Journal of Informetrics 18 (2024) 101502

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Heterogeneity of national accounting systems, world-class universities and financial resources: What are the links?

Alessandro Avenali [a], Cinzia Daraio [a], Simone Di Leo [a,*], Joanna Wolszczak-Derlacz [b]

[a] Department of Computer, Control and Management Engineering (DIAG), Sapienza University of Rome, Via Ariosto 25, Rome 00185, Italy
[b] Faculty of Management and Economics, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk 80-233, Poland

ARTICLE INFO

Keywords:
University rankings
ARWU
Financial data
Predictive machine learning-based tools
XGBoost
Accounting systems

ABSTRACT

This study investigates the relationship between university financial resources, applied accounting systems, and the place of a university in the Shanghai Ranking. We find a strong relationship between the financial resources under the control of a world-class university and the position of that university in the highest tier of the global ranking. We propose a model (available online) to predict a university's tier in the ranking through the financial resources it employs. A critical condition for making a university a world-class university could be to provide it with a sufficiently high level of financial resources, and its efficiency could play an important leverage role. In view of the results, policymakers are challenged with a drastic choice: to increase international competition among universities, it is necessary to concentrate a huge amount of resources on a few universities that are already in the ranking. In contrast, the policy of the proportional distribution of resources does not affect international competition and may be inefficient. Furthermore, financial data are not easy to gather homogeneously for universities across countries, due to the existence of different national accounting systems. Finally, we discuss several critical issues associated with the measurement of specific accounting data of universities.

1. Introduction and research questions

Data are crucial to analyse, describe and evaluate institutions of higher (tertiary) education and research, understand how universities produce their services, evaluate their performance, and identify proposals and suggestions for university reforms. In general, both private and public universities must be transparent and accountable to ensure that they provide a quality education to their students, maintain the trust of their donors and stakeholders and be held accountable for the use of funds, especially when it comes to public funds or public Higher Education Institutions (HEIs). A review on the accountability of HEIs can be found in Macheridis and Paulsson (2021). However, a performance evaluation of universities is far from simple. The multiple activities that universities carry out, which include teaching, research and services (the so-called "third mission"), interact with one another and with the objectives of policymakers and institutional missions calling this interaction "triple helix" (see, e.g., Etzkowitz and Zhou 2017). Triple helix refers to the university's roles including teaching, research and providing services to the public and private environment. University rankings usually aim to assess the performance of HEIs and their different roles. The different university's activities (including providing services) impact rankings. There can be also a vice-versa relationship, the public and private actors might be more eager to collaborate with universities higher in rankings. University rankings usually aim to assess the performance of HEIs by collecting and elaborating

* Corresponding author.
E-mail address: dileo@diag.uniroma1.it (S. Di Leo).

https://doi.org/10.1016/j.joi.2024.101502
Received 28 September 2023; Received in revised form 15 December 2023; Accepted 15 January 2024
Available online 19 January 2024

data related essentially to teaching, research, knowledge transfer, and the global outlook, such as university descriptors, academic degrees delivered in distinct fields of education, students enrolled, graduates, research activities, or the number of academic and administrative staff. Even though rankings were introduced as a tool to inform university governance, they influence institutional behaviour and increase competition in the higher education system, requiring policy responses to rankings. Interest in the theory and methodology of rankings and in their impact and influence is currently high, as remarked in the recent book by Hazelkorn and Mihut (2021). Notwithstanding, rankings have also been damaged by several and various criticisms so that a few institutions have begun to ignore them. Daraio et al. (2015) discuss more thoroughly the main criticisms addressed to university rankings (monodimensionality, lack of statistical robustness, dependence on university size and subject mix, and a lack of consideration of the input-output structure). On the other hand, global rankings in the media receive a great deal of attention and are of great importance for university prestige. In fact, all university stakeholders should use the information provided by these global rankings in a more conscious and responsible manner. To do so, rankings should be studied by adding further points of view and other categories of data which can help institutions and policymakers to better understand what a tier in the ranking means for a university. During the last decade, financial data have attracted a lot of attention. The financial sustainability of universities has become a more and more crucial issue, as only institutions that know the full costs of their activities and projects (i) can judge if they are able to operate on a financially sustainable basis, (ii) can compare themselves with other universities in order to find processes where efficiency can be improved, and (iii) can define a focused strategy to stop spending beyond their means (Di Carlo et al., 2019).

In this work, we show that specific financial data related to resources under the control of a university can be strongly predictive of the outcome of competition among universities in the worldwide higher education industry. In other words, we find that a critical condition for an HEI to become a world-class university might be to provide the university with a sufficiently high level of resources. For instance, abundant financial resources would allow HEIs, on one hand, to purchase the best equipment and to finance laboratories to do high-quality research and, on the other hand, to employ the best professors in the world to do applied and/or frontier research and to teach students. In addition, we develop a machine learning-based tool to effectively predict the tier of an HEI in a selected global ranking. In our analysis we consider the 2019 Academic Ranking of World Universities (ARWU, 2023), also known as the Shanghai Ranking (see Cai 2009, Docampo 2011). We chose the year 2019 to avoid the possible bias of the impact of Covid-19 on investment choices made by universities in the subsequent period. Incidentally, we observe that the Shanghai Ranking can be rounded by a Lorenz curve (Lorenz, 1905)[1] where 1 % of the universities in the ranking obtain the largest part of the overall marks of all the rankings and which thus signals the presence of large asymmetries in the distribution of the marks (most likely due to a high concentration of resources among a few universities).

Furthermore, even if providing a university with significant financial resources is most likely a necessary condition to make it a world-class university, we also show that certain universities are more efficient in using these resources.

It might be argued that it is the reputation (i.e., ranking) that affects the ability to attract funds. On the other hand, reputation is also affected by available resources, although only a very little share of universities has a high rank (reputation) with respect to the whole set. In other words, the relationship between reputation and the availability of financial resources represents a canonical 'chicken-and-egg' problem. In fact, we understand the relationship between the availability of financial resources and reputation is problematic because the direction is unclear. That is, it is unclear whether it is funding that generates reputation, enabling the purchase of state-of-the-art equipment and the recruitment of star scientists, or, whether it is reputation, through cumulative effects, the so-called "Matthew effect", that enables ever-increasing funding. There is a strong confounding effect, as the reputation affects both the resources and ranking. It is highly probable that it is not because of the financial resources, but the reputation itself affects the overall ranking. Therefore, we might ask why give so much importance to financial resources? The answer depends on the perspective of the analysis we adopt. If we place ourselves in a current perspective of analysing the reality of universities, the analysis of financial resources probably loses some of its interest, since few are the world's universities that concentrate most of the resources and are at the top of the rankings. If, on the other hand, we take a future perspective, the one we adopt in this paper, it can be extremely useful to provide a tool for policymakers to have an order of magnitude about the financial resources that would need to be invested to try to enter international competition and be present in international rankings.[2] Obviously, financial resources are a necessary but not sufficient condition for entering the international arena and building a reputation to enter the international competition. For that reason we also analyse the efficiency in the use of resources.

It would be interesting to analyse how the universities that are at the top of the rankings have reached these places. However, reconstructing each university's development and history of funding and reputation acquisition is outside the scope of this work. We instead approach the topic in a pragmatic way, with a perspective of predicting the possible ranking also for universities that are not currently present in the rankings, but which aspire to enter it, starting from the financial resources they have at their disposal. Our

[1] The Lorenz curve is an instrument usually used to measure the "inequality" in the distribution of a variable within a dataset (see Section 5).

[2] A recent important example is represented by several Chinese universities (Fedasiuk et al., 2022). In last decade, Chinese government has selected a limited group of universities and provided them with a huge amount of resources and ambitious development objectives. For instance, Tsinghua University is one of the selected universities and it is now in the top 50 group with position 22 in the 2023 ARWU; however, in 2015 Tsinghua University was in 101–150 group and in just one year (in 2016) it climbed the ARWU by reaching position 58. Similarly, Zhejiang University, another university of the selected group, is now in the top 50 group with position 33 in the 2023 ARWU; however, in 2017 Zhejiang University was in 101-150 group while in just one year (in 2018) it climbed the ARWU by reaching position 67. Nowadays, the six best-financed universities in China receive each year comparable resources as the top U.S. ones, which allow them to dedicate significant efforts to S&T and hosts an array of R&D centers.

perspective may be particularly interesting for those countries that would like to strengthen their international presence but do not have many universities in the top rankings. For example, Chinese government funding has seen a huge increase since 2017. China is concentrating a huge amount of funding in its 34 most elite "Double First Class" universities, and this is beginning to raise some concerns for other global competitors, primarily the United States (see Fedasiuk et al. 2022).

Our prediction model available online can be useful for carrying out simulations of possible policy alternatives, evaluating the potential effect of the concentration of resources on a few institutions that can reach the critical mass of financial resources to try to enter international competition. As the analyses carried out in this paper show, in some circumstances, the availability of substantial financial resources can be crucial to develop an initial reputation or to implement a big leap of the reputation.

In this paper, we investigate the following research questions:

1. Can we identify a relationship between the positions in the ARWU of a class of universities (e.g., top ten HEIs) and the financial resources of these universities? If so, can we state that the availability of significant funding is a critical condition to make a university a world-class HEI? If this is true, can we provide an effective estimate of the position in the ARWU (in terms of specific leagues) of a given university by considering mainly financial data?
2. Can we assess the efficiency of use of the allocated financial resources, considering the ARWU scores attained by universities?
3. Given the importance of financial data, what are the implications of the differences existing among the national accounting systems of university balance sheets?

To answer these research questions, we will analyse 318 universities from different countries in the ARWU 2019 edition. This sample contains the maximum number of universities in the ARWU 2019 for which we found the financial information necessary to conduct the analysis. For many of these universities, we downloaded university balance sheets and performed the appropriate restatement to extract the relevant items for analysis.

The analyses to answer the first question will be geared towards describing the financial data collected to identify the possible relationship with the ARWU score. To define the ranks (i.e., scores given to the classes to which a university belongs), we will use the percentiles of the ARWU Score of 2019. Then, to estimate the score using only the financial data, we will use the machine learning method called eXtreme Gradient Boosting (XGBoost, Chen and Guestrin 2016). Subsequently, to understand whether certain universities may use the financial resources with different levels of efficiency, we will use a Data Envelopment Analysis (DEA, Banker et al. 1984) approach. Finally, in the discussion, we will analyse the potential bias and implications of the heterogeneity of accounting systems for international comparison between universities.

The paper is organised as follows. In the next section the relevant literature is described, while Section 3 introduces the methodology underlying the analysis. Section 4 describes the data used for the analysis and Section 5 reports the results obtained. Section 6 discusses the results and their implications at the policy level and Section 7 concludes the paper.

