EVALUATING THE EFFICIENCY OF THE RESEARCH SECTOR IN RUSSIAN REGIONS: A DYNAMIC DATA ENVELOPMENT ANALYSIS

T. R. Gareev¹ 🆔
I. Yu. Peker² 🆔
T. Yu. Kuznetsova² 🆔
N. A. Eliseeva³ 🆔

¹Skolkovo Institute of Science and Technology,
30/1 Bolshoy bulvar, Skolkovo, Moscow,121205, Russia

²Immanuel Kant Baltic Federal University,
14 A. Nevskogo St., Kaliningrad, 236041, Russia

³Kaliningrad State Technical University,
1 Sovetsky prospekt, Kaliningrad, 236022, Russia

The nonparametric method of dynamic data envelopment analysis (DDEA) has become increasingly popular for conducting comparative efficiency evaluations. In recent years, dynamic data envelopment analysis (DDEA), a variant of this method, has gained significant attention. This article applies dynamic analysis to evaluate the efficiency of the research sector in Russian regions. Traditional input variables such as the number of research staff and R&D expenditure are considered, while publication and patent metrics serve as output indicators. The analysis covers a substantial time period, spanning from 2009 to 2020. Notably, the proposed evaluation method incorporates publication quality measures as a carry-over variable, in addition to accumulated R&D expenditure. The study employs dynamic data envelopment analysis to compare the obtained results with previous evaluations of the research and technology sector in Russian regions. The findings demonstrate that the proposed method serves as a valuable ranking technique, enhancing existing evaluations of regions' research and technology potential in terms of efficiency. The article concludes by discussing the prospects and limitations of the method in evaluating and forecasting research and technology profiles of regions.

Keywords:
dynamic data envelopment analysis (DDEA), efficiency, ranking, research sector, region, panel data

Introduction

Data envelopment analysis (DEA) spans a wide range of nonparametric methods routinely used for *ranking* study objects according to their production efficiency [1—3]. A comprehensive review of modern types of DEA is the work by Chiang

Kao [4]. This widely used method continues to develop, which is especially true for its dynamic [5] and stochastic [6] applications. In Russian literature, the first studies into the use of DEA appeared at the turn of the 21st century [7—9]. Particularly, the method was employed in comparative efficiency assessments [10—17].

DEA builds on the universal notion of *efficiency* defined as the ratio between outputs and inputs for a certain object. Since it does not presuppose functional dependence between inputs and outputs, this method can be classified as non-parametric. The principal tool of DEA is optimisation. The essence of the method is that it solves as many optimisation problems as there are objects of observation. Moreover, the objective function reflects in one form or another the efficiency of objects, one by one, whilst the system of constraints includes data on all the objects [18]. All the objects are optimally assessed using available data, and this makes it possible to compare the objects, ranking them according to their capacity to transform resources into results.

A strength of DEA is the possibility to include in the analysis several outputs varying in scale and units of measurement. This makes the method universal: the selection of study objects (for instance, regions) determines the disciplinary area, and the choice of resources and results (for example, research) determines the subject area of the analysis.

This study aims to create a new type of ranking for regional research, using dynamic DEA as a comparative efficiency assessment method.

DEA has gained wide acceptance in regional studies as a well-established research method for the multi-factor comparison of regions [19]. It is also extensively used in assessing technological capabilities within selected industries [20; 21]. Following several earlier studies [22—25], we employ DEA to assess the efficiency of regions in terms of research and technology. The two latter publications are based on Russian regional data allows comparing the findings.

We view DEA as an approach to evaluate the efficiency of objects' performance, as we do regression analysis, the index method and stochastic production frontier analysis. The results of such evaluations, as a rule, underpin guidelines for increasing the efficiency of objects' performance and help identify inefficient objects of systems at macro-, meso- and micro-levels by building the so-called 'efficient frontier' determined by the performance of the most effective objects from the array. The method can also be employed in analysing the influence of various territory-specific factors that are at odds with the characteristics of the regional system. For instance, this can be done to assess the efficiency of regional research and education systems [26], the energy sector [27], environmental management systems [28], and innovative production [29]. DEA does not focus on processes within the system but aids in analysing inputs and outputs, as well as providing recommendations on optimizing them to increase the system's efficiency [30].

Using DEA to study the system of knowledge production offers helps make informed personnel and financial decisions at a regional level to ensure the effi-

ciency of the innovative system. As mentioned above, the usual result of employing DEA is an analysis of deviations from the efficient frontier in the performance of some objects and recommendations for improving their efficiency.

Following the target-setting method, Svetlana Ratner [31] proposes calculating projections for the inefficient objects of regional systems within the space of inputs and outputs on the efficiency frontier. This would make it possible to determine input reduction and output increase targets, thus ensuring the object's efficiency. Window analysis has been employed for the dynamic assessment of objects' efficiency. This method involves selecting an observation window of a certain length for each production facility, which ensures the robustness of efficiency assessment and reveals trends in efficiency. Olga Komarevtseva [32] uses DEA to assess the intensity of the link between the financial and social efficiency of municipalities in the Orel region. The methods can also be employed to assess the efficiency of organizations, using financial indicators of financial and economic performance rather than focusing solely on the amount of resources expended or outputs [33]. Aliya Alimkhanova and Artur Mitsel emphasise the applicability of DEA to assessing efficiency regardless of the combination of resources and rank objects according to efficiency. Therefore, inefficient objects can be identified and recommendations produced for improving their efficiency.

The method helps describe the states of efficiency and changes in it, which are not often predetermined by the amount of resources in regions. A great strength of DEA is that its results often challenge the stereotypes about how well regions can fare in research and technology. Yet, these results are to be taken with a grain of salt as they heavily depend on the quality of the data. For example, if regional statistics systematically misrepresent the amount of expenditure or personnel capacity in a region, the resultant assessment may be incorrect.

Overall, the methodological problems of DEA are a by-product of its advantages. Firstly, DEA is very sensitive to data quality, including outliers, sudden shifts, and other factors. Secondly, the traditional tools of statistical assessment of results are not applicable to the classical variation of DEA. However, the literature has described some approaches to address this problem (see [6]). Therefore, if it is necessary to take into account the randomness of data evaluation, alternative classes of models should be used such as stochastic frontier analysis (SFA; see [1; 34]).

