

ISSN 2281-4299



DIPARTIMENTO DI INGEGNERIA INFORMATICA
AUTOMATICA E GESTIONALE ANTONIO RUBERTI



SAPIENZA
UNIVERSITÀ DI ROMA

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

Alessandro Avenali
Cinzia Daraio
Simone Di Leo
Joanna Wolszczak-Derlacz

Technical Report n. 05, 2023

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

Email: avenali@diag.uniroma1.it [Alessandro Avenali], daraio@diag.uniroma1.it [Cinzia Daraio, ORCID: 0000-0002-4825-0071], dileo@diag.uniroma1.it [Simone Di Leo], jwo@zie.pg.gda.pl [Joanna Wolszczak-Derlacz, ORCID: 0000-0002-3392-5267]

Address: ^aDepartment of Computer, Control and Management Engineering (DIAG), Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy.

^bFaculty of Management and Economics, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

*Corresponding author: dileo@diag.uniroma1.it

Abstract

This study investigates the relationship between university financial resources, applied accounting systems, and the place of a university in the Shanghai Ranking (one of the most known global rankings). 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) for predicting a university's ranking from 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 confronted with a drastic choice: in order to increase international competition among universities, it is necessary to concentrate a huge amount of resources on a few universities that

are already high in the rankings. In contrast, a 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.

Keywords: University rankings, ARWU, financial data, predictive machine learning-based tools, XGBoost, accounting systems

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 & 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 (called triple helix, see, e.g., Etzkowitz and Zhou, 2017). University rankings usually aim to assess the performance of HEIs by collecting and elaborating 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 & 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, despite their methodological shortcomings, 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. In particular, 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), 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. The ARWU is based on a few indicators: alumni of an institution winning Nobel Prizes and Fields Medals, staff of an institution winning Nobel Prizes and Fields Medals, highly cited researchers, papers published in Nature and Science, papers indexed in Science Citation Index-Expanded and Social Science Citation Index, per capita academic performance (i.e., the weighted scores of the previous five indicators divided by the number of full-time equivalent academic staff). 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

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).

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 could be more or less efficient in using these financial resources.

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 ranking (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, taking into account 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. In order to define the ranks (i.e., scores given to the classes to which a university belongs), we will further investigate the ARWU score by considering different editions of the ARWU. Then, to estimate the score using only the financial data, we will use the machine learning method called eXtreme Gradient Boosting (XGBoost, Chen & Guestrin, 2016). Subsequently, to understand whether certain universities may use the financial

resources with different levels of efficiency, i.e., to assess the efficiency in the use of the resources across universities, 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

Performance evaluation regarding universities has become a fundamental practice in the context of higher education. We can define the performance evaluation of universities as a systematic, structured and objective process used to assess the overall performance of an academic institution. This evaluation could involve various aspects, such as research, teaching, technology transfer, social impact, and governance, among others. It is necessary to assess the performance of universities for several reasons. First, evaluation helps to guarantee the quality of higher education by ensuring that universities meet acceptable standards of teaching, research, and student services. Secondly, evaluation promotes accountability and transparency by providing an independent mechanism to assess the performance of universities and disclose the results to students, public funders, and society at large. This definition of performance evaluation is not unambiguous and there is much debate in the literature about what it is necessary to "*measure*" and "*evaluate*" regarding university performance (for more details see Asif & Searcy, 2014).

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. 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). This transformation has sparked debate about the limitations of rankings as comprehensive indicators of academic quality and the possible influence they may have on the strategies and priorities of universities. Several criticisms 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 & Bonaccorsi, 2017; Fauzi *et al.* 2020; Olcay & Bulu, 2017; Moed, 2017). In addition to methodological and data criticism of the rankings, the literature also assessed their impact on the finances of universities and that of finances on the rankings. Benito *et al.* (2019) study the influence of funding on the position achieved by the top 300 universities in the QS ranking in 2018. They show that public funding is critical for 84 percent of the top universities and that funding explains up to 51 percent of the variability in the positions achieved by universities in the QS ranking. Similarly, Berne (2020) shows that the performance of HEIs, as measured by the ARWU, is essentially related to the wealth of institutions (expressed in terms of annual budget per student) and the amount of tuition paid by students. Lepori *et al.* (2017) analysed the financial environment of HEIs as a possible selection mechanism. They identified a small group of research universities characterised by a high level of research volume and a much higher amount of funding than all other HEIs in the sample; thus, they suggest that the emergence of these HEIs is critically related to resource concentration. Additionally, Lepori *et al.* (2019) compare U.S. and European HEIs and show that the group of U.S. universities, which receive a much larger amount of resources, has a much higher number of publications and citations than the corresponding European universities; therefore, they empirically demonstrate that international 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 (2017) 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, including the escalation of spending on academic and student services 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. 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. Despite this, no model has yet been proposed to estimate the possible place or score in a ranking given the financial resources available at university level. This work contributes to the existing literature by proposing a model for estimating a university's place in the Shanghai Ranking on the basis of 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 applied has two components. In part one, we first analyse the relationship between the positions in the ARWU of a class of universities and the financial resources of these universities. After that, we provide an effective estimate of the position in the ARWU (in terms of the class) of a given university by considering only financial data. In part two, we show how allocating a similar level of funding to a different university may change the ranking position depending on the level of efficiency of that university.

For the score prediction (part one), we use a machine learning technique called XGBoost, while for part two we use a DEA approach.

3.1 XGBoost

The eXtreme Gradient Boosting algorithm (XGBoost, Chen & Guestrin, 2016) 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 latter 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. Overfitting is a common problem in predictive models, where the model fits the training data too well but has poor generalisation capability. The different hyper-parameters available in the XGBoost algorithm allow overfitting to be mitigated. To determine the optimal combination of hyper-parameters, it is common to use k-fold cross-validation. Cross-validation is a technique for evaluating model performance on training data by dividing the dataset into k equal parts (folds). In each iteration, one of the k parts is used as the test set, while the remaining k-1 parts are used as the

training set. This process is repeated k times, so that each part of the dataset is used as a test set exactly once.

During k -fold cross-validation, the different training sets are used to train the model with different combinations of hyper-parameter values. Subsequently, the model is evaluated on the corresponding test sets. At the end of the k iterations, a reliable estimate of the model's performance on unseen data is obtained. The grid search examines all possible combinations of specified values for the hyper-parameters. 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 i -th observation and \hat{y}_i is the value predicted by the model for the i -th observation. The model that obtained the lowest possible RMSE in cross-validation was chosen according to the various hyper-parameters considered.

The main hyper-parameters considered for our model are:

- rounds: specifies the number of iterations, i.e., the number of trees to be trained.
- max depth: specifies the maximum depth of the decision tree.
- eta controls the learning rate: scales the contribution of each tree by a factor of $0 < \text{eta} < 1$ when it is added to the current approximation. Used to prevent overfitting by making the boosting process more conservative.
- 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 of hyper-parameters was used. We searched the optimal value of eta from 0.005 to 0.5 with a step of 0.005; gamma equal to 0 or 1, max depth with values equal to 2, 3 and 4, and a very large number of rounds (1500) was considered,

but stopping during cross-validation automatically if the RMSE value on the testing set 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. 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 given their resources. 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 a number of non-parametric techniques to estimate the efficient frontier, with Data Envelopment Analysis (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.