2. Literature review

University rankings have received wide interest from both the academic community and policymakers as tools to assess the reputation and performance of universities. Initially conceived as information tools to provide a general indication of the quality of universities, rankings such as the ARWU or THE have transformed over time into performance assessment tools. In this sense, rankings are "performative" (Dahler-Larsen, 2011) because University' stakeholders consider them meaningful, and therefore influence their opinions, decisions, and actions. Rankings are essentially based on information elaborated according to a set of criteria to highlight real or perceived differences in performance or reputation (Merisotis & Sadlak, 2005). Due to their ability to provide classifications (ranks) and because of the increasing competition in the higher education sector and the growing demand for transparency and accountability, rankings have become increasingly used as tools to assess performance and become an essential part of decision-making processes for various stakeholders (Hazelkorn, 2018). The main advantage of rankings is the transparency about universities systems (Rauhvargers, 2013 pp. 21–23) in a competitive world market. Several criticisms on rankings have been raised concerning the selection of indicators, the weighting assigned to each indicator and the inherent limitations of the data used (see, e.g., Daraio et al. 2015, Daraio and Bonaccorsi 2017, Fauzi et al. 2020, Olcay and Bulu 2017, Moed 2017). Despite all these criticisms, however, these instruments are still used and seems to be here to stay. The literature also assessed the relation between universities' funding and rankings. Benito et al. (2019) investigated the influence of financial resources on the positioning of the top 300 universities as per the QS ranking in 2018. They found that public funding plays a key role for 84 % of the top universities, attributing up to 51 % of the variance in rankings attained by these institutions within the QS ranking. At the same time, Berne (2020) states that the performance of HEIs, gauged through the ARWU, highlights a significant correlation with the annual budget allocated per student and the tuition fees. Furthermore, Lepori et al. (2017) scrutinized the financial environment of universities as a potential determinant in the selection process. In their analysis, they identified a small group of research universities characterized by a high research volume and a much higher amount of funding than all other HEIs in the sample and thus suggest that the emergence of these HEIs is critically related to the concentration of resources. Additionally, Lepori et al. (2019) undertake a comparative assessment of American and European universities, demonstrating that a group of U.S. universities, with a considerably larger resource allocation, exhibits a markedly amplified volume of publications and citations respect to their European counterparts. This empirical evidence that global academic rankings can represent measures of wealth and therefore should be interpreted by introducing a measure of resources. For the effects *on* the financial resources, Baltaru et al. (2022) studied the resource inequalities among 102 English universities from 2008 to 2017 in relation to their positions in the Complete University Guide (CUG). The authors show how the position occupied by a university in the ranking affects all universities except those with a historically established reputation (elite universities), controlling for previous levels of financial sustainability and differences at the institutional level. They also show that the relationship is partly explained by university revenues from tuition fees. Similarly, Kim (2018) analysed the U.S. News and World Report's Best Colleges rankings to assess the relationship between rankings and the use of financial resources. The author shows how rankings led to an expansion of spending on teaching and non-teaching activities in response to the distinctive nature of the ranking system. The various criticisms drawn by rankings and the need for performance evaluation tools for universities has led to alternative methodologies being proposed in the literature (e.g., composite indicator as proposed by El Gibari et al. 2018). One of the most popular methodologies in the literature for evaluating universities is Data Envelopment Analysis (DEA, see, e.g., Emrouznejad and Yang 2017, for a literature review of DEA usage and Wolszczak-Derlacz 2017, for a review of DEA used in the efficiency analysis of HEIs).

Despite the new models proposed and the various criticisms received by various rankings, this is still used today by universities and, in some cases, even to influence decision-making in this sector.

There are several contributions in the literature analysing rankings and university performance, but few papers delve into the relationship between performance or ranking and the resources available to universities. Although some authors, such as Kim (2018) have addressed the issue of the relationship between financial resources and ranking position, none of the existing works, to the best of our knowledge, has proposed a useful tool for policymakers to predict the score in the ARWU from the financial resources available to universities even those not in the ARWU. The lack of tools for policymakers is a serious limitation of the current literature, as prior to our work it was not possible to analyze the potential on rankings, which represent international competition, of injecting financial resources into university systems.

Our work contributes to the existing literature by proposing a model for estimating a university's place in the Shanghai Ranking based on its available financial resources. We provide a machine learning-based tool for estimating a university's place in a ranking using mainly financial data. Since a significant financial provision is a necessary but not sufficient condition for high performance, we carry out an efficiency analysis to assess whether financial resources are efficiently used at university level. Finally, we discuss the implications of the existence of different accounting systems of university balance sheets for the collection and international comparison of financial data.

3. Methodology

The methodology developed has two parts. In part one, we use a machine learning technique called XGBoost to predict a university's ARWU score by financial data. In part two, we use a DEA approach to analyse the efficient use of resources considering the ARWU score obtained.

3.1. XGBoost

To predict the score, we apply the XGBoost machine learning technique (Chen & Guestrin, 2016). We highlight that the XGBoost technique allows us to develop a model which can be applied for purely predictive purposes, while it cannot be used to identify the distinct causal effect of every regressor considered in the model.

XGBoost is a machine learning technique based on ensembles of decision trees, which combines several iterations of trees to build a predictive model. Unlike other estimation methods, such as linear regression, XGBoost can model complex non-linear relationships. Moreover, XGBoost is generally robust to noisy data or the presence of outliers. An XGBoost model is trained by randomly dividing the dataset under consideration into two sub-datasets called the training dataset and the test dataset. The training dataset is used to train the model while the test set is used to evaluate the model. For our work, in line with the Pareto principle, commonly adopted in this context, the training dataset contains 80 % of the total observations of the initial dataset while the test dataset contains 20 %. To obtain a good estimation model that is robust and does not present the problem of overfitting, care must be taken when choosing the hyper-parameters of this algorithm. To determine the optimal combination of hyper-parameters, it is common to use k-fold cross-validation (i.e., dividing the **training** data into k parts).

To evaluate the model, the Root Mean Square Error (RMSE) was considered as the evaluation metric.

The RMSE formula is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}},$$

where, n is the total number of examples in the dataset; y_i is the actual value of the ith observation and $\hat{y}_i$ is the value predicted by the model for the ith observation. The model that obtained the lowest possible RMSE in cross-validation was chosen according to the various hyper-parameters considered. After the training, the RMSE obtained during the cross-validation is compared to the RMSE obtained using the test set.

The main hyper-parameters considered for our model are:

- nrounds: specifies the number of trees to be trained.
- max depth: specifies the maximum depth of the decision tree.
- eta: scales the contribution of each tree by a factor of $0 < \text{eta} < 1$ when it is added to the current approximation.
- gamma: indicates the minimum reduction in the loss function required to perform an additional division on a tree node.
- subsample: specifies the fraction of samples to be used to train each tree.
- colsample by tree: specifies the fraction of columns to be used to train each tree.
- min child weight: specifies the minimum number of samples required to split a node.

To identify the optimal combination of hyper-parameters, a grid search on 5 k-folds (of the training dataset) of hyper-parameters was used. We searched the optimal value of eta from 0.1 to 1.5 with a step of 0.1; value of gamma from 1 to 4 with a step of 1, max depth with values equal to 3 and 4, min child weight with values equal to 1, 2 and 3 and 1500 nrounds were considered, but stopping during automatically if the RMSE value did not decrease after 50 iterations. To assess the accuracy of the model, it was validated using the test dataset, which was not used in the cross-validation and training of the model itself. At the end of the testing, the final model was trained on the whole dataset.

Although the objective of the XGBoost model is the ARWU score, the goal of the developed tool is the placement of universities in one of the possible ARWU league. Estimating the ARWU score therefore allows us to subsequently assign the league to which it belongs based on the ARWU 2019 score percentiles presented in Section 4. After estimating the ARWU score of a university through the proposed XGBoost model, the tool assigns one league to the university as follows:

$$HEI\ league = \begin{cases} Diamond\ (\mathrm{T10}) & 49.76 \leq if\ ARWU\ score \leq 100 \\ Gold\ (\mathrm{T50}) & 31.72 \leq if\ ARWU\ score < 49.76 \\ Silver\ (\mathrm{T100}) & 25.34 \leq if\ ARWU\ score < 31.72 \\ Bronze\ (\mathrm{T250}) & 16.39 \leq if\ ARWU\ score < 25.34 \\ Copper\ (\mathrm{T500}) & 10.63 \leq if\ ARWU\ score < 16.39 \\ Iron\ (\mathrm{T1000}) & 0 \leq if\ ARWU\ score < 10.63 \end{cases}$$

The R package XGBoost version 1.7.5.1 (Chen et al., 2023) was used to perform the analysis.

3.2. DEA

An efficiency analysis is performed to estimate the different levels of efficiency of universities in using their financial resources. The idea is to assess how different universities have deployed their resources and whether there are universities that have deployed their financial resources better than others by scoring as high as possible in the ARWU. To do so, we estimate an *efficient frontier* against which to compare the performance of a unit with respect to its comparison set. To estimate this frontier, we decided to employ a non-parametric approach as it makes no assumptions about the distribution of inefficiencies and does not impose a functional form for the frontier (for an introduction see, e.g., Coelli et al. 2005). There are several non-parametric techniques to estimate the efficient frontier, with DEA being one of the most used. Introduced by Charnes et al. (1978), the main objective of DEA is to evaluate the performance of a given sample of Decision-Making Units (DMUs) to operate close to the efficient boundary of the production set. DEA is a non-parametric approach that can manage multiple inputs and outputs (as in the case of HEIs), but it requires the assumptions of free disposability (i.e., the possibility to destroy assets without cost) and convexity.

A DEA analysis can be conducted in an input or output orientation. Input-oriented DEA aims to analyse how many inputs can be contracted to reach the efficient frontier, while output-oriented DEA aims to maximise the outputs of each DMU given the level of inputs used.

We consider a vector of inputs $x \in R_+^p$ and a vector of outputs $y \in R_+^q$ and we can define the production set Ψ as:

$$\Psi = \left\{(\mathrm{x}, \mathrm{y}) \in R_+^{p+q} \middle| x\ can\ produce\ y\right\}$$

Ψ is the true but unknown production set. The production set estimated through DEA is denoted by $\widehat{\Psi}$. To determine the efficiency scores, we adopted an output-oriented DEA, based on a linear programming approach, thus considering, in our case, which universities, given their resources (x), scored highest in the ARWU (y). Considering this then, a university represented by the vector (x, y) can be assessed by:

$$\widehat{\lambda}(x,y)_{DEA} = \{\sup\lambda > 0 | (\mathrm{x}, \lambda\mathrm{y}) \in \widehat{\Psi}_{DEA}\},$$

Table 1
ARWU Indicators with descriptions and weights.

Indicator	Description	Weight
Alumni	total number of alumni of an institution that has won Nobel Prizes and Fields Medals	10 %
Award	total number of the staff of an institution winning Nobel Prizes in Physics, Chemistry, Medicine and Economics, and Fields Medals in Mathematics	20 %
HiCi	number of highly cited researchers selected by Clarivate Analytics	20 %
N&S	number of papers published in Nature and Science. Only "Article" type publications are considered	20 %
PUB	total number of papers indexed in the Science Citation Index-Expanded and Social Science Citation Index	20 %
PCP	weighted scores of the above five indicators divided by the number of full-time equivalent academic staff	10 %

where, $\widehat{\lambda}(x,y)_{DEA}$ is the output efficiency score of a university *(x,y)* which is the maximum radial expansion of its output (y) which is feasible given its inputs level (x). We apply Kneip et al. (2016) tests and performed the analysis using FEAR (Wilson, 2008).

4. Data

We analyse all available information from the 2019 edition of the ARWU. We focused on the ARWU 2019 edition combining it with financial data of 2019 to avoid including data on extraordinary investments and funding of universities due to the Covid-19 period, which could bias the analyses. The ARWU is one of the best-known university rankings, which analyses more than 2000 universities worldwide. The ARWU is based on six parameters, described in Table 1.

For the final position, the ARWU calculates the total score (henceforth referred to as the ARWU score) as the weighted sum of the indicators reported and described in Table 1, with their respective weights.

The ARWU scores are distributed between 0 and 100, where 100 identifies the first position in the ranking. The ARWU scores are not available on the ARWU website for universities from 100th place onwards. We have calculated the scores for the other positions based on the assigned weights reported in Table 1. As shown by the indicators used, the ARWU can be considered as a ranking that attempts to measure the excellence of research. In this sense, the ARWU can also be considered as a proxy of the research reputation of HEIs.

From the ARWU we derive some *leagues*, namely, sets of universities with similar ARWU positions. To do this, we selected the minimum ARWU score thresholds based on the percentiles of the 2019 ARWU[3]. In particular, we defined the *Top 10* (or *Diamond* or *T10* for short) league as the 99th percentile of the scores, *Top 10–50* (or G*old* or *T50* for short) league as the (99,95] percentile range, *Top 50–100* (or *Silver* or *T100* for short) as the (95,90] percentile range, *Top 100–250* (or *Bronze* or *T250* for short) as the (90,75] percentile range, *Top 250–500* (or C*opper* or *T500* for short) as the (75,50] percentile range, and *Top 500–1000* (or *Iron* or *T1000* for short) as the (50,0] percentile range. The calculated threshold values are shown in Table 2 and in Fig. 1. Any university that has reached or exceeded the calculated percentile threshold is assigned to the corresponding league.

We select 318 universities included in the ARWU 2019 for which we were able to gather specific financial information for the year 2019, from the Balance Sheets and Income Statements of universities downloaded from university websites. The final dataset includes universities from the USA (171 universities), the UK (61 universities), Italy (46 universities), Australia (32 universities), and Poland (8 universities).