The structure of this article takes the following form. It begins with a characteristic of the *dynamic* model employed, which builds on the ideas of Chen et al. developed in [22]. Then follows a description of data on research[1] conducted in Russian regions in 2009—2020, which are used along with publishing perfor-

---

[1] We use the term 'research' since we use as the input variables the number of publications and patent applications. We do not, however, evaluate economic inputs that can shed light on the condition of the fields of research and technology and/or innovations.

mance data, 2008 internal R&D expenses figures and patent statistics since 2007. Further, cut-off values are given that were used when forming the final sample of regions. The next section contains the main results of the calculations and offers their visualisation, comparing them to earlier data. Discussion and conclusions round off the article.

Dynamic DEA model

DEA is used to assess *efficiency* defined as the *ratio* between effects (results) to resources expended. Within DEA, efficiency assumes, as a rule, the form of *technical efficiency*, i.e., the ratio between the optimal amount of necessary resources per unit of output and resources actually expended (for more detail, see [1]).

When using DEA, the optimisation problem can be viewed as the primal or dual problem, i.e. based on an envelope or a multiplier. We use a *dynamic* model [18; 22], which, in turn, is a modification of the relational analysis model proposed by Kao [18], who adheres to the multiplier form. And we will do the same.

So, n objects of observation (regions) are given, for which panel data have been collected for T periods and three groups of variables: inputs Y_r, outputs X_i and carry-over (transitioning from one period to another), Z_p. In the context of dynamic DEA, particular attention should be paid to carry-over variables Z_p (Fig. 1) since, without them, the model would be a simple aggregate of T statical models.
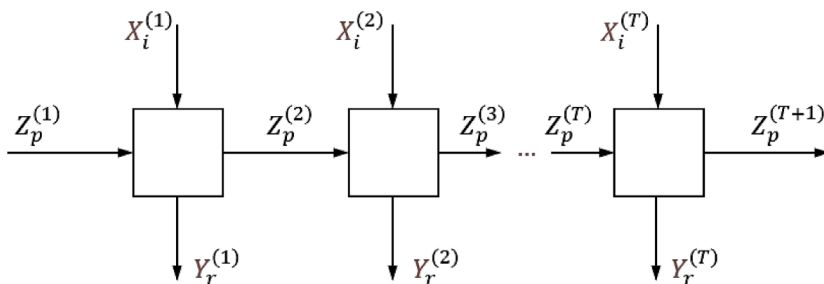

Fig. 1. The principal scheme of dynamic DEA (adopted from [22, p. 107])

The scheme considers one variable of each type for one object of observation. Extending this scheme to all the objects requires collating *panel* data, which will be described in the next section, carry-over variable should be collected for $T+1$ periods.

Unfortunately, matrix notation is not suitable for the description of this problem, and thus the cumbersome index notation is used. The problem of maximising individual efficiency can be expressed as follows:

$$\max E_k = \frac{\sum_{t=1}^{T} \sum_{r=1}^{s} u_r Y_{rk}^{(t)} + \sum_{t=1}^{T} \sum_{p=1}^{q} w_p Z_{pk}^{(t+1)}}{\sum_{t=1}^{T} \sum_{i=1}^{m} v_i X_{ik}^{(t)} + \sum_{t=1}^{T} \sum_{p=1}^{q} w_p Z_{pk}^{(t)}}$$

$$s.t. \quad \frac{\sum_{t=1}^{T} \sum_{r=1}^{s} u_r Y_{rj}^{(t)} + \sum_{t=1}^{T} \sum_{p=1}^{q} w_p Z_{pj}^{(t+1)}}{\sum_{t=1}^{T} \sum_{i=1}^{m} v_i X_{ij}^{(t)} + \sum_{t=1}^{T} \sum_{p=1}^{q} w_p Z_{pj}^{(t)}} \leq 1, \quad j = 1,2, \dots, n$$

$$\frac{\sum_{r=1}^{s} u_r Y_{rj}^{(t)} + \sum_{p=1}^{q} w_p Z_{pj}^{(t+1)}}{\sum_{i=1}^{m} v_i X_{ij}^{(t)} + \sum_{p=1}^{q} w_p Z_{pj}^{(t)}} \leq 1, \quad j = 1,2, \dots, n; \quad t = 1,2, \dots, T$$

$$u_r, v_i, w_p \geq \varepsilon.$$

Optimisation occurs over positive multipliers u_r, v_i, w_p, with ε being some small positive value. Looking ahead, we will consider six multipliers: two per each groups of variables. Optimisation will result in six sets of multipliers. Multipliers do not change across periods.

As can be seen, the maximum value of the objective function (the first expression, E_k) is the indicator of interest, i.e. that of the efficiency of some object k from a set n. The expression of the objective function differs from the first group of constraints (n each) only in that the index j replaces the index k, where $j = 1, 2, \dots, n$. This means that we have to solve a *separate* problem form each k, with the restriction imposed that the efficiency of no object from the group can exceed one.[1]

Finally, there is the second group of $n \times T$ constraints. The expression for it differs from that for the previous one in that it lacks external summation over T (in our case, there will be 804 such constraints for each k). To ensure a more efficient numerical solution, the expressions in the model are modified in such a way as to remove cumbersome ratios (see [18, p. 327]). Since efficiency is a ratio, the denominator can be normalised to 1.

By expressing the numerators and denominators from the second group of constraints as $B = \sum_{r=1}^{S} u_r Y_{rj}^{(t)} + \sum_{p=1}^{q} w_p Z_{pj}^{(t+1)}$ and $A = \sum_{i=1}^{(m)} v_i X_{ij}^{(t)} + \sum_{p=1}^{q} w_p Z_{pj}^{(t)}$, a simplified scheme of the problem can be written as follows[2]:

---

[1] Here is a subtle point: depending on whether the constraint on k is included in the group of constraints, the models will differ slightly. If it is not, efficiency can be above 1 numerically, which is indicative of superefficiency (for more detail, see [1]). This article considers a version without superefficiency.

