Bornmann, Lutz; Gralka, Sabine; de Moya Anegón, Félix; Wohlrabe, Klaus

Working Paper

Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers

CESifo Working Paper, No. 8157

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bornmann, Lutz; Gralka, Sabine; de Moya Anegón, Félix; Wohlrabe, Klaus (2020) : Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers, CESifo Working Paper, No. 8157, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/216553

CESifo WORKING PAPERS

8157
2020

March 2020

Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers

Lutz Bornmann, Sabine Gralka, Félix de Moya Anegón, Klaus Wohlrabe

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8157*

Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers

Abstract

One of the core indicators in the field of scientometrics is the number of papers published by a unit within a given period. However, such indicators can only be assessed properly by considering the unit's available resources. When evaluating the efficiency of institutions worldwide, the problem concerning the availability of internationally standardized data arises. While on the output side consistent publication indicators are available, these data are frequently not available on the input side. We therefore introduce a new input indicator based on the authors' mentions in the institutions' papers. We calculate efficiency scores for more than 4,800 universities and other research-focused institutions worldwide. "Harvard University" is the best performing institution (in all years) followed by many other institutions from Northern America or Europe. The results of the study show that institutions in the Pacific region have the highest average efficiency scores, followed by Northern America and Western Europe. While many results of this study are scarcely surprising, it is the first time that an efficiency analysis is being performed for a multitude of institutions worldwide using a standardized input indicator. It seems that the new proxy indicator based on co-authors is suitable for reflecting institutional staff numbers.

JEL-Codes: I210, I230, D610, H520.

Keywords: bibliometrics, efficiency, data envelopment analysis, universities, research-focused institutions, world.

Lutz Bornmann
Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8
Germany – 80539 Munich
lutz.bornmann@gv.mpg.de

Sabine Gralka
TU Dresden
Münchner Platz 2-3
Germany – 01069 Dresden
Sabine.Gralka@tu-dresden.de

Félix de Moya Anegón
CSIC, Institute of Public Goods and Policies (IPP), Consejo Superior de Investigaciones Cientificas, c/Albasanz 26-28
Spain – 28037 Madrid
felix.moya@scimago.es

Klaus Wohlrabe
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
wohlrabe@ifo.de

Introduction

One of the core indicators in the scientometrics field is the number of papers published by a unit (e.g., an institution or a country) within a given period. According to Auspurg et al. (2015) and other researchers in scientometrics, such indicators can only be assessed properly by considering the unit's available resources: the more the available resources, the better the performance is expected to be. In the present time, which is characterized by the wide popularity of performance-based funding models in science, one can expect that units (e.g., institutions and research groups) competing for funding are especially focused on efficiency (Bloch and Schneider, 2016). The term efficiency is defined by Rhaiem (2017) as follows: "within education settings, technical efficiency is defined in terms of the units of analysis (universities, departments, faculties, countries, researchers, etc.) which use the least inputs to produce a given level of output or alternatively produce the most output for a given bundle of inputs" (pp. 582-583). Similar definitions can be found in Moed (2017) and other publications in scientometrics. For Moed and Halevi (2015), typical examples of metrics for measuring efficiency are "the number of published articles per full time equivalent (FTE) researcher, or the number of citations per Euro spent on research" (p. 1991). A high efficiency score implies, for instance, that a university fully utilizes its employed researchers and produces the maximum number of papers (Agasisti and Gralka, 2019).

Various indicators can be used for measuring efficiency in the higher education sector. In an overview of studies measuring academic research efficiency, Rhaiem (2017) identified as typical groups of output indicators: (1) research productivity indicators (including quality of research indicators), (2) teaching output indicators (including quality of teaching indicators), and (3) entrepreneurial output indicators. For the input indicators, Rhaiem (2017) distinguishes six groups (see also De La Torre et al., 2017): "firstly, human capital category refers to academic staff and non-academic staff; secondly, physical capital category refers to productive capital (building spaces, laboratories, etc.); thirdly, research funds category encompasses budget funds and research income; fourthly, operating budget refers to income and current expenditures; fifthly, stock of cumulative knowledge regroups three sub-categories: knowledge embedded in human resources, knowledge embedded in machinery and equipment, and public involvement in R&D; sixthly, agglomeration effects category refers to regional effect and entrepreneurial environment" (p. 595).

Literature overviews of efficiency studies in the higher education sector can not only be found in Rhaiem (2017) but also in Johnes (2006), Berbegal Mirabent and Solé Parellada (2012), Clermont (2016), De Witte and López-Torres (2017), and Gralka (2018). Against the backdrop of the high number of existing overviews, we abstain from presenting an additional overview in this study. Recent examples of primary efficiency studies are Agasisti and Wolszczak-Derlacz (2015), who analyzed the efficiency of Italian and Polish universities, and

Edquist and Zabala-Iturriagagoitia (2016) analyzing the performance of national innovation systems in the European Union from an efficiency perspective. Abramo et al. (2011) measured the efficiency of universities' research activities in Italy, while De Fraja et al. (2016) measured the relationship between professorial pay and performance in the UK research excellence framework (REF). Efficiency evaluations comparing the university landscape of various countries, in both cases on the European level, were published by Bolli et al. (2016) and Herberholz and Wigger (2020). Based on previous primary efficiency studies, Rhaiem (2017) emphasized the need not only to measure efficiency by relating input to output measures, but also "to identify the environmental variables that may explain the differences among the efficiency scores" (p. 597).

Although many efficiency studies have been conducted in the higher education sector, the realization of these studies is not without problems. The first problem concerns the time dimension: it is not clear whether a certain time lag should be considered between input and output measurements and – if so – how large this lag should be (Aksnes et al., 2017). The second problem concerns the relationship of input and output indicators: according to Aagaard and Schneider (2015), one cannot explain outputs as a linear function of inputs in every case. The third problem is related to the heterogeneity of science units: institutions (and other units in science) have different missions and institutional contexts, which is why one can question the use of the same input and output indicators for efficiency measurements of all institutions (see here Waltman et al., 2017). The fourth problem concerns the availability of internationally standardized data: whereas on the basis of publication data (from literature databases such as Web of Science, WoS, Clarivate Analytics), worldwide standardized indicators are available on the output side, these data are frequently not available on the input side (Glänzel et al., 2016; Moed, 2017; Waltman et al., 2017). According to Aksnes et al. (2017), "countries have adopted different criteria for defining a researcher, thus it is difficult to make cross-national comparisons" (p. 249).