The collected financial variables (converted into Euros at the exchange rate provided by the Central Bank of Europe as at 31/12/2019, BCE Exchange Ratio, 2019) are:

- The net value of long-term physical capital. It includes the net value of property, plant, equipment, machinery, vehicles, intangible assets and goodwill. It is measured by the net value of intangible and tangible fixed assets of accumulated depreciation and amortisation, and it is taken from the Balance Sheet of each university. From now on it is denoted by LtPhCa.
- The total amount of the operating expense. It is taken from the Income Statement of each university. From now on it is denoted by ToOpEx. In the case of universities with hospital activities or extra activities (e.g., due to external laboratories), costs due to these other activities are detected from the Income Statement and removed from the total operating expense to have homogeneous data.

We collect the number of students enrolled at universities (further on, NuSt for short). These data were taken from various official national databases, such as, the Integrated Postsecondary Education Data System (IPEDS) (2023) (https://nces.ed.gov/ipeds/) for the USA case, the Experts in Higher Education Data and Analysis Database (2023) (https://www.hesa.ac.uk/) for universities in the UK, USTAT (2023) (http://ustat.miur.it/) for Italian universities, the database from the official website of the Department of Education of Australian Government (2023) (https://www.education.gov.au/higher-education-statistics/resources/2019-section-2-all-students), and the European Tertiary Education Register (ETER) (2023) (https://www.eter-project.com/) for Polish universities. We also considered estimated academic staff (full professor, assistant professor, and tenure track researcher) full-time equivalent (later, FTE for short) using the method proposed by Docampo (2013). As the adopted Benchmark value for Docampo's method, CalTech's full-time equivalent staff in 2019 was taken from the IPESD database (294.5). Following that, we calculated the teaching load of academic staff

[3] We also analyzed the percentiles by considering the years between 2018 and 2021 together. The percentile are very similar.

Table 2
Thresholds to define the leagues in the ARWU.

Leagues	ARWU's percentile considered	ARWU score thresholds
T10 (Diamond)	99	49.76
T50 (Gold)	95	31.72
T100 (Silver)	90	25.34
T250 (Bronze)	75	16.39
T500 (Copper)	50	10.63
T1000 (Iron)	0	0

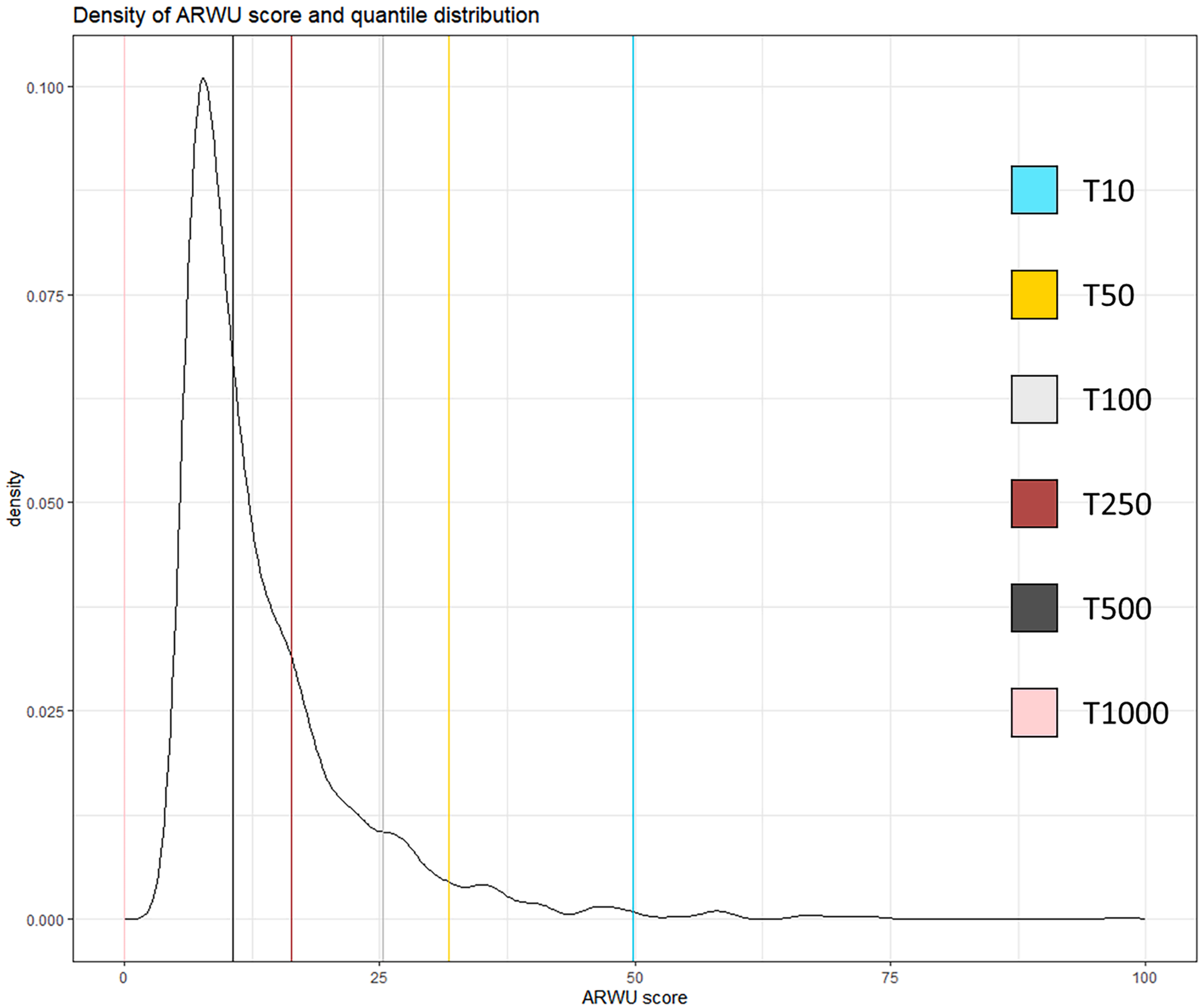

Fig. 1. Density and percentile distributions of 2019 ARWU scores.

(TL later) by considering the ratio of NuSt to FTE.

5. Results

5.1. *The relationship between funding and university performance*

As shown in Table 3, we observe there is a high correlation in our dataset between the ARWU score, ToOpEx, and LtPhCa . The ARWU score is 80 % correlated with LtPhCa, 78.8 % correlated with ToOpEx, 59 % with the FTE and is negatively correlated (−30.9 %) with the TL. It is also important to point out that the NuSt (which is often considered as a proxy for the size of the university) has a low correlation with the ARWU score, LtPhCa and ToOpEx. For this reason, we decided to not consider the NuSt in the main part of the analysis (while we still consider this variable in an easier-to-apply, but less precise, model presented in the Supplementary Materials).

Table 3
Correlation matrix between the ARWU score, ToOpex, LtPhCa, NuSt, FTE and TL.

	ARWU score	ToOpEx	LtPhCa	NuSt	FTE	TL
ARWU score	1					
ToOpEx	0.800	1				
LtPhCa	0.788	0.864	1			
NuSt	0.184	0.282	0.245	1		
FTE	0.590	0.765	0.673	0.419	1	
TL	−0.309	−0.305	−0.297	0.511	−0.365	1

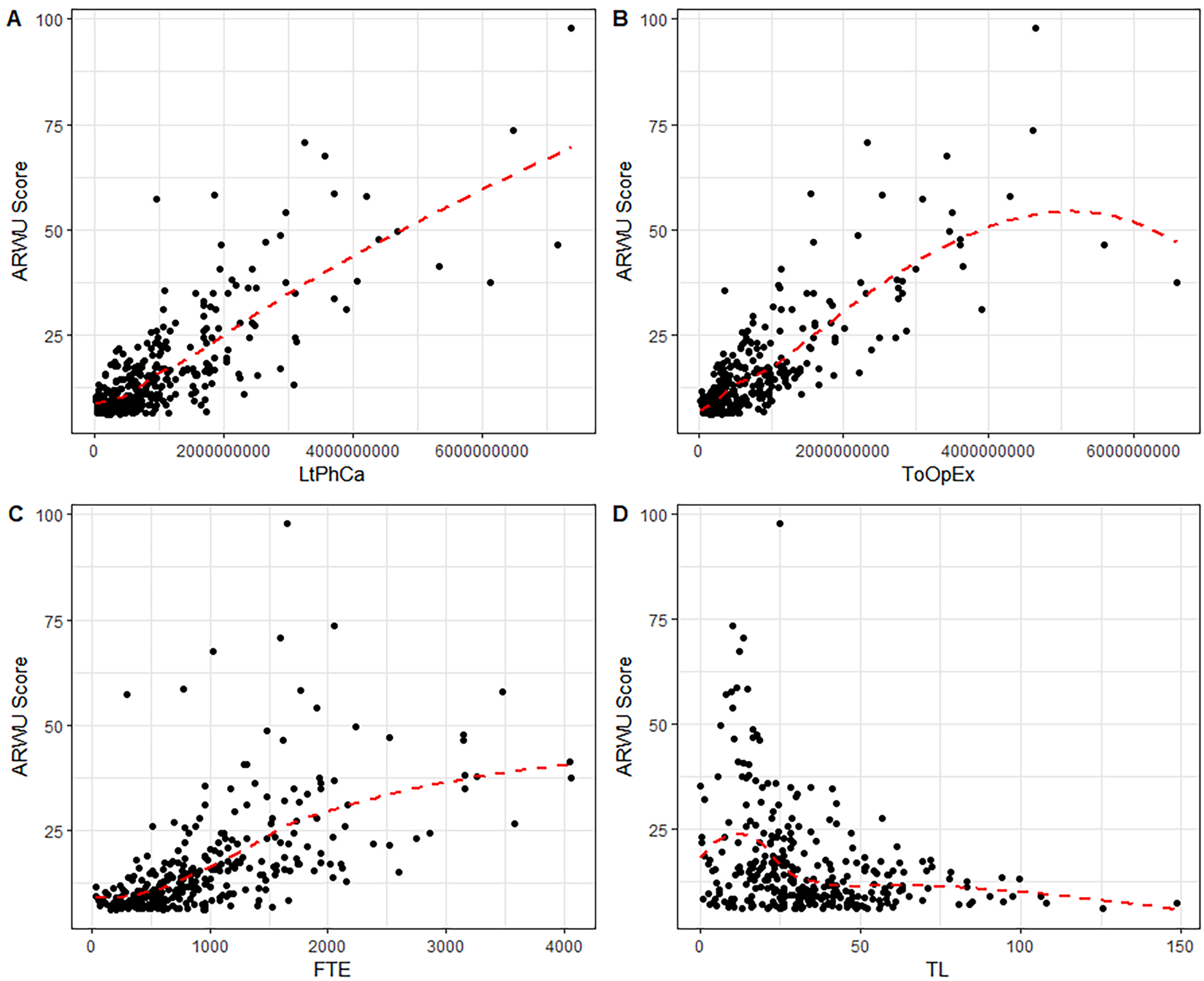

Fig. 2. Dispersion plots of all variables considered. The red line represents a loess regression line.

This high correlation could signal that there exists a relationship between financial resources and university placement into ARWU leagues. More details on the variables analysed are shown in the Appendix, which reports their kernel distribution plots. The observed correlation shows us that there is a relationship but despite that the variables are not perfectly linear as shown in the dispersion plot in Fig. 2[4]. This led us to use other method that can manage easily nonlinear relationship like XGBoost as mentioned in Section 3. Incidentally, we also observe that the score threshold related to the T10 league (reported in Table 2) is quite low (about 49.76), which means that a few HEIs get the largest share of the overall marks of all the rankings. Moreover, a similar result can be noted for the considered financial data.

Indeed, we apply the Lorenz curve (Lorenz, 1905) and the Gini index (Gini, 1936) to obtain a more in-depth description of the

[4] For further confirmation to the above, we did the linear regression, and the results are less imprecise. The final RMSE is equal to 7.15.

