodological parameters are reflected in the choice of matching variables (Table 1). The choice of variables is based on the policy design and is prescribed by the application requirements. According to the ExIn design, the variables represent two key areas of university activity—internationalization and research productivity. Structural characteristics are used as additional covariates to check the biases.

We conduct the PSM using data collected before the actual implementation of Project 5–100, that is, we use the data for the 2012/2013 academic year. Following López-Torres et al. (2016), PSM can be applied to calculate the ATT. For our topic of interest, the selection of universities as participants of Program 5–100 is the "treatment," and selected universities are "treated":

$$\tau_{ATT} = E(D = 1) = E[D = 1] - E[Y(0)|D = 1], \tag{2}$$

where τ is a treatment indicator, and $[D = 1]$ and $Y[(0)|D = 1]$ are means in potential outcomes for entities of the treated group if they receive treatment and do not receive treatment, respectively. The latter is not observed and the methodology for the selection of a substitute is needed. The outcome in the PSM model also reflects the framework of the Russian ExIn and is expressed by the number of publications indexed in Web of Science/Scopus.

3.2. Step 2. Measuring university efficiency

Measuring university performance using DEA is a common practice. Worthington (2001), Johnes (2006), and De Witte and López-Torres (2017) provide an overview of using frontier efficiency applications in education by means of useful surveys of the literature.

DEA, a nonparametric linear programming method, provides a measurement of the efficiency and productivity scores of a decision making unit (DMU)—in our case, a university. DEA is based on a programmed envelopment of observed multiple input–output vectors (Boussofiane et al., 1991) without additional issues of data distribution. DEA is suitable for estimating the efficiency of a multi-input and multi-output production function in the absence of all the market prices of the components (Ray, 2004). The efficiency of each DMU is measured through the changing proportion of inputs or outputs. A DEA model can be input- or output-oriented, depending on whether a minimization or a maximization problem is being solved. It can also be modified depending on a constant or a variable return to scale. An output-oriented DEA model is produced to test whether a DMU is capable of increasing its outputs with the same inputs. In our research, an output-oriented DEA will be used for the calculation of efficiency scores.

The universities represent a technology that can be modeled through a combination of inputs (x) and outputs (y), or a production possibility set (PPS):

$$PPS = \{(x,) : (x; y) \in R_n^+; \quad y \in R_m^+. \tag{3}$$

The combination of inputs and outputs is feasible only for the efficiency frontier of the production possibility set. For the output-oriented model, the technical efficiency is as follows:

$$\max \varphi k + \varepsilon \sum_{r=1}^{s} s_r + \varepsilon \sum_{i=1}^{m} s_i \tag{4}$$

under the condition

$$\varphi_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0, \qquad r = 1, \ldots, s, \tag{5}$$

$$x_{ik} - \sum_{r=1}^{n} \lambda_j x_{ij} - s_i = 0, \quad i = 1, \ldots, m, \tag{6}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \tag{7}$$

$$\lambda_j, s_r, s_i \geq 0 \quad \forall j = 1, \ldots, n; \; r = 1, \ldots, s; i = 1, \ldots, m, \tag{8}$$

where s are outputs and m are inputs; y_{rk} is the volume of output of type r, belonging to university k; x_{ik} is the volume of input of type i, belonging to university k; s_i and s_r are the slack in outputs and inputs, respectively. The efficiency rate of university k is defined as $\varphi_k^* = 1/\varphi$; university k is k efficient, if the efficiency rate $\varphi_k^* = 1$ and there is no slack in the volumes of inputs and outputs. If $\varphi_k^* = 1$, then the university under evaluation is a frontier point, that is, there are no other universities operating more efficiently than this particular one. This analysis provides efficiency scores in terms of input used to produce outputs as efficiently as possible. A variable return-to-scale assumption is implemented for both DEA efficiency estimations and the Malmquist productivity index, as previous research indicates that Russian universities operate under variable return to scale (Abankina et al., 2013; Gromov, 2017). We conducted an additional analysis of the return to scale in our sample, which also operates under a variable return to scale. It should be noted that we use a bootstrap DEA estimator (Simar and Wilson, 1999) to provide the estimates of the confidence intervals.

The most debated point in measuring the efficiency of universities and schools deals with different ways to define the variables used as inputs and outputs (De Witte and López-Torres, 2017). In this study, we select inputs that reflect total available capital (total income) and human resources. The first input indicates the amount of financial resources available for the university and is widely used in universities' efficiency assessments (e.g., Agasisti and Perez-Esparrells, 2010; Veiderpass and McKelvey, 2016; Wolszczak-Derlacz, 2017). Human resource indicators reflect not only the number of faculty staff but also the student ability proxy (the average USE score of incoming freshmen), both used as inputs in university production function (e.g., Sarrico and Dyson, 2000; Ray and Jeon, 2008; Agasisti and Pohl, 2012; Johnes, 2013). The output variables are reduced to two main outcomes of research universities: teaching and research productivity. We use the total number of students to measure teaching activities as we believe it can substitute a more widely used graduation rate (Thanassoulis et al. 2011; Agasisti and Johnes, 2015), because in Russia, dropout rates are moderate (Gorbunova, 2018). The latter is estimated by the maximum number of publications in journals indexed Web of Science/Scopus, which is a target indicator for the ExIn[2] and measure research activity (e.g., Wolszczak-Derlacz, 2017). Descriptive statistics of variables used are reported for the first and the last years of the analysis (see Tables 4 and 5). We measure the efficiency in a metafrontier framework, that is, both groups of universities are included in one model.

