NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Tommaso Agasisti, Aleksei Egorov,
Margarita Maximova

DO MERGER POLICIES INCREASE UNIVERSITY EFFICIENCY? EVIDENCE FROM A FUZZY REGRESSION DISCONTINUITY DESIGN

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 226/EC/2020

Tommaso Agasisti[1], Aleksei Egorov[2],

Margarita Maximova[3]

DO MERGER POLICIES INCREASE UNIVERSITY EFFICIENCY? EVIDENCE FROM A FUZZY REGRESSION DISCONTINUITY DESIGN

This paper studies the effect of merger policies in Russia on university efficiency. We consider the non-voluntary merger policy conducted by the Ministry of Education and Science based on university performance indicators. First, the efficiency scores of universities are estimated using a bootstrapped DEA non-parametric technique. The efficiency scores were evaluated for universities that were merged and for a control group formed through propensity score matching before and after the implementation of the policy. Then, a fuzzy regression discontinuity design was implemented in order to reveal the causal impact of mergers on efficiency levels. We find a positive, statistically significant effect of merger policy on university efficiency. The results of the analysis suggest that merged universities experience greater efficiency gains (or smaller efficiency declines) after mergers.

Keywords: Universities, mergers, fuzzy regression discontinuity design, efficiency, DEA, Malmquist index.

JEL codes: I21, I23, I28

[1] Politecnico di Milano, School of Management, Milan – Italy, tommaso.agasisti@polimi.it
2 Laboratory for University Development of Institute of Education, National Research University Higher School of Economics, Moscow – Russia, * Corresponding author, aegorov@hse.ru
3 Laboratory for University Development of Institute of Education, National Research University Higher School of Economics, Moscow – Russia, * Corresponding author, mmaksimova@hse.ru

1. Introduction

University mergers are often considered a possible policy instrument for responding to key challenges facing higher education (HE) systems: massification, position in international rankings, and more efficient use of scarce resources. There is evidence in the literature that mergers may lead to desired positive outcomes. Economic theory predicts that the establishment of an organization of relatively larger size consolidates resources and improves efficiency when universities can be take advantage of increasing returns to scale (Papadimitriou and Johnes, 2018). Mergers may also cause positive outcomes for the strategic development of universities because the increase of an institution's size may improve its global competitiveness and positioning in international rankings, enhancing its reputation (Valimaa et al., 2014; Harman and Harman, 2003). These considerations about the possible effects of university mergers have consequences for public policy in HE, and this policy tool has been used in many contexts in recent years. Despite universities often being considered as organizations that are "resistant to change" (Rantz, 2002; Clark, 1998) mergers of higher education institutions (HEI) have been implemented in many countries – Australia (Goedegebuure et al., 1993; Harman, 2000), Canada (Eastman and Lang, 2001), South Africa (Sehoole, 2005), Hungary (Morgan, 2000), and China (Wang, 2001) to name a few.

The Russian case is interesting because the HE system has been subjected to an intensive merger policy. Many Russian universities have been merged since the 1990s for different reasons, but the peak of merger activity was in 2012-2016. In 2012, the Russian Ministry of Education and Science (the Ministry) introduced a special management tool called Monitoring of Performance of Higher Education Institutions (MoP) to gather 162 different indicators of university performance which form the basis for calculating 7 indicators that are used for the evaluation of university efficiency (Sokolov and Tsivinskaya, 2018). These indicators represent six dimensions of university activities: educational, research, international, finance, academic staff, and the employment of graduates. For each of the selected indicators, the Ministry sets threshold values representing "expected minimum performance". If a university reaches the threshold values for less than 4 indicators, it is considered "low-performing". Low-performing public HEI have a high probability (although not certainty) of being merged with other universities. Low-performing private universities usually have their licenses withdrawn. What is less clear, however, is whether institutions resulting from merger are more efficient and better performing than the previous inefficient ones.

This paper investigates whether merges resulted in efficiency gains for the universities subjected to the policy. We consider the mergers that happened in 2013-2015, when they mostly were based on MoP. The paper concentrates particularly on how the merger policy influenced the efficiency level of the low-performing HEI. From such a perspective, the results can be informative not only for Russia, but can also provide insights for similar policies in other parts of the world.

The data used in the study contains information on all 475 Russian public HEI without branches. The sample contains 13 universities which were established through merging two or more HEI. These organizations are considered in 2013 and 2017, since the decisions about mergers were made by the Ministry based on the performance indicators for 2013, and all these mergers were completed before 2017. In order to reveal the effect of merger policy on efficiency an empirical strategy that consists of three steps is employed. In the first step, a control group is created, for this purpose, a Propensity Score Matching (PSM) algorithm is applied to observable characteristics. Next, using the PSM sample, the efficiency of universities is estimated using a bootstrapped data envelopment analysis (DEA). The efficiency estimation for the two years is accompanied by the Malmquist productivity index (MPI) calculation (Fare et al., 1994) in order to analyze the efficiency change during the period and its decomposition. Finally, a fuzzy regression discontinuity design (FRDD) is employed in order to investigate the causal effect of the merger policy on university efficiency. A fuzzy instead of sharp RD is applied in order to deal with endogeneity rooted in partial compliance (for example, low-performing universities which were not merged).

The innovation of this study is twofold. First, this paper evaluates the consequences of the merger policy for the efficiency of Russian universities. This is in a context where the debate in this area is central to the policy discourse. Second, this study uses a regression discontinuity design in order to identify a causal effect of public policy on university efficiency, providing more robust evidence than that reported by previous studies (which is, in many cases, just descriptive). To anticipate our main results, the treated universities (i.e. universities that received “low-performing” status by the Ministry and were merged) demonstrate greater efficiency gains between 2013 and 2017 compared to the control group.

The paper includes seven sections, of which this is the first. Section 2 presents the literature review. Sections 3 and 4 describe the economic framework for the analysis of how mergers can affect efficiency levels and a background of the merger policy in Russia. Section 5 describes the empirical strategy used in the paper. Section 6 is devoted to the illustration of our results. Section 7 presents the discussion, concluding remarks, limitations of the study and possible directions for further investigation.

2. Literature Review

Studies on mergers in HE are mostly focused on the motivation for merger activities (Rowley, 1997; Johnes and Tsionas, 2018), factors affecting mergers process (Harman and Meek, 2002; Locke, 2007), and the short-term and long-term effects of merger policies on different aspects of university activities (Valimaa et al., 2014; Wan, 2008). The literature also includes case studies of mergers in specific HE systems (Harman and Harman, 2003; Harman and Meek, 2002; Aagaard et al., 2016).

Mergers in HE are studied from different policy and managerial perspectives. Goedegebuure (1994) noted that most studies of mergers can be classified into one of three categories: macro-, meso- and micro-level studies. Macro-level studies analyze the consequences of merger policies for national HE systems (Kyvik, 2004). Meso-level studies are focused on interactions between different organizations during mergers – how mergers are negotiated and how the decisions regarding mergers are made (Rowley, 1997). Micro-level studies analyze mergers from the viewpoint of an individual organization involved in the process (Hay et al., 2001).

The definition of the term "merger" varies significantly across different studies. Most authors consider mergers as a form of collaboration between HEI. Harman and Harman (2003) highlighted six models of collaborations between universities: informal collaboration, affiliation, consortium, joint departments, merger with a federal structure, merger with a unitary structure, i.e. mergers are considered as the most complete and profound form of collaboration. Wan (2008) concludes from an analysis of literature that the term "merger" is used to describe a wide and heterogenous range of organizational arrangements in HE. The most common definition is a process during which two or more organizations are structurally or functionally combined into one entity (Harman and Harman, 2003; Valimaa et al., 2014). However, even within this definition mergers may differ significantly in their specific features, the literature suggests many different classifications of mergers. Skodvin (1999) proposes the dichotomy between "integration" and "diversification", where the former is the merging of universities with the same academic profile and the latter is the merger of universities with different educational programs. Harman (1991) suggested the classification of mergers based on the sector where it took place and highlighted two types of mergers: cross-sectoral mergers and mergers of similar institutions. Other classifications are based on the number of universities involved in the merger, whether the merger has the form of consolidation or take-over and so on (Meek, 2002; Harman, 1991).

Most studies focused on motivations for HE mergers. Meek (2002) says that most mergers were motivated by pressures to increase the efficiency and effectiveness of universities, to resolve the problem of non-sustainable universities and institutional fragmentation, to improve the quality of teaching and research, and to increase government control over national HE systems. Johnes and

Tsionas (2018) split motives for mergers into two groups: strategically motivated mergers and mergers that are driven by the efficiency theory. Motivations rooted in efficiency theory suggest that universities can be run more efficiently together rather than separately. This effect can be achieved by spreading administrative activities over a larger output, lower maintenance costs for fixed assets, the spread of academic staff over a larger number of students (Johnes and Tsionas, 2018; see also Fielden and Markham, 1997). However, mergers may cause efficiency gains only when the production process of the universities can be described by a production function with increasing returns to scale (in other words, in cases in which there is still a room to produce higher levels of output with the same available resources). In this vein, the motivations for mergers of a purely economic nature are discussed in the next section (see the economic theory discussed there). Strategic motivations for merger activities may have different particular considerations. Relatively large universities tend to be more competitive nationally and globally and, consequently, achieve better positions in international rankings (Valimaa et al., 2014). University mergers with this particular aim were implemented, for instance, in Denmark (Aagaard et al., 2016) and Australia (Harman, 1991). Mergers may also be motivated by strategic considerations which are related to the objectives of public policy in HE. In this case, mergers are implemented in order to achieve financial and management optimization in the whole HE system, by means of larger, more structured institutions. Examples of such mergers can be found in China (Wang, 2001; Zhao and Guo, 2002). Lastly, smaller universities may want to merge with other organization in order to survive and expand (Johnes and Tsionas, 2018). Wan (2008) divided all strategic motivations into two groups: mergers for survival and mergers for mutual growth. However, despite this variety of motivations for mergers, all of them are based on a particular set of assumptions made by the policymaker.

The literature especially important in the context of this study is on the effects of mergers on the efficiency of universities involved. The number and quality of existing empirical works of this kind are limited. Papadimitriou and Johnes (2018) derived efficiency scores for English universities for a 17-year period. The authors, using a regression model, investigated the effects of mergers on efficiency rates. The regression model included the efficiency score as the dependent variable, and dummy variables reflecting the year of merger, each of the four years following the merger, each of the three years prior the merger and control characteristics of universities as independent variables. The purpose was to investigate the difference in the merger effect over the period. The authors conclude that merged HEI demonstrate higher levels of efficiency, but this effect is only short-term. Johnes (2014) also studied English universities using panel data over 13-year period and found statistically significant differences in efficiency between merged and non-merged universities. The analysis shows that merged universities demonstrate higher levels of efficiency. This was the first attempt to estimate the effect of merger activity on efficiency level, where the author compares mean efficiency scores for three groups of universities: pre-merger, post-merger and non-merger. However, these findings cannot be a proof of merger effect, because merging and non-merging universities may differ in some other underlying and unobservable characteristics. One further study on the English HE sector is Johnes and Tsionas (2018) where the authors employ Bayesian techniques organized around the use of Markov chain Monte Carlo methods. The main finding of this study is similar to the results of Papadimitriou and Johnes (2018): efficiency gains may be observed only in the first year after the merger. Hu and Liang (2008) studied changes in the research productivity of Chinese universities before and after mergers using MPI and found that mergers do not have the significant scale effect on institutions involved. Overall, the number of previous academic contributions that assess the effect of mergers on university performance is very limited and almost exclusively focused on English universities.