Charnes *et al.* (1978), in their DEA version, have modelled the frontier with what is commonly called Constant Returns to Scale (CRS), which assumes that entities operate at optimal scale. To model the cases in which DMUs are allowed to adapt their production to achieve maximum efficiency, the DEA Variable Returns to Scale (VRS) was later introduced by Banker *et al.* (1984). Formally, 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:

$$\Psi = \{(x, y) \in R_+^{p+q} | x \text{ can produce } y\}.$$

Ψ is the true but unknown production set. The production set estimated through DEA is denoted by $\hat{\Psi}$. 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 financial resources (x), scored highest in the ARWU (y). Considering this then, a university represented by the vector (x, y) can be assessed by:

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

where $\hat{\lambda}(x, y)_{DEA}$ 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). To assess whether the dataset analysed is convex, the test proposed by Kneip *et al.* (2016) was used. In addition, we employed the return to scale test proposed by the same authors to determine the DEA version (CRS or VRS) to be adopted. The tests were performed using the R package FEAR, version 3.1 (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 and publishes its annual assessment of the top 1000. The ARWU was initially published in June 2003 by the Center for World-Class Universities (CWCU) at Shanghai Jiao Tong University and 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.

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%

Table 1 ARWU Indicators with descriptions and weights. Source: <https://www.shanghairanking.com/methodology/arwu/2019>

The ARWU score is 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 ranking we derive some *leagues*, namely, sets of universities with similar ARWU positions. To do this, we selected minimum ARWU score thresholds based on the percentiles of the 2018, 2019, 2020, 2021 and 2022 ARWUs. 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 *Gold* 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 *Copper* 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. Any university that has reached or exceeded the calculated percentile threshold is assigned to the corresponding league.

Leagues	ARWU score thresholds
T10 (Diamond)	52.62
T50 (Gold)	31.86
T100 (Silver)	25.23
T250 (Bronze)	16.43
T500 (Copper)	10.63
T1000 (Iron)	0

Table 2 Thresholds to define the leagues in the ARWU. Source: own elaboration based on ARWU score percentile, 2018-2022

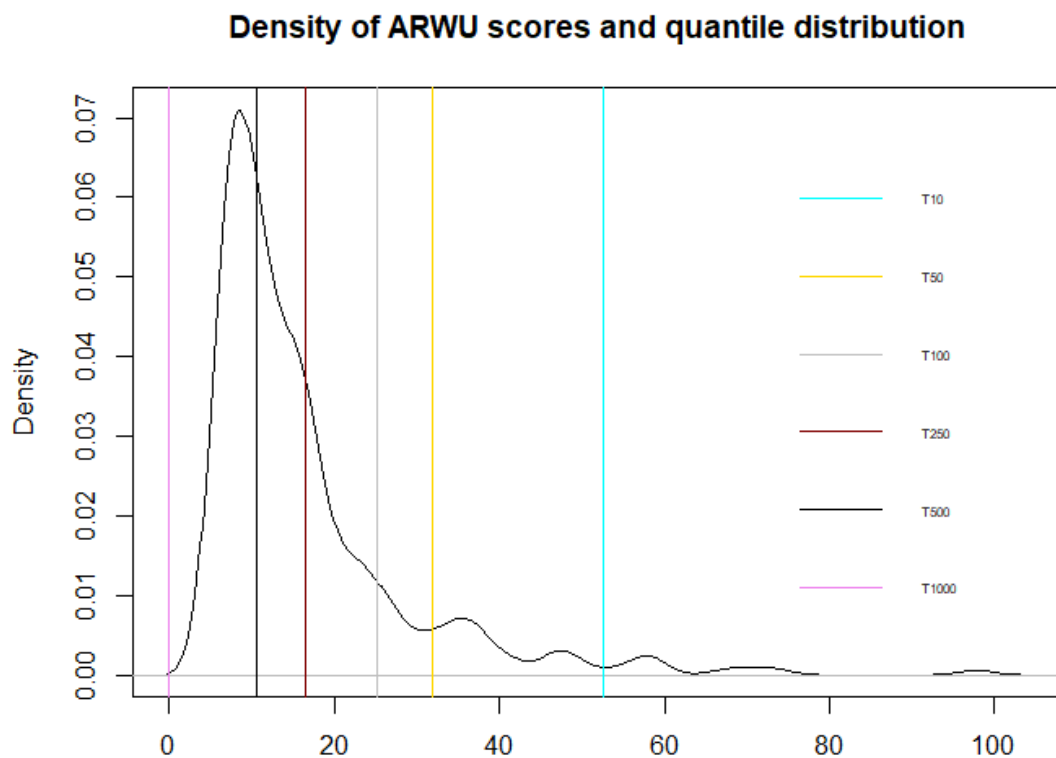


Figure 1 Density and quantile distributions of ARWU scores. Source: own elaboration based on ARWU score percentile, 2018-2022

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 are:

- The net value of long-term physical capital, namely, the long-term assets that assist the production processes of the university to run their institutional mission. Essentially, it includes the net value of property (e.g., land, buildings and real estate), plant, equipment, machinery, vehicles, intangible assets (e.g., software licenses, databases) 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. It is also called net capital assets and from now on it is denoted by LtPhCa. All values were converted into Euros at the exchange rate provided by the Central Bank of Europe as at 31/12/2019.
- The total amount of the operating expense, namely, the ongoing costs for traditional mission-related operational activities which the universities incur in a year to run their institutional activities. It is taken from the Income Statement of each university. It is also called total operational expenditure and 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 in order to have homogeneous data. All values were converted into Euros at the exchange rate provided by the Central Bank of Europe as at 31/12/2019.

We also 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, <https://nces.ed.gov/ipeds/>) for the USA case, the database by the experts in UK higher education data (HESA, <https://www.hesa.ac.uk/>) for universities in the UK, USTAT (<http://ustat.miur.it/>) for Italian universities, the database from the official website of the Department of Education of the Australian Government for Australian universities

(<https://www.education.gov.au/higher-education-statistics/resources/2019-section-2-all-students>), and the European Tertiary Education Register (ETER, <https://www.eter-project.com/>) for Polish universities.

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, total operating expense, and net long-term physical capital. The ARWU scores are 80% correlated with net capital assets and 78.8% correlated with total operating costs. It is also important to point out that the number of students (which is often considered as a proxy for the size of the university) has a low correlation with the ARWU score, LtPhCa and ToOpEx.

	ARWU score	ToOpEx	LtPhCa	NuSt
ARWU score	1			
ToOpEx	0.788	1		
LtPhCa	0.800	0.854	1	
NuSt	0.18	0.282	0.245	1

Table 3 Correlation matrix between the ARWU score, ToOpex, LtPhCa and NuSt. Source: own elaboration based on ARWU score and balance sheets data

This high correlation could signal that there exists a strong linear relationship between financial resources and university placement into ARWU leagues. Incidentally, we also observe that the score threshold related to the T10 league (reported in Table 2) is quite low (about 52.62), 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 both the considered financial data, namely, total operating expense and net long-term physical capital. More details on the variables analysed are shown in Appendix A, which reports their kernel distribution plots.