[2]To provide a robustness check for our model, we replace the absolute numbers of total income and publications with the same variables, scaled to the number of teaching and research staff in each university. The DEA scores of both models correlate substantially and significantly (see Table 8).

Table 4
Descriptive statistics: DEA and Malmquist variables in 2012/2013

		Control group			Treatment group		
	Variables	*N*	Mean	SD	*N*	Mean	SD
Input variables	Total income (thousand rubles)	15	2,690,340.09	2,076,047.229	15	4,612,090.396	2,099,072.516
	Total number of staff	15	1073	541.63	15	1318.20	729.079
	Unified state exam	15	69.585	5.28	15	72.729	7.686
Output variables	Total number of students	15	16,249.867	6422.789	15	16,653.067	10,062.537
	Number of publications	15	130.728	107.535	15	393.257	193.982

Source: Authors' calculations.

Table 5
Descriptive statistics: DEA and Malmquist variables in 2017/2018

		Control group			Treatment group		
	Variables	*N*	Mean	SD	*N*	Mean	SD
Input variables	Total income (thousand rubles)	15	3,523,824.047	2,802,832.193	15	6,498,210.487	3,520,835.135
	Total number of staff	15	1038.933	464.095	15	1178.133	558.138
	Unified state exam	15	72.083	6.235	15	81.891	7.918
Output variables	Total number of students	15	16,371.400	5006.680	15	16,378.533	9261.235
	Number of publications	15	490.801	338.356	15	1956.598	719.223

3.3. Step 3. The Malmquist productivity index

We measure the Malmquist productivity index showing the change in university productivity over time in order to trace the mechanisms of change in universities' efficiency. Based on nonparametric estimation of the efficiency frontier, this index explains the total productivity explained through a combination of

- efficiency change, that is, the change in how far the observed production is from maximum potential production (Färe et al. 1994, p. 71);
- technological change, that is, the shift of the frontier.

In higher education research, there are several attempts to develop Malmquist index calculations for British universities (Flegg et al., 2004; Johnes, 2008), Australian HEIs (Worthington and Lee, 2008), and cross-country comparisons (English and Italian HEIs in Agasisti and Johnes, 2009; Italian and Spanish in Agasisti and Pérez-Esparrells, 2010; European countries in Parteka and Wolszczak-Derlacz, 2013). The methodology is based on DEA and can be modified for input orientation or output orientation. This index is measured as a ratio of two distance functions representing efficiency performance in two different time periods (Lee et al., 2011):

$$MPI_i = \left(\frac{E_i^t\left(x^{t+1},\ y^{t+1}\right)}{E_i^t\left(x^t,\ y^t\right)} \left(\frac{E_i^{t+1}\left(x^{t+1},\ y^{t+1}\right)}{E_i^{t+1}\left(x^t,\ y^t\right)} \right) \right)^{\frac{1}{2}}, \quad (8)$$

where x and y are the levels of output produced by university i and the levels of inputs used, respectively, in period t or $t+1$. $E_i^{t+1}(x^t,\ y^t)$ is the production frontier that could be achieved by the combination of inputs used and outputs produced in period t if operating under the technology in period $t+1$. $E_i^t(x^{t+1},\ y^{t+1})$ is the maximum output that could be produced in period t given the outputs and the inputs of period $t+1$. $E_i^t(x^t,\ y^t)$ is the actual production combination of inputs and outputs in period t under the technology of the same period, as $E_i^{t+1}(x^{t+1},\ y^{t+1})$ is for period $t+1$.

$$MPI_i = \left(\frac{E_i^{t+1}\left(x^{t+1},\ y^{t+1}\right)}{E_i^t(x^t,\ y^t)}\right) \times \left(\frac{E_i^t\left(x^{t+1},\ y^{t+1}\right)}{E_i^{t+1}\left(x^{t+1},\ y^{t+1}\right)} \times \frac{E_i^t(x^t,\ y^t)}{E_i^{t+1}(x^t,\ y^t)}\right)^{\frac{1}{2}}, \tag{9}$$

or productivity change = efficiency change × change in technology (Färe et al., 1994).

Measuring the index, we can show whether the total productivity, technical efficiency (or the pure efficiency change), and technological change (or the technical change) of a certain university is increasing, decreasing, or stagnating over time as the Malmquist index will be greater than, less than, or equal to unity, respectively (9).

4. Results

4.1. Control-group selection

We conducted the PSM to ensure that the treated and the control groups of universities are similar when considering their observable characteristics. We did not observe any statistically significant difference between the two groups and consider the PSM to have been conducted successfully. Tables 1 and 2 show the descriptive statistics for matching and additional balancing covariates before and after the matching, respectively. The latter table also shows tests of means between the two groups. No structural difference was observed after the matching, and this will be additionally proved later by the fact that the participants and the nonparticipants were not significantly different from each other in terms of their relative efficiency before the actual changes in the strategic and operational development, that is, they were not different in terms of mean DEA scores in 2012/2013. Thus, we can claim that we have reduced possible biases in our estimations and achieve a higher level of similarity between treated and control groups in terms of observable characteristics prior to the ExIn. The fulfillment of the common support condition was the final step in the control-group selection procedure. The obtained region of common support is [0.022, 0.996].