Our paper analyzes how the merger policy in the Russian HE sector influenced the efficiency of universities, filling some of the gaps in the literature reviewed in this section.

3. An economic framework for analyzing the impact of mergers on efficiency

This section briefly outlines the main channels through which mergers can have an effect on university performance. It is difficult to decompose into different channels the total effect of a merger on efficiency; such channels should be investigated by means of detailed survey data (not available in this case) or case studies. This paper identifies the total effect that occurs through all possible channels. Despite the fact that our study does not explicitly disentangle the impact of each channel, a comprehensive theoretical framework is needed to derive policy and managerial implications from the key findings, and for developing hypotheses for subsequent studies.

3.1 Economies of scale and scope effects

The main explanations for potential positive effects of mergers on university efficiency in terms of production theory are economies of scale and scope effects. The economies of scale effect means that the cost per unit of output decreases when the total amount of output produced increases. This phenomenon may also be described in terms of increasing returns to scale: a small percentage increase in input may lead to a large percentage increase in output. Economies of scope mean that producing several outputs jointly (for example, joint production of teaching and research in the context of this study) may generate cost savings. Such an eventuality can be realized when the same inputs are used for jointly producing different outputs, without increasing unit costs. This is the case, for example, of academic staff employed in both teaching and research. A merger always implies the increase of an organization's size, and due to economies of scale, higher efficiency of merged institutions can be expected.

When the merger occurs between universities with different missions, for example, between teaching-oriented and research-oriented universities, it may lead to economies of scope, since it might be the case that producing research and teaching jointly is more efficient due to, for example, the involvement of students in research activities, the higher quality of teaching provided by researchers, or the utilization of research equipment for teaching purposes. Finally, for diversification mergers (Cai et al., 2016), i.e. when a merger occurs between institutions that operate in different academic fields, efficiency gains may be generated through the development of interdisciplinary research and more flexible curricula and study trajectories for students.

The literature on economies of scale and scope in HE suggests that increasing returns to scale may be observed in particular countries, for example, UK (Papadimitriou and Johnes, 2018), US (Koshal and Koshal, 1999), Japan (Hashimoto and Cohn, 1997), Germany (Olivares and Wetzel, 2014), Australia (Worthington and Higgs, 2011), Switzerland (Filippini and Lepori, 2007) and Italy (Abramo, D'Angelo and Di Costa, 2014). Given this international evidence, it could be the case that there is room for increasing efficiency through scale and scope effects in Russian HE.

3.2 Managerial changes and quality of implementation

The economic effects mentioned above are just one explanation for possible post-merger efficiency. Another important factor determining the efficiency of HEI is the quality of management (Agasisti, 2017). Mergers may cause substantial changes in management practices because of the merger of two or more corporate cultures and management styles. The literature suggests that the quality of management may be an important determinant of university performance (Shattock, 2010). Certain managerial practices and leadership styles may be associated with higher or lower levels of efficiency, and the final effect of a merger on efficiency may be determined by the set of managerial practices transferred to the new organization from merging ones.

A policy-related issue which can determine the results of mergers is how the merger is implemented. If a merger of universities provokes the conflict and opposition among students or faculty, it may lead to declines in efficiency. The management literature confirmed that conflicts among the top management (Weber and Schweiger, 1992) and conflicts in the corporate cultures (Weber and Camerer, 2003) of merging organizations may cause significant problems, including declines in performance.

4. Policy background

In the Russian HE sector, MoP was introduced in 2012, and in 2013 it became obligatory for all universities[4]. The tool is a database containing information on different performance indicators of participating universities. The indicators are structured into seven blocks – educational activity, research activity, international activity, financial activity, infrastructure, employment of graduates, human resources. Each block contains from 1 (employment of graduates) to 16 (research activity) specific indicators. There are also more than 50 additional performance indicators not distributed across the seven blocks. The data is gathered by the Ministry and officials directly from the university. Every year each university fills special statistical form and transfers it to the Ministry. However, MoP is not only a database and the main data source on Russian HEI, it is an important tool for managing the HE system. The aim of this policy instrument is twofold. First, it is aimed at increasing the transparency in HE, because all the results of MoP are available on the internet. Second, the aim of this instrument is to optimize HE, identifying universities with low levels of performance and withdraw their license or merge them with better performing one(s).

Universities with weak performance were identified based on seven performance indicators which is denoted by "principal performance indicators". These indicators are:

- average unified state exam (entrance exam) (USE)[5] score (educational activity) (e1);
- total value of R&D projects per faculty (research activity) (e2);
- share of foreign students in total number of students (international activity) (e3);
- total income from all sources per faculty (financial activity) (e4);
- total area of training and laboratory facilities per student (infrastructure) (e5);
- share of students employed 1 year after graduation (employment of graduates) (e6);
- faculty with PhD degrees per 100 students (human resources) (e7).

For each indicator, the Ministry sets a desired threshold value. The threshold values may vary across different types of universities and across universities located in different regions. If a particular university fails to achieve the threshold values for at least four indicators, it is recognized as a "low-performing" university.

After a list of low-performing universities is formed, a special commission analyzes the consequences of license revocation/merger and if the commission decides that it is possible to stop the activities of particular university without significant negative consequences, the Ministry issues the

[4] This study is focused on mergers that were based on the results of Monitoring of Performance, therefore here we provide just a description of Monitoring of Performance policy. However, since the beginning of 1990s, there were a lot of other mergers in Russian higher education sector. The detailed description of higher education merger policies in Russia can be found in the Annex 5.
[5] The exam that is taken by all school graduates in Russia. The university's admission is based on the results of this exam.

order for merger or license revocation. In other words, receiving low-performing status increases significantly the probability that university will be closed or merged. The particular treatment depends on the type of HEI. Low-performing public universities were usually merged with stronger institutions, while low-performing private universities usually have their license withdrawn.

We take advantage of the policy's peculiar features. Specifically, we use universities which received low-performing status based on MoP for 2013 for a quasi-experimental evaluation of the merger policy. In total, 38 public universities received this status. However, only 11 out of 38 were actually merged with other HEI. In addition, there were 2 mergers which were not formally based on the MoP results.

The mean average USE score for the treated group is 62.33 which is less than mean value for whole sample (64.79). Moreover, mean total value of R&D projects per faculty is also less than mean value for whole sample (172,000 and 206,000 rubles, respectively). Treated universities receive less income from all sources per faculty (1.85 million rubles) than the mean value (2.03 million rubles). The concentration on mergers that were based on MoP makes the sample of "treated" universities used in the study homogenous. All mergers included in the sample were based on the same mechanisms and principles, and all involved universities were similar in terms of their performance indicators. A regression discontinuity design (RDD) is used to explore the causal effect of the policy, by deriving a threshold of performance – unknown to the universities – which determines a substantial increase in the likelihood of being merged – see the econometric details in the next section.

5. Empirical strategy

5.1 Control group selection using propensity score matching

Before estimating the efficiency of Russian universities and its relationship with the merger policy, the group of universities under scrutiny is restricted by limiting the analysis to a sample of comparable institutions to avoid the risks associated with the heterogeneity of initial sample (i.e. looking at the performance of universities which are structurally different). In order to select the control universities PSM is conducted. To compare universities on their observable characteristics, the seven performance indicators are considered. Universities are judged by these indicators to be low-performing, so this set of indicators can be a good instrument to create a control group.

PSM was proposed by Rosenbaum and Rubin (1983) and developed by Heckman (1997). The PSM algorithm includes the calculation of conditional probabilities $e(x_i)$ of being in the treatment group on the condition of a given characteristic vector x_i. The conditional probabilities are computed according to Equation (1):

$$e(x_i) = Pr(z_i = 1|x_i), \quad (1)$$

where z_i is a dummy variable indicating the belonging to the group of low-performing universities. The joint conditional distribution of z_i is given by Equation (2):

$$Pr(z_1, \dots, z_n|x_1, \dots, x_n) = \prod_{i=1}^{N} e(x_i)^{z_i}\left(1 - e(x_i)\right)^{(1-z_i)} \quad (2)$$

Like the probability measure, the PS value varies from 0 to 1. In randomized experiments the values of PS, by definition, are equal to 0.5, i.e. each observation has an equal chance to be in the treatment and control group. If the study has a quasi-experimental design, the original PS distribution is unknown and needs to be evaluated. The main method of PS evaluation is logistic regression, in which the dependent variable is z_i is a binary variable with a unit value if it belongs to the study group.

The PS values are used for the matching procedure, i.e. for each observation from the treatment group one or more observations belonging to the control group are selected. Selection is carried out in such a way that the distribution of covariates is the same in the two groups. This achieves a balance ensuring that organizations from the control group had the same probability of being included in the treatment group, but for some reason were not.

In our study, PSM was performed using logit model with low-performing status as the dependent variable and the seven principal indicators as independent variables which allow us to obtain a control group which contains universities with similar performance characteristics but were not assigned as low-performing. PS were estimated based on the model reported in Equation (3):

$$\Pr(ineff = 1) = f(\beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3 + \beta_4 e_4 + \beta_5 e_5 + \beta_6 e_6 + \beta_7 e_7) \quad (3)$$

$where\ f(x) = \frac{1}{1+e^{-x}}$;

e_1 is the average unified state exam (entrance exam) score (educational activity);

e_2 is the total value of R&D projects per faculty (research activity);

e_3 is the share of foreign students in the total number of students (international activity);

$e_4$is the total income from all sources per faculty (financial activity);

e_5 is the total area of training and laboratory facilities per student (infrastructure);

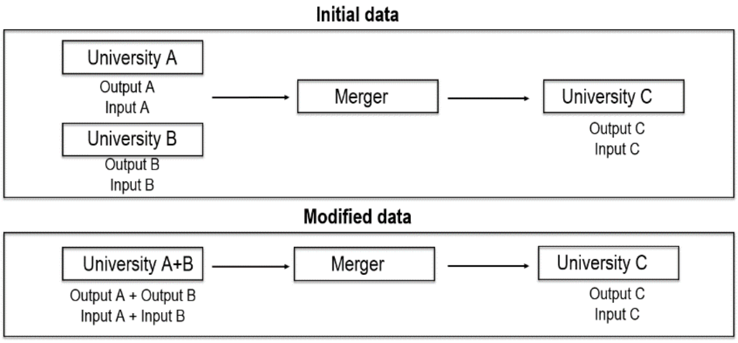
e_6 is the share of students employed 1 year after graduation (employment of graduates);

e_7 is the number of faculty with PhD degrees per 100 of student (human resources);

$\beta_1, \dots, \beta_7$ are the regression coefficients to be estimated.

Regarding the representation of treated universities in the data from MoP, a university that is merged with another based on the results of MoP is observed in the data till the year when the order to merge is issued. After that, it is reported as missing data for the year when the merger is in progress, and then, a year later, the new merged organization is observed in the dataset. In order to trace the dynamics of the efficiency of merged organizations we create "synthetic organizations" for the years before treatment. These synthetic organizations represent universities in the years before mergers as if they were already merged. In order to construct these organizations, weighted averages of inputs and outputs are calculated for the universities were merged. This principle of the initial data transformation is represented in Figure 1.

Figure 1. Modification of the initial sample, creating "synthetic organizations" to consider universities before and after merger



To perform PSM we use the nearest neighbor approach to take, for each element of the treated sample, three comparable elements from non-treated sample. The control group was formed from the whole sample of public universities without branches. The treated sample consists of 38 universities which are "synthetic" organizations. 152 observations were included in the PSM sample. After having restricted the sample to comparable institutions, on the basis of their PS, we proceed with measuring their efficiency.