In this study, we especially target the fourth problem: we measure the efficiency of worldwide universities and other research-focused institutions by using institutional staff numbers on the input side which are estimated from authors' mentions on the institutions' papers. Thus, we abstain from using institutional staff numbers as has been done in previous studies. With our study, we follow the invitation by Waltman et al. (2016) to "investigate more deeply what types of input data are needed to construct meaningful productivity indicators" (p. 673).

This paper is organized into six main sections. The methods of efficiency measurement and their application in the present study are described in the second section. The third section addresses the specification of higher education inputs and outputs and explains the used data. The suitability of the newly introduced input indicator is discussed in the fourth section. We discuss the suitability in a separate section, since the use of staff numbers which are estimated

from the number of authors mentioned on publications is the core of this study. The fifth section presents the main results. The paper ends with a discussion and concluding remarks.

Methods

In the economic literature, efficiency commonly refers to the evaluation of an institution's input used relative to its obtained output. An institution is classified as efficient if it produces the largest possible output from a given set of inputs (or vice versa). For institutional evaluations in the higher education sector, it is commonly assumed that institutions dispose over a given amount of inputs and attempt to maximize their output (output-oriented model). To have a benchmark for evaluating the largest possible output, a group of institutions is assessed relative to a frontier. While the identification of the frontier varies with the chosen method, it is always dependent on the sample of institutions considered. Hence, the resulting efficiency values are relative measures. The values lie between zero and one, with higher scores indicating higher efficiency.

Two standard methods exist for estimating efficiency: the non-parametric Data Envelopment Analysis (DEA) and the parametric Stochastic Frontier Analysis (SFA). The respective advantages and disadvantage are well discussed in the literature (for an overview of both methods applied in the higher education sector, see Johnes et al., 2005). The final choice of one or the other method is mainly dependent on the dataset, the model to be considered, and the preference of the researcher. Because both methods are extensively used in the literature, we refrain from a thorough discussion. For more detailed explanations, we refer to Coelli et al. (2005) for the DEA and Kumbhakar and Lovell (2003) for the SFA.

Given the dataset to hand, with one output and one input variable at our disposable, the DEA is the method of choice. The advantage with the DEA is that no assumptions must be made on the functional form of the frontier. The need to specify a functional relationship between the input and output factors (through the type of function) and to limit one side of the equation to one factor, is the major disadvantage of the SFA. Particularly in the case of this study, with one output and one input variable, it would be presumptuous to choose the appropriate type of function (production, cost, profit or distance function) and the right functional form (linear, quadratic, CES or translog specification). The DEA is based on linear programming; efficiency is measured by calculating the ratio of (weighted) outputs over (weighted) inputs. The frontiers constituting the benchmark are given by efficient institutions. Inefficient institutions are enveloped by them. The distances of the institutions to the frontiers reflect the extent of inefficiency.

In this study, we use the DEA specification by Banker, Charnes and Cooper (1984), assuming variable returns to scale. This approach allows us to take the relative size of institutions into consideration. Following the literature, we choose an output-oriented model. This implies that

universities aim to maximize their output with the given resources (see above). Since institutional efficiency can vary over time in the science sector, we considered the time perspective in the statistical analyses.

Data

The dataset covers the time span from 2003 to 2018. It originates form the SCImago Institutions Ranking (SIR) which is produced by the SCImago Research Group and based on Scopus data (Elsevier). While the SIR provided the necessary input and output variables for numerous universities and research-focused institutions worldwide for this study, several steps were necessary to retrieve a suitable dataset. In a first step, we downloaded the information for all institutions with at least 1,000 (substantial) publications in total for this period. This amounts to 5,254 institutions. As the input variable, we used the indicator 'Scientific Talent Pool' (STP) from the SIR as estimates for the institutional staff numbers at worldwide institutions. The SCImago Research Group defines this indicator as follows: "total number of different authors from an institution in the total publication output of that institution during a particular period of time" (see the definition at https://www.scimagoir.com/methodology.php). For counting the institutional number of authors, the group uses disambiguated author names from Scopus. As an output variable, we considered the number of publications in this study. Worldwide institutions as a rule not only focus on research output, but also quality of research. In order to consider the quality dimension in the efficiency analyses, we use the $P_{top\ 10\%}$ indicator. This is the number of institutional publications that belong to the 10% most frequently cited papers in their subject category (journal sets defined by Scopus subject areas) and publication year. Such percentile-based citation impact indicators are recommended in the Leiden Manifesto in order to measure the research performance of institutions (Hicks et al., 2015).

In a second step of the data preparation, we balanced the data set, i.e. we kept only those institutions that have data for SPT over the whole timeframe. This procedure rules out institutions that were founded after 2003 or closed before 2018. Together with the decision to only consider institutions that publish more than 1,000 papers over the full period, this could lead to a bias of our data towards large and well-performing institutions. Balancing the data set allows us to conduct a consistent analysis of time trends.

In a last step of the data analysis, we excluded all institutions which were type-classified in SIR as "Others" (n=59 institutions, as for example the "Australian Museum") or as "World" (10 institutions, as for example the "Dow Chemical Cooperation"). "World" refers to geographical institutional classifications.

Our final data set consists of 4,857 institutions from 134 countries spanning all continents.[1] The majority of institutions are located in the USA (n=829), followed by China (n=463) and Japan (n=309). For about 60% of the countries, we have less than 10 institutions listed, for 38 countries only one. In addition to the country classification, the institutions can be categorized according to larger geographical areas and sectors. Table 1 shows the corresponding distributions. Most institutions belong to the higher education category, which are mainly universities. With respect to the location, most institutions in our sample are located either in Western European or the Asiatic region.