ATT is 262.53 ± 57.27 publications, reported by the nearest neighbor matching in favor of the participant group. Thus, on average, universities participating in the ExIn had better research output than the nonparticipants in 2013–2018 (Table 3).

4.2. Efficiency analysis of Russian universities after the policy

We consider 2012/2013 as the zero year (time $t = 0$) in which the policy had not yet been implemented. The following dynamics can be explained through the ExIn effect. At the ExIn's starting point, a quick positive effect of the ExIn is observed in the participants. In the 2012/2013 academic

Table 6
Bootstrap DEA descriptive statistics. 2012–2018

	Control group			Treatment group		
Year	*N*	Mean	SD	*N*	Mean	SD
2012/2013	15	0.780	0.112	15	0.731	0.128
2013/2014	15	0.780	0.112	15	0.731	0.128
2014/2015	15	0.783	0.129	15	0.736	0.138
2015/2016	15	0.782	0.112	15	0.791	0.109
2016/2017	15	0.783	0.097	15	0.812	0.120
2017/2018	15	0.821	0.102	15	0.821	0.112

Source: Authors' calculations.

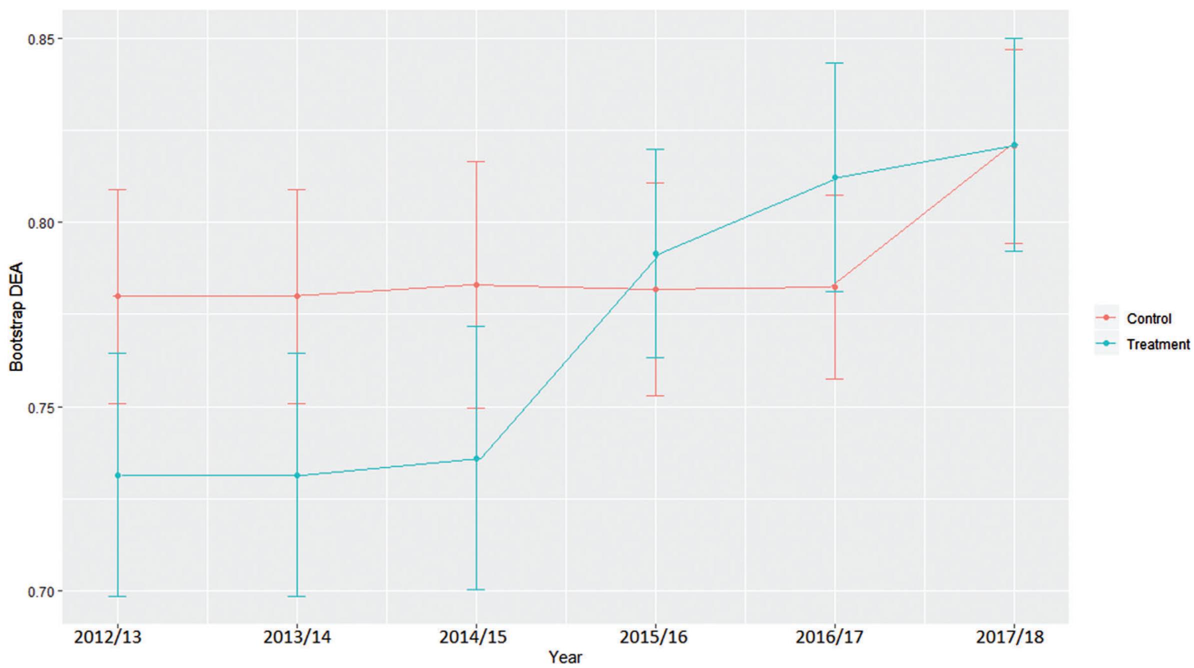

Fig. 1. DEA scores by groups.

year, participants and nonparticipants did not differ significantly in relative efficiency. We do not observe a major effect of the ExIn in terms of efficiency of the participants compared to the nonparticipants over the whole period (Table 6). However, we can see that the participants clearly managed to improve their efficiency scores somewhat (Fig. 1). The most important highlight of the efficiency analysis is related to the fact that we conduct a short-term measurement of a long-term program, and substantial efficiency change is time-demanding.

First, all possible expected effects of the ExIn have delayed effects and require time to settle. The ExIn can be regarded as a specific reputational signal in the higher education system: participation

Table 7
DEA models for robustness check

	Original model	Relative model
Input variables	Total income	Total income per staff capita
	Total number of staff	Total number of staff
	Unified state exam	Unified state exam
Output variables	Total number of students	Total number of students
	Number of publications	Number of publications per staff capita

Table 8
Pearson correlations for DEA robustness check models

Year	Pearson correlation: original model vs. relative model
2012/2013	0.823***
2013/2014	0.934***
2014/2015	0.971***
2015/2016	0.910***
2016/2017	0.940***
2017/2018	0.889***

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Source: Authors' calculations.

in the project can improve a university's status but it can also increase the competition among the universities, including nonparticipants.