Indeed, we apply the Lorenz curve (Lorenz, 1905) and the Gini index (Gini, 1936) to obtain a more in-depth description of the possible concentration of the ARWU scores as well as of financial resources among a few universities.² Halffman & 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, the result is a value of 0.35 for the ARWU score, a value of 0.50 for ToOpEx and a value of 0.48 for LtPhCa. As shown in the Lorenz curves in Figure 2, and by the previously reported Gini indices, there is a big concentration of the ARWU scores (red curve in the graph) as well as of both the considered financial resources (blue and green curves in the graph for total operating expense and net long-term physical capital, respectively). These resulting similar trends could represent further evidence of a possible relationship between the financial resources and the ARWU scores of the universities (as previously pointed out by the correlation matrix in Table 3).

² 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.

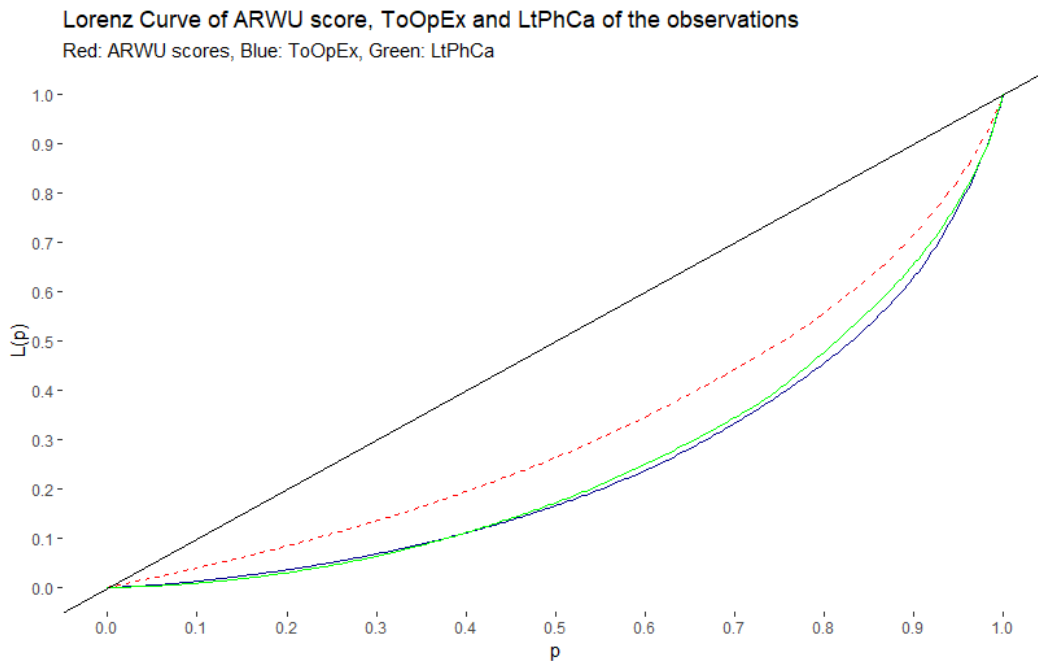


Figure 2 Lorenz curves of the universities analysed. Red: ARWU score, blue: ToOpEx, green: LtPhCa. Source: own elaboration based on ARWU score and balance sheets data

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 is able to effectively predict the league for a university principally on the basis of 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 ultimate goal of the developed tool is actually 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 on the basis principally of the available financial resources, we apply the XGBoost algorithm to detect possible complex non-linear relationships between the ARWU score and the predictors LtPhCa, ToOpEx and NuSt:

$$ARWU \text{ score} = f(ToOpEx, LtPhCa, NuSt) .$$

Despite the low correlation, the number of students is considered in the predictive model as a proxy for university size. The optimal model is characterised by the following hyper-parameters: max depth = 4, eta = 0.23, min child weight = 1, colsample by tree = 0.8, subsample = 0.8, nrounds = 62, gamma = 0. Such hyper-parameters are obtained by applying a cross-validation with 3 folds (resulting in an RMSE of 6.92). After plotting the model on the training dataset, we obtain an RMSE of 1.46, which rises to around 8 in the test phase. In addition, over the whole dataset, we have an RMSE of 3.84.

After estimating the ARWU score of a university through the proposed XGBoost model, the tool assigns one league to the university as follows:

$$HEI \text{ league} = \begin{cases} \text{Diamond (T10)} & 52.62 \leq \text{if } ARWU \text{ score} \leq 100 \\ \text{Gold (T50)} & 31.86 \leq \text{if } ARWU \text{ score} < 52.62 \\ \text{Silver (T100)} & 25.23 \leq \text{if } ARWU \text{ score} < 31.86 \\ \text{Bronze (T250)} & 16.43 \leq \text{if } ARWU \text{ score} < 25.23 \\ \text{Copper (T500)} & 10.63 \leq \text{if } ARWU \text{ score} < 16.43 \\ \text{Iron (T1000)} & 0 \leq \text{if } ARWU \text{ score} < 10.63 \end{cases}$$

Referring to the developed predictive XGBoost model, to assess the role of the predictors, we use the SHapley Additive exPlanations³ (SHAP) value shown in Figure 3. 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. It is worth noting that the ToOpEx variable has a higher weight than the other variables (4.64), followed by LtPhCa (3.384) and NuSt (1.333). Although our model is purely predictive, the importance of the considered regressors shows us that financial variables (ToOpEx, LtPhCa, NuSt) have in general a larger impact on the final score than size (NuSt).

³ 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.

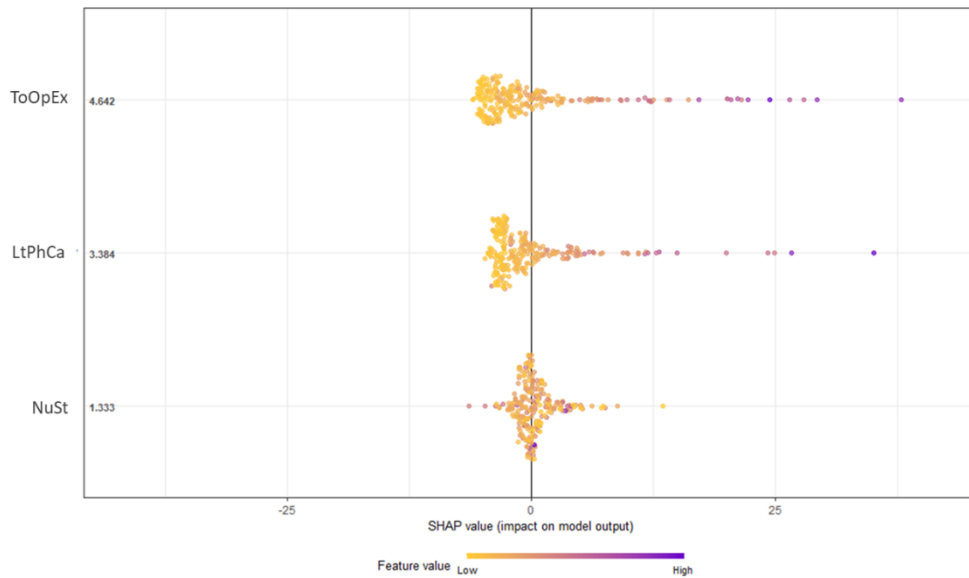


Figure 3 SHAP value (impact on model output) of the input variables used in the XGBoost model. Source: own elaboration.