5.2 Bootstrap Data Envelopment Analysis for measuring the efficiency of universities and the Malmquist productivity index for the analysis of the change in efficiency over time

In the context of production theory, universities are considered as multi-product organizations which transform a vector of inputs into a vector of outputs. Technical efficiency is the production of the maximum output at a given level of resources, and the level of inefficiency is calculated by constructing the boundary of production possibilities and the distance to it. There are two approaches to measure efficiency: parametric (Stochastic Frontier Analysis) (Aigner et al., 1977) and non-parametric (DEA) (Charnes et al., 1978). A more detailed discussion on the advantages and disadvantages of parametric and non-parametric approaches to efficiency estimation is beyond the scope of this paper and can be found, for example, in Murillo-Zamorano (2004). In this paper, the choice is to use a non-parametric technique which does not require assumptions about the distribution of the inefficiency and the functional form of the production function, and which is convenient for multi-input multi-output production functions. However, a robustness check is performed by estimating efficiency through a parametric approach (see Annex 2; the results hold also with the different methodology employed).

In order to estimate the efficiency of universities a bootstrapped DEA non-parametric technique is employed. Particularly, the following model is considered. Assume that university k uses the vector of resources $X_k = (x_{1k}, \dots, x_{Mk}) \in R_+^M$ in order to produce a vector of outputs $Y_k = (y_{1k}, \dots, y_{Mk}) \in R_+^S$. Then the following problem is solved:

$$\max_{\theta_k, \lambda_k, k.i \ = 1, \dots, N.'} \theta_k \qquad (4)$$

$$\theta_k y_{sk} \leq \sum_{i=1}^{N} \lambda_i y_{si}\, , s = 1, \dots, S;\ S = \#\{outputs\}, \qquad (5)$$

$$x_{mk} \geq \sum_{i=1}^{N} \lambda_i y_{si}\, , s = 1, \dots, S;\ S = \#\{inputs\}, \qquad (6)$$

$$\lambda_i \geq 0$$

where θ_k is the efficiency of university k measured by the distance to the production frontier.

The efficiency estimates obtained based on this model are consistent (see the proof in Kneip et al., 1998), but at the same time they are likely to be biased (Simar and Wilson, 1998). DEA suggests that the true production frontier is not known, and an estimate of this frontier is considered based on some finite sample of observed production units. Since the results of the frontier estimation and, consequently, the inefficiency estimates are based on the particular sample, it may be sensitive to the variations in the sample. The bootstrap procedure is a way to analyze the sensitivity of efficiency scores relative to the sampling variations in the estimated frontier (Simar and Wilson, 1998). Bootstrapping in non-parametric frontier models proposed by Simar and Wilson (1998) is employed in order to obtain bias-corrected efficiency scores.

In order to estimate the efficiency model described above, the inputs and outputs should be defined. We assume university production function with three inputs and three outputs. The first input is income from all sources at constant prices measuring the amount of financial resources available to the university. This input is used, for instance, in Agasisti and Johnes (2009) and Agasisti and Perez-Esparrells (2010). The second input is the average unified state exam (USE) score measuring the quality of entrants. There is much literature in the economics of HE which stresses that the quality of students can be considered as inputs for the university production function (see, for example, Hoxby, 1997 and Rothschild and White, 1995). The third input is the total number of faculty which measures the human resources available (Wolszczak-Derlacz and Parteka, 2011; Agasisti and Pohl, 2012). The outputs reflect the three missions of university: teaching, research and knowledge transfer. Teaching is measured by the total number of students. A more widespread indicator for teaching activity in the literature is the total number of graduates (Bonaccorsi, 2007), however, since data on the number of graduates is not officially released by the Ministry, the total number of students is used as a reasonable proxy for the teaching volume. The variation of drop-out rates across different universities is low, therefore, we can expect a high correlation between number of students in the university and number of graduates. Research is the total number of publications in journals indexed in Scopus and Web of Science, which is chosen with the aim of measuring scientific productivity (Parteka and Wolszczak-Derlacz, 2013; St. Aubyn et al., 2009). Knowledge transfer is the total income from grants obtained for applied research (from non-public sources) carried out by the university. In the Russian context, this variable is a good proxy for the intensity of cooperation between universities and industry (Agasisti et. al., 2018).

Efficiency scores for 2013 and 2017 are estimated. 2013 is assumed as a starting point for the analyses before the policy implementation, since the decisions about mergers included in the sample were made on the basis of the results of MoP for 2013. The efficiency estimated in 2013 can be regarded as the performance of universities before the implementation of the policy. The Ministry's orders about mergers were issued only in 2015, and all mergers were completed in 2016. 2017 is used for the analysis of the policy's effects, allowing two years for the policy to realize its potential. The effect estimated is necessarily short-term. It must be kept in mind here that some previous literature argues that policy effects are gained after a longer period (Papadimitriou and Johnes, 2018).

In order to analyze the change of university efficiency between 2013 and 2017 we employ MPI (Fare et al., 1994). This index measures the total productivity change within a particular time period, and decomposies this change into efficiency change ("catching-up") and changes in the frontier ("technological change"). MPI is calculated using Equation (7).

$$MPI_i = \left(\frac{E_i^{t+1}(x^{t+1},y^{t+1})}{E_i^{t}(x^{t},y^{t})}\right) \times \left(\frac{E_i^{t}(x^{t+1},y^{t+1})}{E_i^{t+1}(x^{t+1},y^{t+1})} \times \frac{E_i^{t}(x^{t},y^{t})}{E_i^{t+1}(x^{t},y^{t})}\right)^{1/2} \quad (7)$$

The transformation procedure of merged universities and the creation of "synthetic" organizations balances the sample size for 2013 and 2017, calculating the efficiency change during this period and proceeds to FRDD in order to identify the causal effect of mergers on efficiency.

5.3 Regression discontinuity design for the identification of the causal effect of mergers on efficiency

5.3.1 General setup of the RDD approach

RDD can be described in terms of a simple Rubin causal model (Rubin, 1974). Assume that there is binary intervention and two potential outcomes for the unit i: $Y_i(1)$ if the unit receives treatment and $Y_i(0)$ otherwise. However, in practice these values are not observable together for each unit but considered the average effect of treatment in some particular sample. Assume also that $W_i \in \{0,1\}$ is a variable reflecting the fact of treatment. Then the observed outcome can be written:

$$Y_i = \begin{cases} Y_i(0)\ if\ W_i = 0 \\ Y_i(1)\ if\ W_i = 1 \end{cases} \qquad (8)$$

We can also observe an additional variable X_i which determines the treatment: for example, if the value of X_i for particular unit i is higher than some threshold value set by policy maker, then this unit is prescribed for the treatment.

There are two main versions of regression discontinuity design – sharp and fuzzy (Imbens, Lemieux, 2008). In the sharp RDD the variable $W_i \in \{0,1\}$ is represented:

$$W_i = 1\ if\ X_i > a \qquad (9)$$

where a is some threshold value. Then the average treatment effect is calculated according to:

$$\tau_{AT} = \lim_{x \to +c} E[Y_i | X_i = x] - \lim_{x \to -c} \mathrm{E}[Y_i | X_i = x] \qquad (10)$$

In FRDD the threshold a does not change the probability of treatment strictly from 0 to 1. In this case if the unit reaches the threshold value a, it experiences a change in the probability of being treated. In FRDD, the treatment effected can be estimated through:

$$\tau_{AT} = \frac{\lim_{x \to +c} E[Y|X=x] - \lim_{x \to -c} E[Y|X=x]}{\lim_{x \to +c} E[W|X=x] - \lim_{x \to -c} E[W|X=x]} \qquad (11)$$

In this paper, the discontinuity around a cut-off point is implemented as how the effect of the policy is identified. We employ FRDD since, if a university is low-performing it does not necessarily mean that it will be merged; this status just increases that probability.

5.3.2 Identification of the threshold

We take advantage of the assignment to the treatment being based on MoP results, meaning that all universities in the treatment have low-performing status given by the Ministry. As discussed in the policy background section, a university receives low-performing status if it fails to overcome threshold values for at least four out of seven performance indicators. However, in order to perform FRDD in this context, we decide to artificially build a continuous running variable. First, the ratios of the observed values of the performance indicators to the appropriate thresholds were calculated. Universities located in different regions may have different threshold values for performance indicators, therefore individual threshold values are used in order to calculate these ratios. In the policy background section it is also mentioned that universities of different types may have different threshold values (medical schools, universities of sport and culture, etc.). However, the final sample obtained through PSM contains only universities “without specific activities”, so controlling for the region where university is located, we have the same threshold values for all universities in the sample. These ratios were averaged in order to obtain an aggregate performance score at university level.

Since the exact potential discontinuity point is known only for a discrete variable reflecting the number of principle performance indicators for which university achieves threshold value, it is necessary to identify the discontinuity point for a continuous running variable artificially. This is identified following the literature on structural brakes (Card et al., 2008). The same approach for identification of discontinuity point in RDD was used, for example, by Steinberg (2014). This approach suggests estimating Equation (12):

$$ineff_i = \beta_0 + \beta_1(Threshold_i) + \epsilon_i \quad (12)$$

where $ineff_i$ is a variable representing the low-performing status received by university i from the Ministry, based on MoP in 2013; $Threshold_i$ is an indicator function:

$$Threshold_i = I\{APS_i < \theta\} \quad (13)$$

where APS_i is an aggregate performance score for university i; θ is the threshold value that is identified from our sample; I is an indicator function.

In order to find the optimal value of threshold θ from the sample, a set of different values is considered and the one that maximizes R^2 of Equation (12) is chosen. Following the literature (Steinberg, 2014), to achieve greater robustness of the estimates in addition to Equation (12), four alternative specifications are considered:

$$ineff_i = \beta_0 + \beta_1(Threshold_i) + \beta_2(APS_i) + \epsilon_i \quad (14)$$

$$ineff_i = \beta_0 + \beta_1(Threshold_i) + \beta_2(APS_i) + \gamma X_i + \epsilon_i \quad (15)$$

$$ineff_i = \beta_0 + \beta_1(Threshold_i) + f(APS_i) + \epsilon_i \quad (16)$$

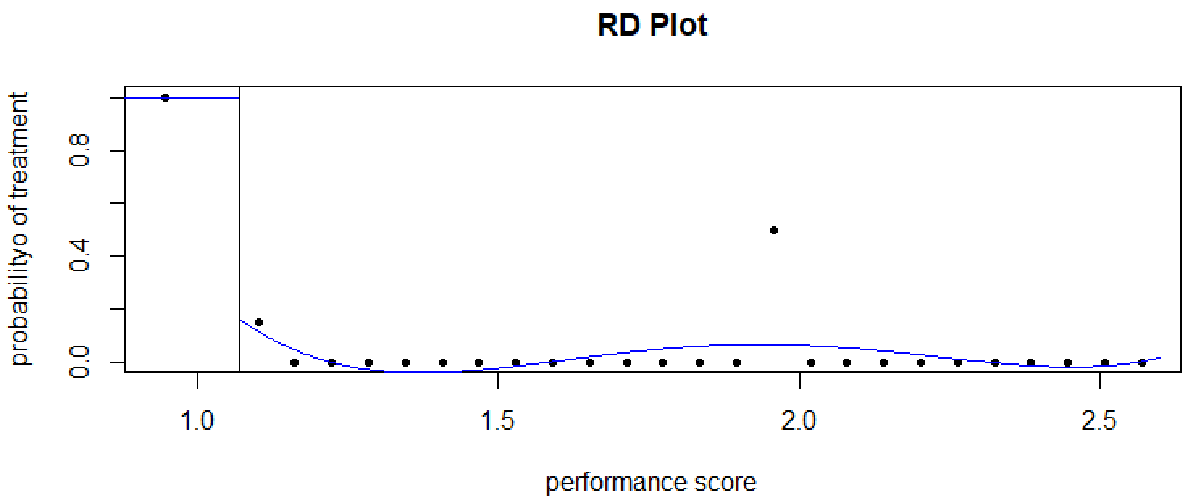
$$ineff_i = \beta_0 + \beta_1(Threshold_i) + f(APS_i) + \gamma X_i + \epsilon_i \quad (17)$$

where γX_i is the set of university characteristics multiplied by the set of regression coefficients, chosen to reflect some important features of the universities that are likely to be correlated with performance: number of branches, share of R&D project income in a total budget, share of students in STEM fields; $f(.)$ is a second-order polynomial function; all other notations remain the same.