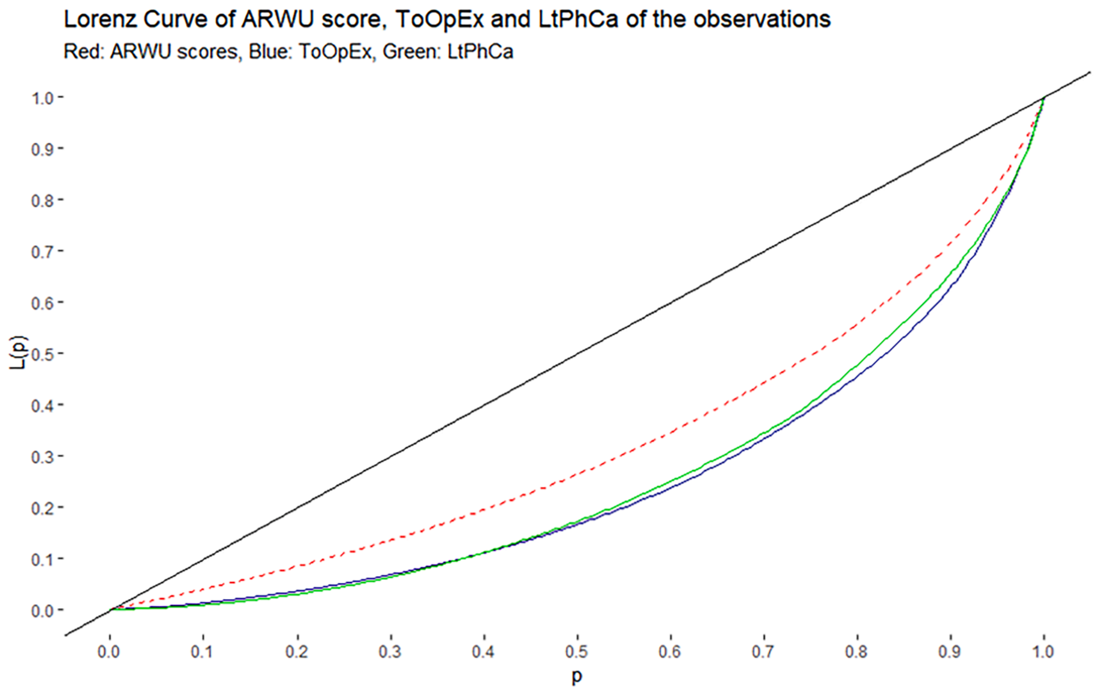

Fig. 3. Lorenz curves of the universities analysed. Red: ARWU score, blue: ToOpEx, green: LtPhCa.

concentration of the ARWU scores and the financial resources among a few universities.[5] The highly skewed distributions of ARWU score and financial resources are evidence of existing inequality among universities in terms of resources and rank (see also Figs. A1, A2, A3, A4 and A5 in Appendix). Halffman and Leydesdorff (2010) analysed the Lorenz curve and the Gini index of the ARWU score, but not for the university financial resources we considered in our analysis. As for Gini indices, for our dataset, the result is a value of 0.35 for the ARWU, a value of 0.50 for ToOpEx, a value of 0.48 for LtPhCa. As shown in the Lorenz curves in Fig. 3, and by the previously reported Gini indices, there is a moderated concentration of the ARWU scores (red curve in the graph) a big concentration of both the considered financial resources (blue and green curves in the graph for ToOpEx and LtPhCa, respectively). This empirical evidence of a highly skewed distributions and existing inequality among universities could represent further evidence of a possible relationship between the financial resources concentration and the ARWU scores of the universities.

5.2. An XGBoost model to estimate the ARWU league: a support tool for the decision maker

We apply a machine learning technique called XGBoost to model the relationship between the ARWU score and financial resources and develop a tool which can effectively predict the league for a university principally based on the allocated financial resources. We highlight that the XGBoost technique allows us to develop a model which can be applied for purely predictive purposes, while it cannot be used to identify the distinct causal effect of every regressor considered in the model.

Although the objective of the estimation model is the ARWU score, the goal of the developed tool is the placement of universities in one of the possible ARWU leagues. Estimating the ARWU score therefore allows us to subsequently assign the league to which it belongs based on the thresholds presented in Table 2. The possibility of estimating the league allows us to obtain an indication that is more useful to policymakers and less susceptible to point estimation errors in the proposed model.

To estimate the ARWU score, in addition to financial resources, we consider two further factors: FTE and TL. FTE is considered as the main productive resource within the activities of universities and is a proxy of the size of an institution, while TL is a mitigating

[5] The Lorenz curve and the Gini index are instruments used to measure large asymmetries in the distribution of a variable within a dataset. The Lorenz curve is a graphical representation that relates the cumulative percentage of the population (on the x-side) to the cumulative percentage of the variable of interest (on the y-side). The perfect equity line represents a distribution in which every individual has the same amount of the variable. Conversely, if the Lorenz curve deviates from the equity line, it indicates the presence of asymmetries in the distribution of the considered variable. In our context, if the Lorenz curve is far from the equity line, it indicates a high concentration of financial resources among a few HEIs, whereas if it is close to the equity line, it suggests a more fair distribution of the resources. The Gini index, derived from the Lorenz curve, provides a numerical measure of inequality in the dataset. It varies between 0 and 1, where, 0 represents perfect equity (all universities have the same amount of the considered resources) and 1 represents maximum asymmetries (a single HEI owns the whole amount of the resources available for the HEI industry). The Gini coefficient can be calculated as the area between the Lorenz curve and the equity line, divided by the total area below the equity line.

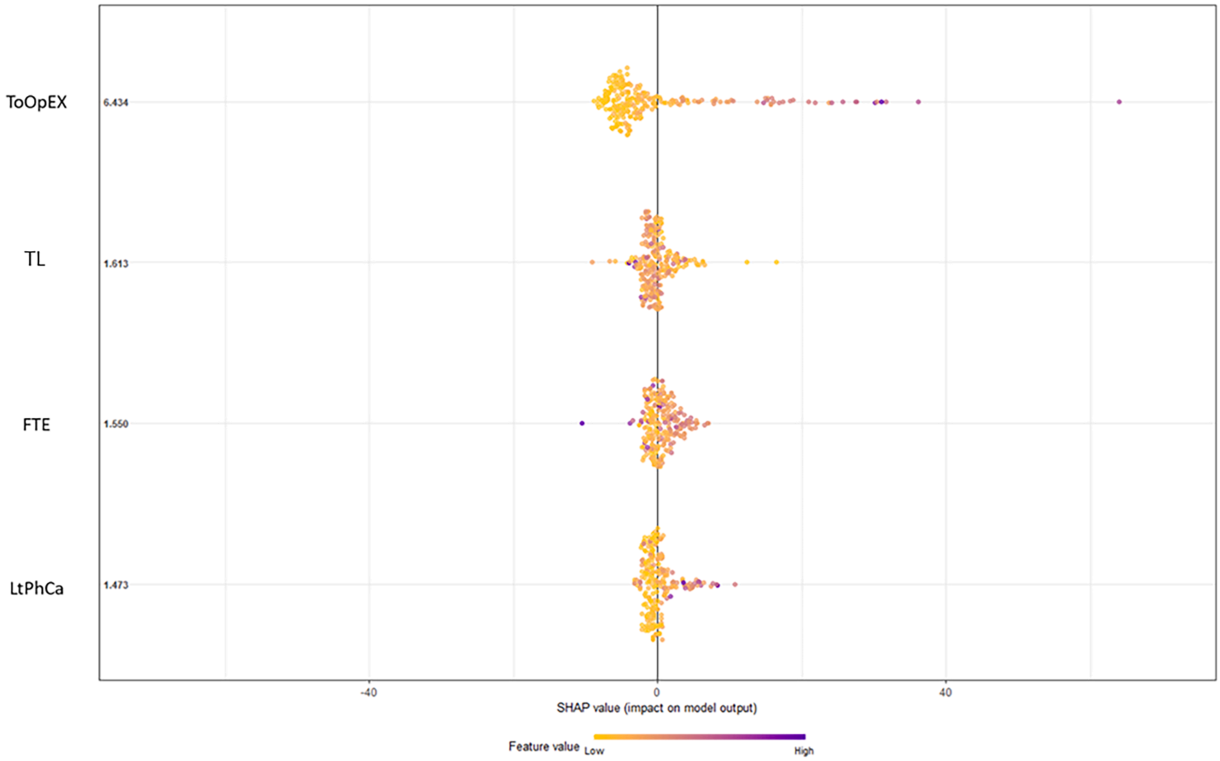

Fig. 4. SHAP value (impact on model output) of the input variables used in the XGBoost model.

variable to consider that a part of the resources (both financial and staff time) must also be devoted to the teaching mission, which is not specifically measured in the ARWU score (see Lukman et al. 2010). Therefore, a part of the resources cannot be fully used for research activities.

Therefore, we apply the XGBoost algorithm to detect possible complex non-linear relationships between the ARWU score and the predictors LtPhCa, ToOpEx, FTE and TL:

$$ARWU\ score = f(ToOpEx, LtPhCa, FTE, TL).$$

The optimal model is characterised by the following hyper-parameters: max depth = 3, eta = 0.7, min child weight = 2, colsample by tree = 1, subsample = 1, nrounds = 1500 (stopped at 116 iteration automatically), gamma = 1. The RMSE obtained during the cross validation was 0.703. During the testing phase we obtain an RMSE around 8. Over the whole dataset, we have an RMSE of 0.986.

Referring to the developed predictive XGBoost model, to assess the role of the predictors, we present the SHapley Additive ex-Planations[6] (SHAP) value in Fig. 4. The SHAP value provides a measure of the relative importance of the different variables in influencing the prediction of a model for a specific university. ToOpEx has a higher weight than the others (6.43), followed by TL (1.613), FTE (1.550) and LtPhCa (1.473).

As shown in the boxplots of the estimated scores by leagues in Fig. 5, the model predicts in general quite well. Moreover, the error made on the value of a certain score does not significantly affect the accuracy of the league assignment (see Fig. 6). The *HEI league prediction tool* is available at this website https://simonedileo.shinyapps.io/ARWU_League_Estimator/ where the reader can carry out his/her own preferred estimation exercises. A policymaker may not have FTE data available as required by our model. For this reason, we have developed a slightly less precise but easier to apply model that only considers economic resources and the NuSt as proxies for size. More technical details on the accuracy of this model and the hyperparameters adopted are reported in the Supplementary Materials.

5.2.1. *Two examples of use of our predictive model*

5.2.1.1. Merging of universities. After obtaining a model to predict the ARWU scores, we can simulate scenarios that policymakers may

[6] To calculate the SHAP values of an XGBoost model, a Shapley linear regression approach is used, which estimates the marginal contribution of each feature. In practice, several permutations of the features are generated, the prediction differences from the reference prediction are calculated and the results are combined to obtain the SHAP values.