As shown in the boxplots of the estimated scores by leagues in Figure 4 (on both the test and training dataset), 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. Indeed, our model predicts university leagues very well, for the majority of the universities predicting the correct league, and, in the case of an error, allocating the universities to an adjacent league (i.e., to the previous or next league). The model is less accurate in assigning universities between rank 100 and rank 250, but it still estimates more than 56% of the universities correctly (see Figure 5). The *HEI league prediction tool* is available on this [website](#) where the reader can carry out his/her own preferred estimation exercises.

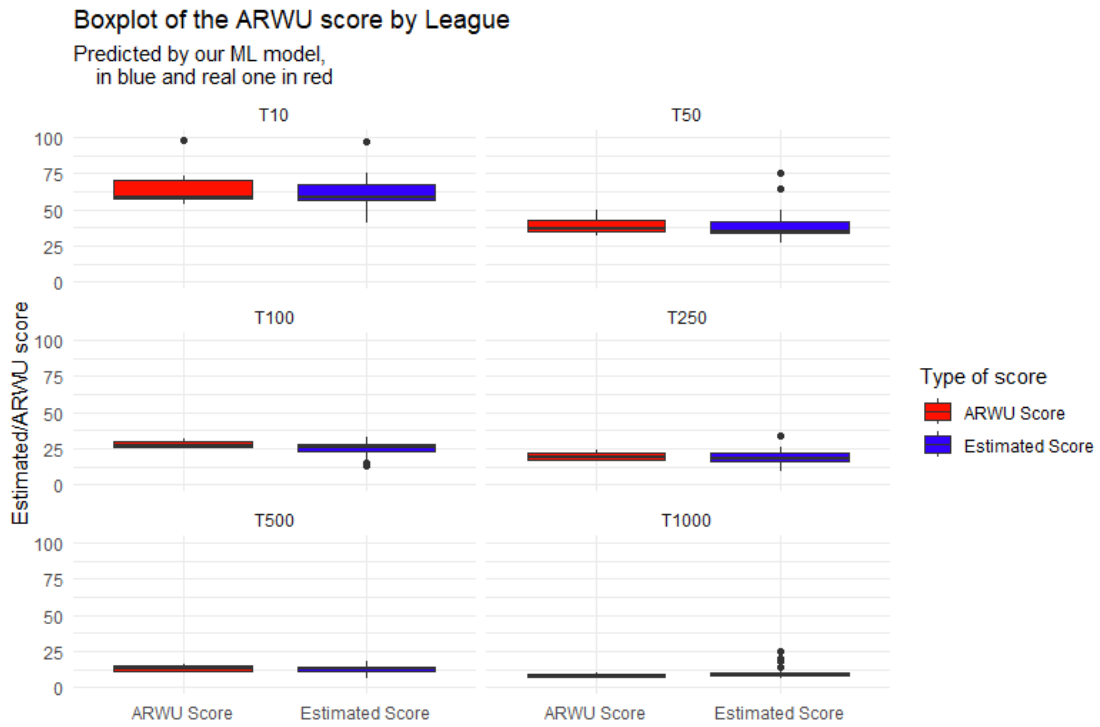


Figure 4 Boxplots of the estimated and ARWU scores by league. From the top: Diamond (T10), Gold (T50), Silver (T100) Bronze (T250) Copper (T500) Iron (T1000): the blue boxplot represents the estimated scores, the red boxplot the actual ARWU scores. Source: own elaboration.

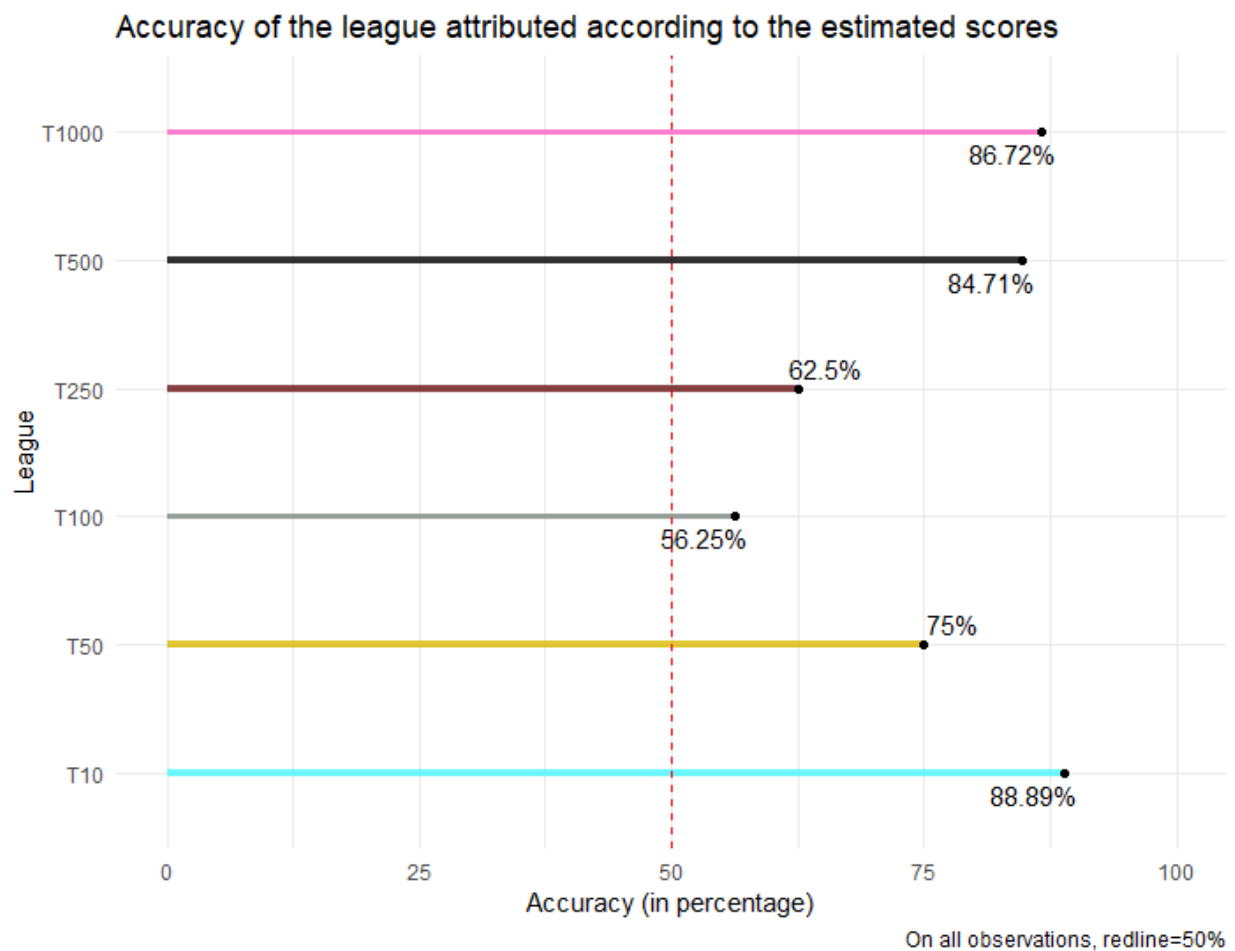


Figure 5 Accuracy of the estimated score to assign the correct league (on all the observations in our dataset). Source: own elaboration.