Each Equation, (12), (14)–(17) is estimated 40 times with different values of θ in Equation (12). These values are taken from the range [0.80;1.20] with steps of 0.01. This range for the search of the discontinuity point seems to be reasonable since all universities with $APS_i < 0.8$ are treated and all universities with $APS_i > 1.2$ are not treated. Annex 1 presents the R^2 for each of these 200 regression equations. The maximum R^2 (from 0.871 to 0.896 depending on the model specification) is achieved for $\theta = 1.07$ and $\theta = 1.08$, the results are identical because these two threshold points provide exactly the same separation (the same number of universities below and above these points).

Next we proceed to RDD with the threshold values equal to 1.07. The relationship between the forcing variable and treatment status is presented in Figure 2.

Figure 2. The relationship between the forcing variable and treatment status



Note: non-parametric estimate based on the uniform kernel.

Bandwidth is chosen in a way that spans the full support of the data.

Figure 2 shows that there is a discontinuous gap in the relationship between the forcing variable and treatment assignment. Therefore, the assignment mechanism can be considered clear and RDD is applicable in this particular context. However, some partial compliance can be observed, for example two units were treated despite a relatively high performance score (these mergers were not formally based on MoP). In order to overcome the endogeneity problem caused by partial compliance, FRDD is used.

5.3.3 Fuzzy RDD for the identification of the causal effects of mergers on efficiency

The constructed forcing variable may be correlated with different unobserved university characteristics such as quality of management, however, these relationships should be smooth and without any jumps at the discontinuity point. Another issue here is that only 11 out of 38 universities assigned to the treatment were actually merged. Moreover, there are two mergers of universities that have a performance score higher than the cutoff point identified in the previous section. This partial compliance to the policy can be explained by two considerations. First, not all universities assigned to the treatment group were merged because of the specific criteria based on which Ministry makes a merger decision. This consists of two steps, first, a special commission of the Ministry forms a list of universities that receive low-performing status according to the MoP results. Second, another commission qualitatively analyzes the risks that can follow the merger for each university from the list and makes a recommendation which universities should be merged. As a result, not all universities deemed low-performing are subsequently forced to merge. The mergers not assigned to the treatment group were formally voluntary and not related to MoP. This partial compliance may lead to an endogeneity problem, related to the unobservable characteristics of the universities, affecting the likelihood of a merger. In order to deal with this problem FRDD is employed, assuming that the probability of receiving treatment jumps discontinuously at the cutoff point.

Following the literature, a parametric approach to the estimation of FRDD treatment effects is used. However, a robustness check is also conducted with a non-parametric estimation that is described in Annex 3. The dependent variable in our FRDD framework is the productivity change between 2013 and 2017 measured by MPI since we are interested in how merged universities improved or worsened their efficiency level after the merger. RDD models are also estimated for the components of MPI – pure efficiency change and the frontier shift. Three control variables are added to the model's specification – number of brunches, share of R&D project income in the total budget, and share of students in STEM fields. These three measures reflect different university structures and allow the capture of residual (after PSM) sample heterogeneity. We can expect the existence of efficiency heterogeneity across universities with different values of these control variables. In a parametric setting, we estimate:

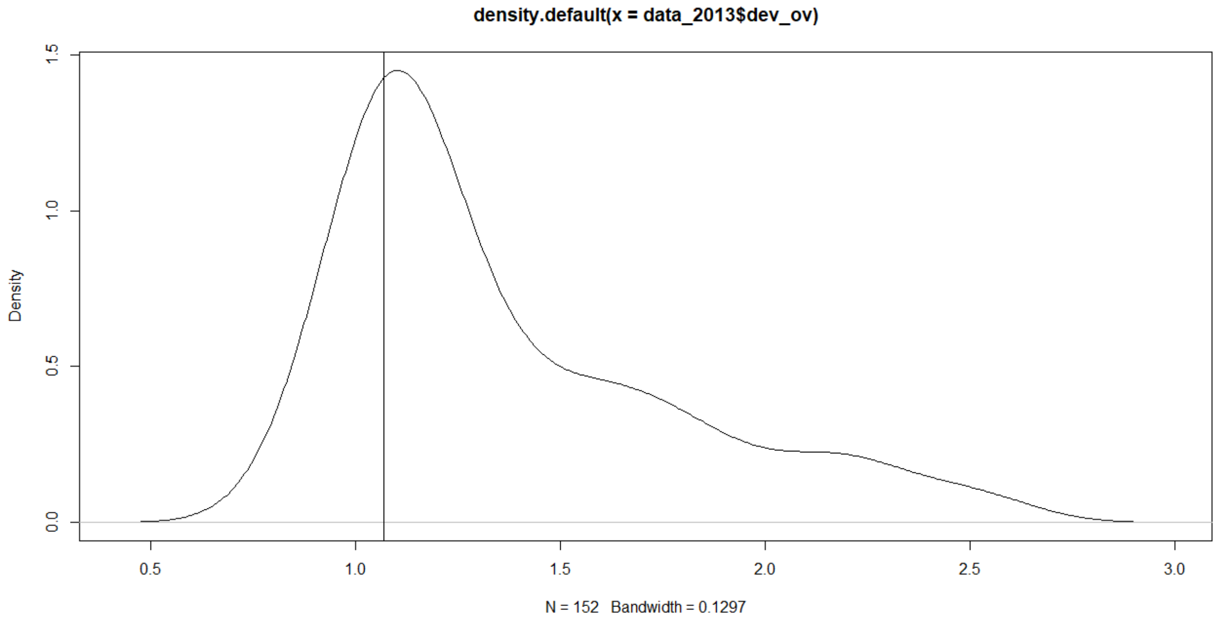
$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 f(PS_i) + \beta X + \epsilon_i \qquad (18)$$

$$T_i = \alpha_0 + \alpha_1 TA_i + \alpha_2 h(PS_i) + \alpha X + \vartheta_i \qquad (19)$$

where Y_i is the productivity change (MPI, efficiency change, frontier shift) of university i between 2013 and 2017; T_i is a dummy variable representing that a university was merged; PS_i is the performance score of university i in 2013; $f(.)$ is a flexible functional form that relates performance scores to the efficiency change; βX is the matrix containing control variables (share of students in STEM fields, number of branches, share of R&D income in total income) multiplied by the vector of parameters; ϵ_i is random error; TA_i is a dummy variable reflecting the university's assignment to the treatment; $h(.)$ is a flexible functional form that relates university performance and participation in treatment; ϑ_i is a random error.

Before performing FRDD estimates, we also check for random assignment near the point of discontinuity (i.e. the threshold defined in this section). Particularly, it is important that university managers cannot manipulate the performance score when it is near the cutoff. The situation when university is above the threshold point can be excluded and its managers rarely manipulate it in order to be assigned to the treatment. This potential behavior of managers is very unlikely, because the assignment to the treatment means negative outcomes for them (i.e. loss of powers). On the other hand, the performance score can be considered as an outcome of university's long-term development in terms of its teaching, research and third mission. If the university is just below the threshold, it is very difficult to manipulate its performance even marginally in short term in order to reach the threshold. Performance depends not only on management efforts, but on a wide set of exogenous factors as well. Therefore, it is safely assumed that universities do not have complete control over the running variable in the short term. These arguments in favor of the validity of RDD in our context are supported by the visual analysis of the density estimate of running variable presented in Figure 3 and by the results of McCrary sorting test[6] (McCrary, 2008).

Figure 3. Kernel density estimate for the running variable (vertical line represents the cutoff point)



The similarity of universities that lie just below and just above the cutoff point is confirmed by further methodological checks. First, we use the sample obtained using PSM algorithm based on the values of performance indicators used for treatment assignment, therefore the whole sample consists of comparable universities. Second, we further control for university heterogeneity by including in the RD models three control variables. These variables are the number of branches, the share on R&D income in total income and the share of students that study STEM subjects as a proxy for the subject mix.

[6] According to the results of this test, the estimated log difference in heights at the cutoff point is equal to 0.263 with the standard error equal to 0.327.

6. Results of the empirical analysis

6.1 Building a sample of universities to be compared with merged ones

We first build a control group based on the seven performance indicators included in the MoP, using PSM. Table 1 represents mean values and standard deviations for the seven indicators for three groups: treated, non-treated and whole sample in 2013 before performing the PSM procedure. Data for 2013 are used in order to perform PSM since the aim is to create a comparable control group before treatment. Treatment indicates that a university had low-performing status in 2014. For all indicators, mean values are higher in the non-treated group than in treated group. This descriptive evidence is coherent with the policy objective because if a university is treated as low performing, it must also report a lower level of performance indicators. These differences are significant for more than half variables[7]. It can also be observed that non-treated universities are less heterogeneous: the standard deviations for most indicators are higher in this group.

Table 1. Mean values of main performance indicators across different groups before and after PSM

	Before PSM			**After PSM**		
	Treated	Non-treated	Whole sample	Treated	Control	Whole PSM sample
Average USE score (e1)	62.33 *(6.50)*	65.06 *(7.95)*	64.79 *(7.85)*	62.33 *(6.50)*	62.62 *(5.94)*	62.55 *(6.06)*
Total amount of R&D projects per faculty (e2)	172 *(178)*	210 *(361)*	206 *(347)*	172 *(178)*	220 *(473)*	208 *(419)*
Share of foreign students (e3)	3.34 *(3.47)*	4.36 *(5.16)*	4.26 *(5.02)*	3.34 *(3.47)*	2.79 *(3.11)*	2.93 *(3.2)*
Total income from all sources per faculty (e4)	1850 *(714)*	2048 *(1330)*	2029 *(1284)*	1850 *(714)*	1856 *(918)*	1854 *(869)*
Total area of training and laboratory facilities per student (e5)	12.68 *(4.14)*	15.86 *(8.63)*	15.54 *(8.35)*	12.68 *(4.14)*	12.82 *(4.64)*	12.79 *(4.51)*
Employment of graduates (e6)	96.75 *(3.5)*	98.13 *(1.86)*	97.99 *(2.12)*	96.75 *(3.5)*	97.22 *(2.48)*	97.10 *(2.76)*
Faculty with PhD per 100 students (e7)	4.27 *(1.44)*	15.33 *(23.73)*	14.24 *(22.77)*	4.27 *(1.44)*	4.31 *(1.31)*	4.30 *(1.34)*

[7] T-tests for comparing means were conducted at the 5% significance level

# of observations	38	357	395	38	114	152

Note: *The table represents mean values for variables used for propensity score matching. Standard deviations are presented in the brackets.*
Source: *Authors' calculations based on the data from Monitoring of Performance.*

Table 1 also gives the means and standard deviations for seven principal indicators in three groups: treated, control and whole sample after performing PSM. There are no significant differences between the means in the treated and control groups[8], therefore PSM is used to create two groups of universities (treated and control) that do not differ significantly in terms of performance indicators.