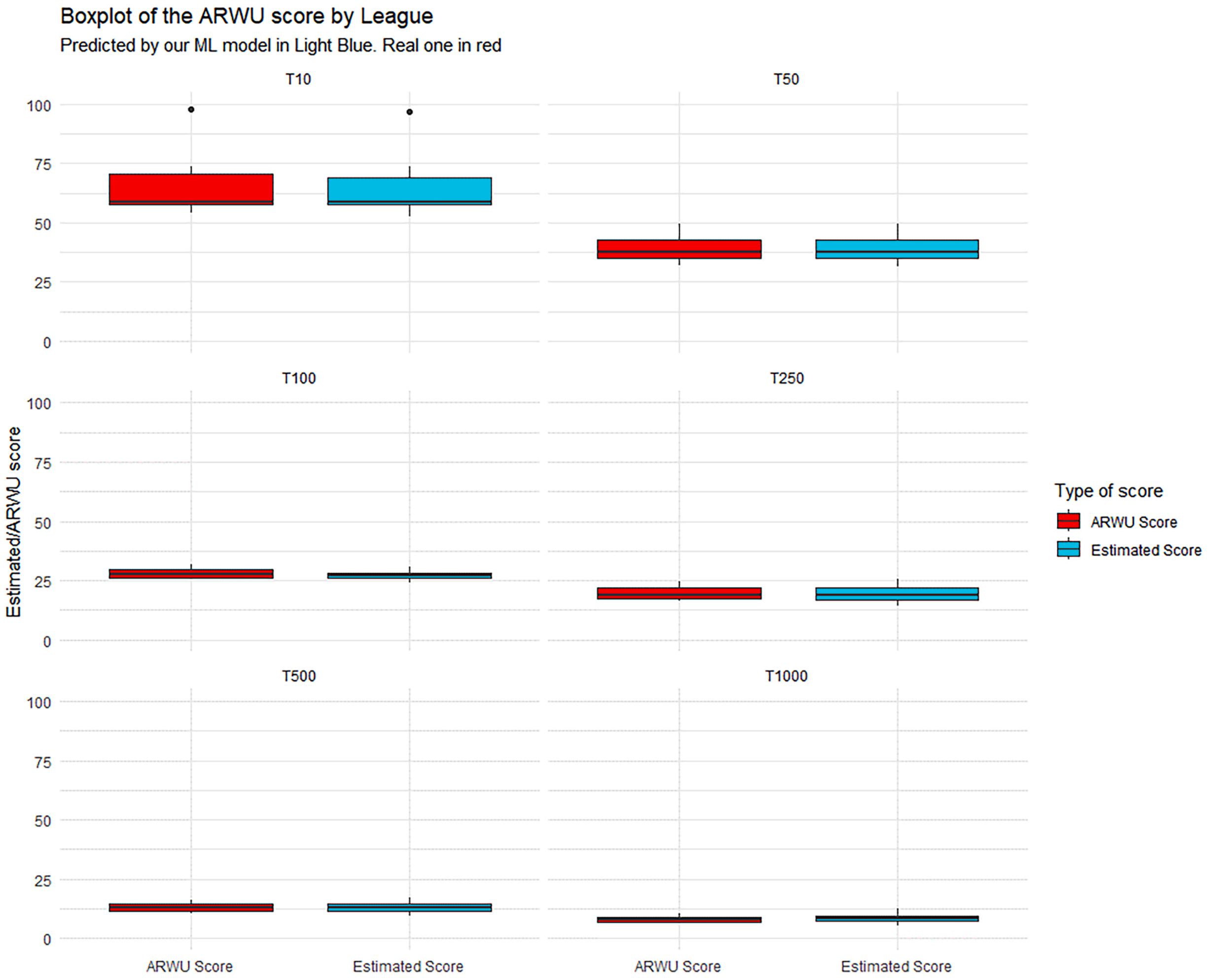

Fig. 5. Boxplots of the estimated and ARWU scores by league. The blue boxplot represents the estimated scores, the red boxplot the actual ARWU scores.

test to improve the ARWU position of desired universities. Aggregation into larger universities can lead to better exploitation of economies of scale (Moed et al., 2011) and increased visibility, which according to Frenken et al. (2017) can improve citation and publication results. An algorithm that estimates the merging effect in the case of ARWU based on the ARWU indicators themselves have been proposed before by Docampo et al. (2015). In our case, we analyze the problem from the perspective of economic resources injected by universities. The advantages of our model, based on economic and dimensional resources, over the one proposed by Docampo et al. (2015), are mainly more simplicity and transparency for the policymakers.

For example, a strategy of the aggregation of universities, such as the one pursued by French policymakers, might be pursued also in Italy, where a policymaker could be interested in assessing the impact of a merger of several universities in the same city into a single city university. Taking the city of Rome as an example, within the ARWU there are three public state universities (Sapienza, Tor Vergata and Roma Tre) that could potentially be merged into a single university. Would this union bring advantages to the Italian university system with respect to the global ARWU? Potentially, if we had access to the source data used by the ARWU, we could work out what the 'original' score of this new hypothetical Virtual University of Rome, further on called VUR, would be. In this sense the Docampo et al. (2015) method can be used to estimate the value based on the ARWU data. However, not having access to the original data but only the re-elaborated data provided by the ARWU, a policymaker can use our model to estimate the ranking score and resulting league by aggregating the financial and personnel data of the various universities. The strength of the proposed model is that by not modelling a simple linear relationship, more accurate results can be obtained. For example, as shown in Table 4, combining the three Roman universities into the *VUR*, this new university would have a score close to 24, placing it in the top 250 (T250). This aggregation would therefore favour the smaller universities (Tor Vergata and Roma Tre) that are in the T1000, but it does not improve Sapienza's league very much, which, at the same time would receive a slight increase in the ARWU score and improve the placement in the T250 without, however, bypassing to the T100. Our model's ARWU score estimates align with those derived from Docampo et al.'s (2015) method, (23.78 vs 24) presented in Table 5.

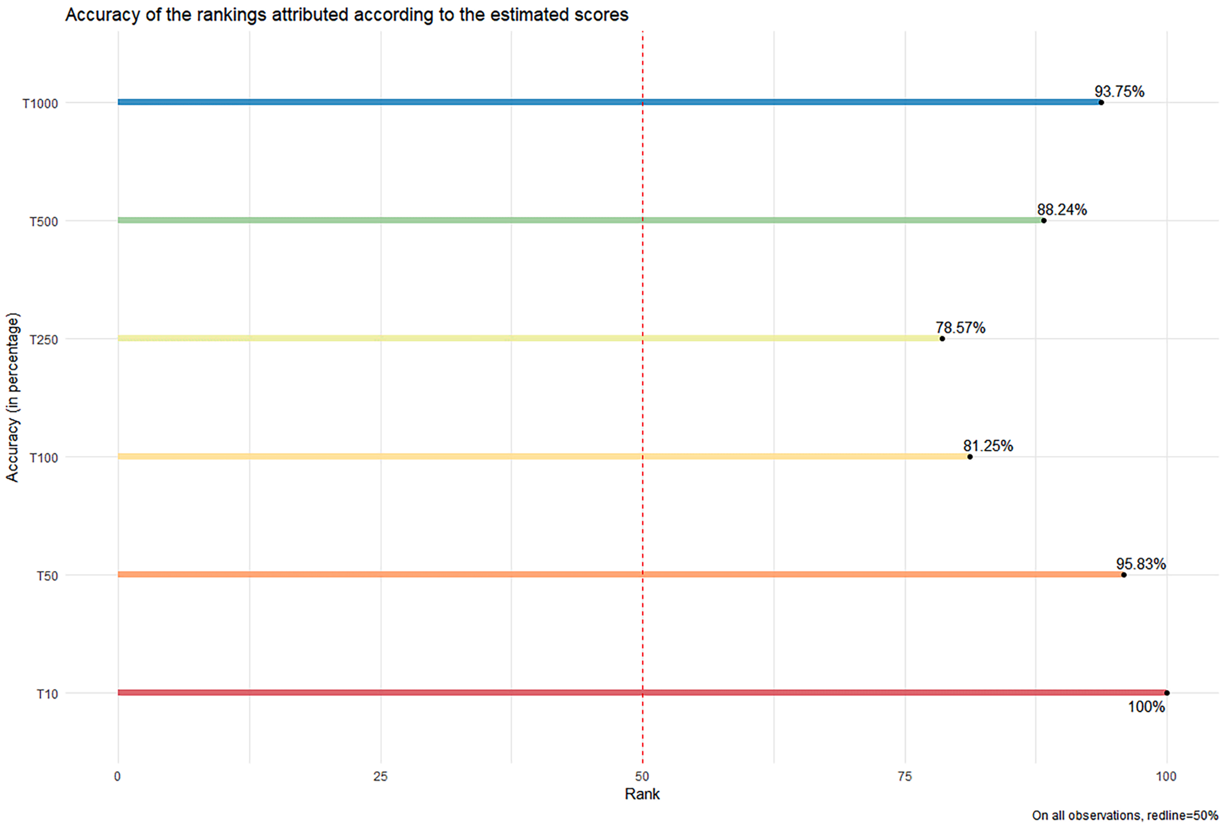

Fig. 6. Accuracy of the estimated score to assign the correct league (on all the observations in our dataset).

Table 4
Simulation of ARWU scores through our model. Aggregation of the existing 3 Roman universities into a VUR characterised by the sum of the various predictors (NuSt, FTE, ToOpEx, and LtPhCa) of the universities Roma Tre, Tor Vergata and Sapienza.

University	TL	FTE	LtPhCa	ToOpEx	ARWU score	XGBoost Score	Real Rank	XGBoost Rank
Tor Vergata	41.53	734.82	602,509,493	287,813,536	10.13	10.71	T1000	T500
Roma Tre	51.93	604.26	243,375,238	180,057,645	6.19	7.35	T1000	T1000
Sapienza	61.49	1828.53	320,524,753	695,173,064	21.03	21.9	T250	T250
VUR	55.04	3167.6	1,166,409,484	1,163,044,246	–	23.78	–	T250

Table 5
Estimation of the ARWU scores and leagues using Docampo et al. (2015).

University	Alumni	Award	HiCi	S&N	PUB	PCP	ARWU score	Estimated rank
Tor Vergata	0	0	0	5.1	35.6	19.9	10	T1000
Roma Tre University	0	0	0	6	18.9	12.1	6	T1000
Sapienza	9.7	13.1	14.7	11.4	51.2	19.8	21	T250
VUR (using Docampo et al. 2015)	9.7	13.1	14.7	13.9	65.2	18.6	24.7	T250
VUR (using Docampo et al. 2015 and considering the degree of cooperation)	9.7	13.1	14.7	12.9	62.7	17.9	24	T250

5.2.1.2. Necessary resources to rank-up a university. Another use case for the Italian policymaker may be to simulate the effect that increasing the resources of just one university has on its own ranking position.[7] Taking the Scuola Normale Superiore di Pisa as an example (henceforth, Normale for short) we can evaluate different scenarios. For example, if FTE and NuSt raise by 10 times while

[7] This simulation could be useful in the case whereby Italian policymaker would implement a strategy like the one applied by Chinese government for Tsinghua University or other selected universities (Fedasiuk et al., 2022).

Table 6
Normale university simulation scenarios.

Case	TL	FTE	LtPhCa	ToOpEx	ARWU score	XGBoost score	Real rank	XGBoost rank
Normale (base case)	9.00	34.75	49,397,017	45,645,073	12	10.76	T500	T500
Scenario 2: increase 10 times of FTE and NuSt, **20 times of** ToOpEx and LtPhCa	9.00	347.5	987,940,340	912,901,460	–	23.20	–	T250
Scenario 3: increase 15 times of FTE and NuSt, 35 times of ToOpEx and LtPhCa	9.00	521.25	1,728,895,595	1,597,577,555	–	30.02	–	T100
Scenario 4: increase 15 times of FTE and NuSt, 40 times of ToOpEx and LtPhCa	9.00	521.25	1,975,880,680	1,825,802,920	–	34.89	–	T50

LtPhCa and ToOpEx by 20 times, then Normale would reach league T250 in the ARWU, (Scenario 2 in Table 6). Instead, in the case that FTE and NuSt increase by 15 times while LtPhCa and ToOpEx raise by 35 times, Normale would reach the T100 league (Scenario 3 in Table 6). In the latter case, if LtPhCa and ToOpEx increment by up to 40 times, Normale could reach league T50 (Scenario 4 in Table 6). This example result can be further investigated by the policymaker to assess whether, based on the resources necessary to be allocated, and considering the improvement in ranking, such a large increase in resources is worthwhile or not. Moreover, the Normale case also shows how the proposed tool could be applied to design policy experiments related to resource allocation for the HEI industry. Indeed, the tool would allow a policymaker to estimate the appropriate level of investment which is required by a university to climb the league ladder. We would like to underline again that ours is a predictive model that may also be subject to errors due to the dataset on which the model was trained. In addition, considering the method used to create the model, it is also possible that an increase in a specific resource may not necessarily lead to an increase in the score.

5.3. *Efficiency and financial resources*

We apply a DEA output-oriented analysis to analyse the performance of universities in obtaining the highest ARWU score (output y), given their financial resources, FTE and TL (considered as inputs, x). We are interested in studying, within the various leagues, which universities are using resources efficiently, maximising their ARWU score. Those efficient universities may exploit better an increase in their resources compared to inefficient universities. We apply a classic DEA VRS analysis (considering the suitable tests performed). The results of the DEA VRS output oriented analysis are shown in Fig. 7. The Figure shows several boxplots coloured according to the leagues. The results mainly show two things: (i) at a higher rank there is also, on average, higher efficiency in the use of resources, and (ii) there are efficient universities in the various rank that have already reached their “maximum” ARWU score with the available resources. Based on this result we then have two possible implications. Firstly, universities with higher efficiency scores are also those that can use their financial resources better. Secondly, some universities in low leagues (e.g., T1000), being already efficient in the use of the resources currently available to them, might reap great benefit from a potential increase in their resources.