The bootstrap DEA-obtained efficiency scores are proved to be robust through Pearson's correlation with another model (Table 7), the specifications of which are reported in Table 8.

4.3. Malmquist index dynamics: how productivity changed after the policy

The Malmquist index can be decomposed into efficiency change, measuring how universities move toward or away from the frontier, and technology change, measuring change in relation to the shift of the frontier. On average, both participants and nonparticipants reach the frontier in terms of the total productivity factor at the beginning of the observed period, and the participants improve their performance over time (mean productivity change exceeds 1, except the control group in 2014/2015–2015/2016; see Table 9 and Fig. 2). The productivity change of participants is higher than nonparticipants, and the difference in pace is significant for year-to-year changes during the period of 2013/2014–2016/2017.

The participants tend to push the technological frontier and expand it. Technological progress is observed over the period, as the mean technical change of the participants consistently exceeds 1, in contrast to nonparticipants, and difference is significant (Fig. 3). Compared to the nonparticipants, the participants are also able to produce more outputs given the resources they have. These effects are constant over time, and we do observe significant difference over the period. Year-to-year

Table 9
Malmquist Index and its decomposition, 2012/2013–2017/2018

Variables		Control group			Treatment group		
		N	Mean	SD	N	Mean	SD
Efficiency change	2012/2013–2013/2014	15	1.056	0.176	15	0.978	0.0889
	2013/2014–2014/2015	15	0.993	0.140	15	1.035	0.0978
	2014/2015–2015/2016	15	0.993*	0.183	15	1.090*	0.121
	2015/2016–2016/2017	15	1.025	0.0538	15	1.046	0.103
	2016/2017–2017/2018	15	1.034	0.0888	15	1.023	0.0697
Frontier shift	2012/2013–2013/2014	15	0.948***	0.0604	15	1.029***	0.0699
	2013/2014–2014/2015	15	1.070***	0.155	15	1.197***	0.110
	2014/2015–2015/2016	15	0.958***	0.0597	15	1.055***	0.111
	2015/2016–2016/2017	15	1.035***	0.0345	15	1.154***	0.105
	2016/2017–2017/2018	15	0.997***	0.0205	15	1.051***	0.0235
Productivity change	2012/2013–2013/2014	15	1.001	0.184	15	1.006	0.114
	2013/2014–2014/2015	15	1.065**	0.237	15	1.245**	0.212
	2014/2015–2015/2016	15	0.955***	0.199	15	1.143***	0.126
	2015/2016–2016/2017	15	1.062***	0.0761	15	1.208***	0.169
	2016/2017–2017/2018	15	1.032	0.0946	15	1.075	0.0729

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
Source: Authors' calculations.

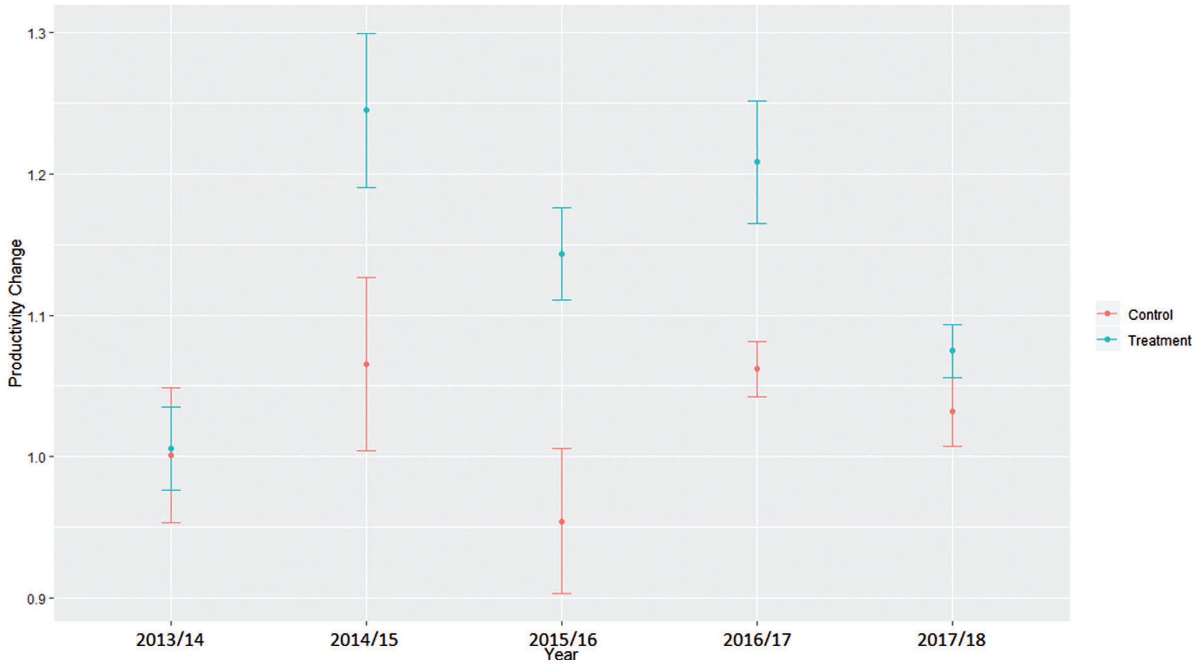

Fig. 2. Productivity change in dynamics.
Source: Authors' calculations.