6.2 University efficiency estimation through bootstrapped DEA and MPI

The descriptive statistics for the data used for efficiency estimation are presented in Table 2. The mean values show that the dynamics of the variables used for efficiency estimation during the period from 2013 to 2017 were approximately the same in the treated and control groups. All universities significantly increased their scientific productivity in terms of the total number of publications, and total value of R&D projects. A decline in total income from all sources measured in constant prices, total number of students and total number of faculty can be observed. Despite conducting PSM to balance the covariates related to university performance, the differences in the mean values of the indicators used for efficiency estimation can be observed. These differences are expected and necessary for efficiency evaluation. The set of variables used for PSM capture performance levels, while those used for efficiency analysis represent university production function.

[8] T-tests for comparing means were conducted at the 5% significance level

Table 2. Mean values of variables used for efficiency estimation

	2013 (Before treatment)			2017 (After treatment)			Change 2013-2017, %		
	Treated	Control	Whole sample	Treated	Control	Whole sample	Treated	Control	Whole sample
Inputs									
Total income from all sources, ml. roubles	3 232 *(2 140)*	1 067 *(1 132)*	1 266 *(1 397)*	2 816 *(1 847)*	1 046 *(1 242)*	1 209 *(1 399)*	-12.8	-2.0	-4.5
Average USE score	71.2 *(7.5)*	65.5 *(9.0)*	66.0 *(9.0)*	71.9 *(8.4)*	64.5 *(9.0)*	65.2 *(9.2)*	+1.1	-1.5	-1.2
Total number of faculty	1208 *(502)*	518 *(387)*	581 *(445)*	1 001 *(359)*	425 *(357)*	478.8 *(393.4)*	-17.1	-17.7	-17.6
Outputs									
Total number of publications	2 537 *(2 215)*	522 *(559)*	707 *(1024)*	4 193 *(3 158)*	1 733 *(1 862)*	1 960 *(2 124)*	+65.3	+231.9	+177.2
Total number of students	12 908 *(4 670)*	5 452 *(3 895)*	6139 *(4508)*	12 804 *(5 065)*	4 928 *(4 384)*	5 653 *(4 627)*	-0.8	-9.6	-7.9
Total amount of R&D projects, ml. roubles	487 *(488)*	108 *(198)*	142 *(261)*	657 *(679)*	129 *(273)*	178 *(362)*	+34.7	+20.0	+24.5
# of observations	38	114	152	38	114	152	38	114	152

Note: *The table represents mean values for variables used for efficiency estimation. Standard deviations are presented in the brackets. All monetary characteristics are in constant prices.*
Source: *Authors' calculations based on the data from MoP.*

The results of the efficiency estimation based on this sample are presented in Table 3. All universities are considered together, so assuming that both the treated and control universities adopt the same "technology" in the production process, and a single frontier of institutional efficiency exists – in other words, no group is assumed to be *a priori* structurally more or less efficient than the other. Table 3 shows that, in general, Russian universities became more productive in 2017 compared to 2013 – the value of MPI is greater than 1. However, the decomposition of the MPI suggests that this productivity increase was due to the frontier shift, while the pure average efficiency declined (0.974). For control universities the same change can be observed – an efficiency change of less than 1 and a frontier shift greater than 1 resulted in an MPI value slightly higher than 1. For treated group both efficiency change and frontier shift are greater than 1.

Table 3. Mean values of the estimated efficiency scores

	Treated	**Control**	**Whole sample**
2013	0.784	0.761	0.767
	(0.120)	*(0.107)*	*(0.111)*
2017	0.733	0.695	0.705
	(0.107)	*(0.127)*	*(0.123)*
MPI	1.215	1.054	1.097
Efficiency change	1.041	0.949	0.974
Frontier shift	1.155	1.119	1.129

Note: *The table represents mean values for efficiency scores. Standard deviations are presented in brackets*
Source: *Authors' calculations based on the data from MoP.*

The results of efficiency estimation and, particularly, the result that in general the treated universities became more productive (in terms of MPI) compared to the control universities, allows us to assume that this treatment may have a positive effect on the level of efficiency. This intuition is preliminary and descriptive in nature. The next section describes the results of FRDD analysis to show the causal nature of this effect.

6.3. The effects of mergers on efficiency: fuzzy regression discontinuity design

In FRDD, a set of additional variables is used to control for sample heterogeneity remaining after PSM – share of R&D income in the total income, number of branches and the share of students in STEM fields. The descriptive statistics for these variables are presented in Table 4. The significant differences in the mean values of these indicators, especially substantial differences in number of branches in treated and non-treated universities suggest that this set of control variables is a useful tool for taking into account sample heterogeneity.

Table 4. Mean values of variables used for fuzzy regression discontinuity design

	Treated	Control	Whole sample
Number of branches	4.2	1.9	2.1
	(3.5)	*(4.3)*	*(4.3)*
Share of R&D income in total income	13.5	7.7	8.2
	(8.2)	*(6.5)*	*(6.8)*
Share of students in STEM fields	0.5	0.4	0.4
	(0.3)	*(0.3)*	*(0.3)*

Note: *The table represents mean values for variables used for fuzzy regression discontinuity design. Standard deviations are presented in the brackets.*
Source: *Authors' calculations based on the data from MoP.*

Table 5 presents the FRDD estimation results. Table 5a corresponds to the FRDD model with MPI as a dependent variable, Table 5b and 5c present the results for FRDD models where the dependent variables are efficiency change and frontier shift between 2013 and 2017. Seven specifications of the RDD model are considered. Model (1) represents the baseline specification with the first-order polynomial (linear regression). Model (2) additionally contains the set of control variables, and model (3) also contains interaction terms[9]. Models (4) and (5) includes second-order polynomials, model (5) additionally contains control variables and interaction terms. Finally, models (6) and (7) have third-order polynomials and differ in terms of controls and interactions.

We can observe a statistically significant treatment effect for all specifications with different functions. This positive effect of the policy is observed also for most FRDD specifications with efficiency change and frontier shift as dependent variables. From an interpretative perspective, this finding indicates that the effect of the policy cannot be detected if we simply observe the average value of the performance indicators before and after the policy, averaging the data from the treated and control universities. When considering the complex nature of the data, instead, the findings demonstrate that mergers causally improved the efficiency of the merged institutions, in the short term. The magnitude of the gap between the treated and control universities near a cutoff point, depending

[9] We include in the model interactions between all three control variables: Number of branches x Share of R&D income, Number of branches x Share of students in STEM and Share of students in STEM x Share of R&D income.

on particular specification lies between 0.354 and 0.625 for MPI, between 0.209 and 0.275 for pure efficiency change, and between 0 and 0.327 for a frontier shift. These findings suggest that the treated universities demonstrate a greater productivity increase compare to the control ones due to pure managerial efficiency advancements and to the improvement of production technology.

In order to check the robustness of these findings two additional robustness checks are implemented. The first robustness check is a non-parametric estimation of the regression discontinuity. For the non-parametric estimation, different bandwidths for local-linear regressions were considered. Models (1)–(3) were estimated using the optimal bandwidth calculated using Imbens-Kalyanaraman method (Imbens and Kalyanaraman, 2009); models (4)–(5) were estimated with half the optimal bandwidth and models (6)–(7) were estimated with double the optimal bandwidth. The specifications considered in this robustness check also differ in terms of the inclusion of control variables and their interactions. For the non-parametric estimation, there is statistically significant treatment effect that is robust for all considered bandwidths, and for the inclusion of control variables and their interactions. The detailed results of non-parametric estimations are presented in Annex 3. As a second robustness check, we implement FRDD with different potential discontinuity points – specifically, the values 1.06 and 1.08 are adopted instead of 1.07 in the baseline model. Since these three points provide approximately the same separation of the sample, a statistically significant treatment effect is also observed in these specifications. The detailed results of these estimations are presented in Annex 4.

Table 5a. Parametric RD estimates for MPI

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Parametric estimation (2SLS)				
Treatment effect	0.625**	0.555*	0.419*	0.613**	0.479**	0.521.	0.354.
(standard error)	*(0.239)*	*(0.220)*	(0.187)	*(0.237)*	*(0.180)*	*(0.280)*	*(0.211)*
Polynomial of performance score	First	First	First	Second	Second	Third	Third
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	152	152	152	152	152	152	152

Note: *The list of control variables contains the total number of students, average USE score, share of students in STEM fields. Presented standard errors are heteroskedasticity robust, ***p<0.01, **p=0.01, *p=0.05, .p=0.1.*
Source: *Authors' calculations based on the data from MoP.*

Table 5b. Parametric RD estimates for efficiency change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Parametric estimation (2SLS)				
Treatment effect	0.209.	0.236	0.221*	0.279*	0.275**	0.234*	0.223*
(standard error)	*(0.112)*	*(0.118)*	(0.109)	*(0.109)*	*(0.099)*	*(0.116)*	*(0.098)*
Polynomial of performance score	First	First	First	Second	Second	Third	Third
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	152	152	152	152	152	152	152

Note: *The list of control variables contains the total number of students, average USE score, share of students in STEM fields. Presented standard errors are heteroskedasticity robust, ***p<0.01, **p=0.01, *p=0.05, .p=0.1.*
Source: *Authors' calculations based on the* data from MoP.

Table 5a. Parametric RD estimates for frontier shift

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Parametric estimation (2SLS)			
Treatment effect	0.327**	0.236*	0.123	0.244*	0.124	0.222*	0.081
(standard error)	*(0.116)*	*(0.118)*	(0.092)	*(0.109)*	*(0.089)*	*(0.148)*	*(0.126)*
Polynomial of performance score	First	First	First	Second	Second	Third	Third
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	152	152	152	152	152	152	152

Note: *The list of control variables contains the total number of students, average USE score, share of students in STEM fields. Presented standard errors are heteroskedasticity robust, ***p<0.01, **p=0.01, *p=0.05, .p=0.1.*
Source: *Authors' calculations based on the data from MoP*

Figure 4a. Regression discontinuity plots for different orders of polynomials (for MPI)

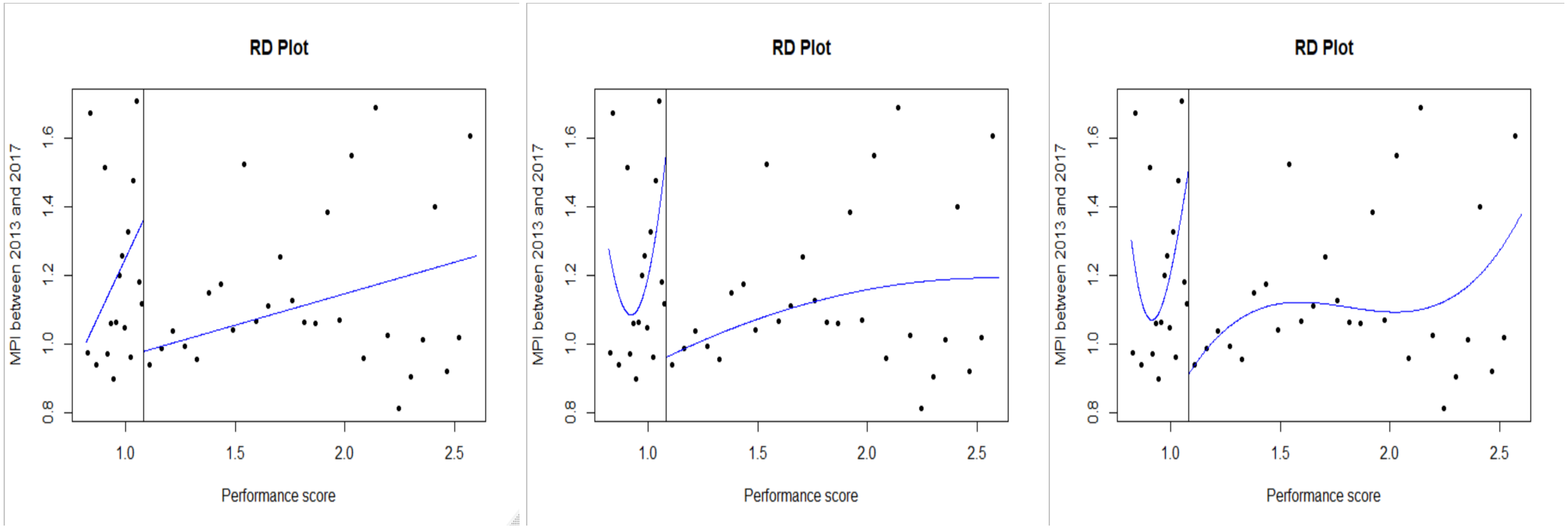


Figure 4b. Regression discontinuity plots for different orders of polynomials (for efficiency change)