6. Discussion

Financial accounting data is crucial for universities and should be systematically collected and analyzed to gain insights into their performance and competitive positioning. In a globalized environment, universities face increasing international competition for students, research funding, and global recognition (Stephan, 2012). To achieve world-class status, universities prioritize key areas such as teaching, research, knowledge transfer, and global engagement. However, the availability of financial resources can significantly impact these key pillars, ultimately determining their competitive standing in the global higher education landscape.

We analysed 318 international universities in the 2019 ARWU edition, identifying a relationship between their economic financial resources and their place in the ARWU. Then, we partitioned the ARWU into a few leagues, and we developed a predictive model to estimate the ARWU score and the corresponding league through the university’s resources. Our findings seem to suggest that a critical condition to make a world-class university is to provide the university with a high level of financial resources. Our predictive model is available online at https://simonedileo.shinyapps.io/ARWU_League_Estimator/.

Subsequently, we evaluated the efficiency of universities in achieving higher ARWU score across various leagues. The analysis revealed that universities with higher scores consistently demonstrate efficient resource utilization. Furthermore, some universities in lower leagues could significantly benefit from increased financial resources, as they are already adept at managing their limited resources effectively.

Our findings have some policy implications. Our predictive model could be used by national agencies or ministries of education to compute the financial resources required to induce an improvement in terms of the ARWU for specific universities. Furthermore, our tool could be used to design a more effective and fairer way of distributing national financial resources among HEIs. As Stephan (2012) points out, the relevant problem of the optimal allocation of resources to universities has not yet been rigorously addressed. Our results could provide conceptual and empirical support for the application of alternative and complementary methods to the currently applied methods regarding university and research funding. Generally, the most widely used methods are based on bibliometric indicators that contribute to increasing disparities between universities due to the so-called “Matthew effect”, a cumulative effect that tends to allocate resources to institutions that already have many resources. The alternative methods identified in Ioannidis (2011) as “egalitarian (fund everybody)” and “aleatoric (fund at random)” should be considered by policymakers more as complementary methods to

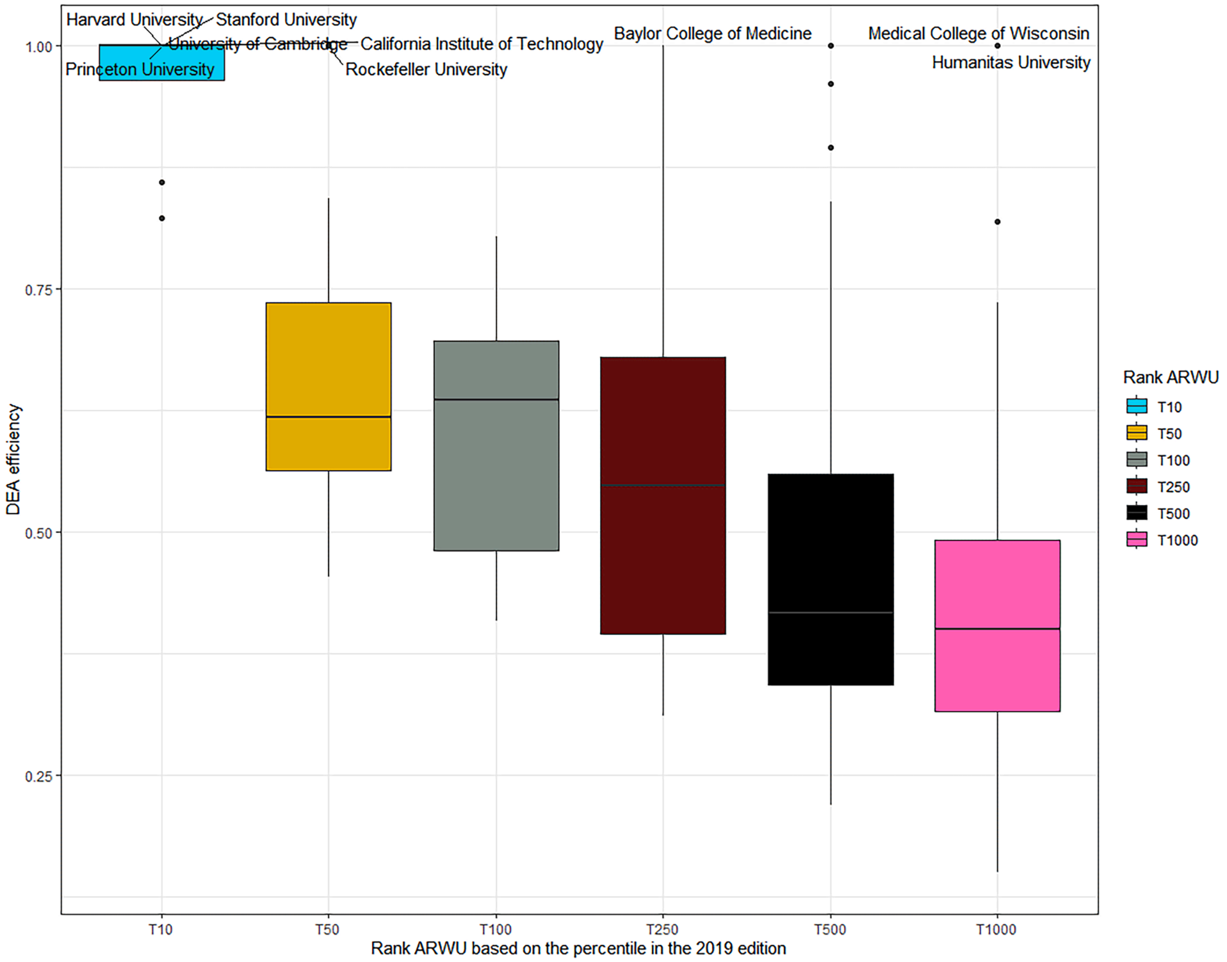

Fig. 7. Boxplots of the efficiency scores of universities grouped by ARWU league. The universities whose names are reported in the top of the boxplots are the efficient universities (whose DEA score is equal to 1).

those generally adopted and based on bibliometric indicators.

However, despite the potential usefulness of our proposed model and predictive tool, and the results obtained, still shares the main limitations of the existing rankings (see, e.g., Daraio and Bonaccorsi 2017, and Fauzi et al. 2020). Furthermore, we do need to consider major limitations, which could imply other policy issues, namely, the availability of university financial data and the lack of harmonisation of accounting standards. Indeed, there are not many international databases available with the financial data of various universities, and national accounting systems can make it difficult and time-consuming to collect many homogeneous financial data to be effectively used in an international comparison. In fact, the selection of the 318 universities analysed has been affected by the availability of financial data.

First, it would be desirable to enrich the current HEI databases with additional financial data (or create new ones from scratch). A relevant example of a dataset that collects financial data is the American Database for HEIs, namely, the IPEDS. Another interesting example, even if not as extensive and detailed as IPEDS, is represented by the European Tertiary Education Register (ETER). Nowadays, ETER financial data are limited only to Expenditures and Revenues, while no data on Assets, Liabilities and Cash Flows are provided. In this sense, ETER could take inspiration from IPEDS and be extended with many other financial data.

Furthermore, to promote the harmonisation of accounting standards, some issues should be addressed while collecting the financial data of HEIs. Here we list the main issues analysed during the work, which may be useful for a discussion concerning the worldwide higher education industry:

1. Some universities do not own all the resources they use. Depending on the accounting rules, the value of not-owned resources may or may not be reported in the balance sheet.
2. Some resources used by universities may belong to the responsible public authority in charge of the university. This authority may provide these resources at the given university's disposal free of charge or at a symbolic low charge. As a result, the corresponding expenditures do not appear in the income statement or appear with a low value. On the contrary, some universities (especially the

old ones) use premises situated in buildings listed as historical monuments and under special protection. In this case, costs in the income statement are much higher than for premises in new buildings.
3. Intangible and tangible fixed assets are stated net of accrued depreciation in the balance sheet. However, certain universities may apply different depreciation methods and/or a different useful life estimate for these fixed assets.
4. Some universities are multi-businesses. There is the main business of providing higher education and research services, but there can also be the business of providing health services or developing specific technologies for a public institution (see, e.g., California Institute of Technology). In this case, assets, funds, revenues, costs, cash in-flows and cash out-flows can be jointly provided by the accounting system and thus impossible to separate ex post.
5. There are two different ways of updating the values of intangible and tangible fixed assets, namely, cost method and fair value[8]. In principle, different universities may apply the cost method or fair value for intangible and tangible fixed assets.
6. Some costs may depend on national laws, such as different tax regimes and compulsory insurances. Therefore, expenses in the financial statement may be partially misrepresented by these items.
7. Under an accrual-based accounting system, a transaction is recorded when it occurs, without waiting until cash is received or paid out. Therefore, cash inflows/outflows and revenues/costs are distinct concepts.[9]

7. Concluding remarks

In this paper, we answered to the research questions described in the Introduction. We investigated the relationship between the positions in the ARWU of top HEIs and their financial resources. We found that the availability of significant funding is a critical condition to make a university a world-class HEI and we developed a predictive model, available online, to estimate the position in the ARWU (in terms of specific leagues) of a given university by considering mainly its financial data. The analysis performed is purely predictive and does not look for causality between funding and reputation. This is a limit of our research. As stated in the paper, the relationship between reputation and availability of financial resources is a 'chicken and egg' problem, i.e., the availability of financial resources can be crucial to develop an initial reputation or to implement a major leap forward. Through some examples, we showed that the efficient use of the allocated financial resources might be a sufficient condition for a policy that aims at improving the ranking score of a university by concentrating financial resources.

Although we have adopted all criteria to avoid overfitting the XGBoost model, one limitation affecting the model is the dataset itself. This, in addition to being a limitation of any predictive model, highlights a problem of international data collection and standardization that can be solved by following what is proposed in Section 6. In fact, comparing financial data from different accounting systems can present some problems and approximations. Therefore, financial data in international official databases should be collected in such a way to mitigate as far as possible the problems described in the previous section, and policymakers should push as much as possible towards a harmonisation of accounting standards within the worldwide higher education industry.

Funding

The financial support of the Sapienza University of Rome, research award no. RM11916B8853C925 and no. RM12117A8A5DBD18, is gratefully acknowledged.

CRediT authorship contribution statement

Alessandro Avenali: Conceptualization, Data curation, Supervision, Validation, Writing – original draft, Writing – review & editing. **Cinzia Daraio:** Conceptualization, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. **Simone Di Leo:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Joanna Wolszczak-Derlacz:** Conceptualization, Data curation, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The author **Cinzia Daraio** is a member of the Editorial Board of Journal of Informetrics.

[8] Under the cost method, assets enter in accounting records at the price paid to acquire them and are not re-evaluated (except for depreciation), while the historical cost of the asset is net of accrued depreciation (as determined at the beginning). Instead, the fair value approach is as follows. At the beginning, the historical cost is applied. Then, at the end of established periods the initial value of the asset is updated with a rational estimate of the potential market price of the asset. Essentially, the last updated value of the asset is net of accrued depreciation.

[9] For instance, payments to suppliers for goods received are not costs, simply cash outflows, while costs are incurred when these goods (as productive inputs) are consumed in production processes. Similarly, clients do not always pay in cash for what they buy. In this case, the organisation registers a revenue, but it is owed money and has an account receivable (it will receive money later). Cash for the purchasing of an asset is not a cost. Costs are the corresponding depreciations. Organisations do not generally pay in cash for materials. If this is the case, the organisation has an account payable (it will pay money later). However, some universities rely on a cash-based accounting system, which records income when cash is received and costs when cash is paid.

Acknowledgments

A previous version of this paper was presented at the XXX Meeting of the Economics of Education Association, 30 June-1 July 2022, Porto (Portugal), and at the XXXIII AiIG Scientific Meeting of the Italian Association of Management Engineering (RSA AiIG 2022), Rome (Italy), 20–21 October 2022. We thank the conference participants for their useful comments and discussions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2024.101502.