[2] The problem has two variants: import-oriented, as is the case here, and export-oriented. To distinguish between them in practice, one should pay attention to the range in which the result falls. If the results of the objective function is $0 < \theta \leq 1$, one is dealing with the import-oriented version; if $\varphi > 1$, export-oriented. Since, in the general case, it is required that $\varphi\theta = 1$, we will not consider this in detail [3].

$$\max \sum B$$
$$\sum A = 1$$
$$\sum B \leq \sum A , \quad j = 1, 2, \ldots, n$$
$$B \leq A, j = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T$$
$$u_r, v_i, w_p \geq \varepsilon.$$

In this representation, summation is conducted over T and t-th indices are omitted for simplicity. This simplified scheme is implemented using the Wolfram functional programming (Appendix).

Data

We collected data on traditional indicators for research. The data on Russian regions are for 2008—2021, with the first and last periods reserved for averaging the research indicators and creating carry-over variables. Therefore, $T = 12$ in the model, i.e., it spans the period from 2009 to 2020.

As resource indicators, or inputs, we used statistical data found in the annual books *Regiony Rossii. Sotsialno-ekonomicheskie pokazateli* [Russian regions. Socioeconomic indicators]:[1]

1. The number of R&D personnel. This metric covers various categories of people employed in the field rather than researchers only, since such an approach is more suitable for efficiency assessment.

2. Internal R&D expenses. All the evaluations, which are based on the regional consumer price index, are normalised to 2010 values.

Resulting indicators (outputs):

1. The number of publications covers research works of all types from the years 2010—2021, indexed in Scopus [5]. It is shifted one period forward since, by convention, there is a time lag of one to three years between expenditure and publication [35].

2. The number of patent applications filed. The statistics on Russian regions are according to Rospatent.[2] The values were averaged over three years as this metric is characterised by sudden annual shifts. In contrast with publication metrics, here we assume that patent applications are filed during its preparation period.[3]

[1] Regiony Rossii. Sotsialno-ekonomicheskie pokazateli [Russian regions. Socioeconomic indicators], 2021, *Rossstat*, URL: https://rosstat.gov.ru/folder/210/document/13204 (accessed 02.09.2022).

[2] Godovye otchety [Annual reports], 2021, *Rospatent*, URL: https://rospatent.gov.ru/ru/about/reports) (accessed 02.09.2022).

[3] We consider patent applications rather than patents, as the preparation of the former is, as a rule, a laborious task regardless of whether they are granted. From the perspective of productive output of research, patent applications are a close substitute for research articles.

The Scopus database by Elsevier was chosen as a source of bibliometric data. Publications' metadata include the affiliation of all authors, making it possible to distribute the array of Russian publications by region. The data were aggregated for Russia's 78 regions and federal cities.[1]

Aggregation was carried out using the full counting method [36]: a publication co-authored by researchers from different regions gives each region a credit. Thus, it was necessary, firstly, to include publications from organisations with a validated Scopus profile into a regional set and, secondly, to search for publications from organisations with such a profile and add them to the corresponding regional sets.

Finally, as carry-over variables, the model used:

— accumulated expenses calculated by aggregating the shares of internal R&D expenses in 2008—2020;

— Scopus field-weighted citation impact (FWCI) of publications for 2009—2021. It is the ratio of citations to the expected worldwide average for the field, type and publication year.[2] A FWCI of 1 means that the citation impact of a work is exactly the global average. We smooth this metric for two consecutive periods.

It has been proposed in the literature to use variables associated with stocks, typically capital stocks, as carry-overs. Statistics of this nature, as is known, are difficult to collect when it comes to research and technology. However, period-wise accumulated internal expenses seem to be a good proxy indicator [22]. In our model, we capitalise 5 % of such R&D expenses annually[3] and use a 20 % depreciation rate.

Citation impact gives insight into one of the aspects of academic publishing quality. We assume that research is closely bound up with the accumulation of experience, which in turn has a bearing on the quality of findings. The accumulated number of publications is, on the one hand, a result of research and, on the other, an input resource to obtain new research results. One of the indicators of academic publishing quality is the FWCI. The indicator is normalised in such a way as to not require taking logarithms.

Since regions differ in scale, it seems reasonable to take *logarithms* of the main variables to obtain comparable time series. Although the final rankings are different between initial and logarithmic data, the logarithmic expression gives a smoother picture with a smaller spread in efficiency by year.

[1] This is a result of consolidating data on the Khanty-Mansi autonomous region and the Yamal-Nenets autonomous region, the Nenets autonomous region and the Arkhangelsk region,as well as the exclusion of the Jewish autonomous region, the Chukotka autonomous region, Crimea and Sevastopol due to a lack of data covering the study period.

[2] What is Field-weighted Citation Impact (FWCI)?, 2022, Elsevier, URL: https://service.elsevier.com/app/answers/detail/a_id/14894/supporthub/scopus/~/what-is-field-weighted-citation-impact-%28fwci%29%3F/ (accessed 02.09.2022).

[3] It is a gross simplification, but, since the available statistical database is not exhaustive, it is impossible to demonstrate changes for each indicator and each region.