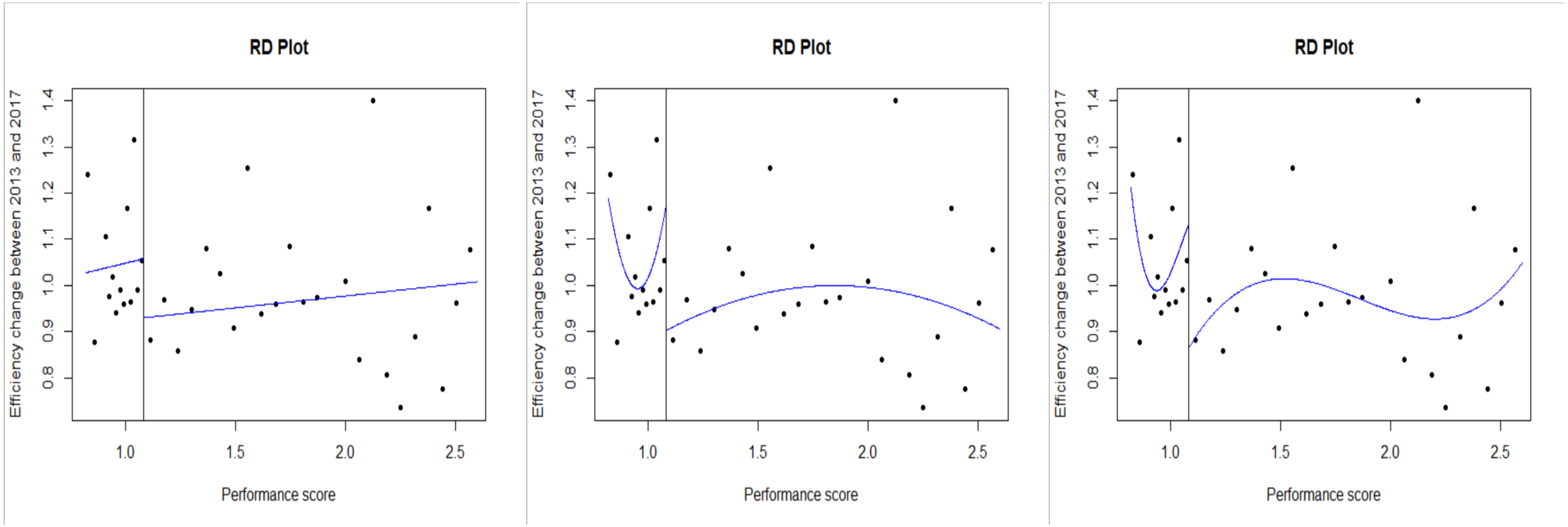


Figure 4c. Regression discontinuity plots for different orders of polynomials (for frontier shift)

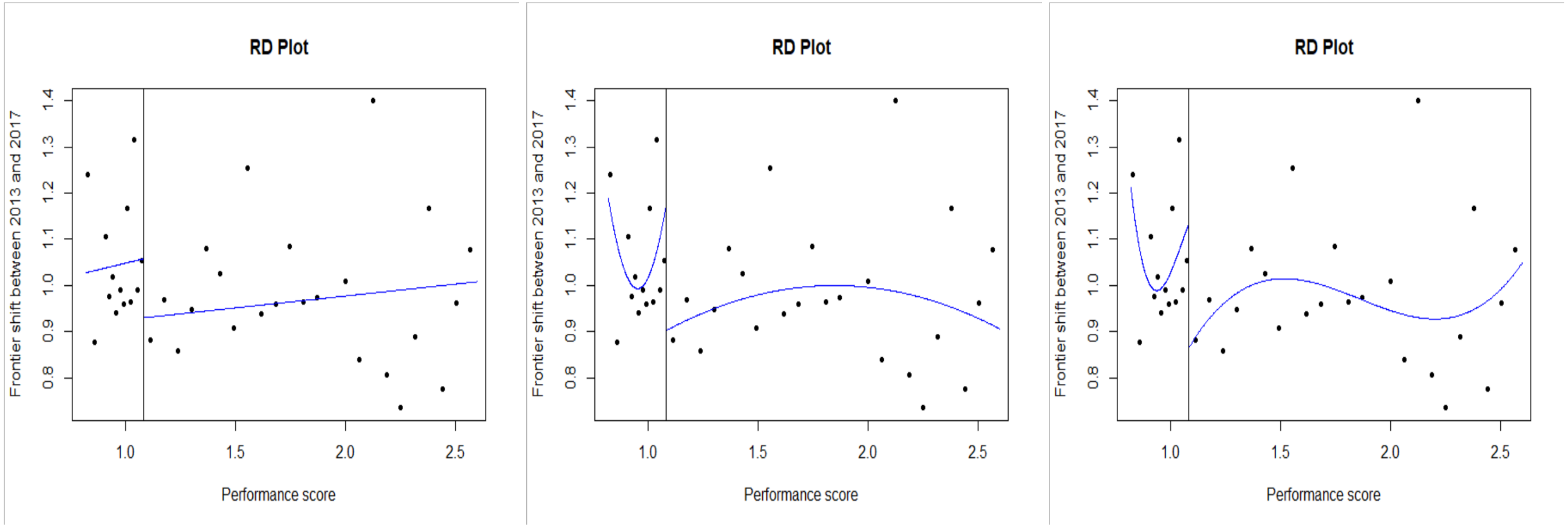


Figure 4 presents the regression discontinuity plots for different-order polynomials. Figure 4a presents regression discontinuity plot for MPI, Figure 4b for the efficiency change and Figure 4c for the frontier shift between 2013 and 2017. Each graph shows a clear and statistically significant point of discontinuity around the cutoff point for performance scores. Allowing for second and third order polynomials makes the gap between productivity change of universities that are just above and just below the discontinuity point more statistically evident compared to the linear regression, and in general statistically significant as demonstrated in Table 5.

7. Discussion and concluding remarks

In this paper FRDD is performed in order to explore how one particular type of merger in the Russian HE system based on MoP causally influence the efficiency level of the universities involved. The results of the analysis show that merged universities experienced greater efficiency gains (or rather, smaller efficiency declines) after the merger. The identification of the significant impact of mergers on efficiency levels presented in the paper is a first step towards understanding of how mergers in HE may influence efficiency levels. This study is the first attempt to analyze quantitatively the impact of the merger policy in Russian HE on the efficiency levels of merged universities. This paper also presents the first attempt to utilize RDD to analyze the relationship between mergers and the efficiency levels of HEI.

We highlight several potential channels through which merger policy may affect efficiency – such as economies of scale, economies of scope, and changes in managerial practices (as suggested by Papadimitriou and Johnes, 2018). Our methodological strategy suggests that the treatment effect identified using FRDD captures all the effects occurring through all possible channels and can be interpreted as the total effect of mergers on efficiency.

However, the research design of this study has several limitations which can be considered as possible directions for further investigation. First, it is important to disentangle the total effect into the different channels highlighted above. Second, we evaluate the effect of merger only on efficiency and productivity. We do not estimate how mergers impact different university stakeholders such as students, professors and administrative staff. We also do not study qualitative changes in different universities activities (teaching, research, etc.) which occur due to the merger. Therefore, we cannot claim whether the policy leads to positive changes in general. We observe the positive effect of mergers only on efficiency, but there could be some negative consequences in the other aspects of university performance. Another important limitation of this study which requires additional investigation is how the efficiency dynamics of merged universities evolve over time. In this study we analyze short-term policy effects. Due to data availability constraints, we cannot measure the long-term effects as the analyzed mergers took place in 2013-2015. Some empirical papers discussed in literature review section show that the positive effect of merger on efficiency level can be observed only just after the merger and disappear in subsequent years (Johnes and Tsionas, 2018; Papadimitriou and Johnes, 2018), therefore it is necessary to check the hypothesis about effect of merger on efficiency in the long term.

Despite these limitations, the results of the analysis suggest that the merger policy can be considered as a good policy instrument for improving university performance. This instrument may be especially relevant for HE systems which consist of relatively small universities or highly specialized universities where economies of scale and scope effects can be strong. These results can be used by policymakers in order to justify merger policy in HE, however, in order to use these findings as a basis for public policy in HE it is necessary to keep in mind the limitations mentioned above.

Acknowledgments

We are very grateful for the fruitful discussion and valuable comments on the previous version of this paper at the following conferences and workshops: 7th International Workshop on Efficiency in Education, Health and other Public Services (Barcelona, 2019), 16th European Workshop on Efficiency and Productivity Analysis (London, 2019), LEER Conference on Education Economics (Leuven, 2019). 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).

References

Aagaard, K., Hansen, H. F., & Rasmussen, J. G. (2016). Mergers in Danish higher education: An overview over the changing landscape. In *Mergers in higher education* (pp. 73-88). Springer, Cham.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2014). Investigating returns to scope of research fields in universities. Higher Education, 68(1), 69-85.

Agasisti, T. (2017). Management of Higher Education Institutions and the Evaluation of their Efficiency and Performance. *Tertiary Education and Management*, 23(3), 187-190.

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., & 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.

Agasisti, T., Shibanova, E., Platonova, D., & Lisyutkin, M. (2018). The Russian Excellence Initiative For Higher Education: An Econometric Evaluation Of Short-Term Results. Higher School of Economics Research Paper No. WP BRP, 201.

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

Ashenfelter, O., & Card, D. (2001). Did the Elimination of Mandatory Retirement Affect Faculty Retirement Flows? (No. w8378). *National Bureau of Economic Research*.

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.

Bonaccorsi, A. (2007). Explaining poor performance of European science: institutions versus policies. *Science and Public Policy*, 34(5), 303-316.

Cai, Y., Pinheiro, R., Geschwind, L., & Aarrevaara, T. (2016). Towards a novel conceptual framework for understanding mergers in higher education. *European Journal of Higher Education*, 6(1), 7-24.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, 123(1), 177-218.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444.

Clark, B. R. (1998). Creating Entrepreneurial Universities: Organizational Pathways of Transformation. Issues in Higher Education. New York, NY

Eastman, J., & Lang, D. (2001). *Mergers in higher education: Lessons from theory and experience*. University of Toronto Press.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American economic review*, 66-83.

Fielden, J., & Markham, L. (1997). Learning lessons from mergers in higher education. *Commonwealth Higher Education Management Service*.