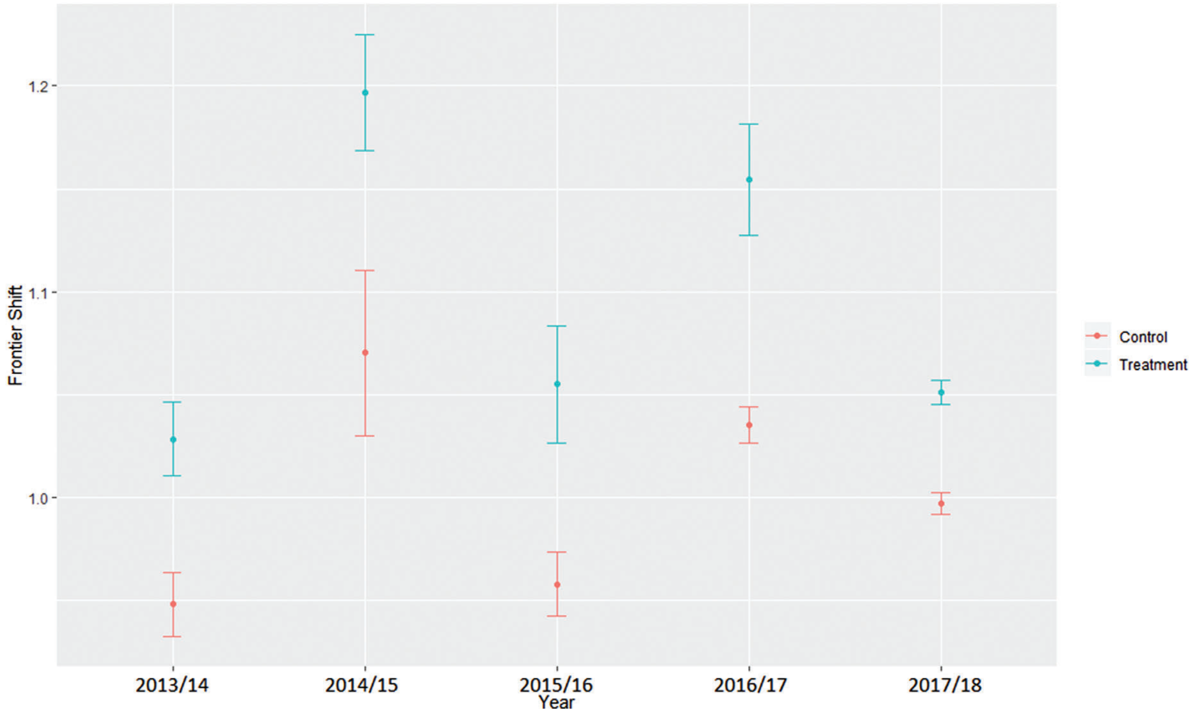

Fig. 3. Technical change in dynamics.
Source: Authors' calculations.

efficiency change is positive for participants, so they are moving slightly closer to the efficiency frontier; however, the difference in case of nonparticipants is not significant here (Fig. 4).

5. Discussion of the results and policy implications

This study contributes to the discussion of the effects of an ExIn on university activities with a special focus on efficiency. Our research is the first attempt to measure the efficiency of universities participating in the Russian ExIn. After control-group selection (PSM), we continue with the evaluation of efficiency (DEA) and its changes (the Malmquist index).

This paper develops a methodology for a complex causal assessment of ExIn effects. The issue of causality is important in the field of educational studies, especially if reform effects are evaluated (Schlotter et al., 2009). We reduce the causality ambiguity by forming a control group of universities, which are as similar to universities already participating in the ExIn as possible. The use of PSM in institutional higher education research is usually limited due to the reasons of high heterogeneity among universities, although there have been attempts in large systems such as China's (see Ou, 2017). The Russian higher education system is large and diverse as well, and our analysis proved that data-driven selection of a control group to assess an ExIn is possible. We found universities similar to the participants of the ExIn in the beginning of the project (2012/2013), if we compared