Table 1: Institutions across sectors and world regions

Sector		**World region**	
Government	562	Africa	144
Health	887	Asiatic region	1,397
Higher education	3,155	Eastern Europe	348
Private	253	Latin America	319
		Middle East	272
		Northern America	945
		Pacific region	121
		Western Europe	1,311
Total	4,857		4,857

Table 2 reports the descriptive statistics for the considered input and output variables. To provide a comprehensive overview, the table shows the total for the whole timeframe and the evolution of the two variables over time. In 2003, the (estimated) institutional staff numbers amount to 440 authors on average per institution. The number then steadily increased to around 1,200 authors in 2018. Similarly, the number of papers published by an average institution has increased, from 60 in 2003 to 170 in 2018. Table 2 also shows that there is a considerable difference between the minimum and maximum in our sample. In 2018, three institutions had only one author ("University of Nordland" from Norway, "The Scottish Agricultural College" from Great Britain and "Senatech" from the US), while the largest institution had 90,934 authors (the "Ministry of Education of the People's Republic of China"). It is not

[1] An overview of all institutions considered can be provided upon request.

at all surprising that the four largest institutions in our sample all belong to the "Government" sector. The largest institution from the "Higher Education" sector is the "University of Chinese Academy of Sciences", followed by "Harvard University". The largest "Health" institution is the US "National Institutes of Health"; the largest institution from the "Private" sector is the "State Grid Corporation of China".

Similarly to the (estimated) institutional staff numbers, there is a large spread between the minimum and maximum number of papers. While there are some institutions that have not published any $P_{top\ 10\%}$ papers (such as the "Nippon Institute of Technology" from Japan) or only one $P_{top\ 10\%}$ paper (such as the "Institute for Health and Consumer Protection" from Italy), both in 2018, one institution in our sample published 10,386 highly-cited papers ("Chinese Academy of Sciences") in the same year.

Table 2: Descriptive statistics of $P_{top\ 10\%}$ and STP over time

	STP									
	2003	2005	2007	2009	2011	2013	2015	2017	2018	**Average**
Mean	437	547	636	712	835	930	993	1,088	1,150	796
Median	131	177	217	251	305	340	369	413	432	292
Minimum	1	1	2	3	11	12	6	2	1	15
Maximum	33,321	41,605	48,638	54,290	59,753	62,320	64,804	76,999	90,934	53,717
Standard deviation	1,046	1,274	1,465	1,643	1,906	2,154	2,311	2,576	2,802	1,814
	$P_{top\ 10\%}$									
	2003	2005	2007	2009	2011	2013	2015	2017	2018	**Total**
Mean	62	79	93	106	128	144	157	174	170	1,943
Median	131	177	217	251	305	340	369	413	432	292
Minimum	0	0	0	0	0	0	0	0	0	10
Maximum	4,393	5,768	6,963	7,981	8,815	9,079	9,201	10,604	10,386	122,165
Standard deviation	179	222	253	285	335	371	401	433	412	4,977

Suitability of the STP input indicator

Previous authors who have investigated institutional efficiency in the higher education sector included the number of researchers, professors or employees (or similar personal numbers) in their studies (see, e.g., De Witte and López-Torres, 2017 and Gralka, 2018). The problem with these data is that they are not comparable across national borders. The definitions that are used in the countries to identify these groups of people working at universities are so different that they cannot be used in cross-country institutional efficiency measurements. Furthermore, if national databases do not exist in countries, comparable data are usually not available for all institutions in single countries. With the STP indicator provided in the SIR, a possible alternative indicator is available reflecting institutional academic staff numbers.

In this section, we present the results of empirical analyses that were intended to analyze whether or not the alternative indicator can be used as a proxy for staff numbers. In order to make the comparison with STP, we collected staff information for two countries from other sources: Great Britain (GB) and Germany. In both cases, we only have information on the "Higher Education" sector and therefore compare institutional numbers from this sector.

For GB, we obtained data for the higher education entities from the Higher Education Statistics Agency (HESA)[1]. HESA provides staff information for "Total academic staff" and "Total staff", which contains the number of employees at the respective institution. For four academic periods (2014/2015 to 2017/2018)[2], information was available for 163 institutions. We were able to match information from SIR and HESA for 109 institutions. Table 3 reports the resulting correlations for three years. For both staff categories, the correlation is quite high, at about 0.9, indicating a close relationship between the two data sources. The "Total staff" variable shows a slightly higher correlation with the STP variable than "Total academic staff".

Table 3: Correlation between SIR and HESA data for Great Britain (n=109)

Staff	**2015**	**2016**	**2017**
Academic staff	0.876	0.889	0.892
Total staff	0.891	0.902	0.905

[1] see http://www.hesa.ac.uk

[2] Relative weights are used to adjust the academic year to the calendar year.

We obtained the data for Germany from the Federal Statistical Office of Germany. In order to conduct an even more thorough check, we use two datasets to compare different staff classifications. First, we compare the STP to the number of full-time equivalents, separated by "Total academic staff" and "Total staff". We report the correlations for the years 2013 to 2015 for 66 institutions in Table 4. Second, we compare the STP to the number of employees, shown in Table 5. This is similar to the measure in the GB dataset. However, the variable for Germany is separated by "Total staff" and "Professors" for the years 2011 to 2013, including 58 institutions.[1] The results in Table 4 and Table 5 show that the correlations are quite high, ranging from 0.7 to 0.8. In contrast to the results in Table 3, "Academic Staff" shows a slightly higher correlation to the STP than the "Total staff" variable. The comparison between Table 4 and Table 5 demonstrates that the count data displays a higher correlation than the variable depicting full-time equivalents. This is to be expected, since the STP measure rather equals a count measure (there is no distinction between full-time and part-time workers).