5.2.1 Two examples of use of our predictive model

After obtaining a model to predict these scores from financial resources and size, we can simulate scenarios that policymakers may test to improve the ARWU position of desired universities. For example, a strategy of the aggregation of universities, such as the one pursued by French and German 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 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 ranking? 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 university of Rome, further on called Virtual Rome University, would be.

However, not having access to this, a policymaker can use our model to estimate the ranking score and resulting league by aggregating the financial and students data of the various universities. The strength of the proposed model is that by not modelling a simple linear relationship (i.e., as resources increase, so does output), more accurate results can be obtained. For example, as shown in Table 4, combining the three Roman universities into a single Virtual Rome University, this new virtual university would have a score close to 15, placing it in the top 500 (T500 league). This aggregation would therefore favour the smaller universities (Tor Vergata and Roma Tre) that are in the T1000 league but would disadvantage Sapienza that is in the T250 league.

University	NuSt	LtPhCa	ToOpEx	Real Score	Estimated Score	Real Rank	Estimated Rank
University of Roma - Tor Vergata	30,519	602,509,493	287,813,536	10	10	T1000	T1000
Roma Tre University	31,381	243,375,238	180,057,645	6	7	T1000	T1000
Sapienza	112,436	320,524,753	695,173,064	21	20	T250	T250
Virtual Rome University (aggregation of the 3 previous universities)	174,336	1,166,409,484	1,163,044,246	-	16	-	T500

Table 4 Simulation of ranking scores through our model. Aggregation of the existing 3 Roman universities in a Virtual Rome University characterised by the sum of the various predictors (number of students, total operating expenses, and net long-term physical capital) of the universities Roma Tre, Tor Vergata and Sapienza. Source: own elaboration based on our predictive model.

Another use case for the Italian policymaker may be to simulate the effect that increasing the resources of one university has on its ranking position. Taking the Scuola Normale Superiore di Pisa as an example (henceforth Normale for short), which is one of the *efficient* universities in terms of ranking position with respect to the financial resources used, as we will see in Section 5.3, we can evaluate different scenarios. For example, what would happen to the ranking score if we doubled the costs and properties of the university but kept the same number of students? Along the same lines, other scenarios can be simulated to see when there is a possible increase in the predicted score. For example, as shown in Table 5, to have an increase in the ARWU league, it would be necessary for the Normale to maintain the same number of students (thus remaining the same size) but increase total operational expenditures and net capital assets by 7 and 14 times, respectively (Case 3 in Table 5). 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 (14 times the original situation) 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 minimum level of investment which is required by a university to climb the league ladder (e.g., to go from Copper to Bronze league).

Case	University	NuSt	LtPhCa	ToOpEx	Real Score	Estimated Score	Real Rank	Estimated Rank
1	Normale (base case)	313	49,397,017	45,645,073	12	11	T500	T500
2	Normale ToOpEx*7 (7 times higher) LtPhCa*7 (7 times higher)	313	345,779,122	319,515,504	-	16	-	T500
3	Normale ToOpEx*7 (7 times higher) LtPhCa*14 (14 times higher)	313	691,558,243	319,515,504	-	17	-	T250
4	Normale ToOpEx*14 (14 times higher) LtPhCa*14 (14 times higher)	313	691,558,243	639,031,008	-	19	-	T250

Table 5 Normale university simulation scenarios. Source: own elaboration based on our predictive model.

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 (considered as inputs, x). We are interested in studying, within the various ranking leagues, which universities are using resources efficiently, maximising their ranking score and reaching a higher league. Those efficient universities may exploit better an increase in their resources compared to inefficient universities. Furthermore, we want to understand whether there is a potential relationship between resource use efficiency and the

ARWU score. For the DEA analysis, the total operating expense and the net long-term physical capital were taken as inputs and the ARWU score as output. We decided not to consider students because, in the considered production process, they are, at least in the short run, not under the control of universities. In any case, we carried out as a robustness check an additional efficiency analysis considering, in addition to financial resources, also students as input, obtaining results similar to those reported in this section (for details on this analysis see Appendix B). We apply a classic DEA analysis as our dataset results are convex (from the suitable tests performed). The Returns To Scale (RTS) adopted is the Variable Returns to Scale (VRS), after having verified through an RTS test that it is the correct choice. The inputs analysed were standardised by their empirical standard deviation. The results of the DEA VRS output oriented analysis (considering universities grouped by league) are shown in Figure 6. The figure shows several boxplots coloured according to the leagues, with the number of universities included in each league written in green in the centre of the boxplot. 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 ranking classes 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 are able to 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.