Figure 2 shows smoothed indicators and their symbolic representations. As can be seen, it is impossible to make any conclusion by visual inspection alone. Moreover, there is more than one indicator for each period in the model. This calls for automatic evaluation methods, such as dynamic DEA; the next section will reveal the results of its application.
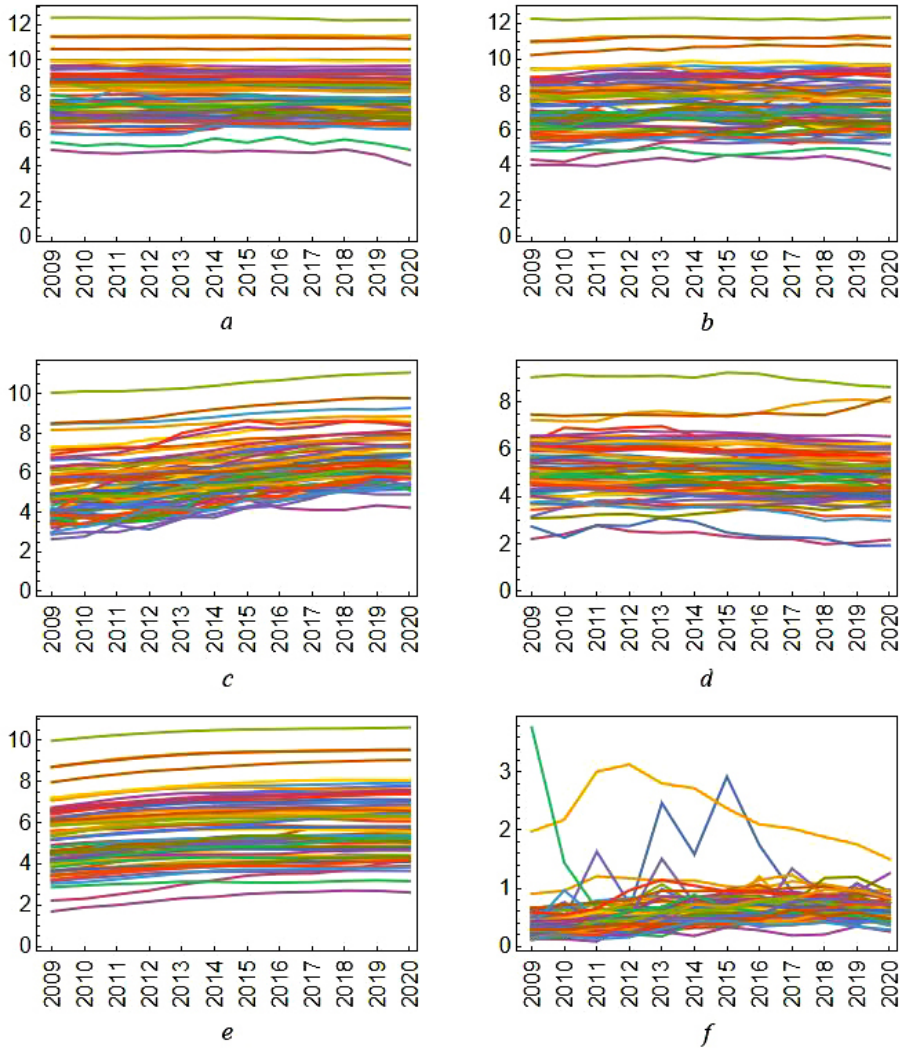
Fig. 2. Changes in model data: a — the logarithm of R&D personnel number, X_1; b — the logarithm of internal R&D expenses, X_2; c — the logarithm of Scopus-indexed publication number, Y_1; d — the logarithm of three-year averaged number of patents, Y_2; e — the logarithm of accumulated data at the end of a period, Z_1; f — two-year averaged FWCI at the end of a period, Z_2

Source: calculated by the authors based on data from Rosstat, Rospatent and Scopus.

Table 1 shows statistical characteristics of the data.

Main statistics based on the available data for the panel $n = 67$, $T = 12$, in logarithmic form, except for Z_2

Symbol	Mean	Standard deviation	Minimum	Quantile 0.25	Median value	Quantile 0.75	Maximum
X_1	7.98	1.44	4.04	6.87	7.77	8.89	12.39
X_2	7.54	1.59	3.82	6.38	7.36	8.57	12.32
Y_1	5.95	1.45	2.64	4.97	5.83	6.74	11.09
Y_2	5.08	1.16	1.92	4.30	5.05	5.78	9.25
Z_1	5.69	1.59	1.69	4.51	5.49	6.72	10.61
Z_2	0.61	0.35	0.08	0.41	0.57	0.72	3.77

Comment: lagged series were used for carry-over variables

For the final ranking, regions with pronounced deviations in the data were excluded from the sample (for example, those with very low publication or patent activity).[1] The final number of regions was $n = 67$.

Results

In basic form, the solution of the optimisation problem, using a Mathematic package, returns indicators of efficiency and corresponding overall ranking shown in Table 2. Columns 5 and 6 demonstrate, for reference purposes, ranking results obtained in earlier works [19; 25].

Results of efficiency assessment of regional research, $n = 67$

Region	Code	Our calculation		For reference	
		Efficiency in 2009—2022	Rank	RIS efficiency, 1998—2012	HSE's Russian Regional Innovation Ranking 2018
1	2	3	4	5	6
Ivanovo region	IVA	0.988	1	7	11
Kostroma region	KOS	0.974	2	3	71
Kemerovo region	KEM	0.967	3	20	22

---

[1] A region was excluded from analysis if, during a least one period, there were fewer than ten publications and five patent applications.

The continuation of the Table 2

Region	Code	Our calculation		For reference	
		Efficiency in 2009—2022	Rank	RIS efficiency, 1998—2012	HSE's Russian Regional Innovation Ranking 2018
1	2	3	4	5	6
Republic of Mariy El	ME	0.964	4	15	8
Novosibirsk region	NVS	0.964	5	10	10
Tomsk region	TOM	0.963	6	4	1
Belgorod region	BEL	0.960	7	43	19
Moscow	MOW	0.960	8	1	6
Lipetsk region	LIP	0.959	9	2	59
Arkhangelsk region	ARK	0.959	10	63	49
Republic of Tatarstan	TA	0.957	11	17	17
Krasnoyarsk Krai	KYA	0.957	12	39	15
St. Petersburg	SPE	0.953	13	5	4
Primorski Krai	PRI	0.953	14	53	23
Irkutsk region	IRK	0.953	15	52	21
Republic of Bashkortostan	BA	0.950	16	14	9
Sverdlovsk region	SVE	0.949	17	28	12
Volgograd region	VGG	0.946	18	44	67
Khabarovsk Krai	KHA	0.946	19	48	56
Saratov region	SAR	0.945	20	41	42
Moscow region	MOS	0.945	21	6	7
Rostov region	ROS	0.945	22	35	25
Republic of Dagestan	DA	0.942	23	33	60
Republic of Karelia	KR	0.942	24	62	30
Republic of Udmurtia	UD	0.941	25	31	63
Vologda region	VLG	0.941	26	12	36
Perm Krai	PER	0.941	27	16	18
Samara region	SAM	0.941	28	27	48
Tyumen region	TYU	0.940	29	23	3