Table 4: Correlation between SIR and data from the Federal Statistical Office of Germany (n=66)

Full-time equivalents	**2013**	**2014**	**2015**
Academic staff	0.861	0.846	0.815
Total staff	0.796	0.783	0.764

Table 5: Correlation between STP and data from the Federal Statistical Office of Germany (n=58)

Number of staff	**2011**	**2012**	**2013**
Academic staff	0.934	0.928	0.920
Professors	0.885	0.897	0.895

Although the correlations between the STP variable and different institutional staff numbers are not perfect, the consistently high correlation coefficients reveal that STP appears to reflect these numbers. The STP variable appears to be an alternative indicator to the staff numbers that are usually used in efficiency studies (which only allow institutional efficiency comparisons within countries). The STP variable in the SIR opens up the unique opportunity to measure the efficiency of universities and research-focused institutions worldwide. Thus, the STP

[1] The differences in the two datasets concerning German institutions (with regard to years and number of institutions) are driven by variances in the datasets provided by the Federal Statistical Office of Germany.

variable seems to be the input equivalent to the number of publications on the output side (publication and citation data can be used without considering national borders).

Results

We present the results of this study in two parts. We start with cross-sectional and average results over the full-time period for various categories stated in the data section. We proceed with results referring to time trends. In the last sub-section, the results of regression models are presented.

Overall results

For 4,857 institutions across 16 years, we obtain 77,112 efficiency scores in total. The average score is rather low at 0.17 and a standard deviation of 0.16. In Figure 1, all scores are plotted in a histogram. It can clearly be seen that the scores are skewed to the right. We counted 1,679 cases with an efficiency score of zero, which implies $P_{top\ 10\%}=0$. In 131 cases an institution was efficient, i.e. it has a score of 1.00. Among all the universities and research-focused institutions included, 34 institutions had a score of 1.00 in at least one year. The only institution that achieved this score across all 16 years is "Harvard University".

Figure 1: Histogram of efficiency scores across institutions and years

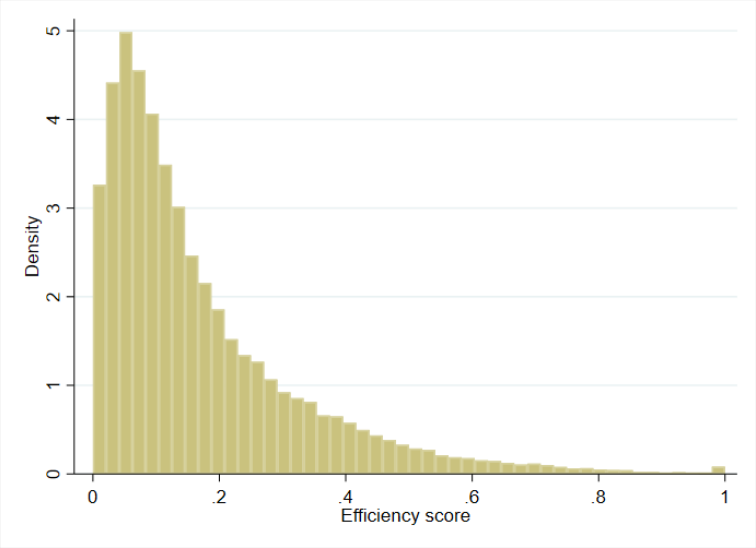


Table 6 lists the 15 institutions with the highest average efficiency scores across all years including the corresponding 95% confidence interval. "Harvard University" is ranked first as it was efficient in every year. Considering the confidence bands in the interpretation of the results, the first five institutions are nearly at the same level as "Harvard University". Most of

the listed institutions are from the US higher education sector. Table 6 lists only institutions from Northern America or Europe.

Table 6: Efficiency ranking for institutions

Rank	Organization	Sector	Country	World region	Mean efficiency score	Confidence bands	
1	Harvard University	Higher education	USA	Northern America	1.000	1.000	1.000
2	Centre National de la Recherche Scientifique	Government	France	Western Europe	0.985	0.967	1.002
3	Massachusetts Institute of Technology	Higher education	USA	Northern America	0.985	0.971	0.998
4	Institucio Catalana de Recerca i Estudis Avancats	Government	Spain	Western Europe	0.984	0.949	1.018
5	Howard Hughes Medical Institute	Health	USA	Northern America	0.970	0.944	0.996
6	Stanford University	Higher education	USA	Northern America	0.967	0.953	0.981
7	University of California, Berkeley	Higher education	USA	Northern America	0.962	0.944	0.979
8	Max Planck Gesellschaft	Government	Germany	Western Europe	0.929	0.916	0.941
9	Princeton University	Higher education	USA	Northern America	0.920	0.882	0.957
10	University of Oxford	Higher education	Great Britain	Western Europe	0.892	0.851	0.934
11	Brigham and Women's Hospital	Health	USA	Northern America	0.887	0.863	0.911
12	University of Cambridge	Higher education	Great Britain	Western Europe	0.886	0.863	0.908
13	University College London	Higher education	Great Britain	Western Europe	0.856	0.819	0.892
14	University of Washington	Higher education	USA	Northern America	0.841	0.826	0.855
15	Imperial College London	Higher education	Great Britain	Western Europe	0.836	0.795	0.877

Table 7 shows the average scores for each region and sector. The highest average score, 0.276, can be seen for the Pacific region. The best institution from this region is the “University of Melbourne” with an average score of 0.706 (institutional rank = 56). Northern America

and Western Europe also have an efficiency level above 0.2. In all other regions, there is an average of below 0.13. With respect to the sector, institutions classified as "Government" have the highest score at 0.197, while the lowest score can be found for the "Private" sector, at 0.138.

Table 7: Average efficiency score across regions and sectors

Sector	Efficiency score	World region	Efficiency score
Government	0.197	Africa	0.088
Health	0.171	Asiatic region	0.107
Higher education	0.170	Eastern Europe	0.083
Private	0.138	Latin America	0.080
		Middle East	0.122
		Northern America	0.255
		Pacific region	0.276
		Western Europe	0.236

The country ranking is reported in Table 8 together with the corresponding numbers of institutions. We only included countries with at least 10 institutions.[1] The results in Table 8 show that, on average, the efficiency level is the highest in Denmark. Most countries (n=7) are located in Europe.