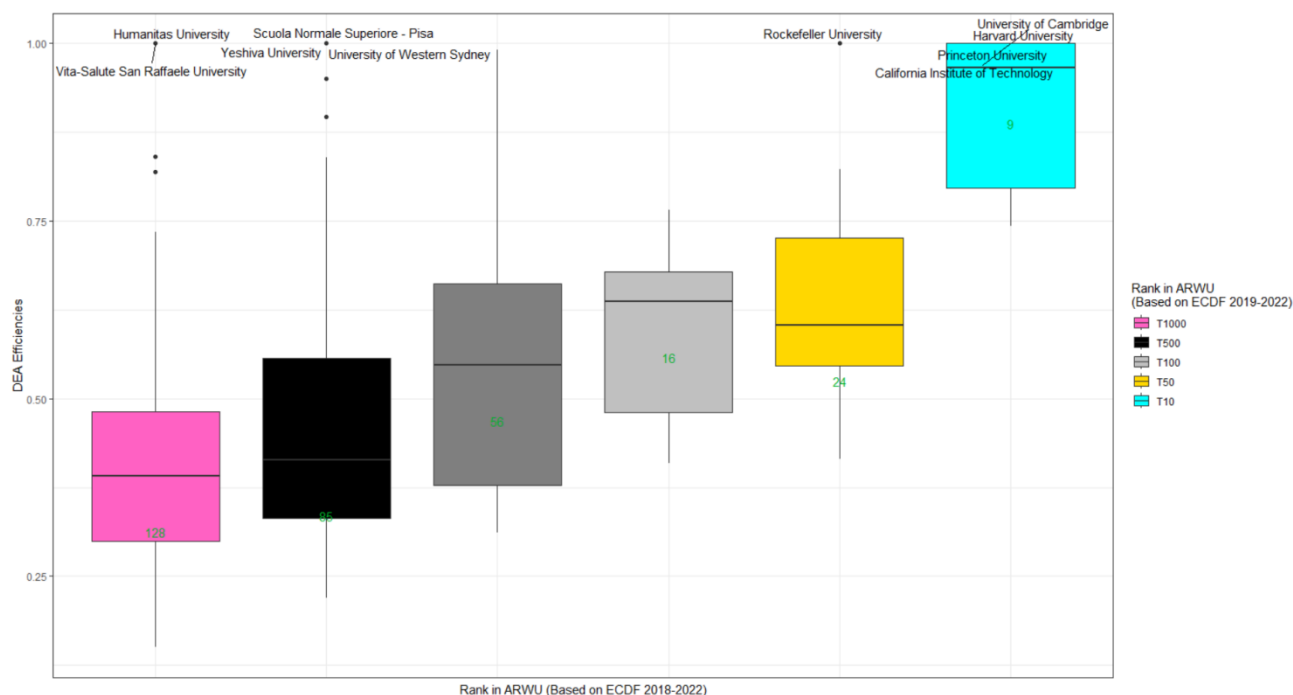


Figure 6 Boxplots of the efficiency scores of universities grouped by ARWU league. From right to left, in light blue the T10 league universities are represented, in gold the T50 league universities, in silver the T100 league universities, in brown the T250 league universities, in black the T500 league universities and in pink the T1000 league universities. The universities whose names are reported in the top of the boxplots are the efficient universities (whose DEA score is equal to 1). Source: own elaboration based on DEA results.

6 Discussion

Data provided from accounting systems are important and should be collected and monitored for each university. In particular, data from financial statements should be taken into account in quantitative analyses on HEIs since they can help to explain differences in outcomes and performances of HEIs. Due to globalisation, stronger and stronger international competition impacts universities regarding attracting students and external funds for large research projects. Therefore, HEIs focus their attention on working on a few key pillars, such as, teaching, research, knowledge transfer and global outlook, to become a world-class university (Stephan, 2012). However, the availability of financial resources could strongly affect these key pillars, as well as, consequently, the competitive placement of the universities in the worldwide higher education industry.

We analysed 318 international universities in the 2019 ARWU edition, identifying a relationship between their economic financial resources and their place in the global ARWU ranking. Then, we partitioned the ARWU ranking into a few leagues, and we developed a predictive (XGBoost) model to

estimate the ARWU score and the corresponding league through the university's financial resources and number of students (using the number of students as a proxy for size). Our findings seem to suggest that a critical condition to make a university great in the world (a world-class university) is to provide the university with a sufficiently high level of financial resources.

Our predictive model is available online at https://simonedileo.shinyapps.io/ARWU_League_Estimator/ as a tool for developing further simulated scenarios for policymakers, in addition to those reported in Section 5.2.1.

We then assessed the efficiency of universities in achieving higher scores in the various ARWU leagues. The analysis shows that universities with higher ranking scores are usually also those that are able to use their resources efficiently. In addition, some universities in the lowest leagues might reap great benefit from a potential increase in their resources because they are already efficient in the use of the limited resources currently available to them.

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 ranking for specific universities. Furthermore, our tool could be exploited by policymakers 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 those generally adopted and based on bibliometric indicators.

However, despite the potential usefulness of our proposed model and predictive tool, and the results obtained, the use of financial data and the estimated ARWU score to predict the leagues still shares the main limitations of other existing rankings (see, e.g., Daraio & 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), even taking inspiration from advanced benchmarks in the higher education sector worldwide. A relevant example of a dataset that collects both important and varied financial data is the American Database for HEIs (<https://nces.ed.gov/ipeds>), namely, the Integrated Postsecondary Education Data System (IPEDS). It is a system of interrelated surveys conducted annually by the U.S. Department of Education's National Center for Education Statistics (NCES). Another interesting example, even if not as extensive and detailed as IPEDS, is represented by the European Database for HEIs (<https://www.eter-project.com/#/home>), namely, 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 (related to operating, investment, and financial activities) 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. For instance, they may not own some buildings for the administrative staff, some facilities for the students, some equipment for laboratories. 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 (especially buildings). 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. In particular, 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. Under the former, 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 (e.g., at the end of every two years) the initial value of the asset (i.e., gross of accrued depreciation) is updated with a rational estimate of the potential market price of the asset (i.e., the price at which the asset can be sold in an orderly transaction to a third party under current market conditions). Essentially, the last updated value of the asset is net of accrued depreciation (as determined after the last update of the value). 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.⁴ Universities applying cash-based accounting might be required to switch from cash-based to accrual-based accounting (as the accrual basis provides a more realistic view of profit and loss during a period of time). For example, in recent years some countries have updated their public law to force the university system to switch from cash to accrual accounting (e.g., Belgium, Italy).

7. Concluding remarks

In this paper, we attempted to provide answers 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 an effective predictive model, available online, to estimate the position in the ARWU ranking (in terms of specific leagues) of a given university by considering mainly its financial data. 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 on it a large amount of financial resources. Finally, 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.

⁴ 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.

Funding:

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

Acknowledgements:

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.

References

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

European Tertiary Education Register (ETER), <https://www.eter-project.com/#/home>, last access 18/05/2023.

The Integrated Postsecondary Education Data System (IPEDS), <https://nces.ed.gov/ipeds/>, last access 18/05/2023.

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.

Experts in higher education data and analysis database, <https://www.hesa.ac.uk/data-and-analysis>, last access 18/05/2023.

USTAT, <http://dati.ustat.miur.it/>, last access 18/05/2023.

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

Asif, M., & Searcy, C. (2014). A composite index for measuring performance in higher education institutions. *International Journal of Quality & Reliability Management*, 31(9), 983-1001.

- 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.
- 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*, vol. 54, pp. 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.
- 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
- 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.
- Emrouznejad A. and 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.
- 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.
- 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) eds, *Research Handbook on University Rankings: Theory, Methodology, Influence and Impact*, Edward Elgar, Cheltenham (UK).
- Ioannidis, J. (2011). Fund people not projects. *Nature*, 477(7366), 529-531.
- Kim, J. The Functions and Dysfunctions of College Rankings: An Analysis of Institutional Expenditure. *Res High Educ* 59, 54–87 (2018). <https://doi.org/10.1007/s11162-017-9455-1>
- Kneip, A. and L. Simar and P.W. Wilson (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), 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). *SWPS*, 1.
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9(70), 209-219.

- 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.
- 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.
- Stephan, P. (2012). How economics shapes science. *Harvard University Press*.
- 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.