Appendix: Kernel density plots

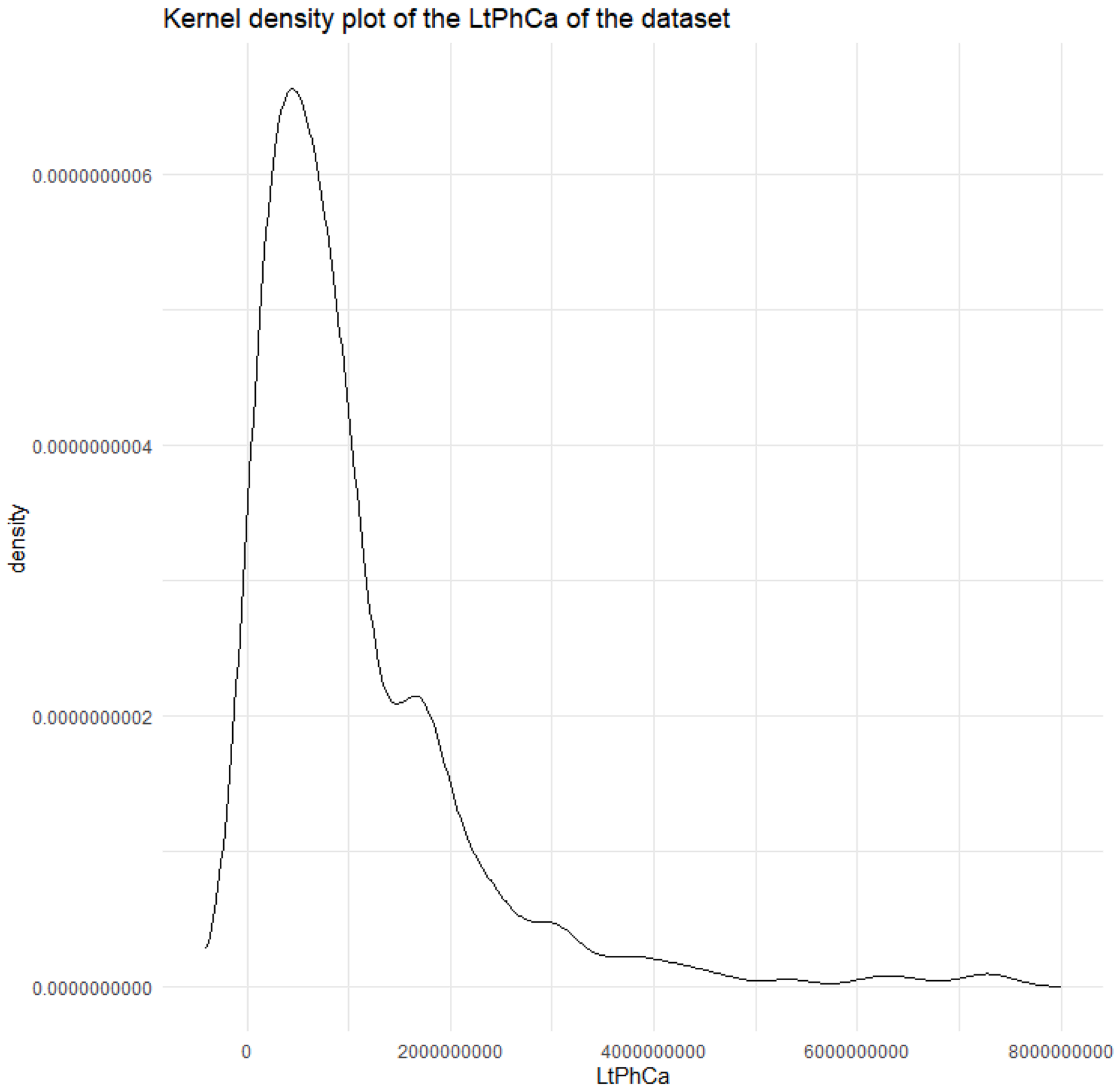

Fig. A1. Kernel density plot of LtPhCa (Number of observations 318).

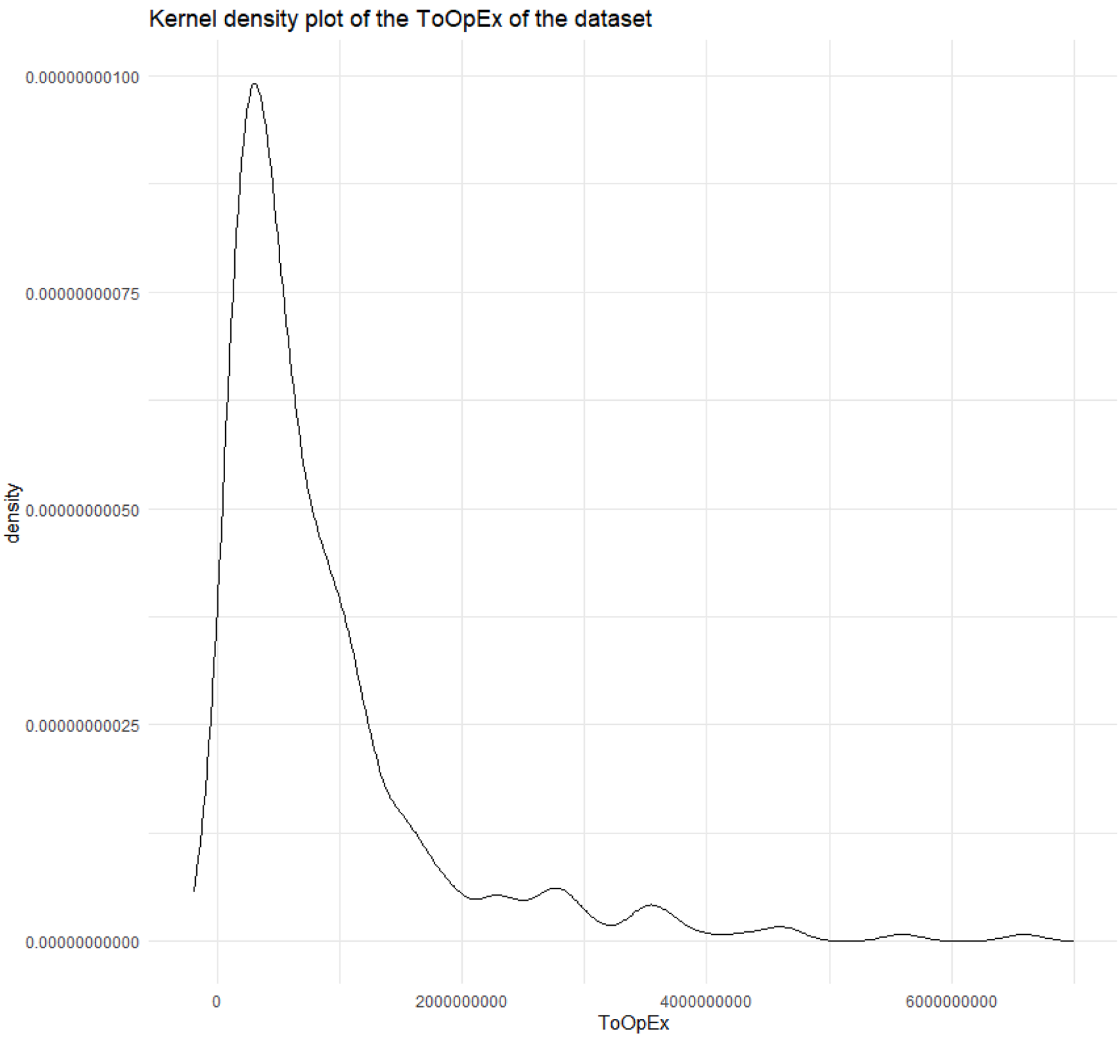

Fig. A2. Kernel density plot of ToOpEx (Number of observations 318).

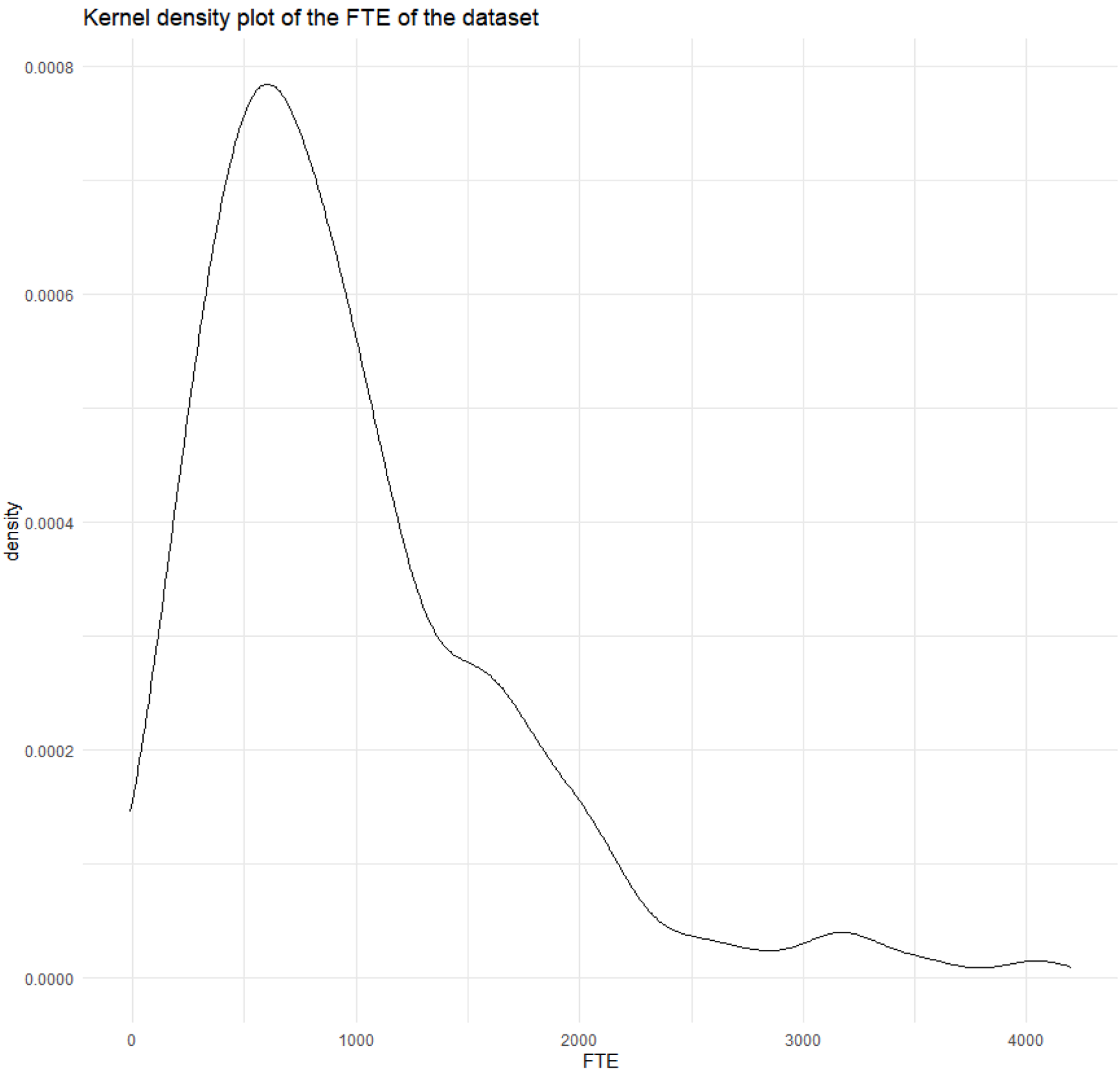

Fig. A3. Kernel density plot of the FTE (Number of observations 318).