[1] Panama would have entered the ten best performing institutions in Table 8 but has only one institution listed in our sample.

Table 8: Country ranking (showing only countries with more than ten institutions)

Rank	Country	Average efficiency score	Number of institutions
1	Denmark	0.355	18
2	Netherlands	0.330	53
3	Switzerland	0.310	16
4	Hong Kong	0.304	43
5	Australia	0.288	102
6	Belgium	0.272	35
7	Great Britain	0.270	228
8	Finland	0.261	47
9	Sweden	0.259	32
10	Canada	0.256	116

Time trends

In the second step of the analysis, we investigated potential time trends in the institutional efficiency scores. In Panel A of Figure 2, efficiency scores are plotted for each year across all institutions. There seems to be a slightly negative trend in the first half of the period. The median continuously decreased from 2003 to 2010. Afterwards it appears to stabilize at a low level. This is supported by the Panel B results with the mean as well as the minimum and maximum efficiency scores in each year. Panel C reveals the numbers of efficient institutions across the years which ranges between 4 and 11. Panel D visualizes Pearson correlation coefficients as well as (Spearman) rank correlations for the relationship between efficiency scores of consecutive years. All correlation coefficients are high (>0.8), but increase slightly over time. This indicates both that the efficiency scores are quite stable across the years and that this stability is consolidated.

Figure 2: Summary statistics for the efficiency scores across time

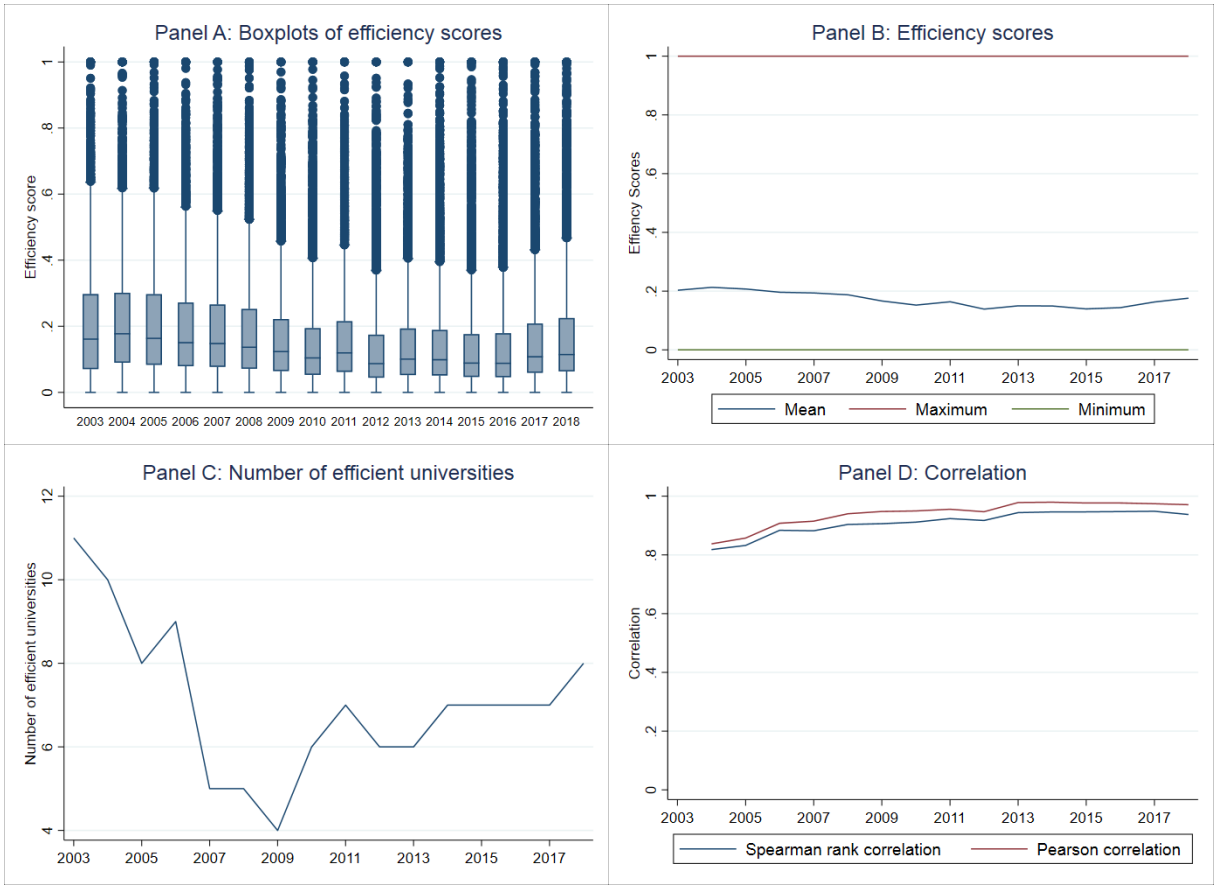
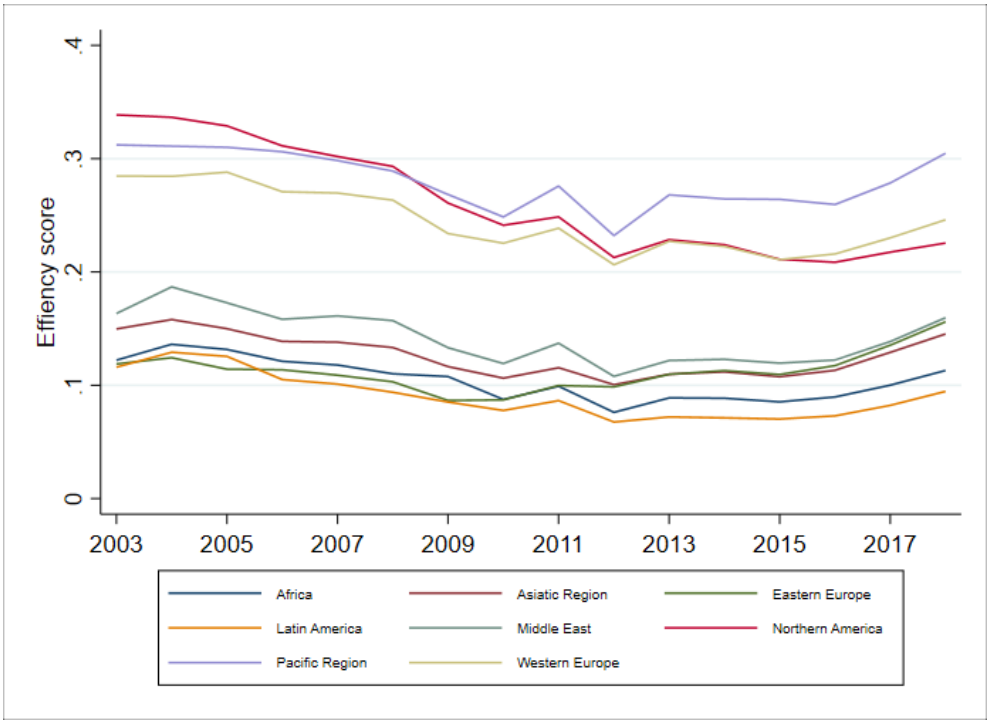