Appendix A: Kernel density plots

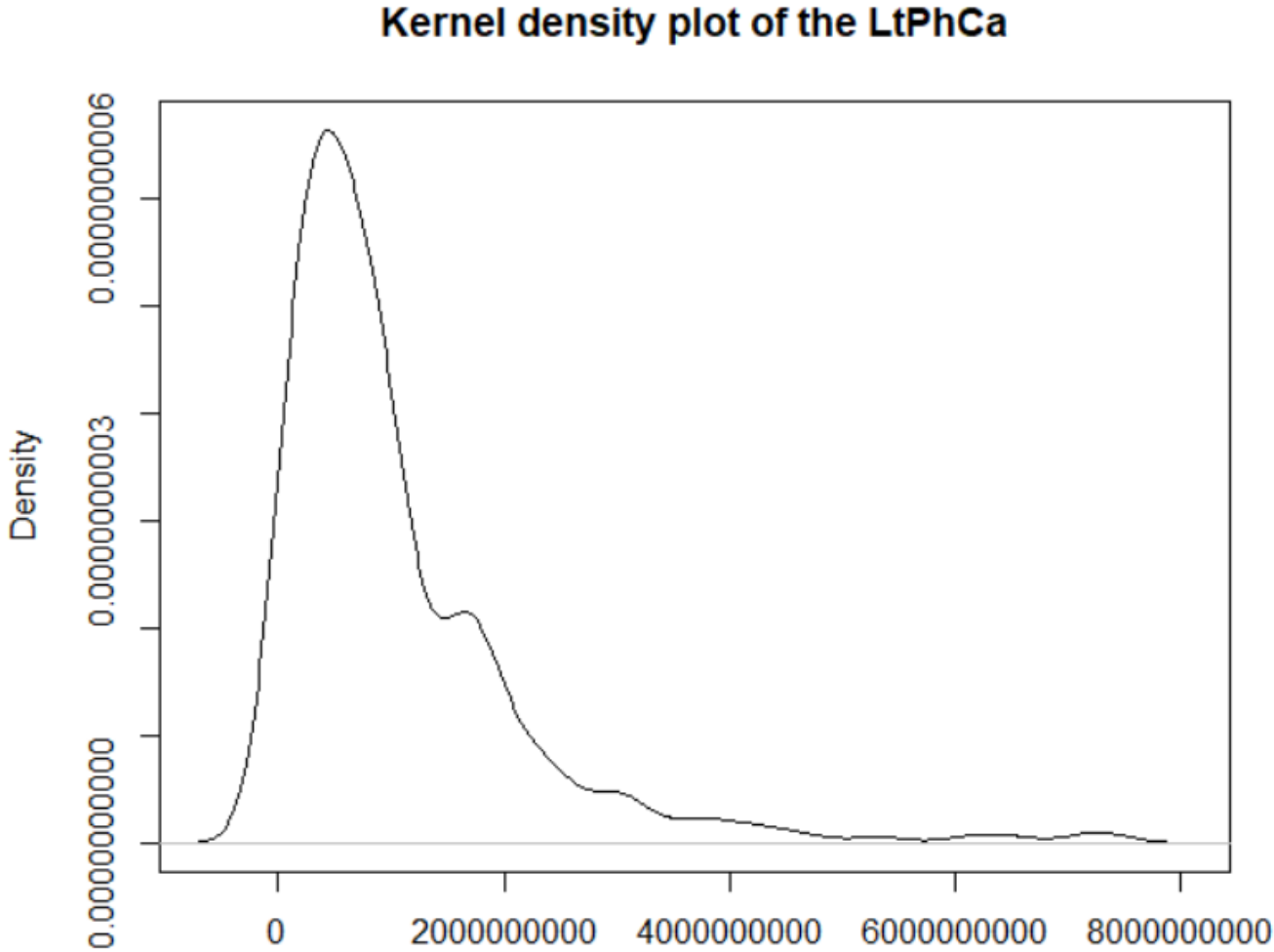


Figure A1 Kernel Density plot of LtPhCa (Number of observations 318). Source: own elaboration based on balance sheets data.

Kernel density plot of the ToOpEx

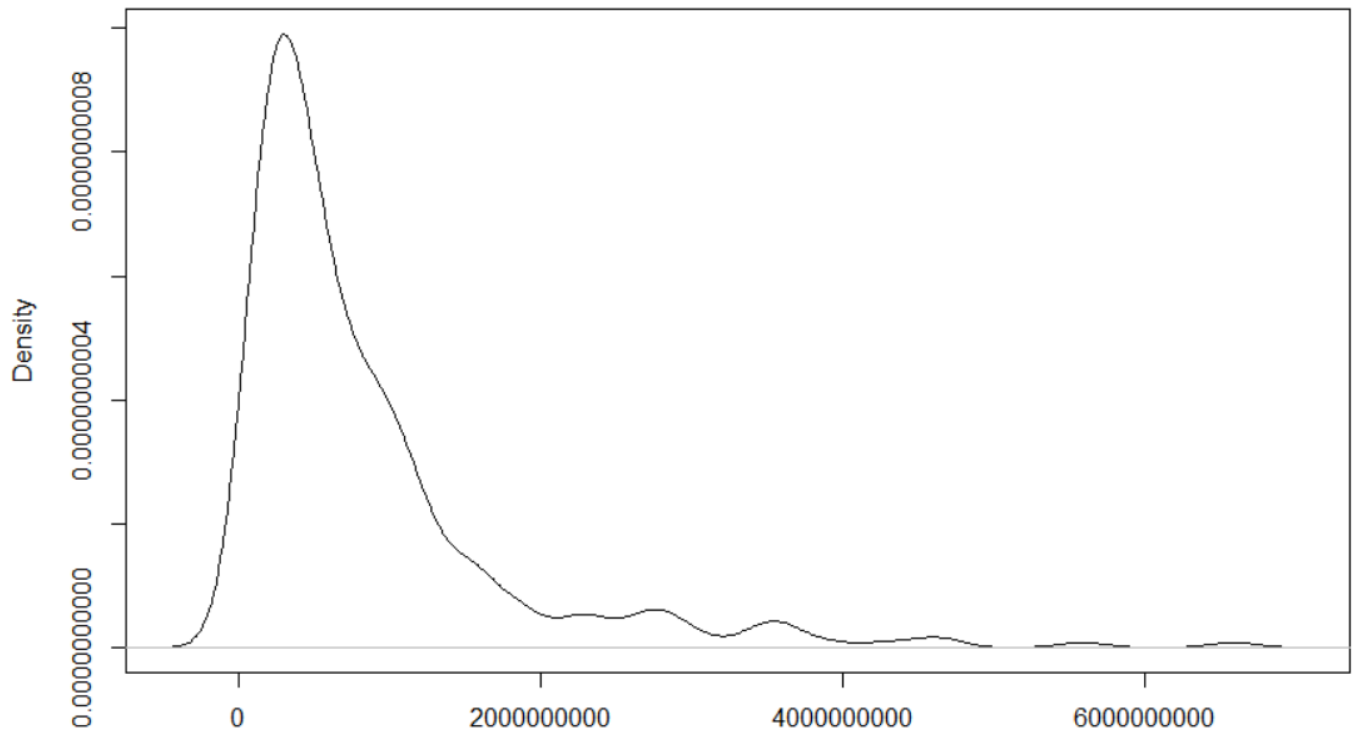


Figure A2 Kernel Density plot of ToOpEx (Number of observations 318). Source: own elaboration based on balance sheets data.