Region	Code	Our calculation		For reference	
		Efficiency in 2009—2022	Rank	RIS efficiency, 1998—2012	HSE's Russian Regional Innovation Ranking 2018
1	2	3	4	5	6
Republic of Chuvashia	CU	0.940	30	30	61
Republic of North Ossetia-Alania	SE	0.939	31	22	76
Republic of Sakha (Yakutia)	SA	0.939	32	58	54
Republic of Buryatiya	BU	0.938	33	64	50
Orenburg region	ORE	0.937	34	36	52
Stavropol Krai	STA	0.937	35	47	58
Karachai-Cherkess Republic	KC	0.937	36	66	62
Nizhny Novgorod region	NIZ	0.937	37	19	5
Altai Krai	ALT	0.937	38	38	39
Chelyabinsk region	CHE	0.936	39	18	28
Kabardino-Balkarian republic	KB	0.936	40	56	66
Krasnodar Krai	KDA	0.936	41	9	47
Kursk region	KRS	0.936	42	26	51
Omsk region	OMS	0.935	43	37	16
Ulyanovsk region	ULY	0.935	44	11	2
Voronezh region	VOR	0.935	45	8	33
Amur region	AMU	0.934	46	50	73
Kaliningrad region	KGD	0.934	47	55	38
Murmansk region	MUR	0.934	48	65	53
Orel region	ORL	0.934	49	13	37
Leningrad region	LEN	0.933	50	45	57
Tula region	TUL	0.931	51	24	65
Republic of Komi	KO	0.931	52	61	20
Transbaikal Krai	ZAB	0.930	53	57	72
Astrakhan region	AST	0.930	54	51	77
Novgorod region	NGR	0.930	55	54	14

The end of Table 1

Region	Code	Our calculation		For reference	
		Efficiency in 2009—2022	Rank	RIS efficiency, 1998—2012	HSE's Russian Regional Innovation Ranking 2018
1	2	3	4	5	6
Tambov region	TAM	0.930	56	42	79
Smolensk region	SMO	0.929	57	60	34
Republic of Mordovia	MO	0.929	58	59	69
Bryansk region	BRY	0.929	59	49	32
Yaroslavl region	YAR	0.928	60	34	26
Penza region	PNZ	0.927	61	40	45
Tver region	TVE	0.924	62	32	27
Kaluga region	KLU	0.924	63	29	13
Kamchatka Krai	KAM	0.923	64	67	41
Vladimir region	VLA	0.923	65	46	31
Kurgan region	KGN	0.921	66	25	80
Ryazan region	RYA	0.921	67	21	46

Comment: in column 5, the last two ranks were added because of the difference in the samples; in column 6, there are gaps in the ranks since our study covers fewer regions.

As Table 2 shows, the results of ranking by efficiency differ greatly from each other and those in column 6, which demonstrated innovation ranking results (19, p. 38—39). This might point to the fact that the link between capacity for innovation and efficiency is rather tenuous: indeed, the former does not necessarily entail the latter, and vice versa.

Figure 3 gives a visual representation of period-wise changes in efficiency. Despite the absence of a steady trend, the efficiency was in the range of 0.9 to 1.0, which is very much in line with the original study [22].

As can be seen, for a large group of regions, the indicator slightly decreased in the first half of the study decade, when the policy to encourage publication activities was introduced [37]. Later, the values grew slowly. The barely visible deceleration towards the end of the study period may be a result of the saturation and exhaustion of opportunities for the extensive growth of publication activities.

One of the most recent publications assessing the efficiency of the Russian region's innovation systems is the contribution by Stepan Zemtsov and Maxim Kotsemir [25]. They examined the period between 1998 and 2012, which we consider only partially, and focused on assessing regional innovations and advances and technology, whilst we concentrate on research. These two areas, however, are closely linked and thus the findings of the two studies may prove comparable.
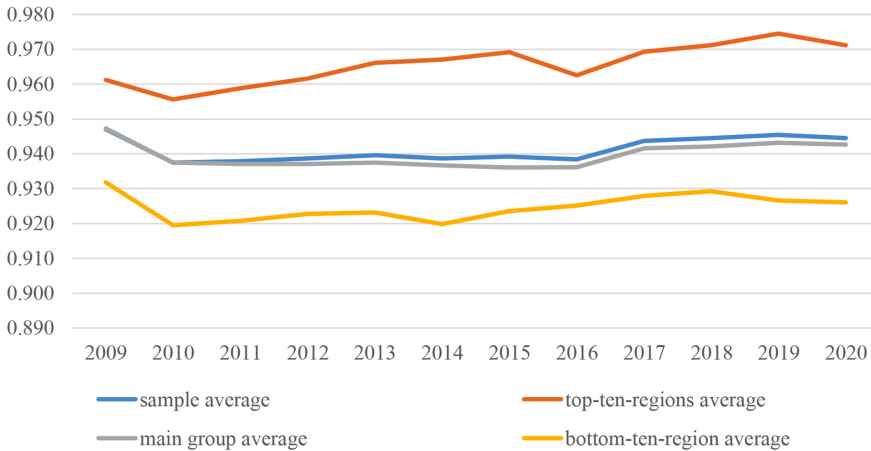

Fig. 3. Changes in the calculated indicators over the study period ($n = 67$)

Overall, there is a statistically significant but weak correlation of 0.39 with the findings of Zemtsov and Kotsemir [25], which may be due to both differences in data and the peculiarities of DEA.

It is worth noting that our calculation has a smaller spread in efficiency values between regions: the difference is 10 %, whilst, the cited work reports an over 20-fold difference between the best and worst-performing regions. Since we examine a narrower range of resources and only nonmonetary results, a slight difference in efficiency might be expected.

Another important difference is that, in the earlier work [25], Moscow is the leader by far, providing a benchmark for comparison. But, according to our calculations, Moscow is not the absolute leader in efficiency despite the considerable increase in publication activities observed over recent years. This decrease in Moscow's relative efficiency is probably a result of its patent activity declining in the second half of the decade,[1] as well as of the capital accounting for the overwhelming proportion of research funding.

[1] Here, we do not consider the criticism of the patent activity indicator, particularly its national version, since the use of this family of indicators is standard in assessing the research (and innovation) sector.

Discussion and conclusions

This study covers a relatively lengthy period, which is not common in the literature on dynamic DEA. As expected, the model is sensitive to the systematic underreporting of resource indicators at the data level. A prime example of this is the Ivanovo region, which typically scores low on resources but ranks high on research indicators.