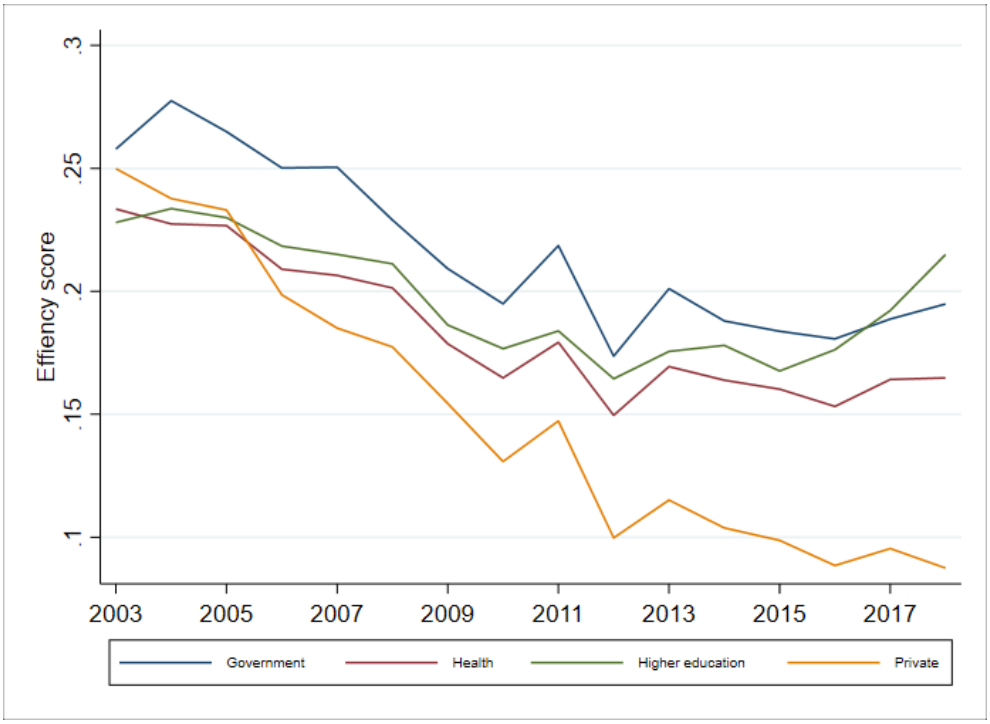
Figure 3 and Figure 4 show the time evolution of the efficiency scores separated by geographic region and sector. Although Figure 3 shows a slightly u-shaped trend, the regional scores resemble the results in Table 6.

Figure 3: Time trends in efficiency scores by region



This is also the case for the results in sectors with one exception (see Figure 4). Institutions from the private sector observed a steady decrease in their efficiency scores. Until 2016, governmental research institutions were on average better than institutions from the other sectors. In 2017, the higher education institutions show the highest average institutional efficiency scores.

Figure 4: Time trends in efficiency by sector



Regression Models

We calculated regression analyses in order to check the partial relationships of both sector and world region (independent variables) and institutional efficiency scores (dependent variable, see Table 7). We estimate a Tobit regression model that accounts both for the panel structure of our data set as well as the fact that the efficiency scores are bounded between 0 and 1. The calculated random effects model is specified as

$$DEA_{it} = \alpha_i D_{it} + \gamma_t + \upsilon_i + \varepsilon_{it}$$

where DEA_{it} is the DEA score of institution i in year t. We account for time fixed effects (γ_t) which capture general time trends in efficiency scores. The dummies D_{it} capture either the sector or the world region. In the regression analysis, we choose the categories with the highest efficiency levels reported in Table 7 as benchmark categories. These are "Government" for the sector and "Pacific region" for the geographical areas. Table 9 shows the results of the regression analyses (three models). In each model, the constant captures the estimated average efficiency score of the base category. The other coefficients cover the difference in relation to this base category.

As there are only dummies in the regression model, the estimated coefficients can be interpreted as differences in the efficiency level. For example, on average, the "Health" sector has a 0.026 lower efficiency score than the "Government" sector. Thus, the "Health" sector (and

the other sectors in the table) is not very different from the "Government" sector in terms of efficiency. Table 9 shows that all coefficients are statistically significant at the 1% level except for "Northern America" which is only statistically significant at the 10% level. Since the coefficient is also very low at -0.021, the efficiency difference in relation to the "Pacific region" appears to be marginal. Although the coefficient for "Western Europe" is also very low (-0.040), the result is statistically significant. If we consider the sector and regional dummies simultaneously, we obtain similar results. The coefficient for "Higher education" becomes non-significant, whereas the coefficient for "Northern America" is now statistically significant at the 5% level. We assume that the general low *p* values result from the very high case numbers considered in the regression models. The differences between sectors and regions in Table 9 are smaller than those reported in Table 7. One reason might be that the regression models also account for general efficiency time trends.