Filippini, M., & Lepori, B. (2007). Cost structure, economies of capacity utilization and scope in Swiss higher education institutions. Universities and strategic knowledge creation: *Specialization and performance in Europe*, 272-304.

Goedegebuure, L. C. (1994). Mergers in higher education: A comparative perspective.

Goedegebuure, L., Lysons, A. & Meek, V.L. (1993). Diversity in Australian Higher Education? *Higher Education*, 25, 395–410

Harman, G. (1991). Institutional amalgamations and abolition of the binary system in Australia under John Dawkins. *Higher Education Quarterly*, *45*(2), 176-198.

Harman, G. (2000). Institutional Mergers in Australian Higher Education, *Higher Education Quarterly*, 54(4), 343–366.

Harman, G., & Harman, K. (2003). Institutional mergers in higher education: Lessons from international experience. *Tertiary Education and management*, 9(1), 29-44.

Harman, K., & Meek, V. L. (2002). Introduction to special issue:``Merger revisited: international perspectives on mergers in higher education". *Higher Education*, 44(1), 1-4.

Hashimoto, K., & Cohn, E. (1997). Economies of scale and scope in Japanese private universities. *Education Economics*, *5*(2), 107-115.

Hay, H. R., Hay, J. F., & Fourie, M. (2001). Are institutional combinations, mergers or amalgamation the answer? An investigation into staff perceptions. *South African Journal of Higher Education*, 15(1), 100-108.

Heckman, J. (1997). Instrumental variables. *Journal of human resources*, 32(3), 441-462.

Hoxby, C. M. (1997). How the changing market structure of US higher education explains college tuition (No. w6323). *National Bureau of Economic Research*.

Hu, Y., & Liang, W. (2008). Malmquist index analysis of the dynamic changes in scientific research productivity of some Chinese universities before and after merger. *Frontiers of Education in China*, *3*(3), 429-447.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.

Imbens, G., & Kalyanaraman, K. (2009). Optimal bandwidth choice for the regression.

Johnes, J. (2006). Measuring efficiency: A comparison of multilevel modelling and data envelopment analysis in the context of higher education. Bulletin of Economic Research, 58(2), 75-104.

Johnes, J. (2014). Efficiency and Mergers in English Higher Education 1996/97 to 2008/9: Parametric and Non-parametric Estimation of the Multi-input Multi-output Distance Function. *The Manchester School*, *82*(4), 465-487.

Johnes, J., & Tsionas, M. G. (2018). Dynamics of Inefficiency and Merger in English Higher Education From 1996/97 to 2008/9: A Comparison of Pre-Merging, Post-Merging and Non-Merging Universities Using Bayesian Methods. *The Manchester School*.

Kneip, A., Park, B. U., & Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric theory*, 14(6), 783-793.

Koshal, R. K., & Koshal, M. (1999). Economies of scale and scope in higher education: a case of comprehensive universities. *Economics of education review*, *18*(2), 269-277.

Kyvik, S. (2004). Structural changes in higher education systems in Western Europe. *Higher education in Europe*, 29(3), 393-409.

Locke, W. (2007). Higher education mergers: Integrating organisational cultures and developing appropriate management styles. *Higher Education Quarterly*, 61(1), 83-102.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2), 698-714.

Meek, V. L. (2002). On the road to mediocrity? Governance and management of Australian higher education in the market place. In Governing higher education: *National perspectives on institutional governance* (pp. 235-260). Springer, Dordrecht.

Morgan, A. (2000). Reform in Hungarian Higher Education, *International Higher Education*, 19, 23-24.

Murillo-Zamorano, L. R. (2004). Economic efficiency and frontier techniques. *Journal of Economic surveys*, 18(1), 33-77.

Olivares, M., & Wetzel, H. (2014). Competing in the higher education market: empirical evidence for economies of scale and scope in German higher education institutions. *CESifo Economic Studies*, 60(4), 653-680.

Papadimitriou, M., & Johnes, J. (2018). Does merging improve efficiency? A study of English universities. *Studies in Higher Education*, 1-21.

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.

Rantz, R. (2002). Leading urban institutions of higher education in the new millennium. *Leadership & organization development journal*, 23(8), 456-466.

Romanenko, K. R., & Lisyutkin, M. A. (2018). University mergers in Russia: Four waves of educational policy. *Russian Education & Society*, *60*(1), 58-73.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rothschild, M., & White, L. J. (1995). The analytics of the pricing of higher education and other services in which the customers are inputs. *Journal of Political Economy*, 103(3), 573-586.

Rowley, G. (1997). Mergers in higher education: A strategic analysis. *Higher Education Quarterly*, 51(3), 251-263.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.

Sehoole, M. T. C. (2005). The politics of mergers in higher education in South Africa. *Higher Education*, *50*(1), 159-179.

Shattock, M. (2010). Managing successful universities. McGraw-Hill Education (UK).

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, 44(1), 49-61.

Skodvin, O. J. (1999). Mergers in higher education-success or failure? *Tertiary Education & Management*, 5(1), 65-80.

Sokolov, M., Tsivinskaya, A. (2018). Monitoring the effectiveness of educational organizations as a source of data on Russian higher education (*HHC-S (1) series of preprints*). St. Petersburg: CIANO EUSPb.

Steinberg, M. P. (2014). Does greater autonomy improve school performance? Evidence from a regression discontinuity analysis in Chicago. *Education Finance and Policy*, 9(1), 1-35.

Välimaa, J., Aittola, H., & Ursin, J. (2014). University mergers in Finland: Mediating global competition. *New Directions for Higher Education*, *2014*(168), 41-53.

Wan, Y. (2008). Managing Post-Merger Integration: A Case Study of a Merger in Chinese Higher Education.

Wang, X. (2001). A policy analysis of the financing of higher education in China: Two decades reviewed. *Journal of Higher Education Policy and Management*, *23*(2), 205-217.

Weber, R. A., & Camerer, C. F. (2003). Cultural conflict and merger failure: An experimental approach. *Management science*, 49(4), 400-415.

Weber, Y., & Schweiger, D. M. (1992). Top management culture conflict in mergers and acquisitions: A lesson from anthropology. *International Journal of Conflict Management*, 3(4), 285-302.

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

Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education*, 61(4), 387-414.

Zhao, J., & Guo, J. (2002). The restructuring of China's higher education: An experience for market economy and knowledge economy. *Educational Philosophy and Theory*, *34*(2), 207-221.

Zong, X., & Zhang, W. (2017). Establishing world-class universities in China: deploying a quasi-experimental design to evaluate the net effects of Project 985. *Studies in Higher Education*, 1-15.

Annex

Annex 1. Identification of the discontinuity point.

Table 1. R^2 for the regression equations used for discontinuity point identification.

θ	Equation (11)	Equation (13)	Equation (14)	Equation (15)	Equation (16)
0.80	0.000	0.263	0.356	0.263	0.356
0.81	0.000	0.263	0.356	0.263	0.356
0.82	0.018	0.269	0.367	0.269	0.367
0.83	0.018	0.269	0.367	0.269	0.367
0.84	0.036	0.276	0.382	0.276	0.382
0.85	0.036	0.276	0.382	0.276	0.382
0.86	0.036	0.276	0.382	0.276	0.382
0.87	0.055	0.284	0.388	0.284	0.388
0.88	0.055	0.284	0.388	0.284	0.388
0.89	0.055	0.284	0.388	0.284	0.388
0.90	0.055	0.284	0.388	0.284	0.388
0.91	0.073	0.292	0.396	0.292	0.396
0.92	0.092	0.301	0.405	0.301	0.405
0.93	0.131	0.320	0.439	0.320	0.439
0.94	0.191	0.352	0.478	0.352	0.478
0.95	0.211	0.363	0.481	0.363	0.481
0.96	0.253	0.387	0.494	0.387	0.494
0.97	0.296	0.413	0.504	0.413	0.504
0.98	0.319	0.427	0.516	0.427	0.516
0.99	0.364	0.456	0.535	0.456	0.535
1.00	0.364	0.456	0.535	0.456	0.535
1.01	0.410	0.488	0.580	0.488	0.580
1.02	0.483	0.541	0.623	0.541	0.623
1.03	0.585	0.619	0.690	0.619	0.690
1.04	0.638	0.663	0.722	0.663	0.722
1.05	0.722	0.734	0.779	0.734	0.779
1.06	0.810	0.814	0.850	0.814	0.850
1.07	**0.871**	**0.872**	**0.896**	**0.872**	**0.896**
1.08	**0.871**	**0.872**	**0.896**	**0.872**	**0.896**
1.09	0.736	0.737	0.753	0.737	0.753
1.10	0.713	0.713	0.731	0.713	0.731
1.11	0.609	0.611	0.651	0.611	0.651
1.12	0.603	0.604	0.640	0.604	0.640
1.13	0.569	0.570	0.612	0.570	0.612
1.14	0.507	0.510	0.564	0.510	0.564
1.15	0.429	0.435	0.505	0.435	0.505
1.16	0.394	0.403	0.477	0.403	0.477
1.17	0.373	0.384	0.466	0.384	0.466
1.18	0.363	0.375	0.460	0.375	0.460
1.19	0.344	0.357	0.450	0.357	0.450
1.20	0.308	0.327	0.423	0.327	0.423

Source: *authors' calculations based on the data MoP*

Annex 2. Efficiency scores robustness check

In order to check the robustness of the efficiency scores that are subsequently used for regression discontinuity analysis, we made two checks – robustness check of efficiency scores when changing the methodology of efficiency estimation and when changing the lists of inputs and outputs.

Firstly, we checked that efficiency scores are robust when changing the method of efficiency estimation. Along with baseline efficiency estimation technique – robust DEA, efficiency estimates based on the stochastic frontier analysis (SFA) with Cobb-Douglas and translog specifications were considered, as well as efficiency scores obtained using quantile regression.

In case of SFA with Cobb-Douglas specification we use the output distance function of the following form:

$$ln(1/rnd_i) = ln(total\_pub_i/rnd_i) + ln(students_i/rnd_i) + ln(use_i) + ln(faculty_i) - u_i + \epsilon_i$$

where rnd_i – total volume of R&D projects (normalizing output); $total\_pub_i$ − total number of publications; $students_i$ − total number of students; use_i − average entrance exam score; $faculty_i$ − total number of faculty; u_i − inefficiency term; ϵ_i − random error term.

In case of translog SFA specification we used output distance function that contains the same list of variables as SFA with Cobb-Douglas specification together with appropriate interaction terms.

Finally, we considered the quantile regression methodology for efficiency estimation, the main idea of which is that quantile regression line for reasonably high quantile represents the production frontier. Particularly, quantile regression for the quantile 0.9 was estimated. Universities' i efficiency in QR setting is then defined as follows:

$$EFF_i = \frac{exp[\hat{f}(p_i, y_i)exp[(ln\widehat{u_{min}})]}{exp[\hat{f}(p_i, y_i)exp[(ln\widehat{u_i})]} = exp[(ln\widehat{u_{min}} - ln\widehat{u_i})]$$

where $ln\widehat{u_i}$ is a residual vector averaged over time; $n\widehat{u_{min}}$- the most efficient university in the sample.

The results of efficiency estimation using methodologies described above are presented in the table 2.

Table 2. Mean values of efficiency scores based on different estimation techniques.