As Figure 4 shows, the regions supersaturated with R&D financing have relatively low efficiency. These are the Ulyanovsk, Samara and Tyumen regions, and Krasnoyarsk Krai, well-resourced but far from excelling.
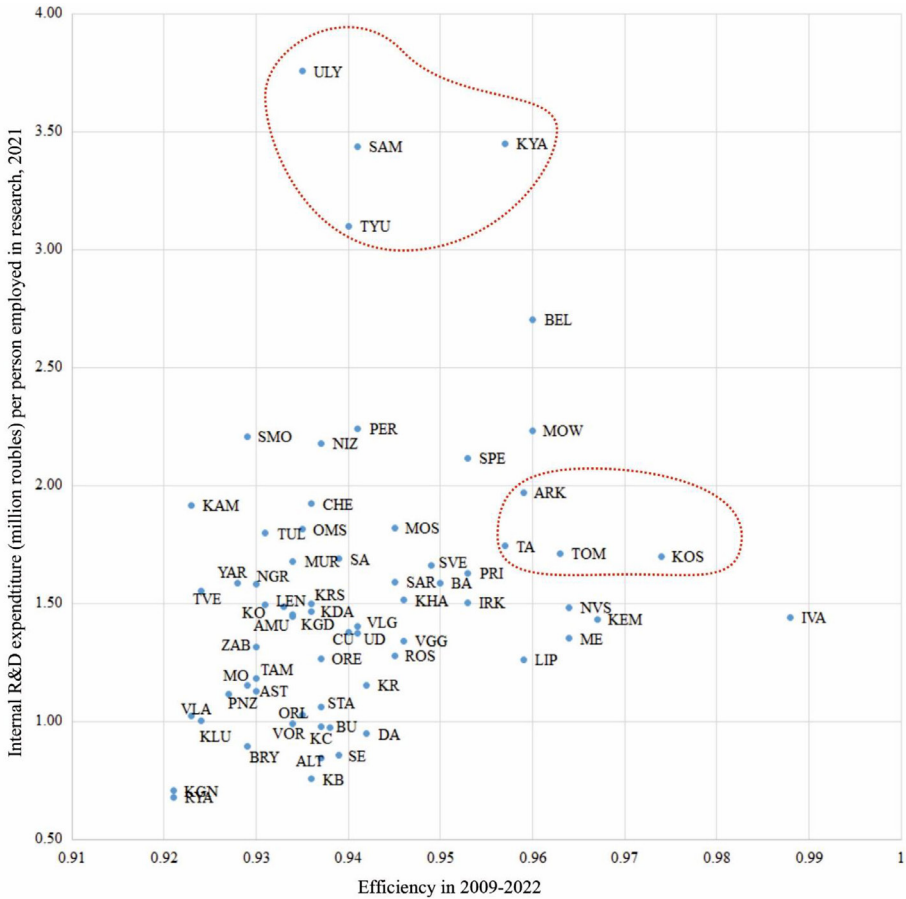

Fig. 4. Correlation between R&D expenses and efficiency in Russian regions

On the other hand, Tomsk and Novosibirsk are ranked as leaders despite relatively generous funding. According to our calculation, Moscow is not the ab-

solute leader, albeit ranked in top ten. This is yet another proof that efficiency is a *relative* indicator, and leadership in absolute values does not mean using the available resources to the fullest.

The opposite situation is also possible, especially when the resource allocation system overrelies on the 'achievement principle'. For example, in the dynamic model (unlike the one described in [24]), the Kaliningrad region is ranked rather low in terms of research efficiency, despite a relatively high effectiveness growth rate.[1] This can be explained by the effect of the border region in a unique geopolitical position receiving substantial research funding.

On average, including in the model carry-over variables increases the efficiency indicators [22]: the distribution of indicators is comparable to that reported in the original study using our own and Kao's methodology [18]. In Chen et al. [22], the overall efficiency is above 0.9 for all the study regions. In this study, period-wise results (the technical efficiency indicators) also vary between $0.9-1.0$.

In general, dynamic DEA yields tangible results and is convenient to use. But we identified the model's heightened sensitivity to expressing the indicators in the logarithmic form (see also [38]). It is also sensitive to changes in the expenses carry-over variable (in this case, it is a 5 % per cent share of internal R&D expense and 20 % depreciation). The model reacts less strongly to the quality variable of FWCI in the overall ranking (the citation impact mostly contributes to the accuracy of the efficiency indicator).

Dynamic DEA seems promising in assessing and comparing the efficiency of regional research. Yet, as in any other case, the ranking must be treated critically and the methodology supplemented with analytical statistics [19]. An important finding of this study is that capacity for innovation and efficiency may be weakly linked. Therefore, capacity-based rankings should be supplemented with efficiency assessments, including those using DEA.

In further research, we will consider the possibility of comparing and combining the results obtained by different methods, such as SFA or the complexity index, in order to rank and cluster regions, using the collected data.

Appendix

Surprisingly, software implementations of dynamic DEA are rare despite the wide selection of libraries modelling static DEA. An exception is the library created based on [5].

---

[1] Following the established practice, we distinguish between efficiency and effectiveness: the former describes the ratio between effects and costs and the latter between the actual and expected effects.

We propose a short code implementation in the Mathematica package, whose functional programming and built-in optimisation functions seem to be adequate for solving the problem.

The minimum code necessary for solving the problem is neither optimised nor suited for scaling: rather, its purpose is to give an accurate picture of a simplified version of the problem.

Having initial data, it is possible to derive several variables:

```
VARS = {u1, u2, v1, v2, w1, w2};
{U1, U2, V1, V2, W1, W2} = ConstantArray[#, {n, T}] & /@ VARS;
OUT = U1 * Y1 + U2 * Y2 + W1 * Z1T + W2 * Z2T;
INS = V1 * X1 + V2 * X2 + W1 * Z1 + W2 * Z2;
```

VARS contains a list of symbolic variables for optimisation. The other rows include symbols for matrices of size $n \times T$ filled with symbols and data. U1, U2, V1, V2, W1, W2 are filled with symbols u1, u2, v1, v2, w1, w2 respectively.

Y1, Y2, X1, X2, Z1, Z2, Z1T, Z2T are filled with numerical values (data).

When forming OUT and INS matrices, the * operator means the element-wise product, i. e., Hadamard rather than the matrix product.