Table 9: Estimation results from three panel Tobit regression models

Variable	Regression analysis including sectors	Regression analysis including regions	Regression analysis including both
Health	-0.026 ***		-0.059 ***
Higher education	-0.028 ***		-0.009
Private	-0.059 ***		-0.091 ***
Government (reference category)	-		
Africa		-0.189 ***	-0.207 ***
Asiatic Region		-0.170 ***	-0.185 ***
Eastern Europe		-0.195 ***	-0.214 ***
Latin America		-0.197 ***	-0.214 ***
Middle East		-0.154 ***	-0.171 ***
Northern America		-0.021 *	-0.025 **
Western Europe		-0.040 ***	-0.044 ***
Pacific Region (reference category)		-	
Constant	0.226 ***	0.305 ***	0.305 ***
Fixed effects of year	Yes	Yes	Yes
N	77,112	77,112	77,112

Note. *** $p<.001$, ** $p<.05$, * $p<.1$

Based on the panel Tobit regression analysis, including both sector and regional dummies (see column 3 in Table 9), we calculate adjusted efficiency scores that account for the influence of the independent variables. These scores are not predicted values from the regression analysis, but institutional scores for which the residuals are added to the mean initial efficiency scores in each year. The adjusted efficiency scores are very similar to the original scores. The average level change is 0.071 and the Pearson correlation between original and adjusted scores is 0.86. The institutional ranking based on the adjusted scores is reported in Table 10. The comparison with the institutional ranking in Table 6 reveals that “Harvard University” is sur-

passed by the "Howard Hughes Medical Institute", but that the confidence bands overlap. Twelve institutions listed in Table 6 are also included in the ranking based on the adjusted scores. The "City University of Hong Kong" is listed in Table 10 ranked 11th; this is 23 positions better than in the ranking based on the original scores.

Table 10: *Adjusted* efficiency ranking for institutions

Rank	Organization	Sector	Country	World region	Mean efficiency score	Confidence bands	
1	Howard Hughes Medical Institute	Health	USA	Northern America	0.918	0.892	0.945
2	Harvard University	Higher educ.	USA	Northern America	0.898	0.897	0.898
3	Centre National de la Recherche Scientifique	Government	FRA	Western Europe	0.893	0.876	0.911
4	Institucio Catalana de Recerca i Estudis Avancats	Government	ESP	Western Europe	0.892	0.858	0.927
5	Massachusetts Institute of Technology	Higher educ.	USA	Northern America	0.882	0.869	0.896
6	Stanford University	Higher educ.	USA	Northern America	0.865	0.851	0.879
7	University of California, Berkeley	Higher educ.	USA	Northern America	0.859	0.842	0.877
8	The University of Hong Kong	Higher educ.	HKG	Asiatic Region	0.844	0.804	0.883
9	Max Planck Gesellschaft	Government	DEU	Western Europe	0.838	0.825	0.850
10	Brigham and Women's Hospital	Health	USA	Northern America	0.835	0.812	0.859
11	City University of Hong Kong	Higher educ.	HKG	Asiatic Region	0.825	0.787	0.862
12	Princeton University	Higher educ.	USA	Northern America	0.818	0.780	0.855
13	University of Oxford	Higher educ.	GBR	Western Europe	0.810	0.769	0.851
14	University of Cambridge	Higher educ.	GBR	Western Europe	0.803	0.781	0.826
15	Chinese Academy of Sciences	Government	CHN	Asiatic Region	0.793	0.698	0.889

Discussion and conclusions

Performance rankings of institutions are popular in post-academic science. Several international university rankings exist, such as the Academic Ranking of World Universities (ARWU) and the Times Higher Education World University Rankings. These rankings are mostly based on output indicators (such as publications or Nobel prizes). In a recent study, Lepori et al. (2019) investigated the relationship between universities' revenues on one side and their number of publications and field-normalized citation impact scores on the other side. The authors included universities from the USA and Europe in their study. The results demonstrated that "international rankings are by and large richness measures and, therefore, can be interpreted only by introducing a measure of resources". Thus, it appears that university rankings cannot refrain from considering input indicators in order to provide a valid measurement of research performance. In this study, we propose a measurement approach that can be used to conceive worldwide institutional rankings based on comparable input and output indicators.

It is a decisive advantage of publication data that they are available in multi-disciplinary databases (such as WoS or Scopus) and can be used for cross-country comparisons. For example, the Leiden Ranking (see www.leidenranking.com) presents bibliometric results for nearly 1,000 major universities worldwide. Similar data are not available on the input side, however: input data (not only staff numbers, but also financial data) could actually be used for comparing institutions within one single country. In this study, we propose the use of the STP indicator as a proxy for the staff numbers on the input side (for worldwide comparisons). We tested the indicator for this use by correlating it with available staff numbers for two countries. The high correlation coefficients indicate that the STP indicator measures something similar as staff numbers. Thus, the indicator seems to be a good alternative for efficiency studies on the input side.

In this study, we calculated efficiency scores for more than 4,800 institutions worldwide. In a similar way to the results published in various university rankings, "Harvard University" is the best performing institution (in all years), followed by many other institutions from Northern America or Europe. The results of our study further show that institutions in the Pacific region have the highest average efficiency scores, followed by Northern America and Western Europe. Denmark is the country with the most efficient institutions on average. The comparison of institutional sectors revealed that institutions in the governmental sector have the highest efficiency scores. One of the reasons for this result might be that researchers in these institutions can be more focused on research than researchers working in institutions from other sectors.