	Robust DEA	SFA Cobb-Douglas	SFA translog	QR
2013	0.781	0.817	0.851	0.817
	(0.102)	*(0.110)*	*(0.088)*	*(0.149)*

2017	0.722 *(0.121)*	0.795 *(0.157)*	0.841 *(0.096)*	0.730 *(0.252)*
Change 2013-2017, %	-7.6	-2.7	-1.2	-10.6

Note: *standard deviations of efficiency scores are presented inside the brackets.*
Source: *authors' calculations based on the data from MoP*

Presented mean values demonstrate that average university included in the sample is characterized by rather high value of efficiency ranging from 0.72 to 0.85 depending on estimation technique and year. However, universities in the sample experienced efficiency decline between 2013 and 2017. Tables 3 and 4 present the correlation coefficients between efficiency scores obtained based on different methodologies.

Table 3. Correlation matrix for coefficients obtained based on different estimation technique (data for 2013)

	Robust DEA	**SFA Cobb-Douglas**	**SFA translog**	**QR**
Robust DEA	1	0.71	0.75	0.74
SFA Cobb-Douglas	0.71	1	0.76	0.97
SFA translog	0.75	0.76	1	0.72
QR	0.74	0.97	0.72	1

Source: *authors' calculations based on the data from MoP*

Table 4. Correlation matrix for coefficients obtained based on different estimation technique (data for 2017)

	Robust DEA	**SFA Cobb-Douglas**	**SFA translog**	**QR**
Robust DEA	1	0.77	0.81	0.79
SFA Cobb-Douglas	0.77	1	0.88	0.95
SFA translog	0.81	0.88	1	0.78
QR	0.79	0.95	0.78	1

Source: *authors' calculations based on the data from MoP*

For both years we observe rather high values of correlation coefficients, so all considered methodologies give consistent results.

The second implemented robustness check is related to the specification of efficiency model. We have considered three alternative lists of inputs and outputs. Considered specifications are described in the table 5.

Table 5. Alternative specifications of efficiency model

	Inputs	Outputs
Model 1 (baseline model)	○ Total income from all sources ○ Total number of faculty ○ Average USE score	○ Total number of students ○ Total number of publications ○ Total volume of R&D projects
Model 2	○ Total income from all sources ○ Total number of faculty ○ Average USE score ○ Total square of buildings used for teaching and research activities	○ Total number of students ○ Total number of publications ○ Total volume of R&D projects
Model 3	○ Total income from all sources ○ Total number of faculty ○ Average USE score ○ Share of faculty with advanced degrees	○ Total number of students ○ Total number of publications ○ Total volume of R&D projects
Model 4	○ Total income from all sources ○ Total number of faculty ○ Average USE score	○ Total number of students ○ Total number of publications ○ Total volume of R&D projects (only private)

Using these specifications we have estimated efficiency scores using baseline methodology – robust DEA. The average efficiency scores obtained based on different specifications are presented in the table 6.

Table 6. Mean values of efficiency scores based on different specifications.

	Model 1	Model 2	Model 3	Model 4
2013	0.781 *(0.102)*	0.788 *(0.090)*	0.791 *(0.090)*	0.788 *(0.088)*
2017	0.722 *(0.121)*	0.787 *(0.119)*	0.784 *(0.119)*	0.788 *(0.119)*
Change 2013-2017, %	92.4	99.8	99.1	99.8

Note: *standard deviations of efficiency scores are presented inside the brackets.*
Source: *authors' calculations based on the data from MoP*

Results presented in table 6 demonstrate that all specifications give approximately the same mean values. As in case of previous robustness check, all specifications show the decline of efficiency during considered period. The correlations coefficients between efficiency scores obtained based on different specifications are presented in the tables 7 and 8.

Table 7. Correlation matrix for coefficients obtained based on different model's specifications (data for 2013)

	Model 1	Model 2	Model 3	Model 4
Model 1	1	0.83	0.85	0.84
Model 2	0.83	1	0.98	0.97
Model 3	0.85	0.98	1	0.96
Model 4	0.84	0.97	0.96	0.99

Source: *authors' calculations based on the data from MoP*

Table 8. Correlation matrix for coefficients obtained based on different model's specifications (data for 2017)

	Model 1	Model 2	Model 3	Model 4
Model 1	1	0.94	0.93	0.93
Model 2	0.94	1	0.98	0.97
Model 3	0.93	0.98	1	0.99
Model 4	0.93	0.97	0.99	1

Source: *authors' calculations based on the data from MoP*

In case of this robustness check we also observe high values of correlation coefficients between different efficiency scores. In overall, we can conclude that the results of efficiency estimation are rather robust both when changing methodology of estimation and specification of the model. Thus, the particular choice efficiency estimation' methodology, as well as the lists of inputs and outputs should not affect substantially the results of regression discontinuity analysis.

Annex 3. Non-parametric regression discontinuity design

Table 9a. The results of the non-parametric RDD estimations for MPI

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-parametric estimation (local linear)							
Treatment effect	0.731*	0.672*	0.544*	0.539.	0.234	0.745**	0.572*
(standard error)	*(0.289)*	*(0.278)*	*(0.256)*	*(0.309)*	*(0.256)*	*(0.278)*	*(0.257)*
Bandwidth	Optimal (0.201)	Optimal (0.201)	Optimal (0.201)	Optimal * 0.5 (0.101)	Optimal * 0.5 (0.101)	Optimal * 2 (0.402)	Optimal * 2 (0.402)
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	85	85	85	57	57	106	106

Table 9b. The results of the non-parametric RDD estimations for efficiency change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-parametric estimation (local linear)							
Treatment effect	0.340*	0.349*	0.318*	0.265*	0.156	0.325*	0.331*
(standard error)	*(0.133)*	*(0.141)*	*(0.137)*	*(0.129)*	*(0.098)*	*(0.129)*	*(0.146)*
Bandwidth	Optimal (0.201)	Optimal (0.201)	Optimal (0.201)	Optimal * 0.5 (0.101)	Optimal * 0.5 (0.101)	Optimal * 2 (0.402)	Optimal * 2 (0.402)
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	85	85	85	57	57	106	106

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-parametric estimation (local linear)							
Treatment effect	0.307*	0.227.	0.155	0.217.	0.089	0.331**	0.173
(standard error)	*(0.131)*	*(0.122)*	*(0.133)*	*(0.132)*	*(0.130)*	*(0.126)*	*(0.124)*
Bandwidth	Optimal (0.289)	Optimal (0.289)	Optimal (0.289)	Optimal * 0.5 (0.145)	Optimal * 0.5 (0.145)	Optimal * 2 (0.578)	Optimal * 2 (0.578)

Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	97	97	97	71	71	118	118

Table 9c. The results of the non-parametric RDD estimations for frontier shift

Annex 4. Estimations with different cut-off points

Table 10. RDD estimation for cut-off point 1.06 for MPI

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Parametric estimation (2SLS)							
Treatment effect	0.649**	0.572*	0.432*	0.641**	0.496**	0.606*	0.430*
(standard error)	*(0.243)*	*(0.222)*	(0.189)	*(0.240)*	*(0.182)*	*(0.274)*	*(0.197)*
Polynomial of performance score	First	First	First	Second	Second	Third	Third
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	152	152	152	152	152	152	152

Table 11. RDD estimation for cut-off point 1.08 for MPI

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Parametric estimation (2SLS)							
Treatment effect	0.886*	0.821*	0.683*	0.937*	0.816**	1.119.	0.943*
(standard error)	*(0.363)*	*(0.358)*	(0.314)	*(0.392)*	*(0.306)*	*(0.613)*	*(0.444)*
Polynomial of performance score	First	First	First	Second	Second	Third	Third
Controls	No	Yes	Yes	No	Yes	No	Yes
Interactions	No	No	Yes	No	Yes	No	Yes
# of observations	152	152	152	152	152	152	152

Annex 5. The history of mergers in Russian higher education sector

According to Russian Civil Code, the notion "merger" means the situation when one new higher education institution is formed by absorbing several existing universities or one or several universities are merged into already existing university. The active merger policy began in Russian higher education sector after the collapse of Soviet Union and still remains rather popular higher education system management tool. During the period from 1991 to 2017 four waves of mergers can be distinguished (Romanenko and Lisytkin, 2018).

The first wave of mergers took place in 1991-2000. Due to the transition from planned to market economy, during that period universities had to adopt to new economic conditions. Most soviet higher education institutions were highly specialized and prepared workforce for particular industries. The distribution of students across different disciplines was determined by the needs of planned economy. Most students studied the programs related to engineering and hard sciences. Liberalization in higher education sector after 1991 dramatically changed this situation: the demand determination shifted from the state to direct consumers of higher education programs – students. In conditions of new market economy the higher education programs in social sciences became the most demanded. Liberalization in higher education was followed by a sharp decrease in amount of public funds for higher education, and the response to the demand of students was the only strategy of survival. In order to meet new demand for higher education, universities started merging with each other in order to overcome the high level of specialization and meet the demand of students. It worth mentioning here that despite the high level of governance centralization in higher education in that period, universities mergers were not the result of special centralized policy. The whole subject of the development of these universities was the work of universities leaders who had the idea of creating mega-university (Romanenko and Lisytkin, 2018). To sum up, the first wave of mergers can be characterized as a shift from highly specialized higher education institutions to comprehensive universities in order to meet the demand of students and to survive in condition of decrease of public funding.

The second wave of mergers took place in 2006-2014 when the Ministry of Education and Science launched special program aimed at creating so-called Federal universities. The aim of this program was to modernize higher education sector and establish new, innovative universities. It was decided that these universities should be created by merging several universities located in one region. The first two Federal universities were established in 2006 by merging together 4 universities. In 2009 another 5 Federal universities were established in different Russian regions. The 8th university of this type was created in 2010, another two federal universities appeared in 2012 and 2014. In total, now there are 10 universities of this type.

The third wave of mergers was associated with the introduction of so-called “Annual Monitoring of Performance of Higher Education Institutions” in 2012. Using this tool Ministry of Education and Sciences gathered data on different aspects of universities activities (education, research, financial indicators and so on) and compared the observed values on each university with specially designed threshold. If the university demonstrated a low level of performance (lower than the threshold), it receives the status “university with the signs of inefficiency”. Universities with this status then were merged to other organizations. Monitoring is not only management tool, but the most comprehensive source of open data on the Russian universities and obligatory for all universities, conducted since 2012. During all five years, 910 universities participated in the Monitoring from 2013 to 2017, including 694 parent organizations and 216 branches (Sokolov and Tsivinskaya, 2018).

Finally, the forth wave of mergers that took place in 2014-2015 was related to the so-called “flagship universities” program. This program introduced by the ministry of Education and Science was aimed at establishing competitive universities in Russian regions (except Moscow and Saint-Petersburg) in order to slow the outflow of school graduates to regions with high concentration of universities that provide high-quality higher education programs (predominantly Moscow and Saint-Petersburg). Another aim of Flagship Universities program is to foster regional economic development through cooperation with industrial partners. Unlike the federal universities program and mergers based on the Monitoring of Performance, these mergers were voluntary. Flagship universities were established through an open competition among HEIs. In order to participate in this competition two or several universities should decide to merge by themselves and prepare a joint strategic plan of development. The universities with the best strategic plans (according to the results of expert assessment) received the status of “Flagship University” and additional funding for implementation of strategic plan. 22 universities were recognized to the winners and merged into 11 Flagship Universities. In 2015 another wave of Flagship universities program was launched, however, the requirement regarding merger was eliminated.

To sum up, during the last decades Russian higher education system experienced very different kinds of mergers. Thus, it is difficult to analyze all these mergers within one framework. In this study we concentrate on the third wave of mergers that are mergers based on the Monitoring of Performance.

Aleksei Egorov
Laboratory for University Development of Institute of Education, National Research University Higher School of Economics, Moscow – Russia, aegorov@hse.ru

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