Kernel density plot of the ARWU Score

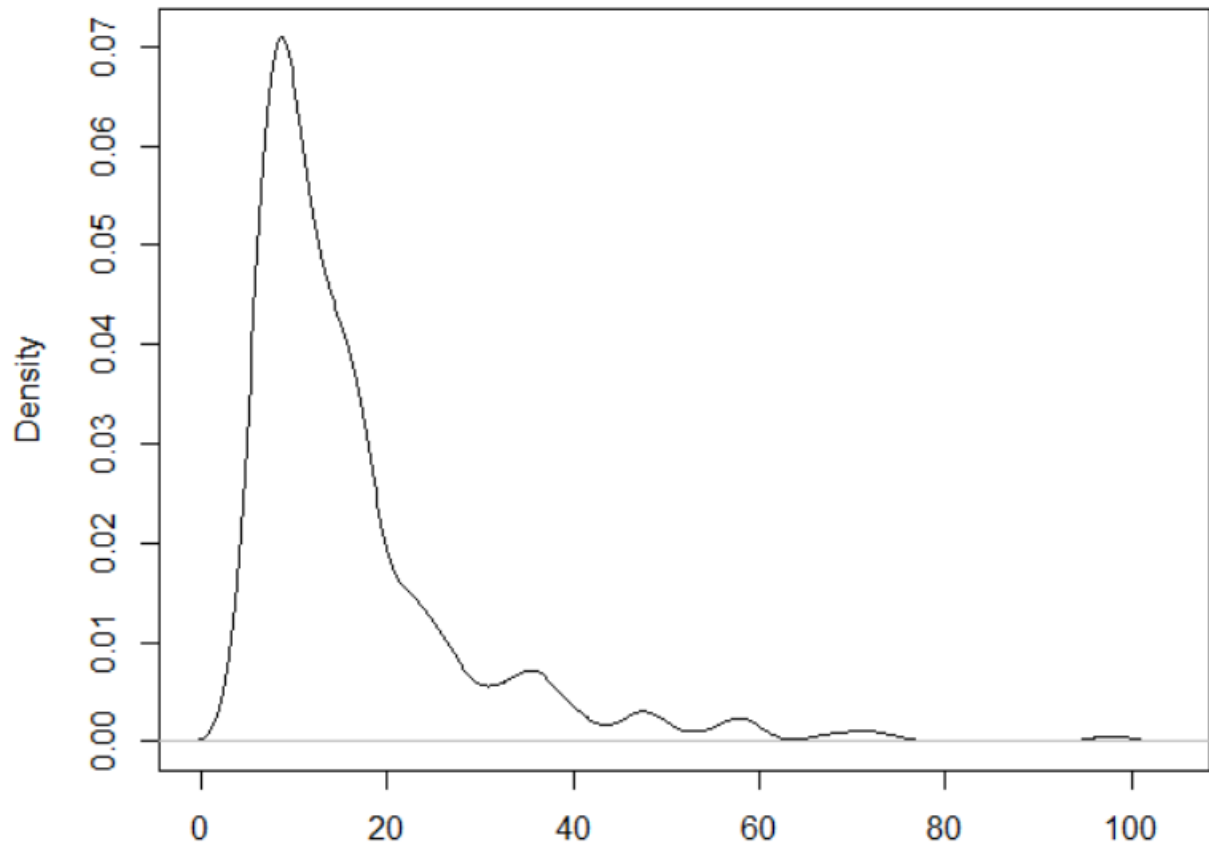


Figure A3 Kernel Density plot of the ARWU score (Number of observations 318). Source: own elaboration based on ARWU data.

Appendix B: DEA also considering the number of students as input

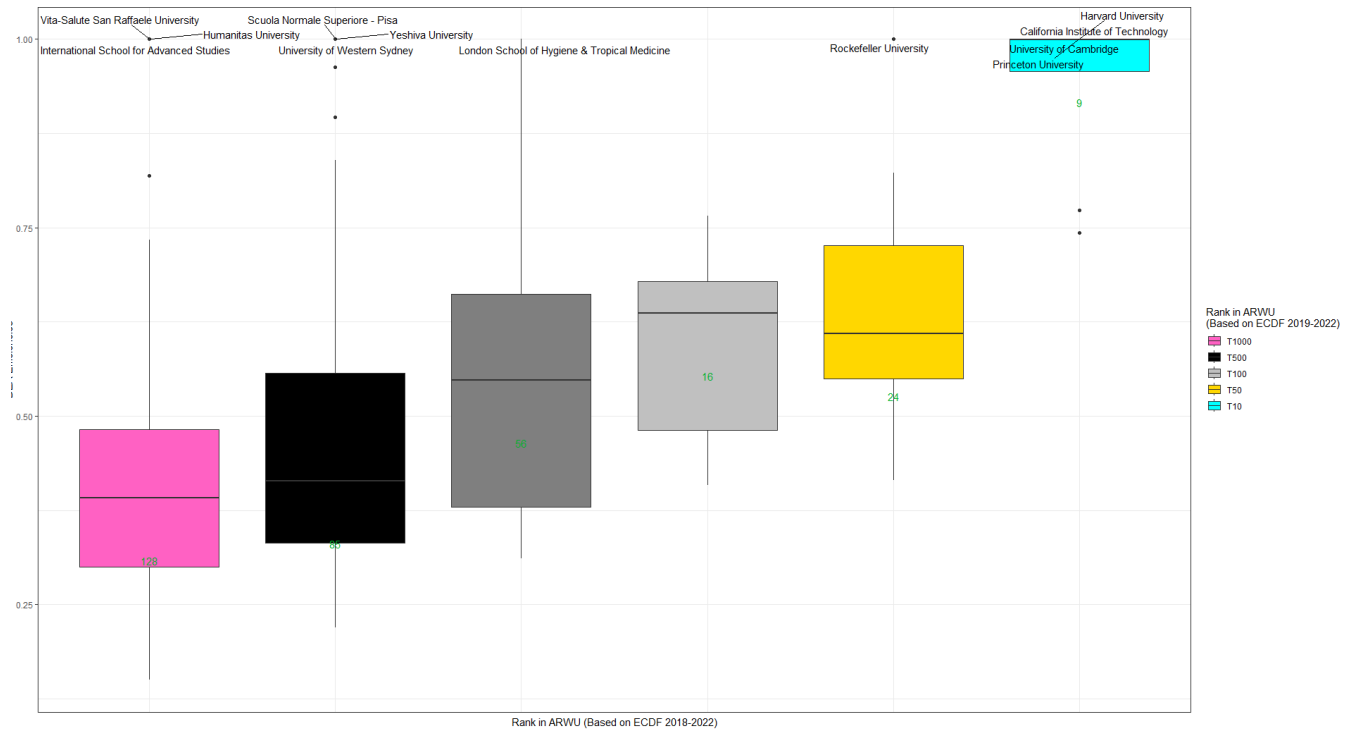


Figure B1 Boxplot of the efficiencies score evaluated considering also the number of students as grouped by league. In light blue the T10 league universities are represented, in gold the T50 league universities, in silver the T100 league universities, in brown the T250 league universities, in black the T500 league universities and in pink the T1000 league universities. Source: own elaboration based on DEA results.