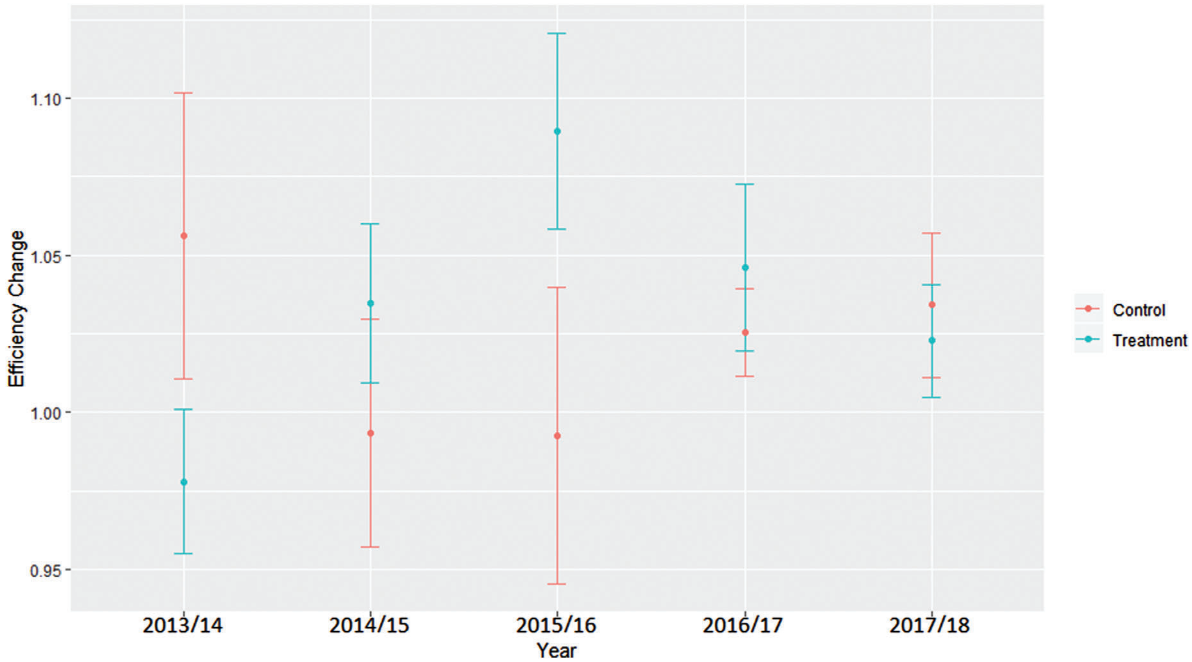

Fig. 4. Efficiency change in dynamics.
Source: Authors' calculations.

them not in absolute values but in relative and ratio measures. One of the main external proofs of PSM quality is the fact that most of the control-group universities applied for the program and were able to compete with those who were finally accepted.

In this study, we derive four main results to contribute to the discussion of efficiency and ExIn effects. First, the universities that participate in ExIn managed to improve their efficiency after the third year of project implementation (mean DEA score for 2013/2014 is 0.73 and for 2017/2018 is 0.82), while efficiency scores of nonparticipants have been stable around 0.78, which corresponds with the DEA results for a larger sample of Russian universities (Gromov, 2017). The selection of variables in efficiency modeling was checked for robustness: the difference in efficiency scores for modeling with absolute and with scaled values is not statistically significant. This result points to positive effects of ExIn on the efficiency of participating universities.

Second, from the first year of ExIn implementation and until 2017, the year-to-year increase of participant productivity was about 20%, while the productivity of nonparticipants changed ±5%. Third, it is the shift of the frontier that is important for productivity changes during the whole period. In contrast to nonparticipants, universities that participate in the ExIn have been transforming the technology. The efficiency change was smaller, and we do not see any significant difference between participants and nonparticipants in their movement toward the frontier. Yaisawarng and Ng (2014) found a positive effect of participation in Project-211 on productivity changes as well as technological advancement. However, the best non-Project-211 universities, despite limited resources, also show productive dynamics that are not so evident in the Russian case.

Finally, both estimations, the DEA score and the Malmquist indices, show unexpected results in 2017/2018. The DEA scores of nonparticipants jumped to the level of participants, and the pace of productivity change of participants decreased to the level of nonparticipants. On the one hand, one of the explanations was on the plane of the changes in the political environment. During 2017/2018 academic year, policymakers actively discussed the future of the ExIn as they developed strategic programs for the higher education system for 2019–2024. This process was linked to the presidential elections in spring 2018. The participant universities might have expected the rules of the game to change and governed with high level of uncertainty. On the other hand, nonparticipant universities might have already been looking ahead to being included in new wave of ExIn and saw participating universities as benchmarks. This supports the idea of the role of an ExIn as a "beauty contest" developed by Menter et al. (2018) with strong competitive basis.

These findings can be used to review approaches to ExIn policy development in general. Usually, ExIns are considered to rapidly change the inputs—upgrade hardware and infrastructure, attract the most talented students, and "star" faculty from around the world (Huang, 2015). But we see that an ExIn also changes processes and standards. For example, a move from research-oriented to international agendas results in more publications in internationally refereed journals. The changes may be due in part to standardized competitive mechanisms (Seong et al., 2008) and the effects of public announcements (Menter et al., 2018).

However, the Russian universities are currently evaluated on a yearly basis, and the projected future has high level of uncertainty. Our study shows that large and stable changes in the production function of universities demand certainty of targets and support, and any possibility that goals may be reconsidered can diminish the changes in technology, as occurred in 2017/2018. Also, the results of the assessment of German ExIn (Gawellek and Sunder, 2016) suggest that applying for the program was expensive for universities, forcing them to risk a large amount of resources. Universities lost considerably in efficiency and productivity, but those that ultimately received the grant successfully recovered in productivity and efficiency during the second period of the analysis.