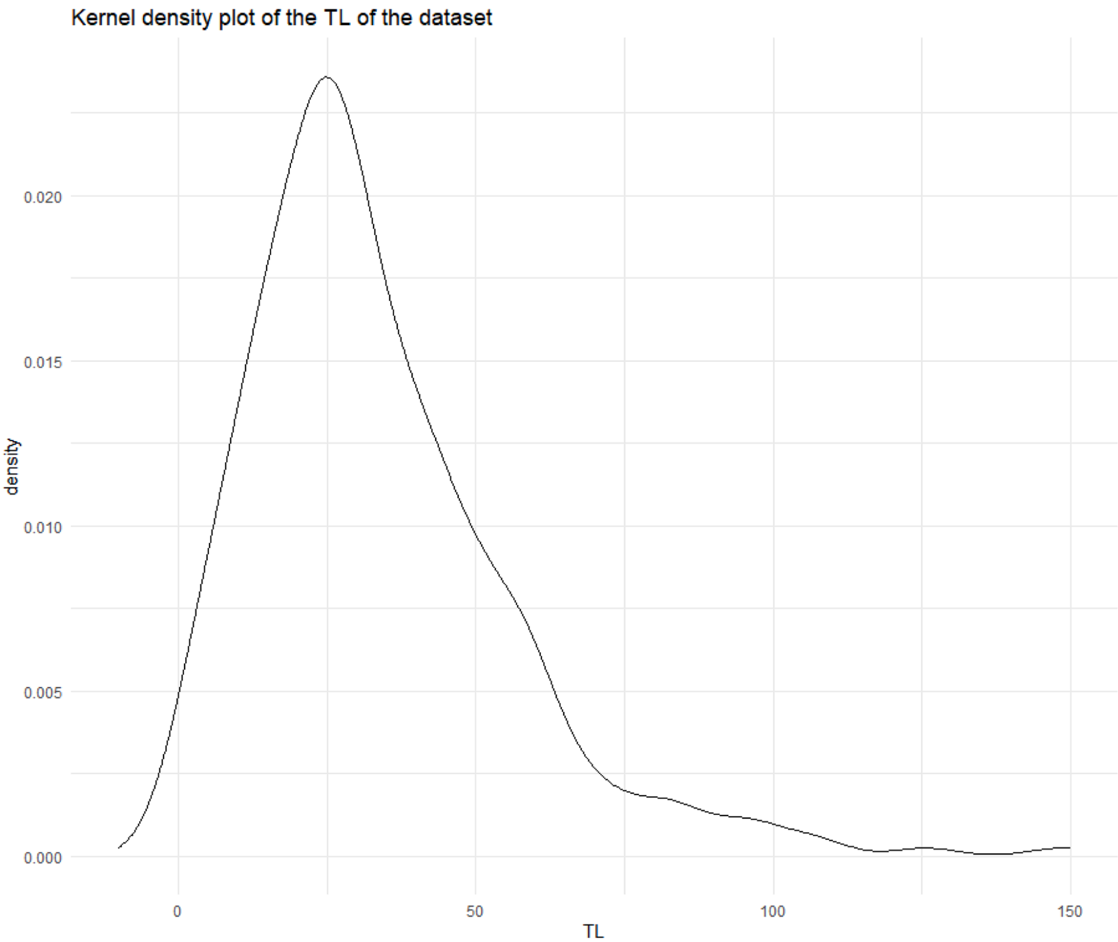

Fig. A4. Kernel density plot of the TL (Number of observations 318).

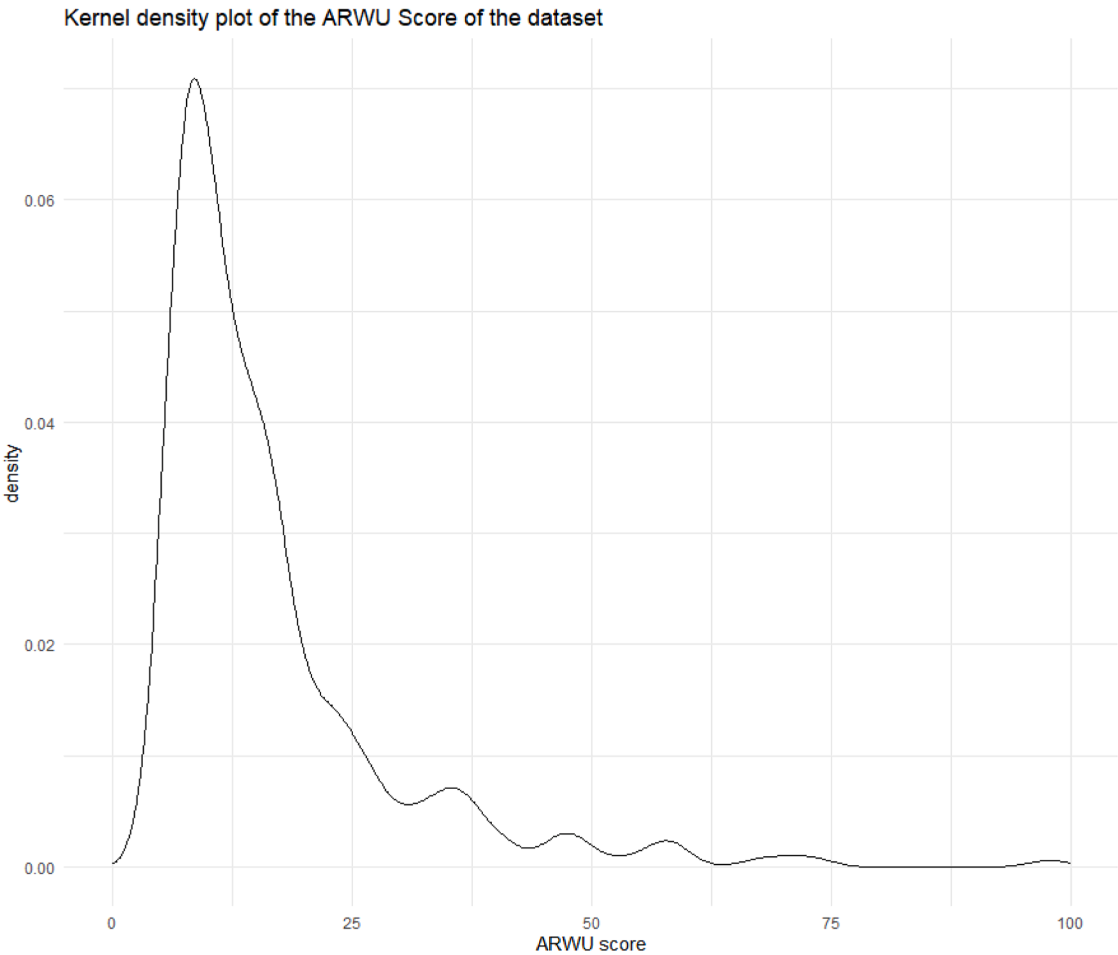

Fig. A5. Kernel density plot of the ARWU score (Number of observations 318).

References

ARWU Ranking 2019 edition, https://www.shanghairanking.com/rankings/arwu/2019, last access 18/05/ 2023.

Baltaru, R. D., Manac, R. D., & Ivan, M. D. (2022). Do rankings affect universities' financial sustainability?–financial vulnerability to rankings and elite status as a positional good. *Studies in Higher Education, 47*(11), 2323–2335.

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.

BCE exchange ratio at 31/12/2019, https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html, last access 18/ 05/ 2023.

Benito, M., Gil, P., & Romera, R. (2019). Funding, is it key for standing out in the university rankings? *Scientometrics, 121*(2), 771–792.

Berne, O. (2020). What does the Shanghai Ranking really measure? Working paper. 2020. ffhal02918290.

Cai, L. N. (2009). The story of academic ranking of world universities. *International Higher Education, 54*, 2–3, 2009.

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

Chen T., He T., Benesty M., Khotilovich V., Tang Y., Cho H., Chen K., Mitchell R., Cano I., Zhou T., Li M., Xie J., Lin M., Geng Y., Li Y., Yuan J. (2023). _xgboost: Extreme Gradient Boosting_. R package version 1.7.5.1, < https://CRAN.R-project.org/package=xgboost >.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer science & business media.

Dahler-Larsen, P. (2011). *The evaluation society*. Stanford University Press.

Daraio, C., & Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology, 68*(2), 508–529.

Daraio, C., Bonaccorsi, A., & Simar, L. (2015). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research, 244*, 918–930.

Department of education of Australian Government, https://www.education.gov.au/higher-education-statistics, last access 30/5/ 2023.

Di Carlo, F., Modugno, G., Agasisti, T., & Catalano, G. (2019). Changing the accounting system to foster universities' financial sustainability: First evidence from Italy. *Sustainability, 11*(21), 6151.

Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of university systems. *Scientometrics, 86*(1), 77–92.

Docampo, D. (2013). Reproducibility of the Shanghai academic ranking of world universities results. *Scientometrics, 94*(2), 567–587.

Docampo, D., Egret, D., & Cram, L. (2015). The effect of university mergers on the Shanghai ranking. *Scientometrics, 104*, 175–191.
El Gibari, S., Gómez, T., & Ruiz, F. (2018). Evaluating university performance using reference point based composite indicators. *Journal of Informetrics, 12*(4), 1235–1250.
Emrouznejad, A., & Yang, G. (2017). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences, 61*(1), 1–5.
Etzkowitz, H., & Zhou, C. (2017). *The triple helix: University–industry–government innovation and entrepreneurship*. Routledge.
European Tertiary Education Register (ETER), https://www.eter-project.com/#/home, last access 18/05/ 2023.
Experts in higher education data and analysis database, https://www.hesa.ac.uk/data-and-analysis, last access 18/05/ 2023.
Fauzi, M. A., Tan, C. N. L., Daud, M., & Awalludin, M. M. N. (2020). University rankings: A review of methodological flaws. *Issues in Educational Research, 30*, 79–96.
Fedasiuk, R., Loera Martinez, A. O., & Puglisi, A. (2022). *A competitive era for China's universities - How Increased funding is paving the way*. Center for Security and Emerging Technology. Data Brief, 2022.
Frenken, K., Heimeriks, G. J., & Hoekman, J. (2017). What drives university research performance? An analysis using the CWTS Leiden Ranking data. *Journal of Informetrics, 11*(3), 859–872.
Gini, C. (1936). On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series, 208*(1), 73–79.
Halffman, W., & Leydesdorff, L. (2010). Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. *Minerva, 48*, 55–72.
Hazelkorn, E. (2018). Reshaping the world order of higher education: The role and impact of rankings on national and global systems. *Policy Reviews in Higher Education, 2*(1), 4–31.
Hazelkorn, E., & Mihut, G. (2021). *Research handbook on university rankings: Theory, methodology, influence and impact*. eds. Edward Elgar. Cheltenam (UK).
Ioannidis, J. (2011). Fund people not projects. *Nature, 477*(7366), 529–531.
Kim, J. (2018). The functions and dysfunctions of college rankings: An analysis of institutional expenditure. *Research in Higher Education, 59*, 54–87. https://doi.org/10.1007/s11162-017-9455-1
Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypothesis in nonparametric models of production. *Journal of Business and Economic Statistics, 34*, 435–456.
Lepori, B., Geuna, A., & Mira, A. (2019). Scientific output scales with resources. A comparison of US and European universities. *PLOS One, 14*(10), Article e0223415.
Lepori, B., Geuna, A., & Veglio, V. (2017). A typology of European Research Universities. Differentiation, layering, and resource distribution. differentiation, layering and resource distribution (December 9, 2016). In *, 1. Proceedings of the SWPS*.
Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association, 9*(70), 209–219.
Lukman, R., Krajnc, D., & Glavič, P. (2010). University ranking using research, educational and environmental indicators. *Journal of Cleaner Production, 18*(7), 619–628.
Macheridis, N., & Paulsson, A. (2021). Tracing accountability in higher education. *Research in Education, 110*(1), 78–97.
Merisotis, J., & Sadlak, J. (2005). Higher education rankings: Evolution, acceptance, and dialogue. *Higher Education in Europe, 30*(2), 97–101.
Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics, 110*(2), 967–990.
Moed, H. F., de Moya-Anegón, F., López-Illescas, C., & Visser, M. (2011). Is concentration of university research associated with better research performance? *Journal of Informetrics, 5*(4), 649–658.
Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change, 123*, 153–160.
Rauhvargers, A. (2013). *Global university rankings and their impact: Report II* (pp. 21–23). European University Association. Brussels.
Stephan, P. (2012). *How economics shapes science*. Harvard University Press.
The Integrated Postsecondary Education Data System (IPEDS), https://nces.ed.gov/ipeds/, last access 18/05/ 2023.
USTAT, http://dati.ustat.miur.it/, last access 18/05/ 2023.
Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-economic Planning Sciences, 42*(4), 247–254.
Wolszczak-Derlacz, J. (2017). An evaluation and explanation of (in) efficiency in higher education institutions in Europe and the US with the application of two-stage semi-parametric DEA. *Research Policy, 46*(9), 1595–1605.