The main procedure is as follows:

```
NMaximize[
    Flatten[{
    Part[ (Total /@ OUT), #],
    Part[ (Total /@ INS), #] == 1,
    Thread[ (Total /@ OUT) < = (Total /@INS)],
    Thread / @Thread[OUT < = INS],
    Thread[ VARS > 0 ]
    }],
    VARS] & /@Range[n]
```

Let us comment on this procedure. In Mathematica, all expressions are contained in lists held by curly brackets {}. For example, the VARS variable includes a list of symbols. The matrix is a list comprising several other lists (rows). Functions take brackets []. Elements in lists and functions are separated by commas.

The notation (Total /@ OUT) for the OUT matrix means summation on rows (in effect, we apply the Total function to each row of the matrix; this operation is represented by an operator of the /@ form. Since OUT has size $n \times T$, summation on rows is the summation over time. The same holds true for INS.

The Part[] function, as its name suggests, highlights a part of the list. For instance, the expression Part[{a, b, c}, 2] will return the second value of b from the list {a, b, c}.

The Thread[] function is auxiliary. An example of the way it works is the expression Thread[VARS > 0], which would return the list of inequalities:

$\{u1 > 0, u2 > 0, v1 > 0, v2 > 0, w1 > 0, w2 > 0\}$.

As its name implies, this function 'sews together' lists of symbols making them lists of expressions. The Flatten[] function removes unnecessary brackets and forms a list with an objective function and constraints, a list that can be processed with the numerical optimisation procedure NMaximize.

Finally, the principal procedure NMaximize[… # …]&/@ Range[n] carries out numerical maximisation with respect to numbers from 1 to n, each number inserted instead of the symbol #. The # symbol stands for a pure variable, always accompanied by the symbol &. In this case, the values substituted for # form Range[]. For example, Range[4] will return a list with the values {1, 2, 3, 4}. Again, the operator /@ means that these values have to be substituted *one by one* but not all at once. This way, *n* optimisation problems are consecutively solved.

The results obtained through this procedure can usually be saved and worked upon later, but we leave this part out for concision. For more detail, see [39].

References

1. Coelli, T., Rao, P., O'Donnel, C., Battese, G. 2005, An introduction to Efficiency and Productivity Analysis. Springer, NY, https://doi.org/10.1007/b136381.

2. Cook, W., Seiford, L. 2009, Data envelopment analysis (DEA) — thirty years on, *European Journal of Operational Research*, vol. 192, № 1, p. 1—17, https://doi.org/10.1016/j.ejor.2008.01.032.

3. Cooper, W. 2013, Data Envelopment Analysis. In: Gass, S. I., Fu, M. C. (eds.), *Encyclopedia of Operations Research and Management Science*, Springer, Boston, p. 349—358, https://doi.org/10.1007/978-1-4419-1153-7_212.

4. Kao, C. 2014, Network data envelopment analysis: A review, *European Journal of Operational Research*, vol. 239, № 1, p. 1—16, http://dx.doi.org/10.1016/j.ejor.2014.02.039.

5. Lee, D-J., Kim, M-S., Lee, K-W. 2022, A revised dynamic data envelopment analysis model with budget constraints, *International Transactions in Operational Research,* vol. 29, № 2, p. 1012—1024, https://doi.org/10.1111/itor.12810.

6. Olesen, O. B., Petersen, N. C. 2016, Stochastic Data Envelopment Analysis — A review, *European Journal of Operational Research*, vol. 251, № 1, p. 2—21, https://doi.org/10.1016/j.ejor.2015.07.058.

7. Krivonozhko, V. E., Utkin, O. B., Volodin, A. V., Sablin, I. A. 2002, Interpretation of modelling results in data envelopment analysis, *Managerial Finance,* vol. 28, № 9, p. 37—47, https://doi.org/doi:10.1108/03074350210768059.

8. Krivonozhko, V. E., Lychev, A. V. 2010, *Analysis of the activities of complex socio-economic systems*, M.: MAKS Press, 208 p.

9. Utkin, O. B. 2001, Apparatus for assessing the effectiveness of the functioning of commercial banks, *Banking services,* № 9, p. 16—19.

10. Gavrilova, A. A., Kolmykov, D. S., Alfeev, A. A. 2006, Multi-criteria evaluation of the efficiency of improved generating equipment of the power system, Journal *Vestnik of Samara State Technical University (Technical Sciences Series)*, № 40, p. 155—161.

11. Latov, A. V. 2003, Computer visualization of the production possibility set in data envelopment analysis, Doklady akademii nauk, vol. 388, № 2, p. 171—173.

12. Solov'ev, M. N., Pestrikov, S. V. 2008, Development of a mathematical model of comparative assessment of the effectiveness of Russia's regions, *Journal of Samara State Technical University, Ser. Physical and Mathematical Sciences,* № 1 (16), p. 175—177.

13. Reiner, G., Hofmann, P. 2012, Efficiency Analysis of Supply Chain Processes, *Russian Management Journal*, vol. 10, № 2, p. 89—116.

14. Fedotov, Y. V. 2012, Organizational Performance Measurement: Principles of DEA (Data Envelopment Analysis) Approach, *Russian Management Journal*, vol. 10, № 2, p. 51—62.

15. Matrizaev, B. D. 2019, Study of the comparative efficiency of the national innovation system and the quality of economic growth: on the example of a comparative analysis of OECD and BRICS countries, *Russian Journal of Innovation Economics*, vol. 9, № 3, p. 673—692, https://doi.org/10.18334/vinec.9.3.40880.

16. Sazonova, D. D., Sazonov, S. N. 2014, The assessment of technical efficiency of productive resources on farms, *Vestnik Cheljabinskoj gosudarstvennoj agroinzhenernoj akademii*, vol. 69, p. 117—125.

17. Fedorova, E. A., Korkmazova, B. K., Muratov, M. A. 2015, Assesing the Effectiveness of Joint Companies in Russia: Industry-Specific Features, *Prostranstvennaya Ekonomika = Spatial Economics,* № 2, p. 47—63, https://doi.org/10.14530/se.2015.2.047-063.