Taken together, the results of this study are to be expected and are hardly surprising. The best performing institutions, sectors, and regions are those that have been established as being sim-

ilarly successful in (many) other studies. This result might be interpreted as disappointing, without new insights in the performance of institutions. However, it was an important goal of the study to deal with a new indicator – the STP indicator – for use on the input side in efficiency measurement studies. The unsurprising results that we obtained can be interpreted as a positive sign for the validity of this indicator and thus for its future use: the indicator appears to be suitable for reflecting (measuring) institutional staff numbers.

What are the limitations of this study? (1) The suitability of the STP input data was only checked for two countries and universities. Thus, suitability was not investigated for other countries and research-focused institutions. We encourage researchers to conduct such investigations in future studies. (2) We used the standard two-step approach for measuring efficiency in this study by calculating DEA scores and performing regression analyses based on DEA scores. Such regression analyses have been criticized by Simar and Wilson (2007) due to a lack of joint data generation processes. This lack may lead to biased DEA scores and regression coefficients. Badunenko and Tauchmann (2019) proposed commands for calculating the Simar and Wilson (2007) efficiency analysis with Stata. However, their proposed framework is only suitable for cross-sectional data, but not for panel data.

References

Aagaard, K., & Schneider, J. W. (2016). Research funding and national academic performance: Examination of a danish success story. *Science and Public Policy*, *43*(4), 518–531.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). A field-standardized application of data envelopment analysis to national-scale research assessment of universities. *Journal of Informetrics*, *5*(4), 618–628.

Agasisti, T., & Gralka, S. (2019). The transient and persistent efficiency of italian and german universities: A stochastic frontier analysis. *Applied Economics*, *51*(46), 5012–5030.

Agasisti, T., & Wolszczak-Derlacz, J. (2015). Exploring efficiency differentials between italian and polish universities. *Science and Public Policy*, *43*(1), 128–142.

Aksnes, D. W., Sivertsen, G., van Leeuwen, T. N., & Wendt, K. K. (2017). Measuring the productivity of national r&d systems: Challenges in cross-national comparisons of r&d input and publication output indicators. *Science and Public Policy*, *44*(2), 246–258.

Auspurg, K., Diekmann, A., Hinz, T., & Näf, M. (2015). Das forschungsrating des wissenschaftsrats für die soziologie in deutschland revisited. *Soziale Welt*, *66*(2), 177–192.

Badunenko, O., & Tauchmann, H. (2019). Simar and wilson two-stage efficiency analysis for stata. *The Stata Journal*, *19*(4), 950–988.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092.

Berbegal Mirabent, J., & Solé Parellada, F. (2012). What are we measuring when evaluating universities efficiency. *Regional and Sectoral Economic Studies*, *12*(3), 32–46.

Bloch, C., & Schneider, J. W. (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the norwegian publication indicator at the individual level. *Research Evaluation*, *25*(4), 371–382.

Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The differential effects of competitive funding on the production frontier and the efficiency of universities. *Economics of Education Review*, *52*, 91–104.

Clermont, M. (2016). Effectiveness and efficiency of research in germany over time: An analysis of german business schools between 2001 and 2009. *Scientometrics*, *108*(3), 1347–1381.

Coelli, T. J., Rao, P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media. Accessed 20 June 2016

De Fraja, G., Facchini, G., & Gathergood, J. (2016). How much is that star in the window? Professorial salaries and research performance in UK universities. *Retrieved September 26, 2016, from https://www.nottingham.ac.uk/gep/documents/papers/2016/2016-13.pdf*.

De La Torre, E. M., Agasisti, T., & Perez-Esparrells, C. (2017). The relevance of knowledge transfer for universities efficiency scores: An empirical approximation on the spanish public higher education system. *Research Evaluation*, *26*(3), 211–229.

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

Edquist, C., & Zabala-Iturriagagoitia, J. M. (2016). Assessing the performance of national innovation systems in Europe. In *Proceedings of the 21 ST International Conference on Science and Technology Indicator*. Spain: Editorial Universitat Politecnica de Valencia.

Glänzel, W., Thijs, B., & Debackere, K. (2016). Productivity, performance, efficiency, impact - what do we measure anyway? Some comments on the paper: A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, *10*(2), 658–660.

Gralka, S. (2018). Stochastic frontier analysis in higher education: A systematic review. *CEPIE Working Paper*, *5/18*, 1–58.

Herberholz, L., & Wigger, B. U. (2020). Efficiency of european universities: A comparison of peers. *CESifo Working Paper Series*, *8044*, 1–46.

Hicks, D., Wouters, P., Waltman, L., Rijcke, S. de, & Rafols, I. (2015). Bibliometrics: The leiden manifesto for research metrics. *Nature News*, *520*(1), 429–431.

Johnes, G., Johnes, J., Thanassoulis, E., Lenton, P., & Emrouznejad, A. (2005). *An exploratory analysis of the cost structure of higher education in England*. UK Department for Education and Skills. Accessed 24 March 2016

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

Kumbhakar, S. C., & Lovell, K. (2003). *Stochastic frontier analysis*. Cambridge University Press.

Lepori, B., Geuna, A., & Mira, A. (2019). Scientific output scales with resources. A comparison of US and European universities. *PloS one*, *14*(10).

Moed, H. F. (2017). *Applied evaluative informetrics*. Heidelberg, Springer: Springer.

Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, *66*(10), 1988–2002.

Rhaiem, M. (2017). Measurement and determinants of academic research efficiency: A systematic review of the evidence. *Scientometrics*, *110*(2), 581–615.

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

Waltman, L., van Eck, N. J., Visser, M., & Wouters, P. (2016). The elephant in the room: The problem of quantifying productivity in evaluative scientometrics. *Journal of Informetrics*, *10*(2), 671–674.

Waltman, L., Wouters, P., & Van Eck, N. J. (2017). Ten principles for the responsible use of university rankings. *URL: https://www. universiteitleiden. nl/binaries/content/assets/algemeen/onderzoek/responsible-use-of-university-rankings. pdf [Accessed on 6th of June 2017]*.