Different universities employed different strategies to fulfill the government's targets and expectations. Chirikov (2018) conducted research based mostly on interviews with government officials as well as administration and faculty in the Russian excellence-driven universities in Moscow. The study revealed that there are four sets of mechanisms for organizational transformation to respond to the different definitions of global competition: paralleling, power play, imitation, and gaming. In the short term, our study shows that all of the strategies that were used can be successful and lead to positive efficiency changes. The main concern, however, is whether the universities will consolidate the achieved results over a longer period.

The limitations of this study are important to mention. First, the general limitations for ExIn's studies are time and attribution challenges (Ou, 2017). Real university modernization takes many years, and several initiatives were implemented only recently. Although the design of the presented research is aimed to diminish attribution challenge, but the wide range of changing contexts and other factors affecting university changes might be taken into account. Second, the sample limitations are important for large-scale interpretations of the results. The ExIns, especially the Russian one, are targeted at the small number of universities and set bounds to implementation of sophisticated quantitative methods for evaluation. Third, the available data limit the analysis of changes in inputs structure such as time of faculty distributed between teaching and research. Different approaches to redistribution of faculty, for example, between full-teaching and research-teaching staff might lead to different teaching and research outcomes due to spillover effects.

Acknowledgments

We are very grateful for fruitful discussion and valuable comments on the previous version of this paper at the following conferences and workshops: IPA Higher Education Symposium for Early-Career Scholars (Moscow, 2017), 5th Workshop on Efficiency in Education (Budapest, 2017; Huddersfield, 2018), CHER Conference (Moscow, 2018), Conference Higher Education in Modern Ecosystems: Efficiency, Society and Policies (Augsburg, 2018). All eventual errors are our own responsibility. This paper is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). We would also like to thank two anonymous reviewers and our editor for their helpful comments.

References

Abankina, I.V., Aleskerov, F.T., Belousova, V.Yu., Zinkovsky, K.V., Petrushenko, V.V., 2013. Ocenka rezul‘tativnosti universitetov s pomoshh‘yu obolochechnogo analiza danny‘x [Evaluation of Universities’ Performance with Data Envelopment Analysis]. *Voprosy Obrazovaniya* 2, 15–48.

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., Johnes, G., 2015. Efficiency, costs, rankings and heterogeneity: the case of US higher education. *Studies in Higher Education* 40, 1, 60–82.

Agasisti, T., Pérez-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.

Altbach, P.G., Hazelkorn, E., 2017. Pursuing rankings in the age of massification: for most—forget about it. *International Higher Education* 89, 8–10.

Altbach, P.G., Salmi, J. (eds), 2011. *The Road to Academic Excellence: The Making of World-Class Research Universities*. The World Bank, Washington, DC.

Boussofiane, A., Dyson, R.G., Thanassoulis, E., 1991. Applied data envelopment analysis. *European Journal of Operational Research* 52, 1, 1–15.

Chirikov, I., 2016. How global competition is changing universities: three theoretical perspectives. Center for Studies in Higher Education, UC Berkeley, Berkeley, CA. Available at https://escholarship.org/uc/item/50g3t797 (accessed July 15, 2018).

Chirikov, I., 2018. The Sputnik syndrome: how Russian universities make sense of global competition in higher education. In Oleksiyenko, V., Zha, Q., Chirikov, I., Li, J. (eds) *International Status Anxiety and Higher Education: The Soviet Legacy in China and Russia*. Center for Studies in Higher Education, UC Berkeley, Berkeley, CA.

De Witte, K.D., López-Torres, L., 2017. Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society* 68, 4, 339–363.

De Witte, K., Rogge, N., Cherchye, L., Van Puyenbroeck, T., 2013. Accounting for economies of scope in performance evaluations of university professors. *Journal of the Operational Research Society* 64, 11, 1595–1606.

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review* 84, 1, 66–83.

Flegg, A.T., Allen, D.O., Field, K., Thurlow, T.W., 2004. Measuring the efficiency of British universities: a multi-period data envelopment analysis. *Education Economics* 12, 3, 231–249.

Froumin, I., Lisyutkin, M., 2015. Excellence-driven policies and initiatives in the context of Bologna Process: rationale, design, implementation and outcomes. In Curaj, A., Matei, L., Pricopie, R., Salmi, J., Scott, P. (eds) *The European Higher Education Area*. Springer, Cham, pp. 249–265.

Gawellek, B., Sunder, M., 2016. The German excellence initiative and efficiency change among universities, 2001–2011 (no. 142). Working Paper, Universität Leipzig, Wirtschaftswissenschaftliche Fakultät, Leipzig.

Gorbunova, E. V., 2018. Vybytiya studentov iz vuzov: issledovaniya v Rossii i SSHA. [Dropout from universities in Russia and the USA]. *Voprosy Obrazovaniya* 1, 110–131.

Gromov, A., 2017. The efficiency of Russian higher education institutions and its determinants. Higher School of Economics Research Paper No. WP BRP 40/EDU/2017.

Hazelkorn, E., 2014. Rankings and the global reputation race. *New Directions for Higher Education* 2014, 168, 13–26.

Hazelkorn, E., 2015. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. Springer, Berlin.

Huang, F., 2015. Building the world-class research universities: a case study of China. *Higher Education* 70, 2, 203–215.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 1, 5–86.