18. Kao, C. 2013, Dynamic data envelopment analysis: A relational analysis, *European Journal of Operational Research*, vol. 227, № 2, p. 325—330, http://dx.doi.org/10.1016/j.ejor.2012.12.012.

19. Gokhberg, L. M. (ed.). 2021, *Russian Regional Innovation Scoreboard.* Issue 7, Moscow: HSE University, 274 p.

20. Lu, W. 2009, The evolution of R&D efficiency and marketability: Evidence from Taiwan's IC-design Industry, *Asian Journal of Technology Innovation*, vol. 17, № 2, p. 1—26, https://doi.org/10.1080/19761597.2009.9668671.

21. Chiu, Y. H., Huang, C. W., Chen, Y. C. 2012, The R&D value-chain efficiency measurement for high-tech industries in China, *Asia Pacific Journal of Management*, vol. 29, p. 989—1006, https://doi.org/10.1007/s10490-010-9219-3.

22. Chen, K., Kou, M., Fu, X. 2018, Evaluation of multi-period regional R&D efficiency: An application of dynamic DEA to China's regional R&D systems, *Omega,* vol. 74, p. 103—114, https://doi.org/10.1016/j.omega.2017.01.010.

23. Belgin, O. 2019, Analysing R&D efficiency of Turkish regions using data envelopment analysis, *Technology Analysis & Strategic Management*, vol. 31, № 11, p. 1341—1352, https://doi.org/10.1080/09537325.2019.1613521.

24. Firsova, A. A., Chernyshova, G. Y. 2019, Mathematical models for evaluation of the higher education system functions with DEA Approach, *Izvestiya of Saratov University. New Series. Series: Mathematics. Mechanics. Informatics,* vol. 19, № 3, p. 351—362, https://doi.org/10.18500/1816-9791-2019-19-3-351-362.

25. Zemtsov, S., Kotsemir, M. 2019, An assessment of regional innovation system efficiency in Russia: the application of the DEA approach, *Scientometrics*, vol. 120, № 2, p. 375—404, https://doi.org/10.1007/s11192-019-03130-y.

26. Levites, D. G., Punantsev, A. A. 2021, Territorial educational segment as an object of assessment of the functioning conditions of regional education systems, *Perspektivy nauki i obrazovania — Perspectives of Science and Education,* № 4 (52), p. 577—593, https://doi.org/10.32744/pse.2021.4.38.

27. Khrustalev, E. Yu., Ratner, P. D. 2015, Analysis of ecological efficiency of Russia's electric companies through the data envelopment analysis methodology, *Economic Analysis: Theory and Practice*, № 35 (434), p. 33—42.

28. Ratner, S. V., Kovalev, A. V. 2021, Assessing the regional environmental management system's efficiency based on data envelopment models, *Economic Analysis: Theory and Practice*, vol. 20, № 6, p. 1014—1042, https://doi.org/10.24891/ea.20.6.1014.

29. Roslyakova, N. A. 2019, DEA tools for assess the role of labor productivity in the innovative output of the regions of the RF Northwestern federal district and Kazakhstan, *Bulletin of the South-Russian State Technical University (NPI) Series Socio-Economic Sciences,* № 6, p. 67—75, https://doi.org/10.17213/2075-2067-2019-6-67-75.

30. Pokushko, M., Stupina, A., Medina-Bulo, I., Dresvianskii, E. 2020, Data envelopment analysis in performance assessment of fuel and energy complex enterprises, *Bulletin of Kemerovo State University. series: political, sociological and economic sciences*, vol. 5, № 2 (16), p. 251—262, https://doi.org/10.21603/2500-3372-2020-5-2-251-262.

31. Ratner, S. 2017, Dynamic problems of estimation of ecological-economic efficiency of regions based on basic models of data envelopment analysis, *Large-Scale System Control,* № 67, p. 81—106.

32. Komarevtseva, O. O. 2016, Development of model of effective management of changes on the basis of DEA technique in economic systems of municipalities of the Oryol region, *Jekonomicheskij zhurnal*, № 3 (43), p. 41—58.

33. Alimkhanova, A. N., Mitsel, A. A, 2019, Evaluation of enterprise performance based on the DEA method, *Proceedings of TUSUR University,* vol. 22, № 2, p. 104—108, https://doi.org/10.21293/1818-0442-2019-22-2-104-108.

34. Greene, W. 2005, Fixed and Random Effects in Stochastic Frontier Models, *Journal of Productivity Analysis*, vol. 23, p. 7—32, https://doi.org/10.1007/s11123-004-8545-1.

35. Fritsch, M., Slavtchev, V. 2007, Universities and innovation in space, *Industry and Innovation*, vol. 14, № 2, p. 201—218, https://doi.org/10.1080/13662710701253466.

36. Sivertsen, G., Rousseau, R., Zhang, L. 2019, Measuring scientific contributions with modified fractional counting, *Journal of Informetrics*, vol. 13, № 2, p. 679—694, https://doi.org/ 10.1016/j.joi.2019.03.010.

37. Dezhina, I. G. 2017, Science and innovation policy of the Russian government: A variety of instruments with uncertain outcomes?, *Public Administration Issues*, № 5, p. 7—26, https://doi.org/10.17323/1999-5431-2017-0-5-7-26.

38. Seiford, L., Zhu, J. 1998, On piecewise loglinear frontiers and log efficiency measures, *Computers & Operations Research*, vol. 25, № 5, p. 389—395, https://doi.org/10.1016/S0305-0548(97)00078-6.

39. Saquib, N. 2014, Mathematica for Data Visualization, Packt Publishing.

The authors

Dr Timur R. Gareev, Skolkovo Institute of Science and Technology, Russia.

E-mail: tgareev@gmail.com

https://orcid.org/0000-0002-6438-7772

Dr Irina Yu. Peker, Immanuel Kant Baltic Federal University, Russia.

E-mail: ipeker@kantiana.ru

https://orcid.org/ 0000-0002-5701-7538

Dr Tatyana Yu. Kuznetsova, Immanuel Kant Baltic Federal University, Russia.

E-mail: TIKuznetsova@kantiana.ru

https://orcid.org/0000-0002-1523-2280

Dr Natalya A. Eliseeva, Kaliningrad State Technical University, Russia.

E-mail: ne2705@gmail.com

https://orcid.org/0000-0003-0651-6692