Johnes, G. 2013. Efficiency in English higher education institutions revisited: a network approach. *Economics Bulletin* 33, 4, 2698–2706.

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.

Johnes J., 2008. Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/2005. *The Manchester School* 76, 6, 653–674.

Lee, K.R., Leem, B.I., Lee, C.W., Thales, S., Lee, C., 2011. Malmquist productivity index using DEA frontier in Stata. Available at https://pdfs.semanticscholar.org/847f/0d96a2b8e5581ba45e32406ec16a8858839c.pdf (accessed March 13, 2018).

López-Torres, L., Prior, D., Santín, D., 2016. Quality improvement programs in public schools. An evaluation of their effectiveness and efficiency. Ph.D. thesis, Universitat Autònoma de Barcelona, Barcelona.

Menter, M., Lehmann, E.E., Klarl, T., 2018. In search of excellence: a case study of the first excellence initiative of Germany. *Journal of Business Economics* 88, 9, 1105–1132.

Möller, T., Schmidt, M., Hornbostel, S., 2016. Assessing the effects of the German Excellence Initiative with bibliometric methods. *Scientometrics* 109, 3, 2217–2239.

Ou, Y., 2017. Evaluating effects of world-class university program using propensity score matching: evidence from China's Project 211. ISSI 2017—16th International Conference on Scientometrics and Informetrics, Conference Proceedings, Wuhan, China, pp. 400–411.

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.

Poldin, O., Matveeva, N., Sterligov, I., Yudkevich, M., 2017. Publication activities of Russian universities: the effects of Project 5–100. *Educational Studies*, 2017, 2, 10–35.

Ray, S.C., 2004. *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. Cambridge University Press, Cambridge.

Ray, S.C., Jeon, Y., 2008. Reputation and efficiency: a non-parametric assessment of America's top-rated MBA programs. *European Journal of Operational Research* 189, 1, 245–268.

Rosenbaum, P.R., Rubin, D.B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39, 1, 33–38.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 5, 688–701.

Salmi, J., 2009. *The Challenge of Establishing World Class Universities*. The World Bank, Washington, DC.

Salmi, J., 2016. Excellence initiatives to create world-class universities: Do they work? *Higher Education Evaluation and Development* 10, 1, 1–29.

Sarrico, C.S., Dyson, R.G., 2000. Using DEA for planning in UK universities—an institutional perspective. *Journal of the Operational Research Society* 51, 7, 789–800.

Schlotter, M., Schwerdt, G., Woessmann, L., 2009. Methods for causal evaluation of education policies and practices: an econometric toolbox. EENEE Analytical Report No. 5, European Expert Network on Economics of Education.

Seong, S., Popper, S.W., Goldman, C.A., Evans, D.K., Grammich, C.A., 2008. *Brain Korea 21 Phase II: A New Evaluation Model*. Rand Corporation, Santa Monica, CA, pp. 64–76.

Shin, J.C., 2009. Building world-class research university: the Brain Korea 21 project. *Higher Education* 58, 5, 669.

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.

Sokolov, M., Tsivinskaya, A. 2018a. *Monitoring the Effectiveness of Educational Organizations as a Tool for Evaluating Russian Universities (NNS-S (2) Series of Preprints)*. CIANO EUSPb, St. Petersburg.

Sokolov, M., Tsivinskaya, A. 2018b. *Monitoring the Effectiveness of Educational Organizations as a Source of Data on Russian Higher Education (HHC-S (1) Series of Preprints)*. CIANO EUSPb, St. Petersburg.

Thanassoulis, E., Kortelainen, M., Johnes, G., Johnes, J., 2011. Costs and efficiency of higher education institutions in England: a DEA analysis. *Journal of the Operational Research Society* 62, 7, 1282–1297.

Turko, T., Bakhturin, G., Bagan, V., Poloskov, S., Gudym, D., 2016. Influence of the program "5–top 100" on the publication activity of Russian universities. *Scientometrics* 109, 2, 769–782.

Veiderpass, A., McKelvey, M., 2016. Evaluating the performance of higher education institutions in Europe: a nonparametric efficiency analysis of 944 institutions. *Applied Economics* 48, 16, 1504–1514.

Wolszczak-Derlacz, J., 2017. An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the US with the application of two-stage semi-parametric DEA. *Research Policy* 46, 9, 1595–1605.

Worthington, A.C., 2001. An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* 9, 3, 245–268.

Worthington, A.C., Lee B.L., 2008. Efficiency, technology and productivity change in Australian universities, 1998–2003. *Economics of Education Review* 27, 285–298.

Yaisawarng, S., Ng, Y.C., 2014. The impact of higher education reform on research performance of Chinese universities. *China Economic Review* 31, 94–105.

Yang, X., You, Y., 2017. How the world-class university project affects scientific productivity? Evidence from a survey of faculty members in China. *Higher Education Policy* 4, 1–23.

Zhang, H., Patton, D., Kenney, M., 2013. Building global-class universities: assessing the impact of the 985 Project. *Research Policy* 42, 3, 765–775.

Zong, X., Zhang, W., 2019. Establishing world-class universities in China: deploying a quasi-experimental design to evaluate the net effects of Project 985. *Studies in Higher Education* 44, 3, 417–431.